CHARACTERISATION OF VARIANT ALLELES AT THE HUMD21S11 LOCUS

Ms. Sarah Louise Robinson

Master of Science (Research)

2011

CERTIFICATE OF AUTHORSHIP/ORIGINALITY

I certify that the work in this thesis has not previously been submitted for a degree nor has it been submitted as part of requirements for a degree except as fully acknowledged within the text.

I also certify that the thesis has been written by me. Any help that I have received in my research work and the preparation of the thesis itself has been acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

Signature of Student

Production Note:
Signature removed prior to publication.

ACKNOWLEDGEMENTS

There are a number of people who I would like to thank and recognize for their tireless support and advice throughout the studying process. I need to thank Mr. Simon Walsh for his assistance in the development of this project.

I would like to thank the Northern Territory Police, Fire and Emergency Services for generously providing me with samples to study. The Division of Analytical Laboratories and the NSW Police Force were also supportive of this research allowing me time and resources to complete this study. In particular I would like to thank Ms. Carmen Eckhoff for her help and advice and Dr. Narindar Bansal who provided an invaluable amount of information and ideas and enabled me to obtain the physical data recorded here.

Dr. John Mitchell of La Trobe University also provided support and assistance throughout this study. His assistance was provided in his own time, and at a time when his own organization was going through a considerable amount of change.

To my family and friends, thank you for all your support and encouragement, especially during those times when I felt that the research was a lost cause.

Finally, this research would not have been completed without Dr. Tamara Sztynda. Tamara stepped up to be my supervisor and mentored and advised me through this time in addition to her already considerable workload. I will never be able to express how much her support and friendship meant to me through this time.

Table of Contents

List of Illustrations and Tables

List of Illustrations

List of Tables

ABBREVIATIONS

Acronym	Definition
DNA	Deoxyribonucleic Acid
dNTP	Dinucleotidetriphosphate
KYA	Thousand Years Ago
$MgCl_2$	Magnesium Chloride
ηg	Nano gram
PCR	Polymerase Chain Reaction
RPM	Revolutions Per Minute
SSM	Slipped Strand Mispairing
SUPAMAC	Sydney University Prince Alfred
UCO	Unequal Crossing Over
µL	Micro Litre

Abstract

Significant genetic substructure within a population can affect the evidential weight of a DNA profile to the detriment of a defendant. To prevent this from occurring, forensic examiners continually look to understand more about the degree and structure of genetic variation within a population. A part of this involves the characterisation of the microsatellite loci applied in forensic testing.

The HUMD21S11 microsatellite is commonly used in forensic examinations in conjunction with a suite of other microsatellite loci in order to identify individuals who may have been present at a crime scene, or who cannot be identified through more traditional non-genetic means (such as visual, dental or medical records).

This research confirmed the existence of an ancestral relationship between the indigenous populations of Australia and Papua New Guinea as well as the presence of regional differentiation within the Australian Aboriginal population. The sequence variation present at the HUMD21S11 microsatellite locus makes it a suitable candidate to further understand and describe the regional differentiation within the Australian Aboriginal population.

This study also confirmed that microsatellites are able to retain their variability after structural change (Möller et al. 1994, Brinkmann et al. 1996, Griffiths et al. 1998, Walsh et al. 2003), and that a single mutation event can involve single repeat units or multiple repeat units.

The structural complexity of microsatellites like HUMD21S11 could be used to further develop mutation models as well as investigate the proposal that the mutation rate of microsatellite loci is be dependent on the DNA sequence present.

The sequence variation at the HUMD21S11 microsatellite is prevalent enough in the Australian Aboriginal population, to warrant a more complete investigation of the genetic variation at this locus. However, in order to better understand the genetic diversity present in the Australian Aboriginal populations, examination of the

population based on traditional tribal boundaries (rather than judicial boundaries) is recommended.

Additional population studies at the sequence level will increase our level of understanding about the genetic relationships of the Australian Aboriginal population on a local and global level. The sequence data obtained will also assist in the understanding of the mutation process and aid in the development of statistical models.

1. Microsatellites in Forensic Science

1.1 Introduction

The Australian Aboriginal population is one of the most genetically diverse today; however the impact of European colonization has made it difficult for geneticists to understand the complex relationships within and between tribal populations. Sequence variation at the HUMD21S11 locus may provide a means of further understanding the genetic diversity within the Australian Aboriginal population.

In order to gain a fuller understanding of the genetic basis of recent human evolution a more extensive investigation of nuclear genome variation is required (Adcock et al. 2001), and studying autosomal variation at the sequence level will provide additional insight into the processes and effects of mutation and migration on human populations.

1.2 Aims of this research

The main aim of the research described in this thesis was to determine if regional differentiation within the Australian Aboriginal population could be confirmed by determining if the sequence variation observed at the HUMD21S11 locus (namely the presence of the HUMD21S11 Type V sequence), between the South Australian and West Australian Aboriginal populations (Möller et al. 1994, Walsh et al. 2003) could be observed in an additional Australian Aboriginal population from the Northern Territory. This additional information would then assist in determining future research directions utilizing this microsatellite.

The HUMD21S11 microsatellite is commonly used in forensic examinations in conjunction with a suite of other microsatellite loci in order to identify individuals who may have been present at a crime scene, or who cannot be identified through more traditional non-genetic means (such as visual, dental or medical records).

Characterization of the microsatellites applied in forensic studies is required to understand more about the degree and structure of genetic variation within a population. If there is significant substructure within the population then the evidential weight (the likelihood that two individuals share the same DNA profile by chance) may be artificially biased towards the prosecution and thus disadvantage the defendant.

Sequence variation at complex microsatellites (such as HUMD21S11) can be used to further describe the level of genetic variation in the Australian Aboriginal population. This can be used to assist in the resolution of questions regarding the initial colonization of Australia. This will provide further information on the local and global relationships of the Australian Aboriginal people prior to European colonization and the subsequent disruption to traditional lifestyle and cultural practices.

However the time and financial cost of sequencing large numbers of samples can be prohibitive, so this preliminary study was made to determine if sequence variation at the HUMD21S11 microsatellite was sufficiently prevalent in the Australian Aboriginal population, to warrant a full scale investigation of genetic variation at the sequence level.

1.3 Forensic Examination of Microsatellites

The knowledge that each person's DNA is unique to them is a powerful concept and the use of its biological features to investigate biological diversity between both people and between continents is not new. Studying features such as stature, hair, eye and skin color have been used over centuries to describe the differences observed between people(s) and nations (Jobling et al. 2004).

The use of DNA profiling is an extension to the study of diversity through phenotypic or physical features. The technology and application of DNA profiling in the judicial system has undergone three major stages of technological advancement since its early beginnings in the mid 1980's when it was used to identify the man who had raped and murdered two girls (Jeffreys et al. 1985).

This technological advancement has focused on further differentiating between individuals by exploiting variable regions of the genome known as microsatellites. Improvements in technology, in particular the Polymerase Chain Reaction (PCR), have increased the range of criminal investigations to which DNA can be applied. These advancements have led to global use of microsatellite analysis and consequently a large reservoir of data which is available to study and understand genetic variation in a population.

Microsatellites, also referred to as Short Tandem Repeats (or STRs), are particularly attractive to forensic scientists because they allow DNA to be examined from highly compromised samples (such as skeletal tissue). Microsatellites which are highly variable in the population can be specifically chosen, they can be multiplexed (where many areas of the DNA genome are examined in a single reaction), and data collection can be automated.

The microsatellites commonly used within the forensic science community are assumed to be selectively neutral. This assumption is because the loci examined are non-coding and so have no subsequent gene product. There is little evidence of any selective pressure on the autosomal microsatellites used in forensic examination even though some of them were discovered because of their closeness to the genes being studied (Balding 2005, Buckleton 2005).

1.3.1 Forensic discipline modifications of molecular biology terms

The majority of microsatellites used by forensic scientists are not genes, and as such the variants observed at a microsatellite locus cannot strictly be classed as alleles. However the terminology used within the forensic science field refers to variants at a microsatellite locus as alleles. The forensic specific terminology will be applied here in this dissertation as the research arose in response to an earlier forensic study to describe rare variants within an Australian Aboriginal population (Walsh et al. 2003).

This is to alert the reader not involved in routine forensic biology work that this 'divergent' use of terms is intentional because it is focused for the readership of forensic biologists.

1.3.2 Microsatellite Structure

Microsatellites are commonly referred to as being simple, compound and complex. Simple loci are the most commonly used in forensic applications. They consist of a single repeated sequence (repeat unit) of four base pairs. For example the HUMTH01 locus which has a repeat sequence of (TCAT) and this sequence has been observed to repeat between 4 and 13 times (Brinkmann et al. 1996b, Griffiths et al. 1998, Buckleton et al. 2005).

Compound microsatellites such as HUMVWF31A, consist of repeated sequences such as $(TCTA)_{1-2}(TCTG)_{3-4}(TCTA)_{3-16}$ (again the number refers to the number of times each repeat unit has been previously observed) (Brinkmann et al. 1996b, Griffiths et al. 1998, Buckleton et al. 2005).

Complex microsatellites such as HUMD21S11are less uniform in the distribution of repeat units, which may be interspersed with non-repeat sequence (Brinkmann et al. 1996b, Griffiths et al. 1998, Buckleton et al. 2005). The sequence and structure of the HUMD21S11 locus is described later in this chapter.

1.3.3 Microsatellites in DNA Profiling

Currently, determining an individual's DNA profile using microsatellites, the DNA is extracted from the source material (for example, blood left at a crime scene) and then amplified using the PCR. This copies a specific region of the DNA (or multiple regions when a multiplex reaction is used) (Kimpton et al.1993). Thus a fragment of amplified DNA (termed an allele) is created for each chromosome a person has (for example in a medical disorder such as a trisomy where an individual has three copies of a chromosome, three fragments will be created).

The amplified DNA fragments are then run through an electrophoretic field and compared against an "allelic ladder". The allelic ladder comprises a set of PCR products (amplified DNA fragments) corresponding to different alleles (or variants) across the known size range of alleles for a microsatellite locus (Buckleton et al. 1995).

The alleles in the allelic ladder are of known sequence, with allelic nomenclature based on the number of repeat units present in an allele (the amplified DNA fragment). For example a HUMTH01 5 allele would be expected to contain five repeat units (of the sequence TCAT). Amplified DNA fragments are assumed to contain the same number of repeat units as the allele it aligns to in the allelic ladder.

Where a partial repeat unit is present, the size of the partial repeat is given in bases after a decimal point. For example a HUMTH01 9.3 allele would have 9 complete repeat units and a three base pair partial repeat sequence.

For microsatellites then, allele nomenclature is based on the size of those alleles in the allelic ladder. While this approach does not take into consideration variations at the sequence level which are known to exist in complex microsatellites (Möller et al. 1994, Brinkmann et al. 1996b, Walsh et al. 2003, de Kock et al. 2006), linking the allelic ladder and the nomenclature of STR loci allows standardization across laboratories to occur. This standardization is required to enable laboratories to compare results across time and space.

Apart from the more prohibitive cost (time and money) of sequencing each allele, comparing DNA profiles based on the size of an amplified fragment allows laboratories a more practical means of searching and databasing DNA profiles (Gill et al. 1995).

1.4 Mutation

To further understand the genetic diversity or variation in and between populations, the processes of mutation and migration and their effect on the introduction and maintenance of genetic variation must also be considered.

For the interpretation and measurement of the effect of mutation on genetic diversity, the physical mechanism needs to be known. There are a number of different mechanisms possible, and data suggests that no one mechanism or model of mutation at work (Brinkmann et al. 1996b, Brinkmann et al. 1998, Amos 2003).

1.4.1 Mutation Mechanisms

It is generally accepted that mutations at microsatellites involve the addition or deletion of one or more repeat units [through the process of slipped strand mispairing (Eisen 2003)]; however there is growing evidence that larger mutation events of multiple repeat units can occur (Brinkmann et al. 1996b, Griffiths et al. 1998, Xiao et al. 1998, Wiegand et al. 2000, Immel et al. 2004, deKock et al. 2006, Tsuji et al. 2006). These larger mutation events may be due to either unequal crossing over (UCO) or slipped strand mispairing (SSM) (Eisen 2003).

Unequal Crossing Over (figure 1) is most often associated with large mutation events. It can occur during the process of recombination (where there is an exchange of genetic information) when the chromosomes are misaligned. Areas of the genome comprised of regions of repeated sequence (such as microsatellites), have an increased likelihood of misalignment occurring (Smith 1973).

UCO is thought to occur most easily for long, tandemly repeated DNA sequences because the recombination process cannot readily determine if the strands are correctly aligned. The end result is a deletion in one chromosome and an insertion in the other chromosome (Hancock 2003).

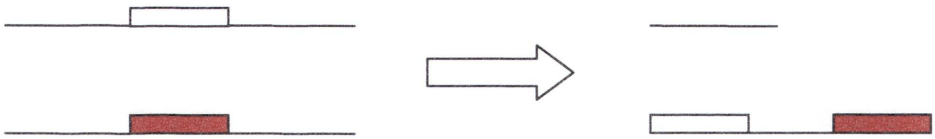

Figure 1 - Unequal Crossing Over - Unequal crossing over (UCO) of chromosomes during DNA replication results in a deletion in one chromosome (the top chromosome) and an insertion of two repeat units from the first chromosome between two repeat units in the second chromosome (the bottom chromosome now has two repeat units).

Slipped strand mispairing (figure 2) is associated with smaller mutation events and occurs when the DNA polymerase slips during replication (Eisen 2003). As for UCO, if the template strand is made up of a repeated sequence, then the new DNA strand (the copy-strand) may reanneal out of phase to the template strand (the unannealed DNA loops out), producing a copied strand which is longer or shorter than the template strand at the end of the replication process (Hancock 2003).

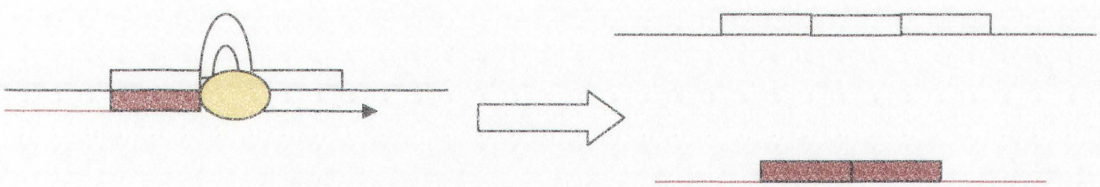

Figure 2 - Slipped Strand Mispairing - The template strand (black) has had the middle repeat sequence looped out by the action of DNA polymerase (yellow) on the strand being copied (red) annealing to the third repeat on the template strand. The resulting copy now has two repeats, while the template strand maintains the original three repeats.

The mispairing of DNA strands is not such a problem when non-repetitive sequences are being copied because the copy-strand can only reanneal to the template strand in one way (Eisen 2003).

Some mutations can be corrected through exonucleolytic proofreading or mismatch repair (Eisen 2003). This is where errors in a newly synthesized DNA strand are degraded by the exonuclease so that the DNA polymerase is able to reverse and recopy the strand correctly. This limits the number of mutations arising from base misincorporation, and is most effective when the error is only a few bases away from the polymerase's active site.

The stabilising effect of proofreading on microsatellites is limited and predominantly effects microsatellites where the repeat unit is small, and only a small number of repeat units are present. Most DNA loops caused by SSM are too far from the replication fork to be corrected through exonucleolytic proofreading (Eisen 2003).

1.5 Migration

In the context of this study, migration refers to the movement of people (s) from one area to another. Migration allows the introduction of new genetic information into a population, as well as limiting genetic divergence between them. When the new migrants successfully introduce new genetic material to an area, this is known as gene flow.

If populations start to diverge genetically, because there is no gene flow between them, then the level of relatedness within each population increases, while the level of relatedness between the diverging populations decreases. This affects the estimate of evidential weight of a DNA profile.

To understand the genetic local and global relationships between populations (locally and globally) we also need to know something of the non-biological factors affecting the interaction of populations needs to be known. These factors can include ethnic, sociocultural [for example the caste system in India (Reddy et al. 2005)] and linguistic differences (Kirk 1989).

Also the more readily observed physical and environmental barriers such as geography (mountain ranges or deserts), ecological and climactic influences should be considered. Climate can limit the opportunity to move away from an area, or alternatively is so congenial there is no need to search elsewhere for resources.

All modern humans belong to the species *Homo sapiens* and are all related genetically. There are two predominant theories available to describe how this has occurred: the 'Multi-Regional' and 'Out-of-Africa' theories. The Multi-Regional theory argues that modern man evolved independently all over the world, while the Out of Africa theory's central concept suggests all modern humans migrated out of Africa and subsequently around the globe (Jobling 2004).

The most scientifically and consistently supported argument is the Out-of-Africa theory (Takahata et al. 2001, Liu et al. 2006).

1.5.1 Application of DNA in Tracing Human Populations

Genetic variation in the population can be used to compare these two theories. Genetic mutations and differences in the distribution of variable DNA regions allow us to trace populations back to their early ancestral and geographic origins. The different ways that DNA can be inherited, determine which form is going to provide the most information for a particular investigation.

Gender specific inherited DNA such as mitochondrial DNA (which is maternally inherited from the ovum cytoplasm) and Y-chromosome DNA (which is paternally inherited from the sex chromosome in male offspring); pass on relatively unchanged between generations. This makes it easier to trace populations because the genetic information remains relatively unchanged between generations.

Autosomal DNA (those excluding the X and Y sex-determining chromosomes) is inherited from both parents. Each child (potentially) then has a DNA profile that is different from their parents, and their siblings. This can make it difficult to use

autosomal DNA as means of tracing genetic pre-history because successive generations will have a different autosomal DNA profile from their parents. However, there is less chance of the parental DNA information being lost to subsequent generations.
For example if a woman only has sons, then her mitochondrial DNA will not be passed on to her grandchildren and it is 'lost'.

1.5.2 Migration and Population Processes

Population processes such as the Bottleneck and Founder Events (Hartl 1999, Jobling 2004, Balding 2005), along with the microevolutionary mechanisms of migration and genetic drift can be used to describe the genetic diversity observed in a modern population. Conversely, the level of genetic diversity in a modern population can be used to make predictions about what has previously happened to a population.

Founder events relate to those instances where a small group of emigrants from an established population founds a new subpopulation, for example, where a large number of people from one population move from one geographic area to another. It is different to a Bottleneck event which is where an established population goes through a severe, but temporary, reduction in population size. Such as when the population is struck by a disease or natural disaster which reduces the number of adults who are able to reproduce (figure 3).

Both Founder and Bottleneck events result in a reduction in the genetic diversity present in the resulting population. This, in turn, reduces the amount of diversity in subsequent generations. The effect of these events can often be observed in modern populations (particularly in Pacific or Remote Oceania populations) where affected populations can show high frequencies of rare alleles and in some instances, high incidences of genetic disease (Lum et al. 2002, Jobling et al 2004, Crawford 2007).

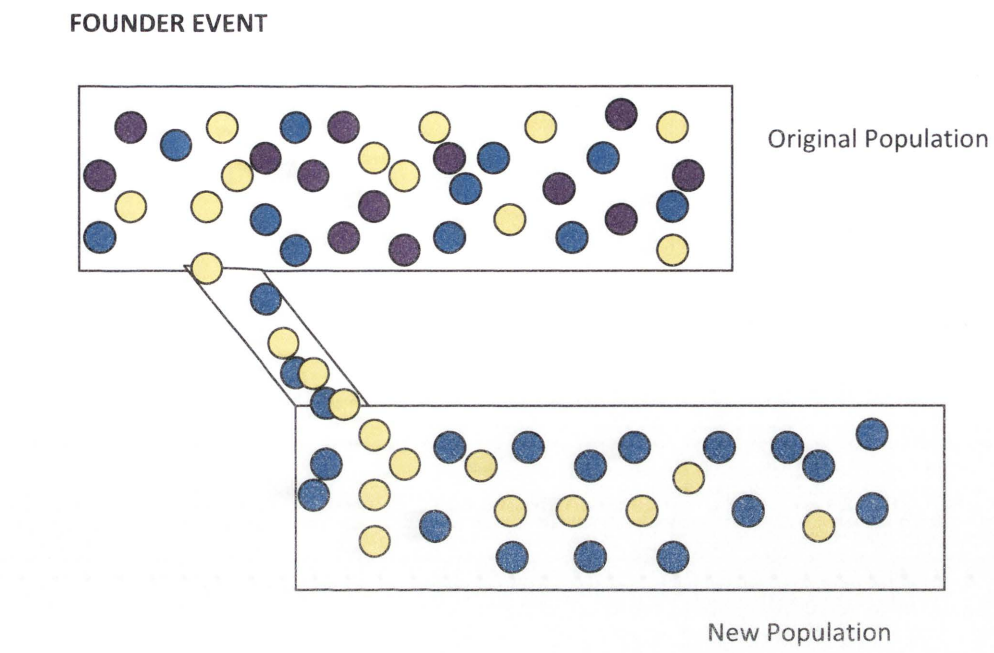

Figure 3 - Founder and Bottleneck Events - Each circle represents different alleles. The Founder effect occurs when a subset of the original population separates and forms its own population. Sometimes not all the genetic information available in the original population is taken in the founding population and so over time the two populations diverge from each other. The Bottleneck event occurs when the population size contracts dramatically and reduces the amount of genetic variation available to subsequent generations.

High frequencies of rare alleles can also be observed in isolated island populations, where with little migration, the population on the island becomes less genetically related to populations on surrounding land masses. This effect can also occur where a

population becomes isolated due to cultural beliefs like religion (Crawford et al. 1989) and does not genetically mix with neighboring populations allowing the populations to genetically diverge.

1.5.3 Genetic Evidence Supporting the Out of Africa Model

Mitochondrial, Y-chromosome, and autosomal DNA examinations all indicate that all non-African populations are derived from the African population. African populations are the most genetically diverse, and all non-African populations (such as those from Europe, Asia, the Americas, and Australasia), exhibit a small subset of the genetic variation observed in Africa (Jobling et al. 2004).

In addition, as the geographic distance from Africa increases, the lower the genetic diversity of the population becomes (Liu et al. 2006), further supporting the argument that there was only one migration occurrence out of Africa, and that the earth was then populated by people who came or migrated away from this initial group.

The Out of Africa model is further supported by the global pattern of genetic differentiation which reflects isolation by genetic drift rather than admixture (Crawford 2007). Genetic drift is observed as a result of random fluctuation in allele frequencies due to random sampling among the individuals of the population. That is, it is only by chance that parental genes are not completely represented in the next generation.

Drift can also play a major role in the amount of variability present if the population has a small number of reproductively active adults (effective population size). This further reduces the amount of genetic information that can be passed on to future generations because fewer individuals are having children.

Admixture occurs where a hybrid population is formed as a result of two ancestral populations coming together (for example where Europeans colonized continents such as Australia and the Americas introducing European genetic information into previously isolated populations).

1.5.4 Migration Into and Out of Africa

Within the African continent, there is evidence for the long-term migration of populations from east Africa (Sudan and Ethiopia) into the south and west of Africa (below what is now the Sahara desert) starting between 160,000 and 135,000 years ago. Populations moved to the Cape of Good Hope, south west to the Congo Basin and west to the Ivory Coast.

The warming of the environment between 135,000 and 115,000 years ago, made the area habitable and enabled people to migrate across the Sahara and into the North of Africa, while the Nile River allowed passage into the Levant (Middle East) (Wells et al. 2003). There is no indication though that these people made a lasting impact on today's genetic record, with a change in climate (a global freeze) turning the Levant and North Africa into extreme desert, resulting in the death of these first migrants.

The next migration out of Africa is believed to have occurred 85,000 to 90,000 years ago, and under the Out of Africa model, it was this group that gave rise to all non-African populations. The archaeological record indicates that these original migrants moved east out of Africa crossing the Red Sea and then following the southern coast of the Arabian Peninsula towards India. The presence of fossilized remains of stone tools and shellfish in middens provide evidence of coastal living along this route (Underhill et al. 2001, Wells et al. 2003).

Genetically, the primary split between African and non-African populations has been calculated as occurring between 44-200 thousand years ago (Tishkoff et al. 1996, Crawford 2007).

There is some evidence (based on Y-chromosome and mitochondrial DNA) to suggest that migration in Africa was not entirely one-sided with gene flow between African and Middle East populations (especially between Yemen and Ethiopia) (Spurdle et al.1996, Thomas et al. 2000, Crawford 2007).

Over the next 10,000 years, people continued to move along the coast towards Sri Lanka, western Indonesia (which was then part of the Sunda landmass and attached to Asia) and then north to Borneo and South China. Shell middens show that populations continued to live along the coast, however the majority of ancient middens are now likely under water (Wells et al. 2003).

The eruption of Mt Toba (India) 74,000 years ago spread volcanic ash up to 5 metres deep over India and Pakistan, creating a world-wide nuclear winter and an instant 1000 year ice-age. The presence of tools in the Toba ash act as an indicator of how far and quickly into South East Asia the initial migrants had come after leaving Africa (Ambrose 1998).

It has been suggested that this eruption produced a dramatic population crash. A dramatic drop in population number has been proposed to explain why there is a genetic break between India and the Far East. While the Indian population still shows the genetic trail of the original population that moved out of Africa, it also has maternal DNA haplogroups that are unique to it, as well as population expansion dates that are younger in India than elsewhere in East Asia and Australasia (Ambrose 1998).

1.5.5 Journeying South into Sahul

After the Mt Toba eruption, it has been suggested that India was repopulated by people moving west along the coast back into India, as well as south east from Timor and Borneo into the continental land mass Sahul (what is now Australia and New Guinea).

The current continent of Australia contains the earliest archaeological evidence for modern human colonization outside Africa with the accepted dates of human arrival being at least 50,000 years ago (Cavalli-Sforza et al. 1994). Climatic and genetic evidence also supports the dates estimated by archaeological data.

Climatically, the northern ice caps locked up large volumes of water which increased the Sahul landscape (much of which is now under water). The sea levels stayed at this level for approximately 5,000 years, until they began to rise as the ice age ended. This time period would have enabled migration into Sahul from South East Asia. Sahul has always been geographically isolated from Asia (which was then part of the larger land mass Sunda) as evidenced by the distinctive flora and fauna. Coastal migration, or 'island hopping', has been proposed as the most likely means of arrival into Sahul (Flood et al. 2004).

Genetically, the estimated date of expansion for Australian populations occurs at approximately 68,000 years ago (Bradshaw 1996). Genetic and archaeological data indicates that Sahul was settled very quickly. Archaeological evidence indicates the existence of people in the north (50,000 years ago), south east (40,000 years ago) and central (20,000 years ago) regions of Australia (Jones 1987, Smith 1987, Roberts et al. 1990, Thorne et al. 1999, Roberts and Gillespie 2000).

While dates derived from the archaeological data may seem difficult to correlate with those estimated by genetic data, it must be remembered that many of the initial landing sites, and living sites of the very first migrants, are now under water.

An argument against an early settlement date of Sahul has been the assumption that the early migrants did not have the technology to construct the type of sea craft required to cross the seas between Sunda and Sahul. However, archaeological evidence and practical tests suggests that this was not the case (Flood 2004). Records made at the time of European settlement of Australia, also describe the use of sea craft designed to travel large distances, and how the indigenous population were able to take advantage of tides, and transport fire as required by the indigenous population (Tindale 1962, Flood et al. 2004).

As sea levels rose and Sahul slowly split into the islands of Australia, New Guinea and the Torres Strait, sea voyages would have become increasingly hazardous and there is

no definitive evidence to suggest that there were further waves of migration into Australia after 8,000 years ago (Flood 2004).

It is unclear if the resulting environmental sea changes completely isolated Australia from further migration. Studies to date have been conflicting in providing clear evidence of additional migration events into Australia (prior to European colonization), particularly between 8 and 10,000 years ago (Kirk and Thorne 1976, Nei and Roychoudhury 1993, Cavalli-Sforza et al. 1994, Redd and Stoneking 1999, Adcock et al. 2001Redd et al. 2002).

There is archaeological evidence of a marine network operating in the Indian Ocean and Bay of Bengal, and this fits in well with observations of genetic admixture in the coastal population of New Guinea (Flood 2004, Kayser et al. 2008). A marine network would have facilitated contact and migration of new people into the area including that of north east Australia. The archaeological and linguistic evidence of contact with Macassan sailors also indicates that the continent of Australia was not completely isolated (MacKnight 1972, Walker and Zorc 1981).

The genetic evidence or otherwise for multiple migratory waves into Australia is examined further in the discussion.

1.5.6 Colonization of the Middle East, Europe and the Americas

From Borneo and South China, migration into Asia continued from the south, containing migrants from the same population that spawned those migrants that went onto Sahul. As the climate warmed the subsequent environmental change allowed the Middle East and Europe to be populated.

Genetic and archaeological evidence suggests there were multiple parallel migrations into the Americas starting 22-25 thousand years ago when the Bering Land Bridge formed between north east Russia and Alaska, when sea levels were at their lowest Migrants came from northern China, southeastern Siberia and Mongolia and tribal

differentiation in mitochondrial DNA indicates that there was still variation in the physical appearance of the founding population (Wallace and Torroni 1992).

The earliest migrations led to settlement in North and South America while later migrations appear to have stayed in North America alone (Wallace et al. 1992). Travel into North and South America appear to have occurred through a mixture of coastal and land routes depending on the environmental conditions at the time

1.5.7 Colonization of Remote Oceania

Remote Oceania (which includes New Zealand, Hawaii, Fiji, Tonga and the Solomon Islands) is believed to be the last geographic region area to be populated and colonized. The genetic diversity of the populations within this area is characterized by a series of sequential bottleneck events rather than founder effects, where ancestral populations had their genetic diversity restricted at a number of times during their migration into the area (Lum et al. 2002).

Genetically, the evidence points to the starting populations of this region originating from South East Asia (Hagelberg et al. 1999, Su et al. 2000). Post-settlement gene flow between populations in this region appears to correlate with geographic distance rather than linguistic affinity, indicating that the interaction between populations and migrants to the region was not restricted by the traditional boundaries of geography or language (Crawford 2007).

Population of these regions was by people who were excellent boat builders capable of successfully navigating and sailing large tracts of water. An earlier hypothesis for the origin of the Remote Oceanic population was that they had sailed from South America. While being physically possible, genetically there is no evidence to suggest that there is any South American genetic influence on the populations of Remote Oceania.

1.6 The HUMD21S11 Locus

Located on chromosome 21, the HUMD21S11 is a polymorphic complex four base pair (tetranucleotide) repeat commonly used in forensic multiplex amplification systems (Sharma and Litt 1992, Walsh et al. 2003).

While microsatellites are not generally considered to be the most efficient means of tracing populations through time (Jobling 2004), complex microsatellites such as HUMD21S11 may be of more use because they are comprised of more than one type of repeat unit and may also have additional variation at the sequence level (Möller et al. 1994, Brinkmann et al. 1996b, Walsh et al. 2003, de Kock and Kebede 2006).

There are four parts to the structure of the HUMD21S11locus: three variable regions and a constant region. Microvariation in the structure of the HUMD21S11locus led to the description of five HUMD21S11 allele types.

HUMD21S11 alleles described as being of Type I have been reported in all populations where the HUMD21S11 locus has been sequenced (Möller et al. 1994, Brinkmann et al. 1996b, Griffiths, Xiao, Wiegand, Grubweiser, Walsh et al. 2003):

Variable Region I	Variable Region II	Constant Region	Variable Region III
$(TCTA)_{4-6}$	$(TCTG)_{5-6}$	$[(TCTA)_3 TA (TCTA)_3 TCA (TCTA)_2 TCCA TA]$	$(TCTA)_{8-13}$

where A, T, C and G represent the four DNA sugars adenosine, thymine, cytosine and guanine; and the numbers after each four base pair repeat unit represents the number of repeat units observed in each of the four regions.

The HUMD21S11 allele nomenclature is based on the combined number of four base pair repeat units within each variable region and within the constant region. The nomenclature closely adheres to the ISFH recommendations (DNA Commission of the ISFH 1994) with the number of complete repeats observed designated by digits and the

number of any additional base pairs indicated after a decimal point (where a repeat unit is one of the four base pair units TCTA and TCTG).

The following is an example of a HUMD21S11 allele with 34 repeats:

Variable Region I	Variable Region II	Constant Region	Variable Region III
$(TCTA)_5$	$(TCTG)_6$	$[(TCTA)_3\,TA\,(TCTA)_3\,TCA\,(TCTA)_2\,TCCA\,TA]$	$(TCTA)_{15}$

Not all HUMD21S11 alleles have a whole number of repeat units, and in most populations, these alleles are identified by the number of complete repeats present, and the number of additional base pairs recorded after a decimal point. Only Variable Region III appears to contain any incomplete repeat units.

Incomplete repeats are generally thought to be created through base pair insertion or deletion, and are most commonly assumed to be of Type II. The Type II alleles share the same structure as Type I, but with an additional dinucleotide unit (TA) in Variable Region III.

Type II alleles have also been reported in all populations where the HUMD21S11 locus has been sequenced (Möller et al. 1994, Brinkmann et al. 1996b, Griffiths, Xiao, Wiegand et al. 2000, Grubweiser, Walsh et al. 2003):

Variable Region I	Variable Region II	Constant Region	Variable Region III
$(TCTA)_{4-6}$	$(TCTG)_{5-6}$	$[(TCTA)_3\,TA\,(TCTA)_3\,TCA\,(TCTA)_2\,TCCA\,TA]$	$(TCTA)_{10-13}TA$ (TCTA)

The Type III and IV alleles have so far only been observed in the African Bantus population (from Namibia in south west Africa). When compared to alleles of Type I and II, Type III alleles are characterised by an increased number of repeats in Variable Regions I and II. While Type IV alleles show an increased number of repeat units in

Variable Region III, and also contain one to two tri-nucleotide units (TCA) in Variable Region III, in addition to a dinucleotide unit (TA) as observed in Type II alleles.

The sequence structure of the Type III and Type IV alleles (Brinkmann et al. 1996b) are shown below, with the range in repeat unit number observed in each Variable Region and Constant Region where variation has been shown to occur:

Type III (Brinkmann et al. 1996b):

Variable Region I	Variable Region II	Constant Region	Variable Region III
$(TCTA)_{10-11}$	$(TCTG)_{5-6}$	$[(TCTA)_3 TA (TCTA)_3 TCA (TCTA)_2 TCCA TA]$	$(TCTA)_{11-13}$

Type IV (Brinkmann et al. 1996b):

Variable Region I	Variable Region II	Constant Region	Variable Region III
$(TCTA)_5$	$(TCTG)_6$	$[(TCTA)_3 TA (TCTA)_3 TCA (TCTA)_2 TCCA TA]$	$(TCTA)_{14}$**TCA** (TCTA)**TCA**(TC TA)**TA** (TCTA)
$(TCTA)_5$	$(TCTG)_6$	$[(TCTA)_3 TA (TCTA)_3 TCA (TCTA)_2 TCCA TA]$	$(TCTA)_{15}$(TCTA) **TCA**(TCTA)**TA** (TCTA)

Type V alleles (the final type described), were originally only observed in a Papuan population sample (Brinkmann et al. 1996b). It has since been observed in an Aboriginal population from West Australia (although it was not observed in an Aboriginal population from South Australia), and while it has been observed in the Caucasian population, it appears to be quite rare (Xiao et al.1998, Walsh et al. 2003).

As in Type III, Type V alleles are characterized by an increased number of repeats in Variable Region I. However Type V alleles also have an increased number of repeats in Variable Region II as well as a fourteen base pair deletion in the Constant Region.

The sequence structure of the Type V alleles (Brinkmann et al. 1996b) is shown below, with the range in repeat unit number observed in each Variable Region and Constant Region where variation has been shown to occur:

Type V (Brinkmann et al. 1996b):

Variable Region I	Variable Region II	Constant Region	Variable Region III
$(TCTA)_{9-11}$	$(TCTG)_{11}$	$[(TCTA)_3$----------TCA $(TCTA)_2$ TCCA TA]	$(TCTA)_{11-13}$

The deletion in the Constant Region has also been observed in alleles which show a typical Type I structure (Xiao et al. 1998, Wiegand et al. 2000). This will be discussed further in Chapter four.

For all Types of the HUMD21S11 alleles described, variation in allele size is a result of either increases or decreases in the number of repeat units present in the three Variable Regions. The sequence of the Constant Region is maintained after mutation has occurred, and retains the same sequence across intermediate alleles.

While there is more support for slipped strand mispairing (or slippage) being the predominant cause of mutation at microsatellites, complex microsatellites such as HUMD21S11 may be affected by both slipped strand mispairing and unequal crossover mutation mechanisms (Brinkmann et al. 1996b).

Unequal cross over could explain the large mutations observed between the Constant Regions observed in Type V and other allele Types, and between the number of repeats in the Variable Region I when alleles of Type I and Type III and V are compared. Slipped strand mispairing is assumed to be the mechanism creating the single repeat unit variations observed in the three Variable Regions (Brinkmann et al. 1996b). The relationship between the different Types will be discussed further in Chapter four.

Early investigations of HUMD21S11alleles (or variants) frequently sequenced the alleles analyzed, and it was observed that variation was present at the sequence level

which was undetectable when looking at the amplified allele fragments alone (Möller et al 1994). Alleles at the one locus could share the same base pair size, and yet have different sequences (Möller et al 1994, Walsh et al. 2003).

This phenomenon is known as size homoplasmy, and occurs where two alleles of the same base pair length are not copies of the same ancestral sequence (said to be identical in state) because they are the same size as a result of coincidental mutation rather than from being inherited from a common ancestor (where alleles are identical by descent).

For calculations of genetic divergence (when calculated from microsatellite data), size homoplasmy underestimates genetic divergence between populations over large time scales. However it does not represent a significant problem for many types of population genetic analyses and the large amount of variability at microsatellite loci often compensates for their homoplasious evolution. Homoplasmy however, may be a problem for those DNA markers with high mutation rates and large population sizes with strong allele size constraints (Estoup et al. 2002).

2. METHODOLOGY

2.1 Selection of samples

Ethics approval for the project was provided by the University of Technology, Sydney, Human Research Ethics Committee (UTS, HREC:2005-009). Approval for examination of samples from the Northern Territory was provided by Acting Director Paul Griggs of the Forensic Services Branch, Northern Territory Police, Fire and Emergency Services (a copy of the letter is held on file at the Faculty of Science at the University of Technology, Sydney).

Samples were chosen from the Northern Territory Aboriginal (self declaration) database. These donors were selected based on their HUMD21S11 alleles as determined against an allelic ladder. The samples covered a range of sizes and repeat units, to determine if the HUMD21S11 Type V alleles could be observed in the Aboriginal population.

It has previously been shown that alleles (amplified fragments) which size as two base pair incomplete repeat units against a commercial allelic ladder, when sequenced exhibit either Type II or Type V characteristics (Walsh et al 2003). To increase the likelihood of observing Type V alleles, samples were selected based on whether they sized as incomplete repeat units against a commercial allelic ladder. Additional alleles were chosen to confirm they were part of the HUMD21S11 locus, and to determine if other previously reported sequence types were present (for example Type III and Type IV), as well as to determine if new mutations were present.

The alleles selected are not considered to be atypical or aberrant. Based on observed fragment size, alleles of similar size have been observed in non-Australian Aboriginal populations (Griffiths et al. 1998, Xiao et al. 1998). The presence of alleles in the Australian Aboriginal population and not in other populations is likely a result of genetic drift and the alleles are not considered to be atypical or aberrant.

Tribal or geographic origin of donors was not recorded in this study which enabled a wider investigation of the variation at the HUMD21S11 locus across the whole Northern Territory Aboriginal population.

2.2 Examination of HumD21S11 alleles

2.2.1 Extraction of DNA

Samples were provided on cotton swabs. The whole swab was placed into a 1.5ml sample tube and 200µL of 5% Chelex was then added to each swab and incubated at approximately 56^0C for approximately 30 minutes. The samples were then incubated at 100^0C for 8 minutes before being centrifuged at approximately 13,000rpm for 3 minutes.

Samples were then refrigerated at approximately 4^0C prior to estimating the amount of DNA present in the extracted sample (quantitation of DNA).

2.2.2 Quantitation of DNA

The DNA in the samples was quantitated on the Applied Biosystems 7500 Real-Time PCR system.

A PCR reaction mix containing 10.5µL of Quantifiler Human Primer Mix and 12µL of Quantifiler PCR Reaction Mix per sample was created. This mixture was vortexed to mix and then 23µL of the mix was added to the appropriate number of wells in a 96 well reaction plate. Then 2µL of extracted DNA sample, standards (containing known amounts of DNA) and a human control were then added to the reaction plate according to a pre-determined layout.

The reaction plate was the sealed with the Optical Adhesive cover. The plate was centrifuged at approximately 3,000rpm for approximately 20 seconds. The plate was then placed in the Real Time PCR system and the analysis started

Once the run was completed, the data was analyzed (against the threshold setting of 0.2 with the baseline start and end cycles of 6 and 15 respectively), and the amount of DNA present in the extracted samples was calculated based on comparison to the known quantities in the standards.

2.2.3 Amplification of HumD21S11 locus

Approximately 2-5ηg of DNA sample was amplified in a 50μL reaction containing

10X PCR Buffer
3mM $MgCl_2$ (stock 50mM)
2.5 U Taq Polymerase (stock 5U/uL)
20mM Forward Primer (stock 20μM)
20mM Reverse Primer (stock 20μM)
200μM dNTPS (stock 2mM)

Samples were then amplified in an Applied Biosystems 96 Well GeneAmp PCR System 9700 thermalcycler for an initial denaturation at 94^0C for 3 minutes, followed by 30 cycles of 94^0C for 40 seconds, 60^0C for 40 seconds and 72^0C for 40 seconds. At the end of the 30 cycles, the samples were given a final extension time of 8 minutes at 72^0C, with a final soak at 4^0C until samples could be removed from the thermalcycler.

Amplified DNA samples were then run on a 12% Bis-acrylamide (in 0.5X TBE Buffer) gel in a Protean II mini gel apparatus with 0.1mm spacers and 10 wells.

Gels were pre-run for at least 45 minutes at constant 250V. Then 10μL of sample (with 2μL of loading buffer) was added to each well and run for three hours at constant 250V against a molecular size standard.

2.2.4 Ethidium Bromide Staining of Gels

After electrophoresis, the gels were covered in distilled water (approximately 20mL) and 10μL of 10mg/mL Ethidium Bromide added. Gels were stained for at 30 minutes on a gently shaking platform, and then destained in fresh distilled water for at least 30 minutes. The gels was then removed the distilled water, and the bands viewed under UV light.

Based on their position in the gel, and by comparison to the molecular size standard, individual alleles (amplified fragments) were cut from the gel using a fine blade scalpel. The gel slices were placed in 1.5mL sample tubes and stored overnight at at least -22^0C.

The gel slices were then ground using a sterile wooden stick. Ground material was re-suspended in 100μL sterile TE overnight at approximately 4^0C.

The next day, the suspension was vortexed and then spun for approximately 1 minute in a centrifuge at approximately 13,000rpm. Then 2μL of the suspension was amplified as described above (2.2.3).

After amplification, the amplified product was run on a gel and alleles visualized as described above (2.2.4). A second electrophoresis was then performed with the amplified product run in duplicate and developed under the same conditions as above to confirm the presence of a single allele.

Where a single allele was observed per sample, the alleles were again cut from the acrylamide gel using a fine blade scalpel. Where two alleles were observed, the original amplified sample was run again and the individual alleles re-sampled.

The gel slices were placed in 1.5mL sample tubes and refrigerated overnight at approximately 4^0C. Gel slices were weighed and 3 gel volumes of buffer QG was added to the sample (1 volume equals 100μL per 100mg). This was incubated at

approximately 50^0C for 1-2 hours, vortexing at various intervals to assist in re-suspending the gel slices.

One gel volume of isopropanol was then added to the gel mix. The tube was vortexed and the solution added toa QIAquick spin column and collection tube. This spin column and collection tube were then centrifuged for approximately one minute at 13,000rpm. The waste in the collection tube was then disposed of.

A volume of 500µL of Buffer QG was then added to the spin column and the centrifugation step repeated. The waste was again removed. Then 750µL of prepared Buffer PE was added to the spin column and the centrifugation step repeated. The waste was again removed. The spin column was then placed in the centrifuge and centrifuged under the conditions previously described to remove all traces of the Buffer PE.

The QIAquick column was then placed into a 1.5ml microcentrifuge tube and 30µL of distilled sterile water was added directly above the membrane. This was then incubated at room temperature for 1 minute prior to centrifuging the column as previously described.

The cleaned sample was then run on 12% acrylamide gel to confirm presence of a single band and estimate the amount of DNA product present by comparison with the molecular weight size standard.

Once the single allele was confirmed and quantitated, it was prepared for sequencing by an external provider.

A sample-primer mix was made containing approximately 20ηg of amplified DNA (the allele) was added to 10pmol of HumD21S11 primer as described above (samples were sequenced in both directions to ensure the sequence obtained was reproducible). Distilled, sterile water was then added to the mix to produce a total volume of 16µL.

The samples were then transported to SUPAMAC where they were sequenced using dye-terminator technology. Laboratory equipment constraints meant that an external supplier was chosen to perform the sequencing process.

Electropherograms displaying the sequence of the alleles examined were then inspected to confirm that the alleles showed the expected HUMD21S11 sequence. The total number of bases sequenced was checked against the expected number of bases predicted based on the reported fragment size (as determined against a commercial ladder) and no discrepancies were observed. The number of repeat units present in each Variable Region were counted, and the sequence of the Constant Region was recorded.

3. RESULTS

3.1 HUMD21S11 sequenced alleles

This study confirmed that two base pair HUMD21S11 alleles can be created through a TA/TC insertion/deletion within Variable Region III, referred to as Type II (Brinkmann et al. 1996b) or, as a result of a fourteen base pair deletion within the Constant Region observed in alleles of Type V (Brinkmann et al. 1996b).

A total of twenty-six HUMD21S11 alleles were sequenced from the Northern Territory Aboriginal population (Table 1). The samples were chosen from the Northern Territory Aboriginal population based on their allelic designation when compared against an allelic ladder. None of these samples had previously been sequenced at the HUMD21S11 locus. In addition, four additional extraordinary alleles were examined to determine if they were the same sequence associated with HUMD21S11, and to determine if they were new sequence variants.

Sequence data for five of the alleles could not be obtained for both directions so the number of repeats in Variable Region I could not be confirmed. These alleles are indicated in Table 3 with an 'n' against the repeat unit of Variable Region I.

The two alleles that had been designated as '12' alleles (against an allelic ladder) shared the same sequence (Table 1) having five (TCTA) repeat units in Variable Region I, six (TCTG) repeat units in Variable Region II and a single (TCTA) repeat in Variable Region III. Within the Constant Region there is twenty-nine base pair deletion when compared to the more common Constant Region (observed in Types I, II, III and IV).

This deletion in the HUMD21S11 sequence structure has not been previously reported. Alleles sizing as '12' have not been described outside the Northern Territory Aboriginal population.

Table 1: Sequence Data Obtained in this study

Twenty-six HUMD21S11 alleles were selected from the Northern Territory Aboriginal population (Aboriginality was based on self declaration) to determine if the Type V sequence (Brinkmann et al. 1996b) was present within the population. Several unusual alleles were also sequenced as part of this study to confirm that they shared the HUMD21S11 allele sequence. Greyed out areas indicate where bases in the sequence are not present (when compared against the sequences described by Brinkmann et al. 1996b). The sequenced data shows that two base pair variants in the Northern Territory Aboriginal population contained Constant Regions associated with Type II and Type V alleles. The sequence of the two alleles sized as 12 repeat units show a 29 base pair deletion in the constant region. The one base pair variant alleles show a Type IV structure.

Allele	Type	VRI	VRII	Constant Region (CR)	VRIII
12	New	(TCTA)$_5$	(TCTG)$_6$	(TCTA)$_3$ TA	(TCTA)
12	New	(TCTA)$_5$	(TCTG)$_6$	(TCTA)$_3$ TA	(TCTA)
24.2	V	(TCTA)$_5$	(TCTG)$_6$	(TCTA)$_3$ TCA (TCTA)$_2$ TCCA TA	(TCTA)$_9$
29.2	II	(TCTA)$_5$	(TCTG)$_6$	(TCTA)$_3$ TA (TCTA)$_3$ TCA (TCTA)$_2$ TCCA TA	(TCTA)$_9$ TA (TCTA)
30.2	II	(TCTA)$_n$	(TCTG)$_6$	(TCTA)$_3$ TA (TCTA)$_3$ TCA (TCTA)$_2$ TCCA TA	(TCTA)$_{10}$ TA (TCTA)
31.2	II	(TCTA)$_5$	(TCTG)$_6$	(TCTA)$_3$ TA (TCTA)$_3$ TCA (TCTA)$_2$ TCCA TA	(TCTA)$_{11}$ TA (TCTA)
31.2	II	(TCTA)$_5$	(TCTG)$_6$	(TCTA)$_3$ TA (TCTA)$_3$ TCA (TCTA)$_2$ TCCA TA	(TCTA)$_{11}$ TA (TCTA)
31.2	II	(TCTA)$_5$	(TCTG)$_6$	(TCTA)$_3$ TA (TCTA)$_3$ TCA (TCTA)$_2$ TCCA TA	(TCTA)$_{11}$ TA (TCTA)
32.2	II	(TCTA)$_5$	(TCTG)$_6$	(TCTA)$_3$ TA (TCTA)$_3$ TCA (TCTA)$_2$ TCCA TA	(TCTA)$_{12}$ TA (TCTA)
32.2	II	(TCTA)$_5$	(TCTG)$_6$	(TCTA)$_3$ TA (TCTA)$_3$ TCA (TCTA)$_2$ TCCA TA	(TCTA)$_{12}$ TA (TCTA)
33.2	II	(TCTA)$_5$	(TCTG)$_6$	(TCTA)$_3$ TA (TCTA)$_3$ TCA (TCTA)$_2$ TCCA TA	(TCTA)$_{13}$ TA (TCTA)

34.1	IV	$(TCTA)_5$	$(TCTG)_6$	$(TCTA)_3$	TA $(TCTA)_3$	TCA $(TCTA)_2$ TCCA TA	$(TCTA)_6$ TCA $(TCTA)_7$ TA (TCTA)
34.1	IV	$(TCTA)_5$	$(TCTG)_6$	$(TCTA)_3$	TA $(TCTA)_3$	TCA $(TCTA)_2$ TCCA TA	$(TCTA)_6$ TCA $(TCTA)_7$ TA (TCTA)
34.2	II	$(TCTA)_5$	$(TCTG)_6$	$(TCTA)_3$	TA $(TCTA)_3$	TCA $(TCTA)_2$ TCCA TA	$(TCTA)_{14}$ TA (TCTA)
36.2	V	$(TCTA)_n$	$(TCTG)_{11}$	$(TCTA)_3$		TCA $(TCTA)_2$ TCCA TA	$(TCTA)_{11}$
37.2	V	$(TCTA)_9$	$(TCTG)_{12}$	$(TCTA)_3$		TCA $(TCTA)_2$ TCCA TA	$(TCTA)_{12}$
38.2	V	$(TCTA)_9$	$(TCTG)_{12}$	$(TCTA)_3$		TCA $(TCTA)_2$ TCCA TA	$(TCTA)_{13}$
38.2	V	$(TCTA)_{10}$	$(TCTG)_{12}$	$(TCTA)_3$		TCA $(TCTA)_2$ TCCA TA	$(TCTA)_{12}$
39.2	V	$(TCTA)_n$	$(TCTG)_{12}$	$(TCTA)_3$		TCA $(TCTA)_2$ TCCA TA	$(TCTA)_{13}$
40.2	V	$(TCTA)_9$	$(TCTG)_{14}$	$(TCTA)_3$		TCA $(TCTA)_2$ TCCA TA	$(TCTA)_{13}$
40.2	V	$(TCTA)_n$	$(TCTG)_{12}$	$(TCTA)_3$		TCA $(TCTA)_2$ TCCA TA	$(TCTA)_{17}$
40.2	V	$(TCTA)_9$	$(TCTG)_{14}$	$(TCTA)_3$		TCA $(TCTA)_2$ TCCA TA	$(TCTA)_{13}$
41.2	V	$(TCTA)_9$	$(TCTG)_{14}$	$(TCTA)_3$		TCA $(TCTA)_2$ TCCA TA	$(TCTA)_{14}$
41.2	V	$(TCTA)_n$	$(TCTG)_{14}$	$(TCTA)_3$		TCA $(TCTA)_2$ TCCA TA	$(TCTA)_{13}$
41.2	V	$(TCTA)_{10}$	$(TCTG)_{14}$	$(TCTA)_3$		TCA $(TCTA)_2$ TCCA TA	$(TCTA)_{13}$
41.2	V	$(TCTA)_{11}$	$(TCTG)_{15}$	$(TCTA)_3$		TCA $(TCTA)_2$ TCCA TA	$(TCTA)_{11}$

The two one base pair variant alleles (sized against a commercial allelic ladder as 34.1 alleles) shared the same sequence with that associated with the Type IV sequence structure described by Brinkmann's laboratory (1996b), and which has previously only been reported in the African Bantu population.

Five (TCTA) repeats were observed in Variable Region I, six (TCTG) repeats in Variable Region II and the Constant Region was as reported by Brinkmann et al. (1996b). The alleles in the Northern Territory Aboriginal population exhibited fewer repeats in Variable Region III than originally reported by Brinkmann et al. (1996b).

These alleles may have occurred as a result of a single base deletion or three base pair insertion within Variable Region III.

The presence of a one base pair variant may indicate a potential ancestral link between the African and Australian indigenous populations. The Indian population shows a high frequency of one base pair variants at the HUMD21S11 locus, and ancestral links between the Australian Aboriginal and Indian populations have been proposed before (Redd et al. 2002).

The remaining twenty-two alleles sequenced had all been previously designated as two base pair variants against an allelic ladder. Twelve of these had the Type V sequence (as described by Brinkmann et al. 1996b), while nine alleles had the more common Type II sequence which has a TA/TC insertion/deletion in VR III (Brinkmann et al. 1996b).

One of the alleles (24.2) had a fourteen base pair deletion in the Constant Region (normally associated with Type V alleles), while the number of repeat units in the three Variable Regions were of a size normally associated with Type I alleles.

3.2 Alleles with Type II sequence

Of the nine alleles with the Type II sequence: the three alleles sized as 31.2 repeats all had the same sequence reported by the Möller (1994) and Griffiths (1998) laboratories.

The allele sized as 30.2 repeats also appears to have the same sequence as that reported by the Möller (1994) and Griffiths (1998) laboratories, however the number of repeat units in Variable Region I could not be confirmed.

The sequence of the two alleles sized as 32.2 repeats were the same and along with the alleles which sized as 33.2 and 34.2 repeats have previously been reported (Möller et al. 1994, Griffiths et al. 1998).

The sequence of the allele sized as 29.2 repeats has not been described in the literature.

3.3 Alleles with the fourteen base pair deletion in the Constant Region

The 24.2 allele sequence observed in this study is the same as that seen by Griffiths et al. (1998). The presence of the Constant Region associated with Type V alleles in an allele which had been sized as 24.2 repeats, replicated previous observations (Griffiths et al. 1998 Xiao et al 1998) that the Constant Region associated with Type V is not restricted to large molecular weight alleles (with a high number of repeat units in Variable Regions I and II) as previously reported by Brinkmann et al. (1996b).

The combination of Type I Variable Regions I and II, (containing four to six repeat sequences), with a Type V Constant Region (fourteen base pair deletion) has so far only been observed in alleles sizing as 24.2 and 25.2 (against an allelic ladder) (Griffiths et al. 1998, Xiao et al. 1998). No intermediate alleles (containing the Type V sequence) have been observed between 25.2 and 30.2 repeats (this study, Griffiths et al 1998, Xiao et al. 1998).

This is the first time an allele with the Type V Constant Region, and Type I Variable Regions I and II, has been described the Australian Aboriginal population.

The remaining twelve alleles were all of the Type V sequence structure previously described by Brinkmann and co-workers (1998). The sequence of the allele sized as 37.2 repeats has been seen before (Walsh et al. 2003). Based on the number of repeat

units in Variable Region II and Variable Region III, the sequence of the allele sized as 39.2 repeats does not appear to have been observed previously (the number of repeat units in Variable Region I was not confirmed in this study), neither has the sequence of the allele sized as 36.2 repeats.

The sequence of the two alleles sized as 38.2 repeats were different to each other. One sequence has not been seen previously while the other has been reported by Walsh and co-workers (2003).

Of the sequences of the three alleles sized as 40.2 repeats, two samples had the same sequence for this allele, while the remaining sequence appears to be new based on the number of repeat units at Variable Region II and Variable Region III. However the number of repeat units in Variable Region I could not be confirmed in this study.

Of the four alleles sized as 41.2 repeats, three previously unreported sequences were recorded. The sequence from one of the samples was unconfirmed, though based on the number of repeat units in Variable Region II and Variable Region III it appears to have been observed in this study twice.

As with previous studies, allelic variation is not restricted to the base pair size of the alleles. Alleles sharing the same number of base pairs also exhibit variation at the sequence level, not only in the Constant Region, but also in the three Variable Regions (Brinkman et al. 1996b, Walsh et al. 2003).

3.4 Range of Repeat Numbers in Variable Regions

Initial studies of alleles associated with the Type V sequence associated it with high molecular weight alleles (those with a high number of base pairs). This study does however show that the Type V sequence may actually apply across a wider range of HUMD21S11 alleles.

Type V alleles were originally described as containing between nine and eleven repeat units in Variable Region I, and eleven repeat units in Variable Region II (Brinkmann et al. 1996b). This study replicated the sequence results observed by Griffiths et al. (1998) and Xiao et al (1998), where as little as five repeats in Variable Region I and six repeats in Variable Region II can be observed with a Constant Region associated with Type V sequence.

No more than eleven repeat units were observed in Variable Region I, although up to fifteen repeat units were observed in Variable Region II which is an increase on that first reported by Brinkmann and co-workers (1996b, Walsh et al. 2003).

A broad range in the number of repeat units in Variable Region III was observed with the number of repeat units ranging in number from nine (this study and Griffiths et al. 1998) up to seventeen (this study). Alleles with the Type II structure can have a range in repeat number between six and fourteen units (Griffiths et al. 1998).

As observed in previous studies, alleles of Type II sequence tended to show high variation in repeat unit number in Variable Regions I and III. As alleles increased in size, Variable Regions I and III showed the most variation in repeat number. The number of repeat units in Variable Region II tended to vary between five and six repeats, staying relatively stable as the allele size increased. Variable Region I [four to thirteen repeats (Griffiths et al. 1998)] and Variable Region III [six to twelve repeats (Griffiths et al. 1998)] both increased with allele size.

Mutations of a single repeat unit or of multiple repeat units can occur in the three Variable Regions occurs across all HUMD21S11 types described (Brinkmann et al 1996), and mutations in these regions appear to be much higher when compared to the apparent rarity of changes to the constant region.

Intermediate alleles were observed for both Type II and Type V alleles. This is where only one Variable Region increased in repeat unit between alleles which were sized as being one repeat unit different (for example allele 40.2 is one repeat unit less than allele

41.2). The increase or decrease in repeat unit could occur in any one of the three Variable Regions, although, most variation was observed in Variable Region III.

It may be that Variable Region II is more subject to mutation based on its position in the overall sequence. Alternatively, the sequence around the repeat region plays a role in how efficiently mispairings are recognized and subsequently corrected (Jeffreys and Neumann 1997, Armour et al. 2003).

Based on the sequence data in this study, mutations of a single repeat unit or multiple repeat units appear possible at the HUMD21S11 locus. Not all intermediate alleles are observed. This may be due to genetic drift, with some intermediate alleles being lost over time. Alternatively we may not have sampled enough alleles to be able to see all the variation possible at this locus.

4. DISCUSSION

4.1 Regional Differentiation in the Australian Aboriginal Population.

The HUMD21S11 Type V sequence originally described by Brinkmann et al. (1996) has now been observed in a second Australian Aboriginal population (from the Northern Territory) and confirms previous data on the existence of regional differentiation within the Australian Aboriginal population. While the data from this study is insufficient to allow for a detailed examination of regional differentiation between Australian Aboriginal tribal populations it does fit with other genetic information available in relation to the Australian Aboriginal population.

Regional differentiation can be utilized to infer how ancestral populations interacted within and between each other. Integrating this information with that of the archaeological record (including anthropology, climate and geographical changes) will assist in developing a more complete view of Australian Aboriginal ancestral life.

Studies examining mitochondrial, autosomal and Y-chromosome DNA all indicate that regional differentiation within the Australian Aboriginal population exists. Mitochondrial DNA data indicates that significant substructure exists between the Aboriginal peoples of the Riverine (south east Australia) and Desert (central and north west Australia) regions of Australia (van Holst-Pellekaan et al. 1998).

4.1.1 Creation of Regional Differentiation

The Riverine and Desert populations contain mitochondrial DNA sequences that distinguish them from each other, but also show some evidence of gene flow between them. The possibility of interaction between these geographically separate populations is supported by archaeological data which show that ancestral tribes travelled large distances for trading and cultural purposes. This contact and possible migration into new territory would have been possible in terms of climate, as central Australia was

quite lush and green during this time, and travel between north and south may have been facilitated through a river network (refer Flood 2004 for a larger discussion).

Ancestral gene flow between populations is also shown using mitochondrial DNA. Mitochondrial DNA from fossilized remains with a gracile skeletal structure in the Riverine region (recovered from Lake Mungo in south east Australia), show a close relationship with the Walbiri tribe found in the Desert region. This may be indicative of a recent expansion of the gracile aboriginal population from the central desert into south east Australia where older populations already existed (peoples with a robust morphology such as those found at Kow Swamp also from south east Australia) (Huoponen et al. 2001).

Genetic examination of DNA from remains at both Lake Mungo and Kow Swamp indicate that they lived in the same geographic area at the same time. So differentiation between tribes occurred after colonization of south east Australia.

A high degree of regional diversity within the Australian Aboriginal population is observed in the Walbiri tribe of the Northern Territory (Huoponen et al. 2001). The study found that the mitochondrial DNA variation observed between tribal areas was along north-south lines. This pattern of variation has also been reported for autosomal DNA microsatellites (Walsh et al. 2007a, Walsh et al. 2007c) as well as body measurements (Macho and Freedman 1987), skin colour and fingerprints (Parsons and White 1973).

A North-South cline in variation could be an indication of a period of separation between the peoples of the north and south of Australia. This time period has to have been of a sufficient length to allow genetic differences to occur.

The cause of separation may have been forced upon the population by changes in the environment, or through natural disasters such as flooding and fire which created a short term drop in population numbers reducing the amount of genetic material available for the next generation.

Climate change at the end of the ice age caused much of inland Australia to become arid, reducing the number of waterways available to facilitate travel over large distances and enable people to interact more easily, thereby allowing gene flow to occur. Waterways also provide an alternative food source for a population which increases their potential for survival and reproduction and as such the amount of genetic material that remains in the population.

Adaption to changes in the different environments across the continent may have given rise to genetic differences. Alternatively, the ancestral populations of the Australian Aborigines of the north and south have different genetic origins (Walsh et al 2007a). This possibility is discussed in section 4.2.

Cultural differences such as language, religion and social organization may also have been a factor in creating the diversity observed between Aboriginal populations of the north and south. The association between linguistic diversity and genetic diversity has been made before (White 1997), where language acts as a barrier to gene flow, however this association is not supported in all studies, so the impact of language on gene flow is uncertain (Kayser et al. 2003).

Given that the HUMD21S11Type V Constant Region was observed within a population from northern Australia, and not found a south Australian population, sequence variation at the HUMD21S11 locus appears to support variation between the north and south of Australia (Möller et al. 1994). However additional samples need to be examined where the geographic, or tribal, origin of the sample donor is known before we can say this with any certainty.

4.1.2 Population examination based on regional boundaries

There is support for examining the Australian Aboriginal population based on geographic or traditional tribal boundaries. Where a population (such as the Aboriginal population of Australia) is examined in datasets based on cultural and/or language

differences (Viard et al. 1998, Walsh et al. 2007a), rather than judicial or political boundaries, an increase in genetic diversity has been found (Walsh et al. 2007a). In many instances, jurisdictional boundaries cover multiple traditional tribal regions and the resulting population dataset is a conglomeration of all of them. This can lead to statistical discrepancies that are difficult to explain or resolve.

Combining autosomal DNA datasets of different (and even disparate populations diminishes the strength of the evidence being presented regarding the genetic composition of the population under discussion (whether for court, or when trying to correlate data for academic purposes).

It must be noted however that the subpopulation model used by forensic scientists in Australia can cope with small discrepancies in databases (such as that created by examining the Australian Aboriginal population based on jurisdictional boundaries) with no disadvantage to the defendant (Walsh et al. 2007b).

However, examining the Aboriginal population on the basis of regional tribes does not correct for all statistical discrepancies (Walsh et al. 2007a). It does support the existence of further substructure existing at the tribal level, as well as the population being more heterogeneous than expected based on fragment analysis alone. This study supports the existence of further substructure at the sequence level of the HUMD21S11 locus, which is hidden at the fragment level (size homoplasmy).

The effect of genetic drift on the Aboriginal population may also be higher than originally thought. This could be a result of the size of the Australian Aboriginal tribes. Tribes are reported to have formed genetic isolates in which gene flow between them was low (Birdsell 1993). If a tribe is small in size then this would create a small effective population size, increasing the effects of genetic drift. This would then lead to statistical discrepancies when the tribes are combined as part of a larger, regional population.

It is unclear if the variation observed in the Australian Aboriginal population is due to a small population size caused by periodic population bottlenecks. The high diversity observed in the mitochondrial and autosomal genome (such as at the HUMD21S11 locus) do not indicate that the Australian indigenous population as a whole has undergone any significant population bottlenecks (even when examined based on regional boundaries).

4.1.3 Effectiveness of examining populations based on regional boundaries

Examining populations based on cultural and regional borders rather than political ones has assisted in resolving questions about Y-chromosome and mitochondrial DNA diversity in the West New Guinean population. By analyzing the data based on regional boundaries rather than political ones, researchers were able to explain previously contradictory data.

Based on regional boundaries, low levels of Y-chromosome diversity have been found in conjunction with high levels of mitochondrial diversity within coastal tribes of the New Guinean population. This is believed to be a result of the social structure present.

The coastal tribes of New Guinea have a patrilineal social structure including high levels of polygyny. In a patrilineal society, all children belong to the father's clan and this produces low Y-chromosome diversity. The males of the family tend to stay in the same village or geographical area. The high mitochondrial DNA diversity is observed as the gene flow is primarily female mediated, with females being introduced into the tribe (Kayser et al.2003).

The opposite is observed in the highland New Guinean population, where gene flow is primarily male mediated (they have a predominantly matrilineal society where males move between tribes). Male mediated gene flow into the highland population is supported by the low frequency of a specific deletion at the DYS390 locus in the highland population which is thought to be due to admixture with the coastal populations where the frequency of this deletion is much higher (Forster et al 1998).

Stemming from this, it has been proposed that for those highland populations that adopted horticulture, the population expanded and diversified (Bayliss-Smith 1996, Forster et al. 1998). Many of the highland populations remained isolated with little contact with the lowland/coastal populations (Hagelberg et al. 1999).

Mitochondrial DNA diversity is high in the Australian Aboriginal population and is at similar levels to that found across Asia (Ingman and Gyllensten 2003), but conversely Y-chromosome diversity is low. If we use the same assumptions as that described for the New Guinean population (Kayser et al. 2003), then it would appear that the Australian indigenous population was primarily patrilocal, with female members moving between the tribal areas.

The Australian Aboriginal population is known to have lived a predominantly hunter-gatherer lifestyle, but there are archaeological records indicating that some tribes also had an agricultural lifestyle. Examination of DNA results by regional boundaries would shed further information on this (van Holst Pellekaan et al. 1998, Huoponen et al. 2001, Flood 2004).

Low Y-chromosome diversity has also been reported in the Australian Aboriginal population (Ingman and Gyllensten 2003). The Australian Aboriginal tribes of the desert region and northern coast show similar levels of polygyny to that observed in coastal New Guinea populations (Schiefenhövel 1998, Birdsell 1993), however further data needs to be collected before predictions about traditional lifestyle and cultural interactions (such as that made for the New Guinean population) can be made about the Australian Aborigines.

4.2 Global relationships

4.2.1 Colonization of Australia and Sahul

The Australian Aboriginal population is one of the most physically and genetically diverse in the world. It is unknown if this occurred as a result of multiple migration events, or as a result of one colonization event.

Based on the anthropological record, Birdsell (1983) proposed three waves of migration into Australia as the means of producing this variation in the population. The first wave of migrants were of gracile build from the Philippines of South East Asia (the Negritos), followed by a migratory wave of people who were of a robust build. Both these waves of people are believed to have populated the whole of the continent, while the third wave – the Carpentarians - populated the north of the continent.

Colonization of the continent could have occurred by following the coast, or alternatively traversing inland utilizing the inland rivers and lakes that were prevalent prior to the sea levels rising at the end of the ice age.

It has also been suggested, based on anthropometric data, that there was an additional wave of migrants from the Indian sub-continent. There is some genetic evidence available to support this hypothesis (Redd et al. 2002, Alfonso-Sanchez et al. 2008). However, the archaeological evidence is not unequivocal, and there is no clear indication of how much contact or migration there was between Australia and other countries after Sahul was divided into the larger islands of Australia and New Guinea.

Migration is not the only means of maintaining and introducing diversity into a population. The heterogeneous nature of the Australian Aboriginal population could be a result of long term isolation (van Holst Pellekaan et al. 2006). Maintenance of genetic diversity through isolation is observed in the Basque population of Spain.

The Basque population of Spain is one of the most genetically diverse in the world (Pérez-Miranda et al. 2005). It has been suggested that this diversity was maintained in the population as a result of isolation during the last mini ice age (between 52,000 and 45,000 years ago) when the Basque area of Spain is believed to have acted as a refuge.

Ice age refuges were isolated from other populations and allowed them to preserve much of their ancestral DNA line and if the population was large enough, maintain their genetic diversity. This is thought to be because the area was less likely to have suffered from a large population crash like those populations further north.

In those areas where there was a significant population crash, there would have been a corresponding drop in genetic variation available for transmission to the next generation, subsequently reducing the diversity of descendant populations as re-expansion occurred. At the end of the ice age, the worldwide population was still in the low millions.

Genetic studies thus far have supported common colonization of Sahul by a heterogeneous population rather than as a result of multiple migrations (Kirk and Thorne 1976, Nei and Roychoudhury 1993, Cavalli-Sforza et al. 1994, Redd and Stoneking 1999, Adcock et al. 2001).

The ability of the Australian Aboriginal people to exploit their environment may have allowed them to maintain sufficient population numbers to preserve genetic diversity in their population. The high percentage of unique genetic diversity present in the Australian indigenous population strongly supports their ancient origin, and long period of isolation. Selection and drift then played a role in the creation of the differentiation we see in today's indigenous population (Huoponen et al. 2001, van Holst Pellekaan et al. 2006).

4.2.2 Relationship between Australian Aborigines and New Guinea

Observing Type V alleles in the New Guinean and Australian Aboriginal population confirms the results of other studies which support an ancestral relationship between these two indigenous populations (Ingman and Gyllensten 2003, Roberts-Thomson et al.1996, van Holst Pellekaan et al. 1998, Walsh et al. 2003).

Geographically, joint colonization by the same ancestral population appears feasible as both islands once formed part of the same land mass, Sahul. Currently there is insufficient HUMD21S11 sequence data available to determine how the HUMD21S11 locus can be used to assist in resolving questions about the colonization of Sahul and then Australia. The HUMD21S11 sequence does however add support to an ancestral relationship between Australia and New Guinea.

Mitochondrial and Y-chromosome DNA data suggest a shared history between the New Guinean and Australian Aboriginal populations since the initial colonization of Sahul (which is compatible with the geographical history of the continent). There is also a closer genetic relationship between these two populations than either of them is to Asian populations, including to populations of South East Asia (Malaysia, Borneo and Philippines) where the Negritos come from (Roberts-Thomson et al. 1996, Ingman and Gyllensten 2003, Huoponen et al. 2001).

Ingman and Gyllensten (2003) recommended examination of additional loci, including autosomal loci to obtain a more balanced view of evolutionary history. Additional sequence data needs to be examined from both populations and compared based on tribal boundaries in order to better understand how the ancestral Sahul tribes interacted. This examination would then provide additional information on the ancestral interactions of tribes in Australia after the land mass was divided.

4.2.3 Relationship between Australian Aborigines and Africa

While there is insufficient data from this study to investigate deeper into the genetic relationships between Australian Aboriginal and New Guinean populations, we can use sequence data from the HUMD21S11locus to determine whether it would prove useful in understanding genetic relationships prior to colonization of Sahul, in particular those relationships at the time of migration from Africa.

Under the Out-of-Africa theory, all world populations share a subset of the genetic variation found in Africa. The similarity in sequence observed between Variable Region I of HUMD21S11 alleles of Type III (found in the African Bantu population) and Type V would support this premise (Brinkmann et al. 1996b, Xiao et al. 1998).

The doubling of the repeat number in Variable Region I may have occurred prior to the split between the population that migrated south into Africa (of which the Bantu population was a part) and those that migrated out of Africa and successfully into the Middle East and Asia. The Type III and Type V alleles then evolved independently after separation of African and non-African populations from the common precursor population (Brinkmann et al. 1996b).

Separate evolution between HUMD21S11 alleles of Type III and V, is supported by phylogenetic tree analysis (Bowcock et al. 1994, Brinkmann et al. 1996a). The relatively high genetic distances in combination with the influence of genetic drift in small populations (Dover 1982) also supports the assumption of independent development of the extraordinary alleles.

However, complex mutational events such as those observed at the HUMD21S11 locus are often best explained by a multistep mutational process rather than single step mutational processes alone (Jeffreys et al. 1994). This can explain the differences observed between the Type I, III and V alleles.

4.2.4 Relationship between Australian Aborigines and Caucasians

Originally the fourteen base pair deletion in the Constant Region was only observed in the Papua New Guinea population (giving rise to the term "Papuan Deletion"), in association with alleles which had a high number of repeats in Variable Regions I and II. The number of repeat units present in these regions is double that observed in alleles described as being of Type I.

Since then, the fourteen base pair deletion, or Papuan Deletion has been observed in an allele from a known Caucasian (Belgian) donor (Xiao et al. 1998), and in association with alleles where the number of repeats in Variable Regions I and II is more commonly associated with alleles of Type I (this study and Griffiths et al.1998).

Unfortunately, the donors of the samples examined in the Griffiths study (1998) are unknown so we cannot say if the presence of the fourteen base pair deletion in the Belgian Caucasian population is a result of recent mutation occurring after the separation of the Caucasian population from Sahul populations (Australian Aboriginal and Papuan populations) as proposed by Xiao (et al. 1998).

However the same deletion is observed in an Australian Aboriginal population, in an allele where the number of repeat units in Variable Regions I and II is normally associated with Type I alleles.

This may mean that the fourteen base pair deletion occurred in the population that left Africa, and before the increase in the repeat unit number in Variable Region II observed in Type V alleles. The deletion may not be observed in similarly sized alleles in other populations due to sampling, or it may have been lost from other populations over time.

There is little evidence for extensive admixture between the Caucasian and Australian Aboriginal populations; however it is possible it is occurring or has occurred relatively recently (Walsh et al. 2007c). Admixture between Aboriginal and Caucasian

populations may explain the presence of a Type V allele in the West Australian Caucasian population (Walsh et al. 2003).

The deletion in the Constant Region may have occurred in both the Caucasian and Australian Aboriginal populations as a result of independent, recent separate mutational events. This would mean that the mutation rate at the HUMD21S11 locus is much higher than expected.

The mutation rate at the HUMD21S11 is currently estimated as 0.19% (www.cstl.nist.gov/strbase/mutation). However, if the mutation rate at this locus is higher, it may affect calculations of evidential weight, as well as for dating genetic events. The high rate of mutation associated with Y-chromosome microsatellites [0.2-0.32% (Heyer et al. 1997, Kayser et al. 1997)] prevents events which occurred more than 5,400 – 8,000 years ago from being traced (de Knijff et al. 1997, Hagelberg et al. 1999, Vishwanathan et al. 2004).

The absence of the fourteen base pair deletion with a high number of repeat units in Variable Region II (as associated with Type I alleles) in the African population, suggests that this deletion occurred after the split between African and non-African populations. A similar situation is observed in the deletion polymorphism of the Human COL1A2 gene which, while found in high frequency in non-African populations, is completely absent from sub-Saharan populations (Mitchell et al. 1999).

At this time, the high number of repeat units in Variable Region II has not been observed outside the Australian Aboriginal and Papuan populations (Möller et al.1994, Brinkmann et al. 1996b, Xiao et al. 1998). This would indicate that the increase in repeat unit number in this Variable Region occurred after the population of the Sahul land mass.

4.3 Mutation events at the HUMD21S11 locus

The large mutation events observed at the HUMD21S11 locus in the Variable Regions I and II, and in the Constant Region could have occurred as a result of a series of single step mutations or as large mutation events involving a number of repeat units.

The HUMD21S11 alleles studied in this and the work of other, support previous findings that microsatellites are able to retain their variability after structural change (Möller et al. 1994, Brinkmann et al. 1996b, Griffiths et al. 1998, Walsh et al. 2003). Further single or multi-step mutation events (occurring to each of the three Variable Regions around the Constant Region) generated a number of successive alleles which all showed the deletion created in the constant region.

For the creation of complex alleles such as those observed at the HUMD21S11 locus, a multifunctional process including unequal crossover and slippage mechanisms must be assumed (Dover 1995). In the case of the HUMD21S11 locus, the sequence data suggests that both unequal crossover and slipped strand mispairing mutations can occur. Having multiple possible mutation mechanisms, acting on the one locus would assist in explaining why the HUMD21S11 locus does not fit within one single mutation model (Brinkmann et al. 1996b).

4.3.1 Single-step or multi-step mutation

Unequal crossover would appear to be the most likely explanation for the large deletion and insertions observed in the constant region (Brinkmann et al. 1996b, Grubweiser et al. 2005) while slipped strand mispairing seems the less likely mutation mechanism given the size of the mutation, multiple repeat unit mutations can occur for both unequal crossover and slipped strand mispairing.

The doubling of repeat units in Variable Region I of HUMD21S11 alleles of Type III and V could be the result of a large mutation event such as a large insertion of bases. Large mutation events have been recorded in the Constant Region (the fourteen base

pair deletion originally associated with Type V alleles) as well as the seventeen base pair insertion in the Constant Region (of a Type I allele) recorded by Grubweiser et al (2005).

Alternatively, this doubling in Variable Region I, could have occurred in a stepwise fashion over a number of intermediate stages which subsequently became lost over time. The apparent absence of intermediate alleles lessens the chance that the Variable Region I of Type III and V alleles occurred as a result of lots of small events with all intermediates being lost over time (Ghiani et al.2006).

The same can be assumed for the doubling of repeat unit number observed in the Variable Region II of Type V alleles. A single large mutation event is the more likely explanation due to the absence of intermediate alleles (Brinkmann et al. 1996b, Griffiths et al. 1998 and this study).

The variation found at the HUMD21S11 locus may indicate that the intermediates haven't been sampled yet. Although, based on the sequence data available at this time, single, large mutation events appear the most probable explanation for differences between the different HUMD21S11 allele types.

4.3.2 Mutation mechanisms

Mispairing of DNA is more likely to occur in the middle of a repeat section as the corrective mechanism is less likely to pick it up due to the surrounding repeat units indicating that replication is occurring as normal. Whereas if the mispairing occurs at a point in the sequence where replication starts on a region that is not repeated, then the corrective mechanism of DNA replication is more likely to detect the error and make the necessary correction (Eisen et al.2003).

Large mutation events, which are restricted to one site of the repeat sequence, have previously been observed in minisatellite polymorphisms (Jeffreys et al. 1994, Monckton et al. 1994) and it has been suggested that similar events can occur in

microsatellites as well (Urquhart et al. 1994, Meyer et al. 1995). This postulate would assist in explaining the large mutation events occurring in the Constant Region.

The Variable Regions I and II may also be more susceptible to large mutation events, as mutations of similar size have not been observed in Variable Region III (though intermediate alleles can be observed in Variable Region III) (Möller et al. 1994, Brinkmann et al. 1996b, Griffiths et al. 1998, Walsh et al. 2003 and this study).

An independent mutation event is also probably true for the Type IV alleles found (those that sized as HUMD21S11 34.1) in the Australian Aboriginal population. On face value the presence of two alleles of Type IV would appear to strengthen the link between the African Bantus and ancestral Sahul population.

However, fragment analysis shows alleles of this type (those sizing as 1 base pair incomplete repeats), are rare within the Australian Aboriginal population. By contrast the Sub-Saharan and south-east Indian populations show a high proportion of alleles with a 1 base pair variant (Steinlechner et al. 2002, Reddy et al. 2005, Walsh et al. 2007a and 2007b).

Sequencing one base pair variants from Sub-Saharan and Indian populations as well as from further Australian Aboriginal populations will assist in further defining the genetic relationship between these three populations.

A large deletion is the most plausible explanation for the HUMD21S11 alleles sized as 12 repeats. Again, no intermediate alleles have been identified, lessening the chance that these alleles have occurred as a result of a number of small, multiple mutation events. If small stepwise mutations have occurred, then it can be assumed that all intermediates between the HUMD21S11 12 allele and the next smallest allele (designated as 24.2 repeats) have been lost over time.

The HUMD21S11 12 allele may also support the proposal that once a critical expansion has been reached or exceeded, subsequent mutations can occur much faster.

Alternatively the 12 allele may be a "broken allele" where it has become so big that it dies (Chambers et al. 2000).

Additional sequence data is required to explain why insertions/deletions of incomplete repeats (that is one, two or three base pairs) in the Type II and IV alleles are only ever followed by a single, complete (that is four base pairs) repeat unit. Understanding the position of a mutation in a sequence and whether the DNA repair mechanism is sequence dependent would assist in further clarifying the mutation process (Estoup et al. 1995, Armour et al.2003).

If the mutation rate is greater for complex microsatellites than predicted, and possibly dependent on the mutation mechanism in operation, this will affect their use in tracing historical events and when making calculations of population relationships and migration history.

Additional alleles need to be sequenced at the HUMD21S11 locus in order to learn more about the mutation rate. Further sequencing will also assist in reviewing the current nomenclature and descriptions originally applied by Brinkmann's group (1996b). A review of the current microsatellite nomenclature has been raised before, and this study reinforces the need for this (Buckleton et al. 2005).

Examining and sequencing more HUMD21S11 alleles and further characterizing other complex loci will provide additional information relating to mutation rates and mechanisms at microsatellite loci. It would also assist in determining if the absence of intermediate alleles is an indication of mutation rate or genetic drift.

5. CONCLUSION

The sequence variation within HUMD21S11microsatellite alleles has confirmed the existence of regional differentiation within the Australian Aboriginal population through the presence of Type V HUMD21S11 alleles being observed in two Australian Aboriginal populations, and not in a third. Variation observed at the sequence level (in addition to that observed at the fragment level) may also assist in explaining departures from statistical equilibrium observed within the Australian Aboriginal population.

Consideration should also be give to examining populations based on traditional, tribal boundaries, rather than based on judicial boundaries. This allows for patterns of regional differentiation to be more readily resolved. Genetic and archaeological data can then be combined to provide a glimpse into the lifestyle of our ancestors, including political and cultural practices in use. A finer sampling approach, including the recording of tribal affiliation and geographic location of donors, is needed in order to resolve the true level of diversity in the Australian Aboriginal population (White 1997, Walsh et al 2007b).

Larger sample sizes will improve calculations of genetic distance between Australian Aboriginal tribes and those of other international populations. This study supports previous work indicating a shared genetic history between Australian Aborigines and the populations of Papua New Guinea. Further samples will define the genetic relationship between these two peoples, and assist in resolving issues regarding their interaction with other peoples after the landmass Sahul was divided by rising sea levels.

The HUMD21S11 microsatellite will also prove useful in further elucidating the genetic relationship between the Australian Aboriginal population and those of Africa and Europe. The fourteen base pair deletion in the Constant Region in alleles of Type I sequence may indicate that either this mutation occurred prior to the split between Sahul and Caucasian populations. The presence or absence of the deletion may assist in tracing the origins of populations after the split from Africa.

Further investigation into the similarities in the number of repeat units in Variable Region I between the Australian Aboriginal population and the African Bantu population will also assist in learning more about the heterogeneous nature of the population that lived in Africa prior to people migrating out and into Asia.

These additional sequences will also aid understanding the mutation rate and mutation (and genetic) processes in operation. Complex microsatellites such as HUMD21S11 are best placed to do this as they have a number of different sequence arrangements present. This will aid in the further development of statistical models.

By continually characterizing and seeking to understand microsatellites, forensic scientists continue to provide increasingly accurate calculations of evidential weight for the judicial system. This can only be of benefit to the defendant, the judicial process, and to society.

APPENDIX A: Glossary

Term	Definition
Adenine	A purine base; one of the four molecules containing nitrogen present in the nucleic acids DNA and RNA; designated by letter A
Allele	One of a series of alternative forms of a gene (or VNTR) at a specific locus in a genome.
Allelic Ladder	A mixture of the most common STR alleles that are used to compare with the amplified samples and infer the number of repeats present in the amplified sample.
Amelogenin	A PCR amplified locus that varies between males (212 and 218 bp) and females (212 bp only)
Amplification	An increase in the number of copies of a specific DNA fragment; see PCR
Autosome	A chromosome not involved in sex determination. The diploid human genome consists of 46 chromosomes, 22 pairs of autosomes, and one pair of sex chromosomes (the X and Y chromosomes).
Base pairs	The building blocks of the DNA molecule. In DNA the bases adenosine (A) always pairs with Thymine (T) and Cytosine (C) always pairs with Guanine (G). Because DNA is a double stranded helix, the size of DNA molecules is often expressed in base pairs.
Bottleneck	Dramatic reduction in population size.
Buccal Sample	See DNA Person Sample
Chromosome	Structures, composed of DNA and protein, found in the nucleus of the cell. Most human cells contain 46 chromosomes; 23 derived from the mother and 23 derived from the father. Gametes (sperm or egg cells) contain 23 chromosomes until conception.
Complementary sequences	DNA sequences that form a double-stranded structure by matching base pairs; the complementary sequence to G-T-A-C is

	C-A-T-G
Crime Scene Profile	A DNA profile recovered from the scene of a crime
Cytosine	A pyrimidine base; one of the four molecules containing nitrogen present in the nucleic acids DNA and RNA; designated by letter C.
Deoxyribonucleic acid (DNA)	The genetic material that contains all the information that determines our inheritable characteristics. DNA is the double helix molecule of hereditary. It is composed of 4 bases or building adenine (A), thymine (T), cytosine (C) and guanine (G). These bases are strung along two opposing strands, like pearls on a string. The A is always bound to T and G is always bound to C by week hydrogen bonds. The specific order of bases is the DNA sequence.
Denaturation	Use of heat or alkali chemical to disrupt hydrogen bonds that hold the two strands of the DNA double helix together.
Diploid	Two complete genomes in every cell, one derived from mom and the other from dad.
DNA profiling	The identification of variable characteristics at one or more loci in an individual's DNA, and the comparison of those characteristics with other DNA samples to determine whether they could have a common origin
DNA sequence	Order of bases in a DNA molecule.
Effective population size	Size of an idealized population that would behave similarly to the real population size
Electropherogram	The pattern of fluorescent peaks produced by alleles during electrophoresis. Collectively, this pattern is the DNA profile.
Electrophoresis	The process of separating charged molecules, such as fragments of DNA, in an electric field. The electric current is passed through an agarose gel, which allows the DNA fragments to separate, by size.
Enzyme	A protein that can speed up a specific chemical reaction without being changed or consumed in the process

F_{ST}	Measures genetic distance between populations using differences in gene frequency caused by genetic drift.
Gene	Unit of heredity, region of DNA that encodes or regulates the expression of a protein.
Gene Flow	Movement of genes and their subsequent successful reproduction into a new population as a result of migration of people between areas/populations
Genetic Distance	Estimate of evolutionary relatedness of two populations
Genetic Drift	Random variation in allele frequencies due to the chance contribution of each individual in the population to the next generation
Genome	All of the genetic material or DNA in a cell.
Genotype	The genetic makeup of an organism
Gracile	Lightly built (in relation to skeletal structure)
Guanine	A purine base; one of the four molecules containing nitrogen present in the nucleic acids DNA and RNA; designated by letter G
Haplotype	Specific combination of allelic states involving a number of polymorphic DNA loci
Heterozygous	An individual that containing two different allele (genetic traits), one from mom and the other from dad, for a given trait. In RFLP this means that an individual produces two different bands on a gel.
Holocene	Time period of the last 11,000 years, characterized by an unusually warm and stable climate
Homoplasmy	Generation of an allelic state by independent events. In relation to microsatellites, two alleles are the same size due to mutation processes rather than ancestry.
Homozygous	An individual that contains two of the same alleles (genetic traits), one from mom and the other from dad, for a given trait. In RFLP this means that an individual produces a single band on a gel.

Identity by Descent (IBD)	Alleles in one or more individuals are identical as a result of being inherited from a common ancestor
Identity by State (IBS)	Alleles in one or more individuals are identical due to coincidental mutational processes rather than inheritance from a common ancestor
Infinite Allele Model	Mutation model where all mutations create novel alleles not previously present in the population
K Allele Model	Mutation model where K allelic states are allowed and mutations create alleles to any of the K allelic states.
Kilobase (kb)	Unit of length for DNA fragments equal to 1000 nucleotides
Ladder	A mixture of the most common STR alleles that are used to compare with the amplified samples.
Linkage	A measure of association between two loci. Loci on different chromosomes are non-linked. Those that are close together on the same chromosome are closely linked and are likely to be inherited together.
Locus	(pl., loci) The site on a chromosome where a gene or other feature of the DNA is located.
Meiosis	The process involving two cell divisions, where diploid cells produce haploid gametes.
Microsatellite	DNA sequence containing a number (≤50) of tandemly repeated short (2-6 bases) sequences. Often polymorphic.
Migration	Movement of people or population from one area to another
Migration Drift Equilibrium	Stable level of population subdivision reached when migration decreases the amount of genetic variation between two populations
Minisatellite	DNA sequence containing a number (~10 to >1000) of tandemly repeated units (10-100bp) of sequences.
Mismatch repair system	Proteins involved in the recognition and removal of errors that occur during synthesis of DNA
Multiplexing	The process of amplifying using the PCR reaction a number of different loci in the one tube;

Mutation	Any inheritable change in DNA sequence
Mutation Model	Description of probabilities of change among assumed, finite or infinite range of possible allelic states
Mutation rate	Mutations per DNA replication or organism generation
Nanogram	ng - One billionth of a gram
Nucleotide	A building block of DNA or RNA
Off-Ladder Alleles	An allele that is not included in the allelic ladder. These can be true, rare alleles or artifacts.
Pleistocene	Time period between the last 2 million years ago and 10 thousand years ago
Polymerase chain reaction (*PCR)*	Polymerase Chain Reaction. A method of amplifying (repeatedly copying) a defined region or target of DNA so that it is present in larger, more detectable amounts. The reaction involves repeated cycles of denaturation (heat), annealing and extension. After 30 cycles, a single target DNA can be amplified over a billion fold.
Polymerase (DNA)	Enzymes that catalyze the synthesis of DNA on pre-existing nucleic acid templates.
Polymorphic	A highly variable genetic trait in a population. An STR locus that is polymorphic will have many possible size variants between individuals.
Polymorphism	Two or more allelic variants present in the population at significant frequencies
Population	A group of individuals residing in a given area at a given time. May also refer to a group of individuals defined by a shared characteristic; in a theoretical sense, it is a group of individuals in which there is random mating
Population differentiation	Process where allele frequencies between populations diverge over time

Population genetics	Study of genetic diversity in populations, and over time
Primer	A short length of synthetic DNA used to initiate PCR. Also known as a probe.
Probe	A fragment of DNA that will hybridize to another complementary sequence of DNA
Profile	A description of an individual or evidence resulting from the examination of one or more polymorphic loci
Recombination	The process of exchange of DNA information between members of a chromosomal pair. Usually occurs during meiosis.
Repeat Number	Or Repeat Count; actual number of repeats at a locus in an allele
RFLP	Restriction Fragment Length Polymorphism; a method in which the target DNA is cut, using a restriction enzyme, at specific sequences resulting in a range of fragment sizes. The DNA fragments are separated by agarose electrophoresis and transferred (blotted) to a nylon membrane. The nylon bound DNA is hybridized with a labeled (radioactive or enzyme) DNA probe (binds to the bound DNA). A DNA profile is produced by exposing the membrane to X-ray film.
Robust	Heavily built (in relation to skeletal structure)
Selection	Tendency of bearers of particular genotypes to reproduce more or less than others in the population.
Short Tandem Repeat (STR)	Short Tandem Repeat; short, repetitive sequence elements of 2-5 bases. STR's used in DNA profiling are polymorphic, which is to say that the number of repeat elements varies between individuals in a population. STR alleles can be sized by polyacrylamide gel electrophoresis (PAGE) or capillary electrophoresis (CE).
Slipped Strand Mispairing	Error during DNA replication where the DNA polymerase
Stepwise Mutation Model	Model of microsatellite evolution where the length of the microsatellite increases or decreases by one repeat unit
Tandem repeat	The end-to-end duplication of short, identical DNA sequences in

	the genome.
Taq polymerase	An enzyme isolated from bacteria that live at high temperature, that catalyses the polymerase chain reaction.
Template	Genomic DNA, from crime scene samples or reference samples that are the targets for PCR amplification.
Thymine	A pyrimidine base; one of the four molecules containing nitrogen present in the nucleic acids DNA and RNA; designated by letter T
Unequal Crossing Over	Unequal exchange of DNA between chromosomes at meiosis. Results in one chromosomal segment being longer or shorter than the original starting DNA
Variable number tandem repeats (VNTR)	Longer, repetitive sequence elements which can range in size from 9-80 bases. VNTR's used in DNA profiling are polymorphic, which is to say that the number of repeat elements varies between individuals in a population. VNTR alleles can be sized by polyacrylamide gel electrophoresis (PAGE) or capillary electrophoresis (CE).
Variant Allele	An allele with an incomplete repeat unit. A common THO1 variant allele is 9.3. This allele has 9 full repeats plus 3 nucleotides.

APPENDIX B: Reagents and Consumables

30% Acrylamide:-

29g Acrylamide (Bio-Rad Cat# 161-0107)

1g N,N'-methylenebisacrylamide (Amresco Cat# 0172-50G)

dH_20 to 100mL

Ammonium persulfate (APS):-

1g ammonium persulfate

dH_20 to 10mL

10% Chelex

1g Chelex (Bio-Rad Cat# 142-2832)

dH_20 to 10mL

dNTPs (dinucleotidephosphates) Fisher Scientific Cat# GZ28406562

Ethidium Bromide – Bio-Rad Cat# 1610433

Loading Buffer

0.25% Bromophenol blue (Bio-Rad Cat# 161-0404)

0.25% xylene cyanol FF (Amresco Cat# 0819-20G)

30% glycerol in water (Amresco Cat# E520-100ML)

EZ Load Precision Molecular Mass Ladder – Bio-Rad Cat# 170-8356

Quantifiler® Human DNA Quantification Kit – Applied Biosystems Cat # 4343895

QIAquick PCR Purification Kit – QIAGEN Cat# 28704

Platinum Taq DNA Polymerase – Invitrogen Cat#10966-018

Primers (Sharma an Litt 1992) - Ordered through Invitrogen (www.invitrogen.com):

Forward primer: 5'-GTGAGTCAATTCCCCAAG-3'

Reverse Primer: 5'-GTTGTATTAGTCAATGTTCTC-3'

SUPAMAC:-

The University of Sydney

Medical Foundation Building - K25

92-94 Parramatta Road

CAMPERDOWN NSW, 2050

10 X TBE Buffer (Tris-Borate-EDTA Buffer):-

108g Tris base

55g Boric Acid

40ml 0.5M EDTA (pH 8.0)

dH_20 to 1000 mL

TEMED – AMRESCO CAT# 0761-25mL

BIBLIOGRAPHY

Adcock. G.J, Dennis. E.S., Easteal. S., Huttley. G., Jermiin. L.S., Peacock. W.J., Thorne A. Mitochondrial DNA sequences in ancient Australians: implications for modern human origins. (2001) PNAS 98: 537-542.

Alfonso-Sanchez. M.A., Perez-Miranda. A.M., Herrera. R.J. Autosomal microsatellite variability of the Arrente people of Australia. (2008) Am. J. Hum. Biol. 20: 91-99.

Ambrose S.H.. Late Pleistocene human population bottlenecks, volcanic winter, and differentiation of modern humans. (1998) J. Hum. Evol. 34:623-651

Amos. W. A.. Comparative approach to the study of microsatellite evolution. Microsatellites: Evolution And Applications (2003) (Goldstein DB and Schlötterer C eds) Oxford University Press

Armour J.A.L., Alegre S.A., Miles S., Williams L.J., Badge R.M.. Minisatellites and mutation processes in tandemly repetitive DNA. Microsatellites: Evolution And Applications(2003) (Goldstein D.B. and Schlötterer C. eds) Oxford University Press.

Ayres K.L., Chaseling J., Balding D.J.. Implications for DNA identification arising from an analysis of Australian forensic databases. (2002) For. Sci. Int.129: 90-98.

Balding D.J.. Weight-Of-Evidence For Forensic DNA Profiles. (2005) Wiley Publishing.

Bellwood P.. A pre-history of Indo-Malaysian archipelago. (1985) Sydney: Academic Press.

Birdsell J.B.. Microevolutionary patterns in Aboriginal Australia. (1983) New York Oxford University Press.

Birdsell. J.B. Microevolutionary patterns in Aboriginal Australia: A gradient analysis of clines. (1993) Oxford University Press.

Brinkmann B., Sajantila A., Goedde H.W., Matsumoto H., Nishi K., Wiegand. P.. Population genetic comparisons among eight populations using allele frequency and sequence data from three microsatellite loci. (1996a) Eur. J. Hum. Genet. 4:175-182

Brinkmann B., Meyer E., Junge A.. Complex Mutational Events At The HumD21S11 Locus. (1996b) Hum. Genet. 98: 60-64.

Brinkmann B., Klintschar M., Neuhuber F., Hühne J., Rolf B. Mutation Rate In Human Microsatellites: Influence Of The Structure And Length Of The Tandem Repeat. (1998) Am. J. Hum. Genet. 62: 1408-1415.

Buckleton J, Triggs CM, Walsh SJ. Forensic DNA Evidence Interpretation. (2005) CRC Press. USA

Cavalli-Sforza L.L., Menozzi P., Piazza A.. The History and Geography Of Human Genes. (1994) Princeton University Press.

Chambers G.K., MacAvoy E.S. . Microsatellites: Consensus and Controversy. (2000) Comp. Biochem. Physiol. Part B. 126: 455-476.

Crawford M.H., Dyles D.D., Polesky H.F.. Genetic structure of Mennonite populations of Kansas and Nebraska. (1989) Hum. Biol. 61:493-514

Crawford M. (ed) Anthropological Genetics: Theory Methods And Applications. (2007) Cambridge University Press. New York

Crow J.F., Kimura M.. An Introduction To Population Genetics Theory. (1970) Harper and Row Publishers.

Dauber E.M., Bär W., Klintschar M., Neuhuber F., Parson W., Glock B., Mayr W.R.. Mutation rates at 23 different Short Tandem Repeat loci. (2003) Intl. Congress Series 1239: 565-567.

de Kock A., Kebede E.. DNA sequence characterization of the FGA STR locus in the Free State population of South Africa. (2006) J. Forens. Sci. 51: 199-200.

DiRenzo A., Peterson A.C., Garza J.C., Valdes A.M., Slatkin M., Freimer N.B. Mutational processes of simple-sequence repeat loci in human populations. (1994) PNAS 91: 3166-3170.

DNA Recommendations – 1994 Report Concerning further recommendations of the DNA Commission of the ISFH regarding PCR-based polymorphisms in STR (Short Tandem Repeat) systems. (1994) Int. J. Leg. Med. 107: 159-160.

Eisen J.A.. Mechanistic basis for microsatellite instability. Microsatellites: Evolution And Applications (Goldstein D.B. and Schlötterer C. eds) (2003) Oxford University Press.

Estoup A., Tailliez C., Cornuet J.M., Solignac M.. Size Homoplasmy And Mutational Processes Of Interrupted Microsatellites In Two Bee Species, *Apis mellifera* And *Bombus terrestis* (Apidae) (1995). Mol. Biol. Evol., 12: 1074-1084.

Estoup A., Jarne P., Cornuet J.M.. Homoplasy and mutation model at microsatellite loci and their consequences for population genetics analysis. (2002) Mol. Ecol., 11: 1591 – 1604.

Estoup A., Cornuet J.M.. Microsatellite evolution: inferences from population data. Microsatellites: Evolution And Applications (Goldstein D.B. and Schlötterer C. eds) (2003) Oxford University Press.

Evett I.W., Weir B.S.. Interpreting DNA Evidence. Sinauer, Sunderland MA.(1998)

Flood J.. (ed.) In Archaeology of the Dreamtime: The Story Of Prehistoric Australia And Its People. (2004) Harper Collins Publishing. Australia

Forster P., Kayser M., Meyer E., Roewer L., Pfeiffer H., Benkmann H., Brinkmann. B.. Phylogenetic resolution of complex mutational features at Y-STR DYS390 in Aboriginal Australians and Papuans. (1998) Mol. Biol. Evol. 15: 1108-1114.

Ghiani M.E., Moral P., Mitchell R.J., Hernández M., Garcia-Moro C., Vona G.. Y-Chromosome-specific STR haplotype data on the Rapanui population (Easter Island). (2006) Hum. Biol. 78: 565-578.

Gill P., Urquhart A., Millican E., Oldroyd N., Watson S., Sparkes R., Kimpton C.P.. A new method of STR interpretation using inferential logic – development of a criminal intelligence database. (1995) Int. J. Legal Med. 109:14-22

Griffiths R.A.L., Barber M.D., Johnson P.E., Gillbard S.M., Haywood M.D., Smith C.D., Arnold J., Burke T., Urquhart A.J., Gill P.. New reference allelic ladders to improve allelic designation in a multiplex STR system. (1998) Int. J. Legal Med. 111: 267-272.

Grubweiser P., Mühlmann R., Niederstätter H., Pavlic M., Parson W.. Unusual variant alleles in commonly used Short Tandem Repeat Loci. (2005) Int. J. Legal Med. 119: 164-166.

Hagelberg E., Kayser M., Nagy M., Roewer L., Zimdahl H., Krawczak M., Lió P., Schiefenhövel W.. Molecular genetic evidence for the human settlement of the Pacific: analysis of mitochondrial DNA, Y Chromosome and HLA markers. (1999) Phil. Trans. R. Soc. Lond. 354: 141-152.

Hancock J.M.. Microsatellites and other simple sequences: Genomic context and mutational mechanisms. Microsatellites: Evolution And Applications (Goldstein D.B. and Schlötterer C. eds) (2003) Oxford University Press.

Hartl D.. A Primer Of Population Genetics. (1999) Sinauer Associates.

Huoponen K., Schurr T.G., Chen Y.S., Wallace D.C.. Mitochondrial DNA variation in an Aboriginal Australian population: Evidence for genetic isolation and regional differentiation. (2001) Hum. Immunol. 62: 954-969.

Immel U.D., Kleiber M., Neerman-Arbez M., Klintschar M.. An isolated exclusion in the FGA system. (2004) Intl Congress Series 1261: 148-150.

Ingman M., Gyllensten U.. Mitochondrial genome variation and evolutionary history of Australian and New Guinean Aborigines. (2003) Gen. Res. 13: 1600-1606.

Jeffreys A.J., Wilson V., Thein S.L.. Individual-Specific 'Fingerprints' Of Human DNA. (1985) Nature 316: 76–79.

Jeffreys A.J. and Neumann R.. Somatic mutation processes at a human minisatellite. (1997) Hum. Mol. Genet. 6:129-136

Jobling M.A., Hurles M., Tyler-Smith C.. (eds) Human Evolutionary Genetics. (2004) Garland Publishing.

Jones R.. Pleistocene life in the dead heart of Australia. (1987) Nature 328: 666

Kayser M., Brauer S., Weiss G., Schiefenhövel W., Underhill P., Shen P., Oefner P., Tommaseo-Ponzetta M., Stoneking M.. Reduced Y-Chromosome, but not mitochondrial DNA, diversity in human populations from West Guinea. (2003) Am. J. Hum. Genet. 72: 281-302.

Kayser M., Choi Y., van Oven M., Mona S., Brauer S., Trent R.J., Suarkia D., Schiefenhövel W., Stoneking M.. The impact of the Austronesian experience from mtDNA and Y chromosome diversity in the Admiralty Islands of Melanesia. (2008) Mol. Biol. Evol. 25:1362-1374

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

Kirk R.L. Aboriginal man adapting. The human biology of Australian aborigines. (1983) Oxford University Press. Melbourne

Liu H., Prugnolle F., Manica A., Balloux F.. A geographically explicit genetic model of worldwide human settlement history. (2006) Am. J. Hum. Genet. 79: 280-287.

Lum J.K., Jorde L.B., Schiefenhovel W.. Affinities among Melanesians, Micronesians and Polynesians: A neutral biparental genetic perspective. (2002) Hum. Biol. 74: 413-430.

MacKnight C.C.. Macassans and aborigines. (1972) Oceania 42:283-321

Macho G., Freedman L., A reanalysis of the Andrew A Abbie morphometric data on Australian Aborigines. (1987) Occasional Papers in Human Biology, Australian Institute of Aboriginal and Torres Strait Islander Studies 4.

Mitchell R.J., Howlett S., White N.G., Federle L., Papiha S.S., Briceno I., Mc Comb J., Schanfield M.S., Tyler-Smith C., Osipova L., Livshits G., Crawford M.H.. Deletion

polymorphism in the human COL1A2 gene: Genetic evidence of a non-African population whose descendants spread to all continents. (1999) Hum. Biol. 71: 901-914.

Möller A., Meyer E., Brinkmann B.. Different types of structural variation in STRs: HumFES/FPS, HumVWA and HumD21S11. (1994) Int. J. Leg. Med. 106: 319-323.

Mullis K.B., Faloona F.A.. Specific synthesis of DNA in vitro via a polymerase-catalyzed chain-reaction. (1987) Methods Enzymol. 155: 335-350.

Parsons P.A., White N.G..Variability of anthropometric traits in Australian Aborigines and adjacent populations: its bearing on the biological origin of the Australians. (1973) Australian Institute of Aboriginal and Torres Strait Islander studies, Canberra.

Pérez-Miranda A.M., Alfonso-Sánchez M.A., Kalantar A., García-Obregón S., de Pancorbo M.M., Peña J.A., Herrera R.J.. Microsatellite Data Support Subpopulation Structuring Among Basques. (2005) J. Hum. Genet. 50: 403-414.

Redd A.J., Stoneking M.. Peopling Of Sahul: MtDNA variation in Aboriginal Australian and Papua New Guinean populations. (1999) Am. J. Hum. Genet. 65: 808-828.

Redd A.J., Roberts-Thomson J., Karafet T., Bamshad M., Jorde L.B., Naidu J.M., Walsh B., Hammer F.. Gene Flow From The Indian Subcontinent To Australia: Evidence From The Y Chromosome. (2002) Curr. Biol. 12: 673-677.

Reddy B.M., Naidu V.M., Madhavi V.K., Thangaraj K., Langstieh B.T., Venkataramana P., Kumar V., Singh L.. STR Data for the AmpF/STR Profiler Plus loci among 27 populations of different social hierarchy from southern part of Andhra Pradesh, India. (2005) For. Sci. Int. 149: 81-97.

Roberts R.G., Jones R., Smith M.A.. Thermoluminescence dating of a 50,000-year human occupation site in northern Australia (1990). Nature 345:153-156

Sharma V., Litt M.. Tetranucleotide Repeat Polymorphism At The D21S11 Locus. (1992) Hum. Mol. Genet. 1: 67.

Smith G.P.. Unequal crossover and evolution of multigene families. (1973) Cold Spring Harbour Symposium on Quantitative Biology 38: 507-513

Smith M.A.. The case for a resident human population in the central Australia ranges during full glacial aridity. (1989) Archaeol. Oceania24:93-105

Spurdle A.B., Jenkins T.. The origins of the Lemba 'Black Jews' of South Africa: Evidence from p12F2 and other Y-chromosome markers. (1996) Am. J. Hum. Genet. 59: 1126-1133.

Steinlechner M., Schmidt K., Kraft H.G., Utermann G., Parson W.. Gabon Black Population Data On The Ten Short Tandem Repeat Loci D3S1358, VWA, D16S539, D2S1338, D8S1179, D21S11, D18S51, D19S433, TH01 And FGA. (2002) Int. J. Leg. Med. 116: 176-178.

Su B., Jin L., Underhill P., Martinson J., Saha N., McGarvey S.T., Shriver M.D., Chu J., Oefner P., Chakraborty R., Deka R.. Polynesian origins: Insights from The Y Chromosome. (2000) PNAS 97: 8225-8228.

Takahata N., Lee S.H., Satta Y.. Testing multiregionality of modern human origins. (2001) Mol. Biol. Evol. 18: 172-183.

Thomas M.G., Parfitt T., Weiss D.A., Skorecki K., Wilson J.F., le Roux M., Bradman N., Goldstein D.B.. Y Chromosomes traveling south: The Cohen modal haplotype and the origins of the Lemba – The 'Black Jews of Southern Africa'. (2000) Am. J. Hum. Genet. 66: 674-686.

Thorne A., Grun R., Mortimer G., Spooner N.A., Simpson J.J., McCulloch M.T., Taylor L., Curnoe D..Australia's oldest human remains: age of the Lake Mungo 3 skeleton. (1999). J. Hum. Evol. 36: 591-612.

Tindale N.B.. Some population changes among the Kaiadilt of the Bentinck Island, Queensland. Records of the S.A. Museum. (1962) 14:297-336

Tishkoff S.A., Dietzsch E., Speed W., Pakstis A.J., Kidd J..R, Cheung K., Bonne-Tamir B., Santachiara-Benerecetti A.S., Moral P., Krings M.. Global patterns of linkage disequilibrium at the CD4 locus and modern human origins. (1996) Science 271:1380-1387.

Tsuji A., Ishiko A., Ikeda N.. The structure of a variant allele which is considered to be 30.3 in the STR locus D21S11. (2006) Leg. Med. 8: 182-183.

Underhill P.A., Passarino G., Lin A.A., Shen P., Mirazon Lahr M., Foley R.A., Oefner P.J., Cavalli-Sforza L.L.. The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations (2001) Ann. Hum. Genet. 65:43-62

van Holst Pellekaan S.M., Frommer M., Sved J.A., Boettcher B.. Mitochondrial control-region sequence variation in Aboriginal Australians. (1998) Am. J. Hum. Genet. 62: 435-449.

van Holst Pellekaan S.M., Ingman M., Robert-Thomson J., Harding R.M.. Mitochondrial genomics identifies major haplogroups in Aboriginal Australians. (2006) Am. J. Phys. Anth. 131: 282-294

Vandenberg N., Van Oorschot R.A.H., Tyler-Smith C., Mitchell R.J.. Y-Chromosome-specific microsatellite variation in Australian Aboriginals. (1999) Hum. Biol. 71:915-931.

Viard F., Franck P., Dubois M.P., Estoup A., Jarne P.. Variation of microsatellite size H homoplasy across electromorphs, loci and populations in three invertebrate species. (1998) J. Mol. Evol. 47: 42-51.

Vishwanathan H., Deepa E., Cordaux R., Stoneking M., Usha Rani M.V., Majumder P.P.. Genetic structure and affinities among tribal populations of southern India: A study of 24 autosomal DNA markers. (2004) Ann. Hum. Genet. 68: 128-138.

Walker A., Zorc D.. Austronesian loanwords in Yolngu-Matha of Northeast Arnhem Land. (1981) Aborig Hist 5:109-134

Wallace D.C., Torroni A..American Indian prehistory as written in the mitochondrial DNA: A review. (1992) Hum. Biol. 64: 403-416

Walsh S.J., Robinson S.L., Turbett G.R., Davies N.P., Wilton A.N.. Characterisation of variant alleles at the HumD21S11 locus implies unique Australasian genotypes and re-classification of nomenclature guidelines. (2003) For. Sci. Int. 135: 35-41.

Walsh S.J., Buckleton J.S.. Autosomal microsatellite allele frequencies for fifteen regionally defined indigenous Australian population datasets. (2007a) For. Sci. Int. 168: 29 – 42.

Walsh S.J., Mitchell R.J., Watson N., Buckleton J.S.. A comprehensive analysis of microsatellite diversity in Aboriginal Australians.(2007b) J. Hum. Genet. 52: 712-28.

Walsh S.J., Mitchell R.J., Torpy F., Buckleton J.S.. Use of subpopulation data in Australian forensic DNA casework. (2007c) For. Sci. Int. Genet. 1: 238-46.

Wells S.. The Journey of Man: A Genetic Odyssey. (2003) Random House.

White N.G.. Genes, languages and landscapes in Australia. McConvell P., Evans N. (eds) Archaeology and linguistics: Aboriginal Australia in global perspective. (1997) Oxford University Press, Melbourne

Wiegand P., Meyer E., Brinkmann B.. Microsatellite structures in the context of human evolution. (2000) Electrophoresis 21: 889-895.

Xiao F.X., Gilissen A., Cassiman J.J., Decorte R.. Quadruplex fluorescent STR typing system (HumVWA, HumTH01, D21S11 And HPRT) with sequence-defined allelic ladder identification of a new allele at D21S11. (1998) For. Sci. Int. 94: 39-46.