University of Technology Sydney
Centre for Forensic Science

EVALUATION OF NOVEL MITOCHONDRIAL DNA PANELS FOR FORENSIC USES

A thesis submitted in fulfillment of the requirements for a Master of Science (Research) degree

Ka Tak Wai

Bachelor of Forensic Biology in Biomedical Science

Supervisors

Associate Professor Peter Gunn, University of Technology Sydney, Centre for Forensic Science, Ultimo NSW 2007, Australia

Dr. Mark Barash, University of Technology Sydney, Centre for Forensic Science, Ultimo NSW 2007, Australia

October 2018

Certificate of Original Authorship

I, Ka Tak Wai declare that this thesis is submitted in fulfilment of the requirements for the award of the Master of Science (Research) degree, in the Faculty of Science at the University of Technology Sydney. This thesis is wholly my own work unless otherwise reference or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

This document has not been submitted for qualifications at any other academic institution.

This research is supported by an Australian Government Research Training Program Scholarship.

Production Note:

Signature …Signature removed prior to publication.……

Date …4th October 2018……...

Acknowledgements

This project would not have been possible without the support of everyone involved in it. Associate Professor Peter Gunn (University of Technology Sydney) and Dr. Mark Barash (University of Technology Sydney) are thanked for their supervision and guidance throughout the project. The volunteers who kindly provided their DNA sample for the project are thanked. The project recognises the funding and support provided by UTS, Centre for Forensic Science. This includes the travel grants, which were provided in the Vice Chancellor's Postgraduate Research Student Conference Award and by Professor Claude Roux for attending the 27th Congress of the International Society for Forensic Genetics (Seoul, South Korea). Researchers from the UTS Forensic Biology Unit, Elisha Prasad, Samara Garrett-Rickman, Andrew Walton, Alicia Khuu, Liam Pullan and Alexandra Summerell are acknowledged for their support and collaboration throughout the project. Furthermore, the technical support provided by Sophie Winiarski and Connie Soilemezis (UTS Science), James Cooke (Bio-Rad Laboratories, Inc.) and Beatrice Goh and Lucy Dagostino (Thermo Fisher Scientific) is greatly appreciated.

Presentation of Thesis

This thesis is presented in six chapters and follows the process of mitochondrial DNA testing in forensics. Chapter 1 introduces the theory of mitochondrial DNA, its testing and forensic uses in cases. Sequencing technologies for mitochondrial DNA are reviewed as well as the limitations in DNA markers for quantification and sequencing, providing a strong justification for this research in developing and evaluating novel mitochondrial DNA markers for human identification. Chapters 2-5 listed below describe the methods and results, and discuss the two projects undertaken in this research. Note that Chapter 3 is prepared as a manuscript for publishing and Chapter 5 is accepted as a research article. As such, the references of Chapters 3 and 5 are independent to the thesis. Both of these manuscripts were written in collaboration with the candidate as the primary contributor.

Chapter 2 introduces the qualities of DNA that can be encountered in compromised forensic samples and the research methods, which were used to mimic these samples.

Chapter 3 describes the development of a quantification assay which was used to assess the mitochondrial DNA quantity. While it does appear prior, it should be noted that the assay in Chapter 3 was developed last and is discussed retrospectively.

Chapter 4 presents the nuclear typing of DNA samples as an indication of DNA quality.

Chapter 5: ‘Performance of the Early Access AmpliSeq™ Mitochondrial Panel with degraded DNA samples using the Ion Torrent™ platform’

Chapter 6 provides a final discussion of the projects and its conclusions as well as recommendations for the research field.

Research Output

Publication

Wai, K.T., Barash, M., and Gunn, P., *Performance of the Early Access AmpliSeq™ Mitochondrial Panel with degraded DNA samples using the Ion Torrent™ platform.* Accepted in Electrophoresis, 2017.

Wai, K.T., Barash, M., and Gunn, P., *Development of a real-time quantification of mitochondrial DNA copy number in degraded samples.* Manuscript in preparation.

Presentation

August, 2017 'Evaluation of the Early Access AmpliSeq™ Mitochondrial Panel utilising Massively Parallel Sequencing'.

Poster Presentation: 27th Congress of the International Society for Forensic Genetics, Seoul, South Korea

Contents

List of Figures

List of Tables

List of Abbreviations

DVI	Disaster victim identification
DNA	Deoxyribonucleic acid
ATP	Adenosine triphosphate
BGA	Biogeographical ancestry
BP	Base pairs
CE	Capillary electrophoresis
DI	Degradation Index
HVR	Hypervariable region
INDEL	Insertion and Deletion
ISFG	International Society for Forensic Genetics
MPS	Massively parallel sequencing
mtDNA	Mitochondrial DNA
nDNA	Nuclear DNA
NUMT	Nuclear inserts of mtDNA
PCR	Polymerase chain reaction
PHR	Peak height ratio
qPCR	Quantitative PCR
RFU	Relative fluorescence unit
S.E.M	Standard error of mean
SD	Standard deviation
SNP	Single nucleotide polymorphism
STD	Standard
STR	Short tandem repeat

Abstract

Mitochondrial DNA testing is expanding the use of DNA as a forensic tool for human identification. The testing of mitochondrial DNA is a common practice for biological samples, which are compromised, degraded and contain limited STR information. In these cases, the amplification and sequencing of mitochondrial DNA becomes a valuable tool in determining the source of DNA samples. A review of mitochondrial DNA testing in forensic science reveals a number of improvements which can be made to this process. The project therefore aimed to improve the process of mitochondrial DNA testing. In particularly, it focuses on the developing and testing of mitochondrial DNA markers involved in the quantification and amplification of samples.

The quantification of mitochondrial DNA is important to optimising the amplification of DNA samples. Methods for quantification of mitochondrial DNA commonly involve estimating mitochondrial DNA quantities from nuclear DNA. This is usually unreliable as a high variability of mitochondrial DNA copy number exists between human cells. Hence, in this study, the development of a specific DNA assay for mitochondrial DNA provides a reliable determination of mitochondrial copy number. The testing of this novel assay has shown it is specific, sensitive and reproducible in DNA samples of artificially degraded qualities. As such, the incorporating of the DNA assay into workflows of mitochondrial DNA testing will improve the overall amplification of samples for downstream processes such as mitochondrial DNA sequencing.

Furthermore, the transition of mitochondrial DNA sequencing from capillary electrophoresis to massively parallel sequencing platforms is increasing the feasibility of typing multiple DNA fragments in a single reaction. This has led to the development of small PCR markers, which are capable of amplifying the entire mitochondrial genome even in challenging forensic samples. While the sequences of these PCR markers and panels are available for use, its full performance in amplifying compromised samples remains unknown due to the limited and usually specialist use of mitochondrial DNA testing and massively parallel sequencing in forensic laboratories. Therefore, the technical work carried out in this study tests the performance of the Early Access AmpliSeq™ Mitochondrial Panel (Applied Biosystems, CA, USA) for amplifying complete mitochondrial genomes in samples of degrading qualities. The testing of this Panel in compromised samples with limited STR success informs the use of the Panel in the mitochondrial testing of DNA in forensic laboratories. In this study, the typing of amplified DNA fragments in parallel reveals the recovery of complete mitochondrial DNA sequences in all samples. These samples were concordant to reference sequences and the HV1 and HV2 sequences provided by the ‘gold standard’ of capillary electrophoresis platforms. Importantly, the analysis of mitochondrial DNA sequences shows a capability to

resolve mitochondrial haplogroups and ancestries for familial matching. Overall, the results of this technical work confirms the Panel is able to amplify the complete mitochondrial genome of compromised samples for sequencing using massively parallel technologies. As such, this contributes to the use and validation of massively parallel sequencing technologies in forensic DNA testing.

Overall, the development of a novel DNA assay for the quantification of mitochondrial DNA and the technical testing of the Early Access AmpliSeq™ Mitochondrial Panel (Applied Biosystems, CA, USA) has contributed to improvement of mitochondrial DNA testing in compromised samples. Past use of mitochondrial DNA testing has provided identifications in cases of mass disasters, missing persons and historical remains. As such, the improvements to this process will continue to assist in these identifications.

Chapter 1: Introduction

General Overview

Disasters are events that significantly disrupt the functioning of a society and are beyond the response capabilities of local authorities. Mass disasters may be either man-made, such as terrorist attacks or due to natural phenomena such as earthquakes, tornados, and floods. Notable recent disaster events include the 2004 Indonesian Tsunami, the MH17 Malaysia Airlines Flight and the September 11 terrorist attacks. These events are characterised by the mass loss of human life and require a multi-disciplinary response of experts in the mass identification of individuals. The disaster victim identification (DVI) response is composed of a combination of dental, fingerprint and DNA profiling methods. Of the three primary methods, DNA is a slower means of identification, though it is more potent in cases of limited dental and fingerprint records [1]. DNA profiling pertaining to the field of forensic biology uses the genetic material found in all cells to identify or exclude individuals directly or in a kinship manner. The genetic material known as deoxyribonucleic acid [2] is predominately confined to the cell nucleus. With the exception of monozygotic twins, certain regions of nuclear DNA (nDNA) contain a number of sequence motifs known as short tandem repeats (STRs) that are unique to every individual. By combining a number of STR loci, a highly discriminated DNA profile can be obtained to either exclude or identify anonyms [3].

Though a highly discriminated profile is useful for human identification, the success of DNA profiling may become limited in cases of compromised biological samples [4]. In these cases, sequencing of mitochondrial DNA (mtDNA) offers an alternative means of identification. In comparison to nDNA, mtDNA exists in mitochondria as a circular, supercoiled and more abundant DNA molecule. These three characteristics allow mtDNA to withstand degradation for longer periods. However, it is important to recognise that the persistence of time will eventually degrade all DNA for sequencing – nuclear or mitochondria. In these cases, whole mitochondrial sequencing will utilise the most amendable DNA fragments for sequencing.

Traditional sequencing via chain-termination and capillary electrophoresis platforms has a limited capacity to sequence complete mitochondrial genome. While possible, the process is expensive, laborious and time-consuming. As a result, sequencing was confined to the HV1 and HV2 region and later expanded to the entire control region to account for phantom mutations. However, both these regions restrict the mitochondrial genome to its full use and provide limited insights into an individual's haplotype. The development of massively parallel sequencing (MPS) technologies in forensic genetics has expanded mtDNA sequencing to the complete genome. Complete amplifications of compromised samples in cases of mass disasters, missing persons and historical remains continue to improve as amplicons for MPS continue to reduce in size.

Starting from several kilobase pairs (bp), amplicons were later reduced to 380 bp, allowing for complete typing of mtDNA from hair shafts [5, 6]. The development of the Early Access AmpliSeq™ Mitochondrial Panel (Applied Biosystems, CA, USA) generates amplicons of 200 bp which has improved the sequencing of whole mitochondrial genomes from DNA samples of ancient skeletal remains [7].

The Panel containing 162 primer pairs amplifies DNA samples for mtDNA sequencing on the Ion Torrent Personal Genome Machine™ (PGM™) System (Applied Biosystems, CA, USA). While the Panel is capable of completely amplifying DNA samples from historical remains for mtDNA-MPS sequencing, evaluation of its performance with degraded samples are limited. The main aim of this research is to evaluate the performance of the Panel to amplify degraded samples for sequencing on the Ion Torrent PGM™ platform (Applied Biosystems, CA, USA). This involves artificially creating a number of DNA samples of progressively degraded qualities, which are amplified with the Panel for mtDNA-MPS sequencing. Evaluations of the Panel focus on the quality of sequencing data for coverage, variant calling and assignment of haplogroups/haplotypes as well as its overall workflow.

Furthermore, assessment of DNA is common practice to determine the quality of DNA and direct appropriate downstream processing for maximum sequence. The assessment usually involves a quantification of intact DNA molecules, which can also be visualised in agarose gel electrophoresis. Quantification of nucleic acids are well developed for the nuclear genome and are commonly used to estimate the copy number of mtDNA. However, the different proportions of the two genomes expressed between cells and individuals often leads to estimations of mtDNA content, which are misleading [8]. There is need to overcome this variability with a method of quantification, which is specific, sensitive and reproducible for mtDNA. This research extends to developing a real-time *in-house* quantitative PCR (qPCR) assay for the quantification of human mitochondrial genomes. The performance of the assay was tested in degraded DNA samples, focusing on its specificity, sensitivity and reproducibility.

Overall, this research reviews the testing of mtDNA and its uses in forensic genetics. It aims to improve the testing of mtDNA through evaluating the performance of novel mitochondrial markers. This includes developing an *in-house* qPCR assay for providing reliable quantifications of mtDNA in compromised samples. It also includes assessing the Early Access AmpliSeq™ Mitochondrial Panel (Applied Biosystems, CA, USA) for amplifying complete mtDNA samples for sequencing and establishing the testing of mtDNA using MPS technologies in forensic genetics.

1. Mitochondrial DNA

Mitochondria are subcellular organelles containing genetic material that is separate and distinct from nDNA. In normal cells, mtDNA exists as a circular molecule of approximately 16,569 bp that is inherited as a single genome via the maternal lineage. Depending on the cell type, the number of mtDNA can outnumber that of nDNA by at least a hundred-fold. For these reasons, recovery and use of mtDNA for human identifications is common in compromised samples.

This chapter reviews mtDNA and its uses as a genetic marker for human identification in forensic DNA testing. The review provides an overview of the origin of mitochondria and its genome, its structure and replication and, its unique mode of inheritance. The forensic uses of mtDNA in kinship and ancestral identifications are discussed by focusing on the mutations of the genome. Case studies highlight the forensic uses of the genome in cases of mass disasters, missing persons and historical remains. It also provides a perspective of mtDNA-sequencing technologies from chain-terminating platforms to transitioning MPS platforms. Overall, the review introduces mtDNA testing and the improvements that can arise from evaluating mtDNA markers for quantification and MPS.

1.1 Origin

The emergence of the mitochondrion organelle is one of the most significant events in evolutionary history as it marks the beginnings of eukaryotic organisms [9]. It is hypothesised mitochondria evolved as an adaptation to the changing conditions on Earth, in particular, the increase in atmospheric oxygen 3.8 billion years ago. The climatic transition pressured early life forms to adapt to a new means of producing energy, which favored a highly oxygenated atmosphere. Consequently, the mitochondrion organelle evolved in eukaryotes to fulfill these new energy requirements. DNA sequencing of eukaryotic genomes has identified mtDNA as a genetic admixture of the α-Proteobacteria, Rickettsiales, providing a phylogenetic link of this organism as the symbiont of mitochondria [10].

The endosymbiosis theory is the most widely accepted theory for the development of mitochondria in eukaryotes. The basis of the theory is that mitochondria arose when a bacterial cell was engulfed and began living as a symbiont in an archaeal host cell [11]. There are two scenarios for the endosymbiosis theory:

1. Archezoan scenario
2. Symbiosis scenario

The Archezoan scenario recognises mitochondria to evolve when the α-Proteobacteria symbiont was endocytosed by archezoan host cells. These pre-eukaryotic archezoan host cells were characterised to have membrane bound organelles such as a nucleus but were amitochondriate [11].

In contrast, the symbiosis scenario recognises the host cell to be more primitive and to have lacked membrane-bound organelles [10]. It views the endocytosis of the α-Proteobacteria symbiont to have catalysed the formation of mitochondria and subsequent membrane-bound organelles [11].

Overall, the archezoan scenario hypothesises that the host cell was a eukaryote with a primitive mitochondria vacuole and the symbiont provided the mtDNA for its evolution. The symbiosis scenario recognises the host cell to be primitive with no membrane-bound organelles and the symbiosis with the mitochondria symbiont catalysed the evolution of eukaryotes with membrane-bound organelles and mitochondria [12].

DNA sequencing of mtDNA and nDNA has found certain homologies that co-exist between the two genomes as a consequence of genetic transfer [10]. More than 200 regions of nDNA contain sequences, which fully or partially resemble those of mtDNA. These pseudogenes or homologies, referred to as nuclear inserts of mtDNA (NUMT) can lead to non-specific amplification of either genome. A common method to exclusively select genomes is to use primers that are specific to the heterologous regions of mtDNA [13].

It is important to note that not all eukaryotes developed the conventional mitochondria found in *Homo* sapiens. Eukaryotes such as *Trichomonas vaginalis* and *Giardia lamblia* have developed derivatives of mitochondria referred to as mitochondria related organelles (MROs) [10]. The fundamental difference between conventional mitochondria and MROs is that MROs lack a genome but still may or may not retain the ability to produce energy in the form of hydrogenosomes and mitosomes, respectively [10].

1.2 Structure and Replication

Having originated from prokaryotes, the structure and replication of mtDNA are similar to that of bacterial plasmids. Structurally the human mitochondrial genome is made of 16,569 bp and organises into a supercoiled double-stranded circular DNA molecule. In contrast to the linear form of nDNA, the rigid plasmid-like structure allows mtDNA to become highly resistant to a range of physical, chemical and biological degradation factors that are common in the natural environment (Table 1).

Table 1: Commonly encountered DNA degradation factors.

Physical Factors
Combustion (e.g. explosives)[14]
Heat (e.g. arid environments, fired cartridge cases, fire investigations)[15-17]
Sonication/ ultrasonic frequency waves [18]
Radiation (e.g. ultraviolet)
Humidity
Chemical Factors
Household cleaners (e.g. bleach, detergent)
Alcohols
Destructive treatment reagents (e.g. cyanoacrylate, ninhydrin)[17]
Petroleum (e.g. fire investigation)
Biological Factors
Microbial activity (e.g. decomposition)
Nucleases, DNases (ubiquitous)

Compared to the highly regulated replication of nDNA where one replication cycle terminates before commencing the next, the genetic material in mitochondrion undergoes a continuous replication, which is independent to the cell cycle. As the primary function of mitochondria regulates cellular processes by producing adenosine triphosphate (ATP), its rate of mtDNA replication responds to the energy demands of the cell. Depending on the cellular activity, mtDNA presents higher copy numbers in cells of greater activities than others. In muscle cells, a high mtDNA number is present to respond to the body's movement and exercise. Contrast to leukocytes, a lower mtDNA content is sufficient for protecting the body from infections [19]. The rate of mtDNA replication also varies between individuals due to the different energy demands of one's age and lifestyle.

The ability to replicate independently of the cell cycle means that each cell has an abundance of mitochondria where each mitochondrion has multiple copies of mtDNA. As a result, mtDNA has a higher copy number than nDNA where the amount of mtDNA is on average 1,000 to 10,000 times to that of nDNA. A high copy number is beneficial for mtDNA typing in cases of highly degraded DNA samples as the recovery of thousands of mtDNA copies is greater than the two copies of nDNA [20].

1.3 Inheritance

In contrast to the bi-parental inheritance of nDNA, mtDNA is normally inherited as a single genome of the maternal lineage. Its inheritance as a single genome is referred to as a haplotype. In mitochondria, uniparental inheritance is essential for survival as bi-parental inheritance promotes an extensive degree of heteroplasmic cells that impairs mitochondria and its function to produce cellular energy. This was observed in a study conducted by Sharpley *et al.,* [21] where mice with dysfunctional mitochondria were deprived of cellular ATP and subsequently developed physiological disorders, mental retardation and reduced behavioural despair. Thus it has been agreed upon that the bi-parental inheritance of mtDNA is unfavourable and overtime a uniparental mode of inheritance was naturally selected to produce physiologically 'fit' organisms with functional mitochondria [22].

The paternal inheritance of mtDNA is a rare condition [21]. The explanation being that the paternal mtDNA located in the midpiece of sperm are oxidised and damaged during fertilisation [23]. The paternal inheritance of mtDNA of such mutations results in offspring of impaired mitochondria function. This is seen in a case of an adult male whose exercise tolerance was limited by a sporadic point mutation from mtDNA of paternal origin [24]. Thus, inheritance of maternal mtDNA is favoured over paternal mtDNA to avoid such disorders.

Knowledge of mechanisms by which maternal inheritance occurs is limited to the two proposed models:

1. The simple dilution model
2. The active degradation model

The simple dilution model proposes that a dilution of low amounts of paternal mtDNA to the excess amounts of maternal mtDNA in oocytes, in an undetectable amount of paternal mtDNA in newborns [25].

The active degradation model proposed that paternal mtDNA is degraded from spermatic mitochondrion either before or after fertilisation. An ubiquitin-proteasome pathway has been proposed for the elimination of paternal mtDNA post-fertilisation. The proposed mechanism is based on self-recognition where during spermatogenesis, paternal mtDNA are marked with ubiquitin for degradation by proteasomes and lysosomes circulating inside zygotes [25, 26]. However, amplification by polymerase chain reaction (PCR) reveals a lack of mtDNA in motile sperm, suggesting an elimination of mtDNA prior to fertilisation as opposed to degrading in zygotes [23]. The study conducted by Luo *et al.,* [23] proposes a passive degradation of paternal mtDNA in spermatogenesis, which leaves a vacuole of mitochondria in the sperm

midpiece to supply the energy necessary for fertilisation. This pre-elimination of mtDNA in spermatogenesis has also been observed in *Drosophila* where EndoG digests are used to degrade mtDNA [27].

2. Heteroplasmy

In theory, the maternal inheritance of mtDNA as a single genome creates a monoclonal population of mtDNA. Heteroplasmy represents the occurrence of two or more variable types of mtDNA within a single individual that arise from base mutations and polymorphisms. Heteroplasmic mutations are common, presenting:

a) Within the one tissue
b) One haplotype in one tissue, another haplotype in the other
c) Heteroplasmy in one tissue sample, homoplasmy in another tissue sample

These mutations can be point mutations that affect the base sequence or length mutations that affect the size of the genome. Point mutations are commonly composed of single nucleotide polymorphisms (SNPs), which are single base pair changes that are highly variable between individual members of a population. To be classified as a SNP, these nucleotide changes need to occur in at least 1% of the population. Once identified, SNPs are classified into two categories: 1) transitions which are nucleotide changes that remain within the same carbon nitrogenous base and, 2) transversions which are nucleotide changes that do not remain within the same carbon nitrogenous base (Figure 1).

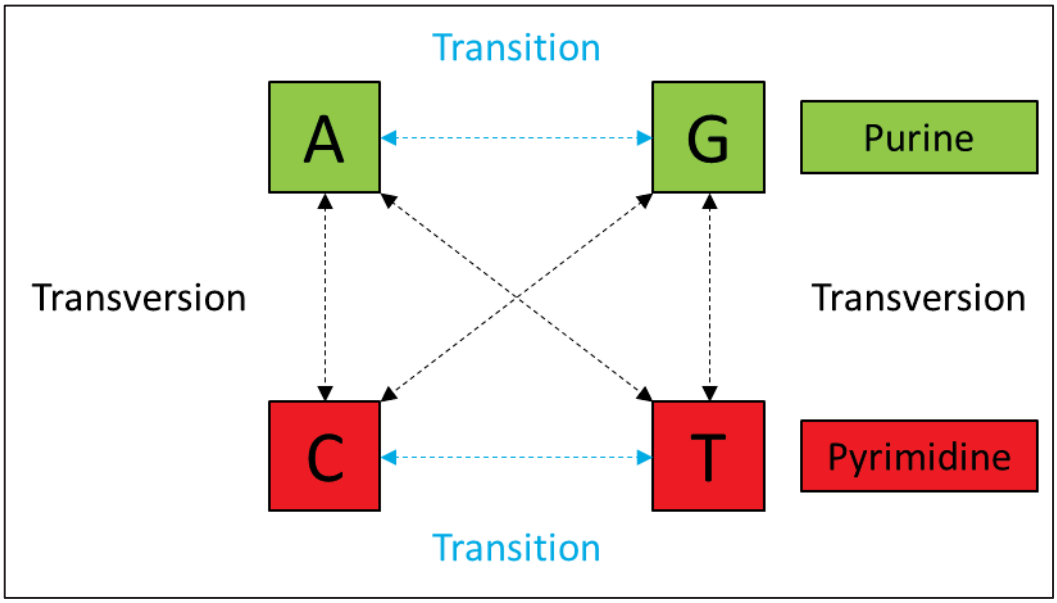


Figure 1: Types of Single Nucleotide Polymorphisms (SNPs). SNPs are classified into two categories: 1) Transitions and 2) Transversions.

Length heteroplasmy are composed of insertions and deletions (INDELs) of DNA bases, which changes the overall length of the genome. In mtDNA, length heteroplasmy is frequently reported between nucleotides 302-310 of the mitochondrial genome as an uninterrupted homopolymeric stretch of cytosine residues [28]. The extent of SNP and INDEL mutations varies between individuals as it is mediated by one's genetic and lifestyle factors such as diet and exercise. However, due to the progressive nature of mutations, higher levels of heteroplasmic cells have generally been observed in aged individuals [19].

While most mutations are silent and harmless, other mutations are harmful especially when they occur in the coding regions of mtDNA. These genetic mutations could lead to the expression of diseases that alters a person's physical appearance. These externally visible characteristics could be of investigative use as they provide a phenotypical profile to either exclude or fail to exclude persons of interest. An example of such an autosomal investigative tool is the HIrisPlex® system. The HIrisPlex® system is a validated sequencing assay that targets the SNPs in nDNA responsible for the pigmentation of an individual's iris and hair. Consequently, the assay provides investigators with phenotypical information such as a person's hair and eye colour [29].

Further to the investigative value of mtDNA, there is the potential to infer one's age from the percentage of heteroplasmy found in an individual. The rationale is that progressive mutations accumulate as one ages contributing to an overall higher degree of heteroplasmy in aged individuals. Further to that, studies have reported that when the level of mutated mtDNA exceeds a certain threshold, a person develops an acquired mitochondria deficiency that accelerates ageing. As it stands, the percentage of mutated mtDNA being representative of the extent of heteroplasmy may be used to infer an individual's estimated age [19].

2.1 Interpretation of Heteroplasmy

Heteroplasmy presents a significant challenge to the interpretation of mtDNA sequences for human identifications as mixed variants between test and reference samples may be explained by mutations. By means of heteroplasmy, sequences may be recognised as being concordant and a 'fail to exclude' result is reported. However, the nucleotide differences can also arise from samples truly originating from different donors, in which case, an exclusion of profiles is reported. The presence of heteroplasmy does not constitute a reason to exclude identity. In the uploading of mtDNA sequences to population databases, it is convention to report the most dominant variant. Current guidelines recommend the comparing of a questioned sample to multiple reference samples of the donor's hair, blood and buccal swab [30]. Multiple samples are required to cover a wide range of reference haplotypes that is

representative of the individual's genome. A comparison to a well referenced sample should provide a result of exclusion, inconclusive or fail to exclude for the overall mtDNA sequence (Table 2).

Table 2: Mitochondria Interpretation Guidelines for Heteroplasmy [30, 31].

	Exclusion	Inconclusive	Fail to Exclude
Samples have the same sequence	No – differ by two or more nucleotides[1]	No – differ by a single nucleotide	Yes
Samples have a common length variant (positions 302-310 rCRS)	Yes or No	No	Yes
Variant nucleotide present in reference samples (heteroplasmy)	No	No	Yes

[1]Excluding length heteroplasmy

3. Mitochondrial DNA Sequencing

The human mitochondrial genome contains SNPs that are highly variable between individuals. SNPs span the whole mitochondrial genome with a significant portion of variants located in the HV1 (16,024-16,365 rCRS) and HV2 (73-340 rCRS) region. MtDNA typing of these highly variable regions produces a sequence that can be used for haplogrouping and ancestry analysis. Variant calling of substitutions, insertions, deletions and heteroplasmy are reported relative to a whole mitochondrial reference genome including the sequence range(s) (Table 3) [32]. The revised Cambridge Reference Sequence (rCRS) (NC_012920.1) is the most widely used reference followed by the Reconstructed Sapiens Reference Sequence for human ancestries [33]. The forensic reporting of mtDNA sequences for identification and population databases are recommended by the International Society for Forensic Genetics (ISFG) (Table 3) [34]. These guidelines include aligning sequences to the closest phylogenetic neighbour, anchoring of the C-tract positions as 310C and 16,189C and reporting of INDELs relative to the 3' of repeat sequences [34].

Table 3: Nomenclature for reporting mtDNA variants. Each type of mtDNA SNP variant is reported with an example. Variants are referenced in relation to the rCRS. Description of variants follows the nomenclature recommended by the International Society for Forensic Genetics (ISFG) [34] [35].

Variant	Example	Description
Substitution	8993C	Substitution of T[1] with C in the 8993 nucleotide position.
Deletion	8993-	Deletion of T at the 8993 nucleotide position.
Insertion	8993.1T	Insertion of T at the 8993 nucleotide position.
Heteroplasmy	8993Y	Mixed population of T or C at the 8993 nucleotide position.

[1]A, T, C, G are representative of the nucleotides bases adenine, thymine, cytosine, and guanine, respectively. This follows the nucleotide code of the International Union of Pure and Applied Chemistry (IUPAC).

The processing of mtDNA samples involves amplifying DNA extracts for specific mtDNA fragments and the sequencing of these amplicons on technological platforms such as chain-terminating sequencing and massively parallel sequencing. The method of chain termination sequencing involves incorporating dye-tagged dideoxynucleotides into a series of single-stranded DNA (ssDNA) amplicons that selectively terminates and build up a chain of DNA fragments. These fragments are resolved using a capillary electrophoresis platform, providing an overall read of the DNA sequence [36, 37]. In 1981, a pre-form of the technology was used to sequenced the whole mitochondrial genome, becoming the first original (Cambridge) reference sequence for recording mtDNA SNP variants [38]. As a single CE sequencing reaction is limited to the typing of amplicons of the same sequence, sequencing of mitochondrial genomes is usually restricted to the HV1 and HV2 regions as well as the entire control region [39]. This becomes impractical with the additional reactions that are needed for both forward and reverse DNA strands.

3.1 Massively Parallel Sequencing

Massively parallel sequencing (MPS) is a collection of second generation sequencing technologies and were developed after the advent of dideoxynucleotide sequencing. In

comparison to CE-based platforms, MPS uses a range of technologies for resolving DNA sequences including an ion semiconductor chip-based platform.

The Ion Torrent PGM™ System (Applied Biosystems, CA, USA) uses a 'sequencing by synthesis' approach, which monitors the release of H^+ ions as nucleotides are incorporated onto a DNA strand. The release of H^+ ions causes a pH change in the sequencing environment. By monitoring DNA synthesis on an ion sensitive semiconductor chip, the pH change is recorded and converted to an electrical signal indicating that a base was called for the respective nucleotide. The system cycles through the four nucleotides until base calling and sequencing is complete for the overall DNA amplicons. With chip output reaching capacities of 5 million reads, targeting of the HV1 and HV2 regions as well as the entire control region and complete mitochondrial genomes can be sequenced multiple times. The ability to sequence genomes multiple times offers a high coverage and sensitivity that has improved the accuracy to call SNP variants and resolve the levels of heteroplasmy in individuals [40]. The Ion Torrent PGM™ System (Applied Biosystems, CA, USA) is a common MPS platform with only a handful of alternatives such as the MiSeq FGx™ Forensic Genomics System (Illumina® Inc., CA, USA) [41].

The MiSeq FGx™ Forensic Genomics System (Illumina® Inc., CA, USA) uses fluorescently labelled nucleotides to amplify and record DNA sequences. Its multiplex dye system allows each nucleotide to be identified by a separate fluorescence (e.g. A-green, T-blue, C-yellow, G-red). These free nucleotides are used to synthesis the DNA strand by DNA polymerase where during sequencing one nucleotide group is used at any given time to extend the complementary strand and produce a fluorescent signal. The MiSeq captures a fluorescent image of this reaction and calls the base in real time. The nucleotide group is washed away and a different DNA group is added to continue extending the strand. This process is cycled until the entire DNA sequence has been synthesised.

Overall, the underlying sequences produced by CE and MPS technologies are of similar quality. However, the different methods of sequencing in MPS technologies is expanding its forensic uses to typing mtDNA in samples of a compromised nature. The amplification of compromised samples continues to improve with smaller primer sets, leading to an increase number of amplified DNA fragments for sequencing. This increase is also facilitated by expansion of mtDNA analysis to the entire genome to better resolve phylogenetic relations. While it is possible to sequence the entire mitochondrial genome using CE systems, the increased number of fragments makes it no longer feasible or practical to do so, especially when there are multiple samples requiring a forward and reverse sequence. The additional indexing of fragments and samples in MPS technologies enables the sequencing of all

amplicons in a single sequencing reaction that is usually made-up of multiple samples (Figure 2). As a result, the overall sequencing times for samples is improving with some platforms completing typing in as little as three hours [42]. The ability to multiplex samples in faster sequencing times continues to become more affordable to forensic service laboratories [43]. Indeed the cost of MPS has been estimated to be $1,700 for 150,000,000 reads compared to Sanger reads costing on average $1.70 per read.

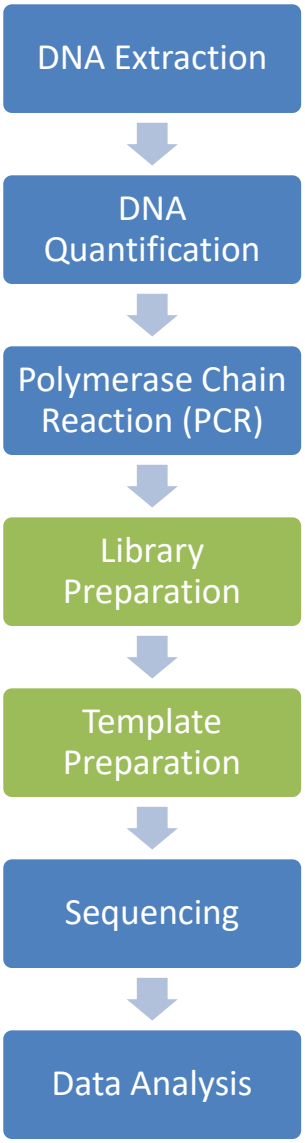


Figure 2: Comparison of a MPS workflow relative to a dideoxynucleotide workflow. Procedures common to both workflows are filled in blue. Procedures exclusive to MPS are filled in green. In chain-terminating sequencing, a separate amplification and sequencing reaction is set-up for each type of amplified ssDNA fragment. A separate reaction is needed for the forward and reverse strand. In MPS technologies, amplification and sequencing is multiplexed in a single reaction. This single sequencing reaction can also be multiplexed with many samples.

4. Maternal Biogeographical Ancestry

By sequencing thousands of SNPs in the human genome and comparing it across 51 populations, geneticists have discovered modern populations to be divided into one of seven continental groups: Africa, Middle East, Europe, Central/South Asia, East Asia, Oceania and Native America. These continental groups are based on the theory that prior to the era of exploration and invasion, groups were geographically isolated from one another for many generations and thereby evolving a set of unique genotypes in their mitochondrial genome [44]**.** By comparing the homology across populations, the study also shows the mitochondrial genome of *Homo sapiens* to have originated from a single source in sub-Saharan Africa before migrating into the different continental groups.

While most of these genome-wide association studies have increased understanding of human migration, evolution and disease, it has also benefitted the forensic field by allowing the potential to predict an individual's maternal ancestry. Through sequencing the SNPs responsible for ancestry, an individual may be assigned to a haplogroup with its biogeographical ancestry (BGA). An example of a number of mtDNA markers used to identify ancestries of an admixture of Asian and European individuals in the Australian population are reported in Table 4 [45].

Table 4: Mitochondrial DNA markers used to infer Asian and European ancestry in an Australian population [45]. The reported markers are part of the GenoCoRe22 SNaPshot assay [46]. Frequencies (%) of defining SNP variants are provided for each ancestry group.

SNP variant position (rCRS)	Haplogroup	Frequency (%)		Biogeographical Ancestry
		Europe	Asia	
4,580	V	0.9 – 26.7	-	European
5,178	D	-	10.0 – 62.0	Asian
7,028	H	22.0 – 52.2	-	European
8,280-	B	-	8.8 – 30.4	Asian
10,400	M	-	6.6 – 29.9	Asian
10,550	K	0.0 – 13.3	-	European
11,467	U	20.0 – 52.9	-	European
12,612	J	4.4 – 17.8	-	European
13,263	C	-	2.0 – 8.0	Asian
13,368	T	4.3 – 8.7	-	European
13,928	R9	-	4.3 – 10.0	Asian
14,766	HV	36.2 – 51.1	-	European

A haplogroup is a phylogenetic term used to define a genetic population that has descended from a single common ancestor. Haplogroups are first assigned a capital alphabet to indicate the main continental group. The alphabetical assignment is based on the order in which the haplogroup was discovered and is unrelated to the haplogroups age or origin. For example, haplogroup A indicates the group to be first discovered by geneticists as opposed to pertaining to the African continent. The nomenclature is highlighted in Table 5 and reports the major mitochondrial haplogroups in modern Europe with respect to its age and maternal descendant [47].

Table 5: The Major Mitochondrial Haplogroups of modern Europe. Haplogroups are reported with their respective matrilineal descendant and the time since formation. Adapted from Sykes [47].

Haplogroup	Matrilineal descendant	Formed (years ago)
U	Ursula	45,000
X	Xenia	25,000
H	Helena	20,000
V	Velda	17,000
T	Tara	17,000
K	Katrine	15,000
J	Jasmine	10,000

In addition to the assignment of the continental group, haplogroups may be further divided into subpopulations depending on the extent of population divergence. These subpopulations or subhaplogroups are represented as lower case letters or numbers and proceed the capital continental assignment in nomenclature (Figure 3) [48].

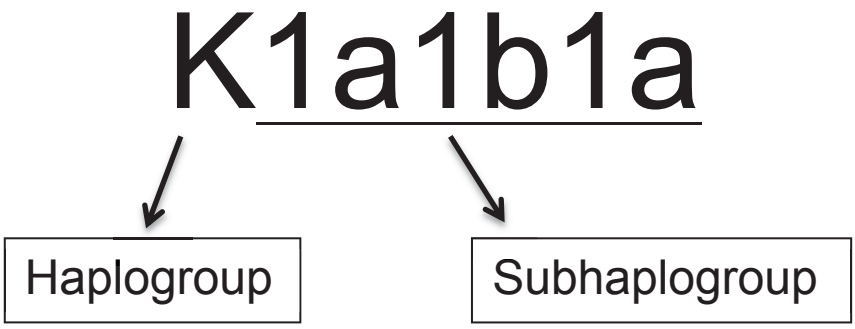


Figure 3: Nomenclature for Haplogroup Assignment.

The defining set of SNP variants and overall haplotypes used to determine each of the haplogroups have been summarised by Van Oven and Kayser [48]. To achieve discrimination, the SNP markers need to be highly specific and dominant in one population group while being absent or recessive in the other [45]. While most of these SNP variants are located in the control region or displace loop (D-loop) of the mitochondrial genome that is composed of hypervariable regions (HVR) 1, 2 and 3, it has been found that other polymorphic sites exist in the coding regions of mtDNA [49] (Figure 4).

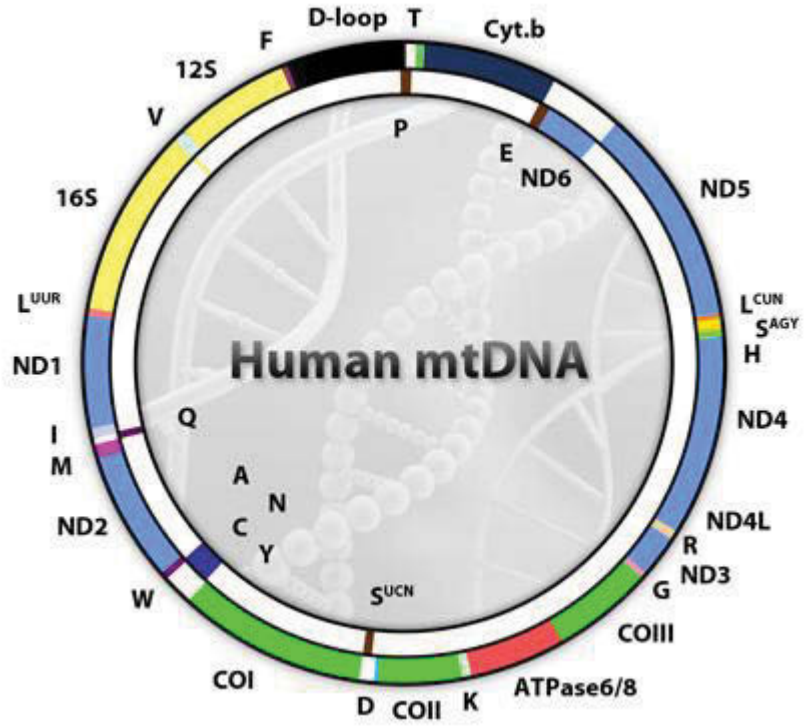


Figure 4: Organisation the Human Mitochondrial Genome. The majority of SNP ancestry markers are located in the D-loop or control region (positions 16024-576 rCRS) of the mitochondrial genome with few reported in the coding region. Image courtesy of http://www.mitopedia.org/eng/mitocondri2_z.html.

The typing of mtDNA in the HV1 and HV2 regions is not an ideal practice for forensic reporting as the resolution of haplogroups and ancestries are limited [34]. While typing of the entire control region improves haplogroup resolutions, sequencing of the whole mitochondrial genome can maximise ancestry estimations to a sub-population level where regions such as Northwest Europe and Northeast Europe can be resolved [7].

Commercial providers such as 23andMe™ (CA, USA) and Family Tree DNA™ (Texas, USA) offer whole mitochondrial door-to-door sequencing services to the public for limited ancestry tracing. In addition to whole mitochondrial sequencing, these companies type for Y-chromosomal markers as well as autosomal ancestry informative markers (AIMs). These additional markers are claimed to provide a comprehensive and full profile of an individual's ancestry as mtDNA typing only provides a partial profile of an individual's maternal ancestry. Such a limitation must be considered when it comes to reporting genetically admixed individuals as a uniparental marker would most likely misrepresent an individual's full ancestry [45].

5. Forensic uses of Mitochondrial DNA

5.1 Kinship Matching

The typing of mtDNA is common in cases of mass disasters, missing persons and historical remains. These cases typically present samples of a compromised or aged nature, from which mtDNA can infer an individual's identity from their maternal lineage. Accordingly, the identification of individuals is usually aided by referencing to samples of maternal relatives. This was seen in historical cases such as the identification of the remains of the Romanovs and King Richard III as well as the mass repatriation of the Australian and British war dead soldiers in the Fromelles Project and victims of the 2004 Indonesian Tsunami [50, 51] [52]. In forensic investigations, the use of mtDNA typing is also routinely performed in the analysis of hair shafts that contain little to no genomic DNA such as in the case of the Eaglesham Murder.

5.1.1 Identification of the Romanov Remains

For over 300 years, the Romanovs ruled the Russian Empire until the year of 1918 when the royal family including their children and servants were captured, executed and buried by the Bolsheviks. The mass graves hidden for over 80 years, led to speculation of the Romanovs escaping execution and living in the community. Over 200 people tried to claim the royal identity of the Romanovs with the most famous case being Anna Anderson whose appearance closely resembled Anastasia Romanov. In 1991, a mass grave containing the remains of nine individuals were discovered in Yekaterinburg, Russia. The remains showed gunshot and bayonet wounds before being burned and doused in acid. These peri- and post- mortem injuries coincided with the historical accounts of the Romanovs' execution. STR testing confirms a family group, which were putative of Tsar, Tsarina and three of their daughters. The identification of Tsarina to these three children, was confirmed with mtDNA sequencing, which showed an exact match of the HV1 (16,020-16400 rCRS) and HV2 (48-408 rRCS) sequences. This mtDNA sequence was compared to a living maternal relative, Prince Phillip, the Duke of Edinburgh (Figure 5).

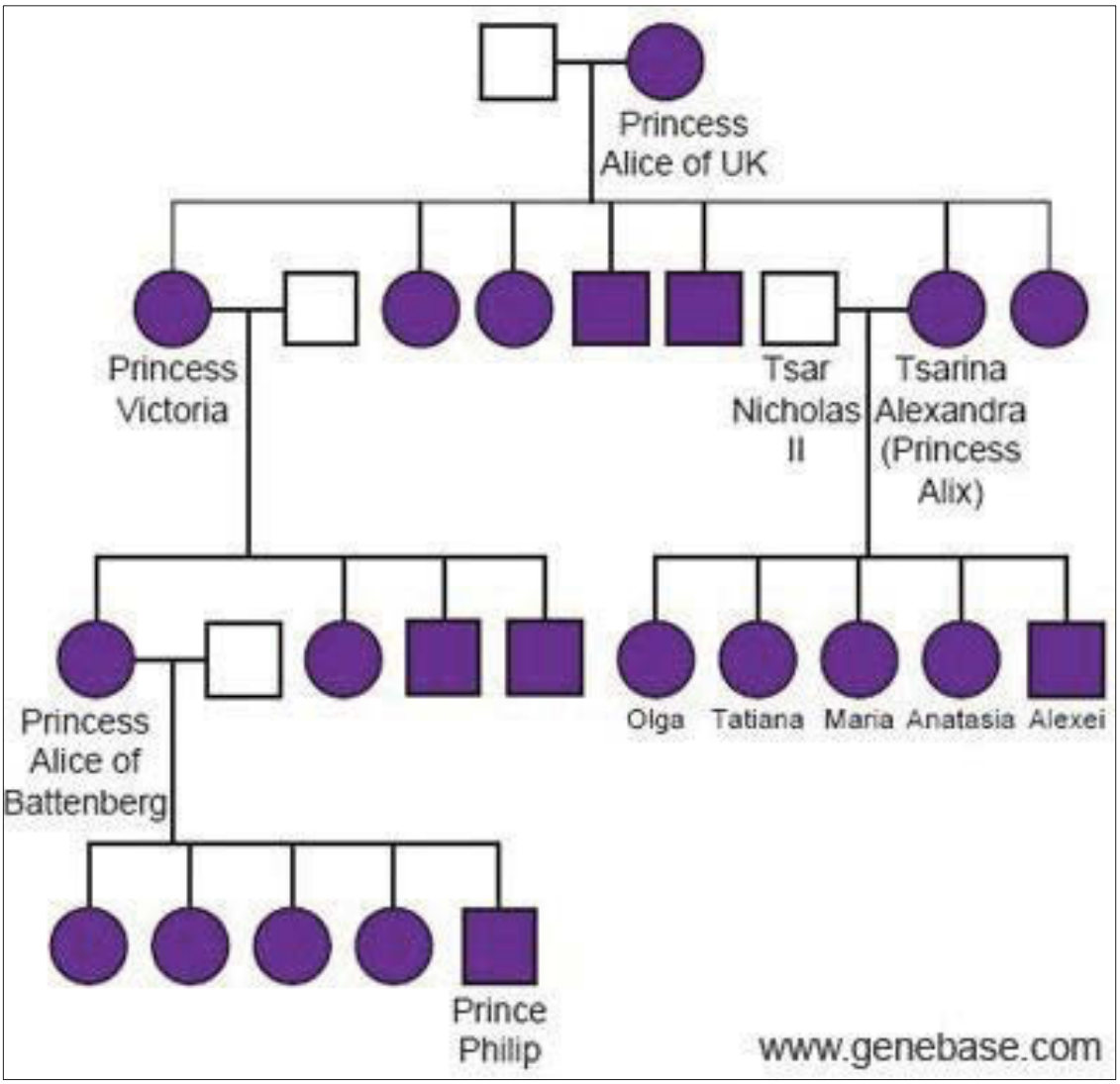


Figure 5: Kinship analysis of the Romanov Maternal Lineage. Image courtesy of www.genebase.com. Maternal inheritances of mtDNA are highlighted in purple. Males are represented as squares. Females are represented as circles. Prince Phillip, Duke of Edinburgh was referenced to be the great-nephew of Tsarina Alexandra, mother of the Romanov line.

MtDNA typing of the putative remains of Tsar reported a single point heteroplasmy (16,169Y) in the mitochondrial control region. Overall, one of these mtDNA sequences was the same as his two living maternal relatives, confirming his identity as Tsar Nicholas II [53].

In 2007, a second grave containing two anonyms of a male and female were discovered in a close-by grave. These remains were petite and exhibited similar peri- and post- mortem injuries to the first grave. Sequencing of the entire mitochondrial control region (16,024-576 rCRS) showed a complete homology of both the remains to Tsarina and concluded a maternal relatedness. Additional STR testing and Y-chromosomal testing (of the male-related individual to Tsar and a living paternal relative), places the two petite remains in a family pedigree to the

first grave. These DNA results conclude the remains as the last two missing children of the Romanov family who were Alexei and one of his sisters [51].

5.1.2 Identification of King Richard III

In 2012, a set of skeletal remains were discovered at a carpark of the Grey Friars friary in Leicester. Osteological examination reveals the anonym to have features that coincided with historical accounts of King Richard III – male, 30 to 37 years old, scoliosis and peri-mortem injuries sustained from the Battle of Bosworth. Radiocarbon dating also placed the remains to a time consistent with his death in the 1485 common era [50].

Using MPS technologies, complete mitochondrial genomes were sequenced and compared to maternal living relatives, Michael Ibsen and Wendy Duldig who were 19 and 21 generations removed from Richard III, respectively (Figure 6). The mtDNA sequences show a complete reference to Michael's sequence while one single base difference (8,994A) was reported in Wendy's sequence. This difference is reconciled with the high substitution rate of mtDNA over time. Further Y-chromosomal testing placed the remains in a northern and eastern European haplogroup (J1c2) that is predominately evident in the English-Wales population. This reconciles with Richard's and his sister, Anne of York's English background. Overall, these DNA results concluded the identification of King Richard III.

The historical identification of Richard III highlights the use of mtDNA testing in degraded samples that are common in missing persons and cold cases. It shows the use of MPS technologies that expands analysis to the complete mitochondrial sequence, providing more discrimination and phylogenetic data. It also demonstrates the importance of living maternal relatives as reference sequences, especially in aged samples.

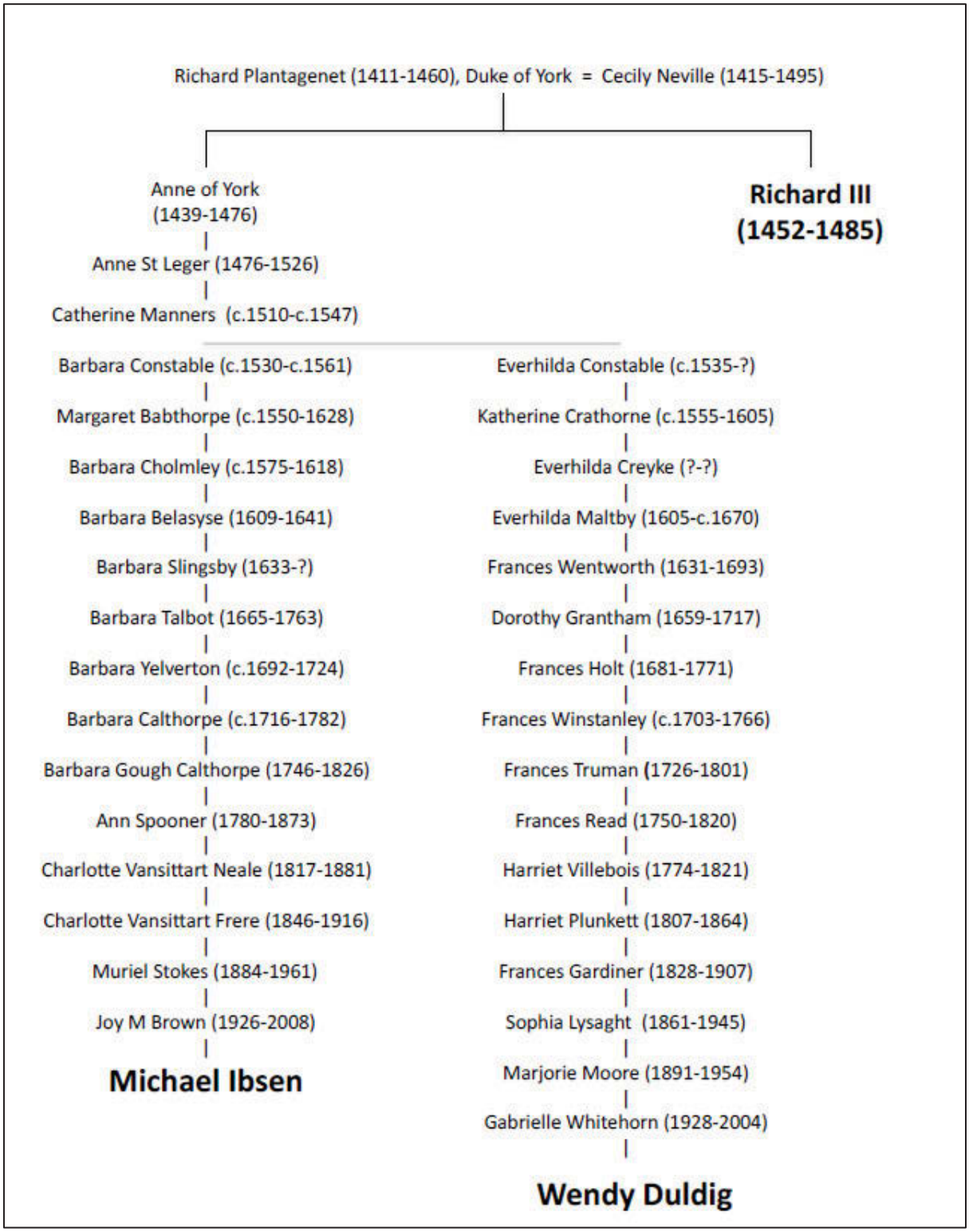


Figure 6: Kinship analysis of the Plantagenet Maternal Lineage [50]. Mitochondrial DNA typing identified Michael Ibsen and Wendy Duldig to be living matrilineal relatives of Richard III's sister, Anne of York and were 19 and 21 generations removed from Richard III, respectively.

5.1.3 The Fromelles Project

In 1916, World War I, up to 2000 Australian causalities alongside British soldiers were recorded in the Battle of Fromelles, Northern France. Many of these soldiers became missing-in-action and remain unidentified in the Fromelles grounds. In 2009, 250 skeletal remains were discovered at the battlegrounds of Pheasant Wood. A joint Australian and British Fromelles project team was formed to aid in the mass identification of these remains as potentially being the missing-in-action soldiers. In many cases, STR profiling was limited due to the compromised state of remains, which only showed PCR amplifications of up to 300 bp. Furthermore, reference samples for autosomal profiles from first-degree relatives were limited provided the historical age of the samples. In contrast, amplification of DNA samples for mtDNA and Y-chromosomal DNA were successful, generating sequences for human identifications [52]. Involvement of the living maternal and paternal relatives of the fallen soldiers in establishing a DNA reference database was essential to comparing DNA sequences. In reconciliation with the database, mtDNA and Y-chromosomal typing has successfully repatriated the remains of 124 Australian soldiers with their respective name, rank, and regiment.

5.1.4 The Indonesian Tsunami Victim Identification

The Thai-Indonesian Tsunami DVI operation involved the identification of the nearly 5,400 people who were killed in the 2004 Southeast Asia Tsunami in Southeast Thailand. DNA identifications accounted for up 799 people while the remaining victims were identified through fingerprints, dental and physical means (e.g. tattoos and personal medical devices such as pacemakers) (Figure 7). MtDNA typing of the HV1 and HV2 regions was successful in 258 tooth samples, proving complete sequences for matching [54]. Familial reference samples were obtained either from living relatives or deceased relatives whose profile was previously lodged into the ante-mortem database [1]. In reconciling with the DNA databases, 3,308 victims of the mass disaster were identified and returned to their countries. In-line with the Fromelles Project, the case study of the identification efforts in the 2004 Tsunami highlights the importance of the public's support in establishing DNA databases for references and identifications.

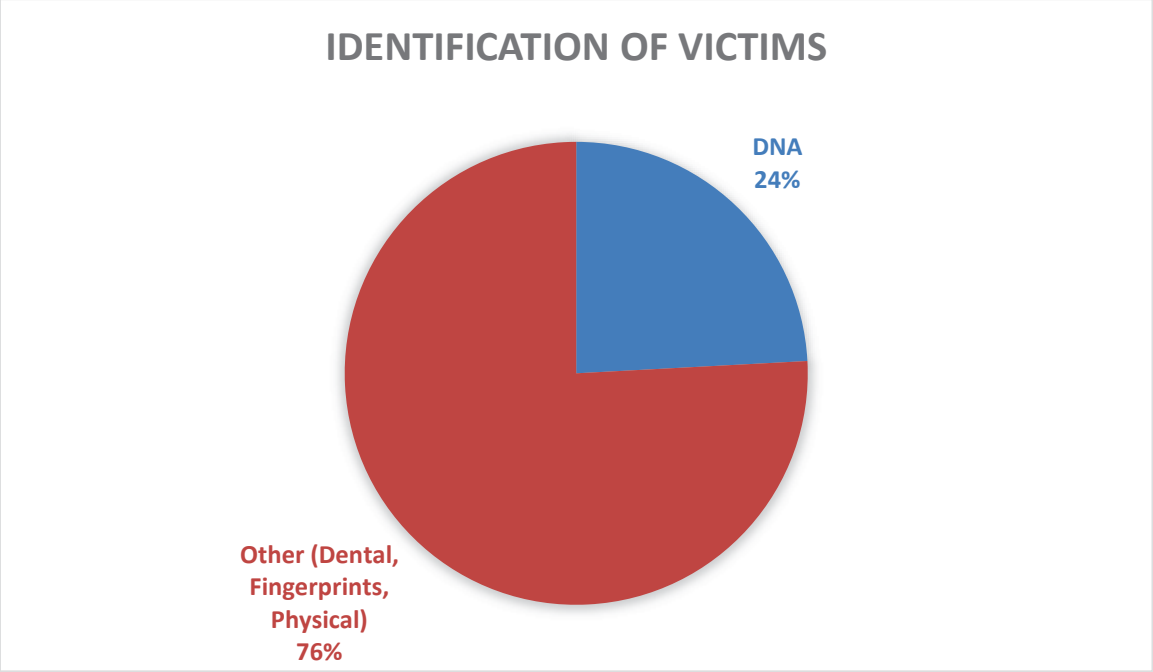


Figure 7: Identifications in the Thai-Indonesian Tsunami DVI Operation. By July 2008, 3,308 victims were identified. DNA profiling contributed to approximately 25% of identifications [1]. A total of 258 identifications were made through typing of the mitochondrial HV1 and HV2 regions.

5.1.5 Eaglesham Murder

In cases where samples are compromised with limited genomic DNA and other primary identifiers are of low probative value, typing of mtDNA can be used as an alternative means of identification. In 2002, Northern New South Wales, Australia, Petrina Eaglesham, 24, was found with her throat slashed several times in the family car. Evidence recovered from the car included bloodied fingermarks and two hair fibres clutched in the victim's hand. The fingermarks covered in Petrina's blood were indistinguishable to the reference prints provided by Petrina's estranged husband, Glenn William Eaglesham, 33. Mr. Eaglesham denied his presence at the crime scene, and his fingermarks were of low probative value as their presence was accounted for by his regular use of the family car. As limited amounts of genomic DNA are recovered from hair shafts, mtDNA typing was performed on the two hair fibres clutched in Petrina's hand for identification. The DNA results revealed that one of the two hairs was indistinguishable at every base position to Mr. Eaglesham's profile. The DNA expert testified that Glenn Eaglesham could not be excluded as the donor of one of two hairs found clutched in the victim's hand. Subsequently, he was found guilty in the case of *R v Eaglesham 2004*, for the murder of his estranged wife, Petrina Eaglesham, and sentenced to 20 years gaol with a non-parole period of 15 years.

5.2 Investigative Use

While direct comparisons of mtDNA sequences with a reference sample can assist in identification of an individual, its use can be limited when an unknown mtDNA profile is located. In these cases, there is value in acquiring intelligence about the unknown source's ancestry and ethnicity. Inference of maternal BGA from mtDNA provides investigative leads to the possible identity of suspects. In line with forensic DNA phenotyping (FDP) techniques, BGA is useful in cases where crime scene samples fail to provide a match to a known person as well as in the re-opening of 'cold cases'. There is also the use of BGA to assist in the facial reconstruction of skeletal remains and assert or refute eye witness accounts, providing investigators with leads to the possible appearance and identity of the anonym.

Currently, the preferred method of reporting DNA evidence follows a likelihood ratio [55] or Bayesian approach, which comparatively weights the DNA evidence under two hypotheses [55]. In DNA profiling, autosomal STR and non-autosomal STR results for sex as well as the relevant population database are used to present its statistics. The inclusion of maternal BGA from mtDNA would add another independent variable to the LR calculation and refine the sample to the relevant maternal ancestry haplogroup. Such use would increase the overall power of discrimination. For these calculations to be made possible, the haplogroup frequencies need to be available. This is developing through worldwide population studies such as the 1000 Genomes Project as well as the establishing of the EDNAP mtDNA Population Database (EMPOP) [56, 57]. The identification of the remains of King Richard III illustrates the use of both maternal lineage and ancestry to infer an individual's identity [50]. Although it is a historical case study, the ancestry approach can easily be applied to indigenous remains, unnamed war dead and missing persons cases.

5.3 Medical Diagnostics

Mitochondria are cellular organelles found in eukaryotic organisms. Their cellular functions include the production of energy, processing of amino acids, fatty acids, and cofactors as well as cell signaling. Polymorphisms of the mitochondrial genome have been related to variances in endurance capacity and trainability along with a number of other mitochondria disorders [58] (Table 6).

Table 6: Clinical Mitochondrial Disorders caused by DNA Mutations. Sourced from [59].

Mitochondrial Disorder	Clinical Phenotype	MtDNA Mutation
Kearns-Sayre syndrome	Progressive myopathy, ophthalmoplegia, cardiomyopathy	Single large-scale deletion
Myopathy and diabetes	Myopathy, fatigue, diabetes	14709T
Pearson syndrome	Pancytopoenia, lactic acidosis	Single large-scale deletion
MELAS	Myopathy, encephalopathy, lactic acidosis, stroke-like episodes	3243A 3271T
Sensorineural hearing loss	Deafness	1555A Individual mutations
Exercise intolerance	Fatigue, muscle fatigue	Individual mutations
CPEO	Ophthalmoplegia	Single large-scale deletion

Medical diagnosis of these disorders is comprised of genetic history (maternal), clinical symptoms, biochemical tests and targeted sequencing assays. While some clinical features are externally visible, their presentation varies between individuals and may not serve as an accurate predictor of disease. In many cases, sequencing of whole mitochondrial genomes acts as a screening tool for candidate genes and confirms disease progression [60].

6. Legal Matters

Forensic science is the application of science to assist the courts in determining the legal or criminal matter at hand. It involves the collection, preservation, analysis and presentation of scientific evidence in a court of law. This process is regulated using a set of guidelines prior to becoming admissible in criminal and civil proceedings. The *Uniform Evidence Act* 1995 (Cth) governs the admissibility of evidence across Australian jurisdictions. The act is founded on the Frye and Daubert standards, which were legal precedents set in the American cases

of *Frye v United States* [1923] and *Daubert v Merrell Dow Pharmaceuticals, Inc* [1993], respectively.

6.1 Frye Standard

The case of *Frye v United States* [1923] involved the admissibility of polygraph tests as a novel technology to infer deceptive answers based on monitoring one's physiological responses such as blood pressure and skin conductivity. The tests became highly controversial and were criticised for the lack of a sound scientific method from the respective scientific community [61]. The lack of acceptance by the scientific community ruled polygraph lie detector tests as inadmissible as expert evidence. The case established the Frye Standard, setting the precedent for the admissibility of expert evidence to only be adduced when the technique/s or technology used to arrive at the opinion has gained acceptance in the relevant scientific community.

6.2 Daubert Standard

The American trial, *Daubert v Merrell Dow Pharmaceuticals, Inc* [1993] involved the admissibility of expert evidence that the alleged morning sickness drugs, Bendectin™ (Merrell Dow Pharmaceuticals, Inc., by The Dow Chemical Company, Michigan, USA) had caused birth defects in the Daubert infants. Evidence in this case was ruled inadmissible due to a lack of a scientific basis to establish or test its claims. The ruling reformed the evidence rules previously set by the Frye case to incorporate the Daubert Standards. The Daubert Standards requires all opinions testified by expert witnesses to be based on sound scientific methods that are:

- Organised and generally recognised to be reliable
- Testable
- Peer reviewed
- Generally accepted among the relevant scientific community

Subsequently, the Frye and Daubert standards have been enshrined into Australian law under section 79 of the *Uniform Evidence Act* 1995 (Cth).

6.3 Landmark Cases

Landmark cases regarding the use and legal issues surrounding mtDNA sequencing, familial searching or kinship matching and, DNA phenotyping are reported in Table 7.

Table 7: Significant mitochondrial DNA cases [62].

Jurisdiction	Case	Significance
United States of America	State v. Ware [1999] WL 233592	First US case of mtDNA profiling. Admissibility of mtDNA profiling.
New South Wales, Australia	R v Eaglesham [2004] NSWSC 747 (27 August 2004)	Admissibility of mtDNA profiling in criminal proceedings.
United Kingdom	R v Harman (Unreported, 2003)	Operational use of familial searching or kinship matching.
United Kingdom	R v Delroy Grant (Unreported, 2011)	Operational use of Forensic DNA Phenotyping to predict race.
High Court of Australia	Aytugrul v The Queen [2012] HCA 15 (18th April 2012)	Reporting of the maternal inheritance of mtDNA and the subsequent larger observed population frequency to avoid unfair prejudicial bias.

Currently, Australian courts recognise mtDNA profiling as admissible identification evidence in cases of familial searching and kinship matching. Privacy concerns have limited the admissibility and full use of DNA phenotyping technologies with only a few jurisdictions in the US, Germany and the Netherlands using it for investigations. In Australia, DNA phenotyping is being examined and researched for use by the Australian Federal Police, Victoria Police Forensic Services Department and New South Wales Police Force while Tasmania, Northern Territory and Western Australian have not considered the technology [63]. The Australian Federal Police has established an MPS workflow for DNA samples. However this workflow is yet to be validated and there has been no updating of the law in the Australian jurisdictions [62]. This also extends to the use of MPS technologies in mtDNA analysis, especially for

compromised samples. As a novel sequencing technology, the transition of MPS platforms into forensics requires validation, peer review and recognition among the scientific community. This is necessary to establishing the Frye and Daubert standards for admissibility as DNA evidence. Few cases have yet to establish the admissibility of MPS for human identification. MPS technologies are widely used for medical, diagnostic and research purposes [50]. The success of MPS in identifying historical and ancient remains in the archaeological fields provides support for a transition of mtDNA-MPS testing in forensic cases [50]. Such tools can assist in the identification of compromised samples.

7. Improving the sequencing of Mitochondrial DNA

The use of mtDNA typing in challenged samples with little to no nDNA or in cases where STR profiles are unmatched to samples on a reference database has expanded the scope of DNA investigations. Over the years, the sensitivity to type mtDNA from degraded samples has increased from DNA fragments of several kilo bp to a couple of hundred bp (Table 8).

Table 8: Development of mtDNA typing kits for control region and whole mitochondrial genomes, using MPS technologies.

Year	Average length of degraded DNA Fragment (bp)	Reference
2013	8,500	Parson 2013 [5]
2015	380	Parson 2015 [6]
2015	200	The Early Access AmpliSeq™ Mitochondrial Panel (Applied Biosystems, CA, USA) [7]
2016	163	Precision ID mtDNA Whole Genome Panel (Applied Biosystems, CA, USA) [64]
2017	70	Eduardoff 2017 [65]

In 2013, Parson *et al.,* [5] evaluated the feasibility of sequencing complete mitochondrial genomes on the Ion Torrent PGM™ System (Applied Biosystems, CA, USA). In direct

reference to the dideoxynucleotide sequencing of partial profiles, Parson *et al.,* [5] proposed the MPS platform to be feasible to sequencing whole mitochondrial genomes from two overlapping DNA fragments of 8.5-kilo bp. By a re-design of PCR primers, mitochondrial genomes could be sequenced from hair shafts from amplicons as small as 380 bp [6]. This was reduced to a smaller amplicon size of 200 bp fragments, which overlap in a tiling design to form the whole mtDNA sequence in the Early Access AmpliSeq™ Mitochondrial Panel (Applied Biosystems, CA, USA). This tiling approach can be multiplexed to a number of samples and has been shown to be amendable to reference and historical remains [7]. The Precision ID mtDNA Whole Genome Panel (Applied Biosystems™, CA, USA) is a commercial DNA typing kit which follows on the tiling approach and includes additional degenerate primers [64].

Novel forensic DNA typing kits undergo an extensive process of organisation, evaluation and peer reviews before becoming admissible as a reliable method in the legal courts. This was evident in the first use of the Profiler Plus™ STR system in the landmark cases, *R v Karger* [2001] SASC 64 and *R v Gallagher* [2001] NSWSC 462. As such, evaluating the performances of panels for amplification and sequencing is necessary to establishing a standard of mtDNA-MPS testing in forensic genetics.

7.1 Degradation Study

Preliminary studies of the Early Access AmpliSeq™ Mitochondrial Panel (Applied Biosystems, CA, USA) shows a high sensitivity to typing degraded DNA fragments [66]. However, there are limited studies indicating the performance of the Panel in other compromised samples which are typical in forensic cases, mass disasters and historical remains. Information on the Panel's performance to different DNA qualities would better inform the processing of samples for mtDNA analysis. The primary aim of this study is to test the performance of the Panel to degrading qualities of DNA. It involves creating a set of DNA samples, which are of progressively degraded qualities, for amplification and sequencing. The quality of DNA is indicated by using agarose gel electrophoresis, DNA quantification and STR profiling of genomic DNA. The testing of the Panel focuses on evaluating the overall mtDNA sequences for coverage, variant calling, phylogenies and familial matching. These sequences are compared between different degraded samples as well as mtDNA sequences typed by CE technologies. Additionally, the workflow of the Panel is reviewed to provide a complete overview of the mtDNA-MPS testing process and its feasibility to operational laboratories.

The results of this study provide a technical guide to the Panel's performance in amplifying DNA samples of various degrading qualities for sequencing on MPS platforms. It is useful for

forensic laboratories which are planning to transition to MPS technologies for DNA testing. Furthermore, the technical work contributes to establishing the forensic use of MPS technologies as a reliable method of mtDNA sequencing.

8. Quantification of Mitochondrial DNA

The quantification of mtDNA determines the amount of 'amplifiable' DNA that is available in an extract (Figure 2). This process optimises the quantity of DNA, which is added to DNA amplification and contributes to its overall success. The under-amplifying of DNA can lead to allele dropouts, heterozygote imbalance and low sequence coverages. While the excess amplification of DNA can introduce stutter peaks, non-specific amplification and reduce the overall representation of coverages between samples [67]. Methods of quantifying mtDNA are limited to estimations provided by nuclear DNA [68]. These estimations are limited as a high variability exists between the copy numbers of the mitochondrial and nuclear genome [8]. While there are many multiple copies of mtDNA in each cell, there is only one copy of nuclear DNA in each cell. Such variability can lead to the under- or over estimation of mtDNA. As a result, there is need to develop a method which is specific to the quantification of mtDNA. There are real-time assays which are available for the quantification of mtDNA [8, 69-71]**.** While these assays can quantify the mitochondrial genome, its primers may also co-amplify sequences in nuclear DNA and result in quantities, which are over-reported. The aim of this study is to address the limitations in quantifying mtDNA by developing a DNA assay, which is specific to the quantification of mitochondrial genomes. The testing of this assay on DNA samples of degrading qualities indicates a robust performance that is sensitive and reproducible for samples, which are typical in mitochondrial cases.

The use of this assay can lead to improving the quantifications of mtDNA workflows. Currently, the workflow of the Early Access AmpliSeq™ Mitochondrial Panel (Applied Biosystems, CA, USA) recommends estimating of mitochondrial quantities from nuclear DNA quantities for its amplifications. The use of the assay developed in this study, which specifically quantifies mtDNA content may lead to improving the overall amplification and workflow of the Panel. Furthermore, potential multiplexing of the mitochondrial assay with other DNA targets would further expand the quantification and assessment of DNA samples and inform the optimal downstream processes.

9. Aims and Hypothesises

Overall, it has been found that mtDNA testing is useful in investigations involving unmatched STR profiles or when genomic DNA is limited. Although the probative value is not as discriminating as genomic DNA profiling, information such as maternal lineage and ancestry can be useful in cases of mass disasters, missing persons and historical remains when there are no references for first-degree relatives. A review of mtDNA testing reveals a progression and research of sequencing technologies from CE to MPS platforms. This transition is leading to the introduction of MPS panels, which are capable of amplifying samples of smaller fragments. This has accommodated a greater sensitivity to sequence mtDNA from degraded samples as well as expanded the feasibility of sequencing complete mitochondrial genomes. While the development of MPS panels continue to develop, the use of MPS technologies in forensic DNA testing is relatively novel, with a limited number of laboratories using the technology. As a result, the testing of MPS panels to indicate its forensic uses remain limited. Importantly, the performance of panels to different samples of different qualities are unknown. Hence, the aim of this project is to evaluate the robustness of the Early Access AmpliSeq™ Mitochondrial Panel (Applied Biosystems, CA, USA), focusing on its performance to amplify a set of compromised samples for sequencing on MPS platforms. This technical work contributes to the quality management and assurance of the Panel and the establishing of mtDNA testing using MPS technologies.

In addition, a review of the mtDNA testing process reveals methods for quantification of mtDNA are lacking. In particular, many methods for mtDNA quantification rely on the estimating of mtDNA quantities from nuclear DNA. These estimations have been reported to be unreliable as the copies of mtDNA are highly variable compared to the two copies of nuclear DNA. As a result, another aim of this project will be to address this limitation by developing a method of quantification that is specific to the mitochondrial genome of compromised samples. The development of such an assay contributes to the overall amplification process in mtDNA testing.

To summarise, the aims of this project are to:

1. Conduct a degradation study to create a set of DNA samples of progressively degraded qualities for testing (Chapter 2).
2. Develop a qPCR assay for mtDNA to overcome the limitations of estimating mtDNA quantities from genomic DNA (Chapter 3).
3. Perform genomic DNA Profiling to assess sample degradation and its suitability for mtDNA sequencing (Chapter 4).

4. Perform mtDNA sequencing with the Early Access AmpliSeq™ Mitochondrial Panel (Applied Biosystems, CA, USA) to test its performance (Chapter 5).

Chapter 2: Degradation of DNA

1. Introduction

By nature, DNA can be damaged and degraded when exposed to a variety of environmental insults. While damage can be natural from oxidative stresses in the body, DNA damage can also occur from biological and physical factors that are commonly encountered in forensic environments. The major degrading factors present in the natural environment include heat, chemicals such as household cleaners, microbial activity and UV radiation (Table 1). While the mechanism of degradation may differ, the end result typically fragments DNA to the extent it becomes unsuitable for forensic applications. Fortunately, advancements in PCR and MPS technologies have expanded the analysis of compromised and aged DNA samples [43, 72-74]. Kits using these technologies are considered to be essential for STR identification in forensic biology laboratories. Even so, the release of novel kits, especially those using MPS technologies are yet to be fully evaluated on samples that are degraded to a similar casework quality [7].

In response, the proposed degradation study aims to create a set of progressively degraded DNA samples that are representative of qualities encountered in casework scenarios. It is expected that in later studies, these samples could be evaluated for genomic DNA quality in STR profiling as well as mtDNA quality using Ion Torrent™ sequencing (Applied Biosystems, CA, USA).

It has been examined that the extent of degradation is controlled by two variables [16]:

1. The duration of exposure to the factor
2. The amount exposed to the factor

Hence, in order to create a set of progressively degraded DNA samples, degradation factors such as sonication and heat treatment were exposed to DNA samples at variable timeframes. Evaluation of factors as a potential degradation method involved fragment and DNA assay analysis.

2. Method

The methodology used to create a set of progressively degraded DNA samples involved the collection of DNA samples that were extracted and quantified. Subsequently, these biological samples were degraded *in vitro* by either heat or sonication at various time intervals. It is noted that the methods employed in this study, closely reflect procedures in operational Forensic DNA laboratories that have been accredited by the National Association of Testing Authorities, Australia [66]. This included the handling of DNA samples in separate pre- and post- DNA amplification areas (where applicable) with dedicated Topsafe™ 1.2 Class II biosafety

cabinets (EuroClone®, Milan, Italy) and protective materials such as gloves and laboratory coats. Prior to use, biosafety cabinets and materials were decontaminated using 10% bleach, followed by 70% ethanol and UV for a minimum of 10 minutes to reduce the introduction of extraneous DNA, especially mtDNA. All experiments were carried out at the University of Technology Sydney. All statistical analysis was completed using Microsoft Excel 2010 (Microsoft Corporation, WA, USA) and Minitab® 17 Statistical Software (Minitab®, PA, USA).

2.1 DNA Collection

Human ethics approval was obtained from the UTS Human Ethics Committee (Approval Number: 2015000296) for the collection of DNA samples from volunteers. Volunteers were informed of ethics approval and consented to providing a buccal swab for the study. A total of 26 volunteers were recruited for the study with each volunteer swabbed twice using cotton swabs (Classiq Swabs™, Copan Diagnostics, CA, USA). The collection procedure involved volunteers swabbing the inner side of their mouth cheek for 30 seconds, providing an abundant amount of buccal cells for DNA extraction. Swabs were collected and stored at -20°C until DNA extraction. Additional, participants provided ancestry details of their maternal and paternal lineages on consent forms as reference information for evaluating mtDNA sequences for haplogroup ancestries. A sample consent form has been attached in Appendix I.

2.2 DNA Extraction

DNA was extracted using the QIAamp® DNA Mini Kit (Qiagen, Hilden, Germany), following the manufacturer's recommended protocol for buccal swabs [75]. The QIAamp® DNA Mini Kit (Qiagen, Hilden, Germany) uses a silica-based method to purify a range of nucleic acids (including mtDNA) from its cellular components. Swabs were buffered in phosphate-buffered saline and proteinase K was added to lyse cellular organelles. Following this, DNA was bound to a silica-based membrane and cellular contaminants were eluted using two stages of wash buffers. Cellular contaminants include proteins, carbohydrates, lipids and any other exogenous contaminants. Once purified from these contaminants, DNA was eluted and stored at -20°C until it was ready to be quantified.

2.3 DNA Quantification

DNA samples were quantified using the Quantifiler™ Trio DNA Quantification Kit (Life Technologies, CA, USA) on the QuantStudio™ 6 Flex Real-Time PCR System (Applied Biosystems, CA, USA). The Quantifiler™ Trio DNA Quantification Kit (Life Technologies, CA, USA) uses real time PCR technology for direct detection of DNA amplification in the

exponential growth phase. Validation studies of the assay has been shown it is able to reliably quantify between 50 to 0.005 ng/µL amounts of DNA [76]. In 11 µL reaction volumes, Quantifiler® Trio Primer Mix (2.5X) and Quantifiler® THP PCR Reaction Mix (2X) was combined with 2 µL of DNA samples. For quantification, PCR reactions were amplified in 'Fast' mode: 2 minutes at 95°C followed by 40 cycles of 9 seconds at 95°C and 30 seconds at 60°C on the QuantStudio™ 6 Flex Real-Time PCR System (Applied Biosystems, CA, USA) [77]. A single reaction quantifies a small (80 bp) and large (214 bp) autosomal target as well as a Y-chromosomal target (when present). Analysis of all targets was performed using the system's supplied software, QuantStudio™ Real-Time PCR Software v1.2 (Applied Biosytems, CA, USA). As recommended by the manufacturer, the small autosomal target was used to determine sample quantity given its abundance in DNA samples [77].

2.4 Degradation

Quantified extracts were heated at either 99°C or 125°C for increasing increments of time. Specifically, equal aliquot volumes of sample DNA were suspended into microfuge tubes and heated in a dry heat block at various timeframes, before being removed. Once removed, DNA was resuspended in TE buffer and stored at 4°C for degradation assessment.

Likewise, equal volumes of DNA samples were aliquoted into microfuge tubes and sonicated in a sonication bath (Branson 2210R-MT Ultrasonic Cleaner, Branson Ultrasonics, CT, USA) for 0, 30, 60, 90 and 120 minutes. Once respective times had elapsed, samples were removed and stored at 4°C until assessed for quality.

To assess degradation, samples were re-quantified with the Quantifiler™ Trio DNA Quantification Kit (Life Technologies, CA, USA). In addition to providing quantities, the Quantifiler™ Kit assesses sample quality using a degradation index of small to large autosomal targets (ng/µL). Specifically degradation is calculated as follows:

$$Degradation\ Index\ (DI) = \frac{small\ autosomal\ concentration}{large\ autosomal\ concentration}$$

Based on the DI, DNA sample quality can be classified into four categories: non-degraded (DI, 0 – 1.5), mildly degraded (DI, 1.5 – 4), degraded (4 – 10) and severely degraded (DI, > 10) [78].

Furthermore, sample quality was evaluated for fragment size using gel electrophoresis and the Agilent 2100 Bioanalyser (Agilent Technologies, CA, USA). For electrophoresis, treated extracts were run on a 1.5% agarose gel at 60V for 5 minutes then 100V for 40 minutes and detected under the InGenius3 manual gel documenter (Syngene™). For Bioanalyser analysis,

DNA extracts were prepared onto High Sensitivity DNA Chips and electrophoresed on the Agilent 2100 Bioanalyser (Agilent Technologies, CA, USA). Electropherograms produced from the Bioanalyser were analysed using the manufacturers supplied software [79].

3. Results

The assessment of suitable degradation methods was evaluated by visualisation on gel electrophoresis and a real-time DNA assay. DNA degradation is detected as the transition of DNA bands to smears.

3.1 Heat Treatment

To investigate the effects of heat on DNA, heat-treated DNA samples were run on agarose gels. At 99°C, smearing of fragments was noticed in samples heated between 5 and 30 minutes (Figure 8). With respect to the 50 bp hyperladder standard, all heat-treated fragments were consistently located in the larger bp regions. At a higher temperature (125°C) and longer treatment time (5, 10, 20, 40 and 80 minutes), smear fragments were shown to become more degraded as band intensities further migrated down to the lower bp regions (Figure 8). This trend was shown to be more evident with longer treatment times (Figure 9).

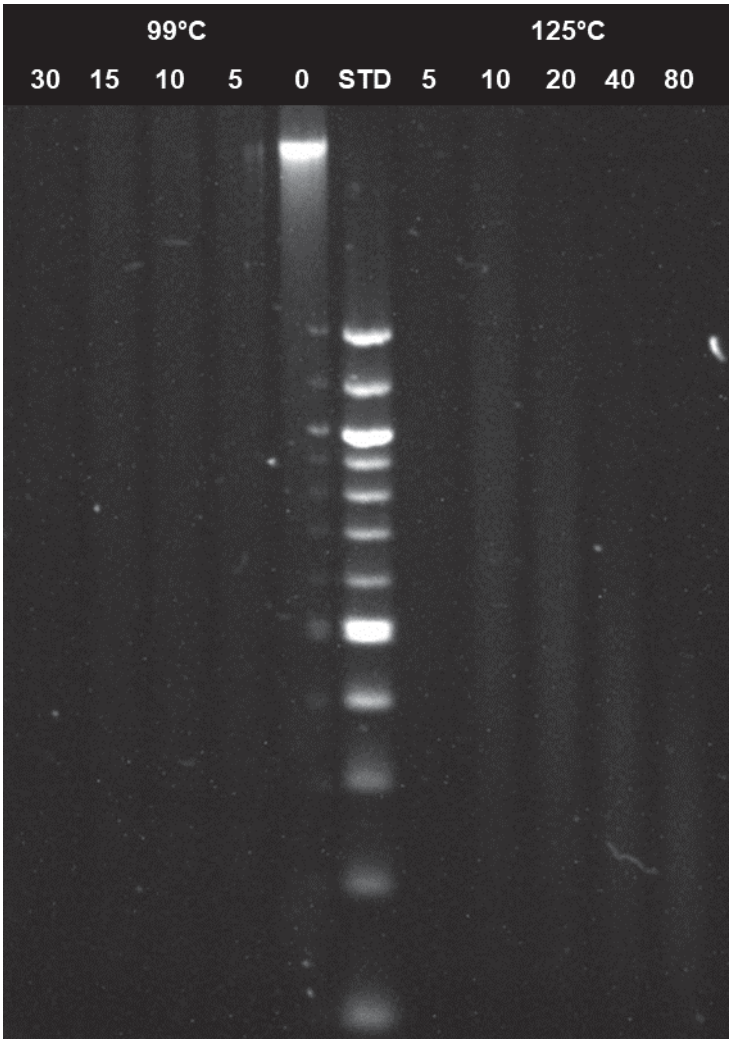


Figure 8: Agarose Gel (1.2%) of Heat Treated DNA Samples. One set of DNA extracts were heated at 99°C for 0, 5, 10, 15 and 30 minutes. Another set of extracts were heated at 125°C for 5, 10, 20, 40 and 80 minutes. Standard (STD) displayed is a 50 bp hyperladder. Electrophoresis was run at 60 V for 5 minutes followed by 100 V for 60 minutes.

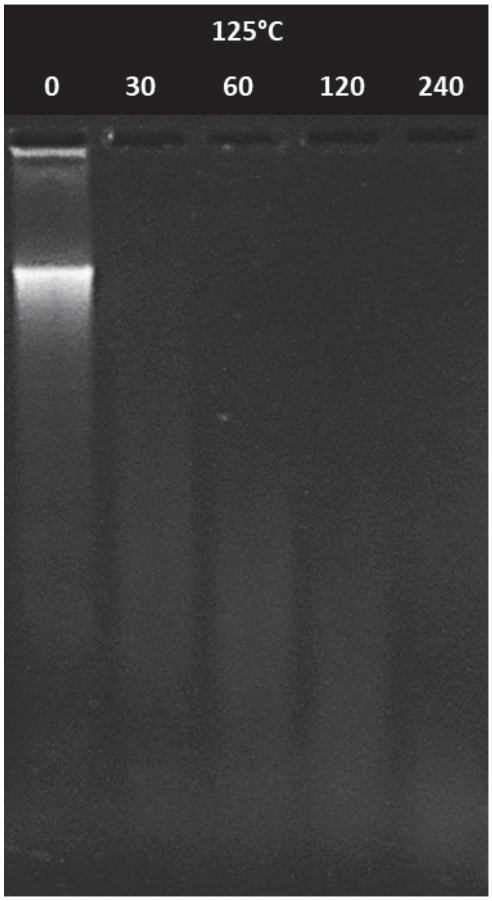


Figure 9: Agarose Gel (1.5%) of Heat Treated DNA Samples. DNA extracts were heated at 125°C for 0, 30, 60, 120 and 240 minutes. Electrophoresis was run at 60 V for 5 minutes followed by 100 V for 60 minutes.

The DNA fragments of heat-treated samples were quantified to indicate the extent of fragments created from the heating process. In untreated control samples (0 minutes), equal quantities at 4 ng/μL were observed for both small and large autosomal targets (Figure 10).

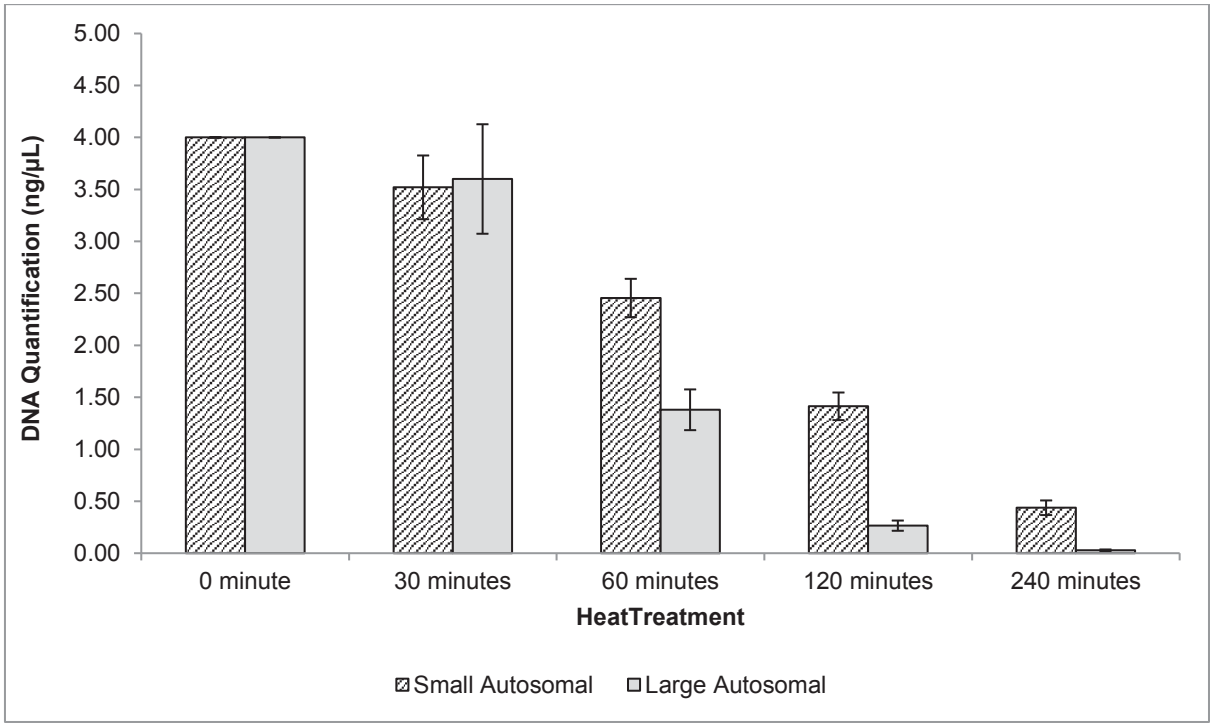


Figure 10: Determination of DNA Concentration in Heat Treated DNA Samples. DNA concentrations (ng/µL) of small autosomal (SA) and large autosomal (LA) targets for each respective treatment group were combined and plotted as an average mean (*n*= 5). Statistical significance ($p < 0.05$) from controlled treatment group, 0 minutes was recorded for treatment groups 30 to 240 minutes. Error bars represent standard error of mean (S.E.M) between replicates.

The mean ± S.E.M concentration (ng/µL) of small autosomal fragments were recorded at 3.520 ± 0.686 for 30 minutes, 2.455 ± 0.413 for 60 minutes, 1.413 ± 0.297 for 120 minutes and 0.4371 ± 0.1572 for 240 minutes. The decrease in mean concentration (ng/µL) for small autosomal targets was measured to be significant across all treatments (ANOVA one way, *p*= 0.000). For large autosomal fragments, the mean ± S.E.M concentration (ng/µL) was measured at 3.600 ± 1.178 for 30 minutes, 1.380 ± 0.438 for 60 minutes, 0.2648 ± 0.1099 for 120 minutes and 0.02773 ± 0.01806 for 240 minutes. In a similar manner, a reduction of large autosomal targets was reported with prolonged treatment times (ANOVA one way, *p*= 0.000). In all cases, it was observed that small autosomal quantities were consistently greater than large autosomal quantities. The only exception was the treatment group at 30 minutes where there was a greater number of large autosomal targets than small autosomal targets (Figure 10). It should be noted that attempts were made to further quantify the mean size (bp) and concentration (ng/µL) of degraded fragments via the Agilent 2100 Bioanalyser (Agilent Technologies, CA, USA). However, it was found that the detection of fragments was beyond the sensitivity of the instrument (data not shown).

Nevertheless, by comparing relative quantities of the two autosomal targets (obtained from qPCR), sample degradation was assessed and reported as a degradation index (DI). The mean ± S.E.M DI (small autosomal concentration / large autosomal concentration) of extracts were calculated to be 0.75 ± 0.03 at 0 minutes, 1.02 ± 0.07 at 30 minutes, 1.74 ± 0.21 at 60 minutes, 5.64 ± 0.67 at 120 minutes and 18.18 ± 3.22 at 240 minutes (Figure 11).

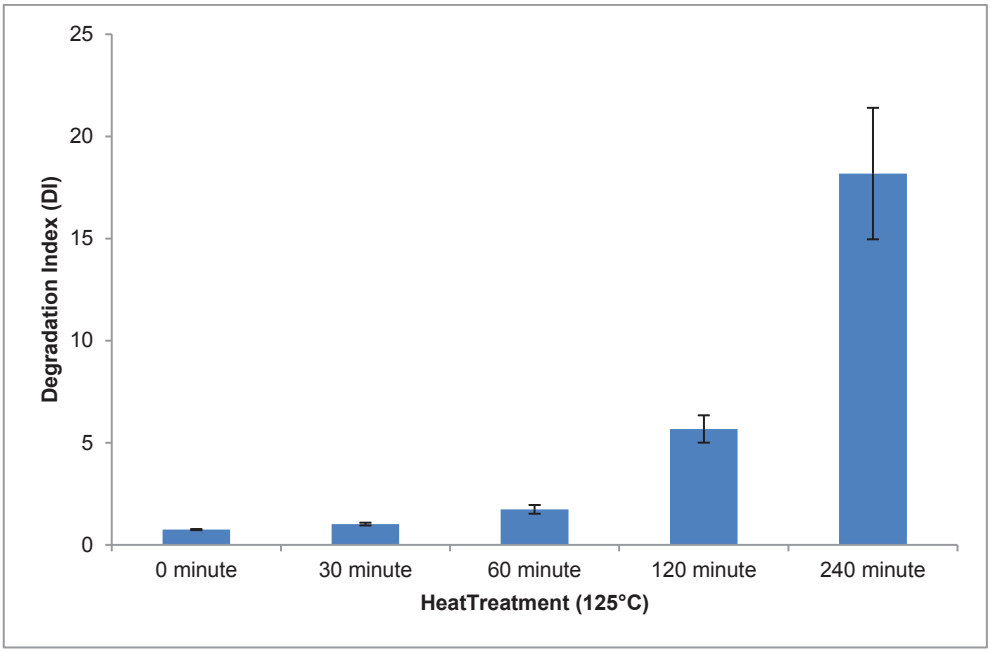


Figure 11: Quality of genomic DNA in heat-treated DNA extracts. Mean (*n*=5) degradation index (DI) of heat-treated DNA samples determined using the Quantifiler™ Trio DNA Quantification Kit (Life Technologies, CA, USA). DI calculated as ratio, small autosomal target/large autosomal target [15]. Error bars represent S.E.M. between replicates.

As observed, the increase in DI indicates a progressive reduction in sample quality for prolonged heating times. With respect to negative controls (0 minutes), significantly higher indexes were shown in treatment groups at 120 minutes (p= 0.003) and 240 minutes (p= 0.008), indicating DNA samples were highly fragmented and severely reduced in quality.

3.2 Sonication

To assess the suitability of sonication as a degradation method, sonicated samples were electrophoresed and visualised on agarose gels. As shown in Figure 12, a clear band of high molecular weight DNA can been seen in untreated control samples at 0 minutes, indicating intact double stranded DNA of high quality. In a similar manner, DNA bands were detected for sonicated samples, showing that sonication had no to very little effect on DNA integrity (Figure 12). The only exception was seen at 5 minutes of sonication, with the particular treatment displaying no to a very faint DNA band.

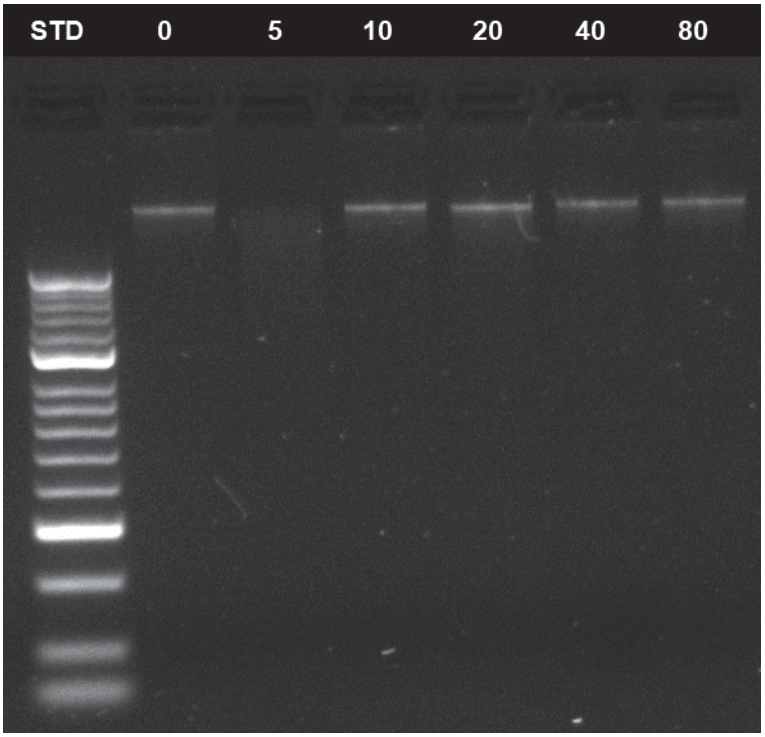


Figure 12: Electrophoresis of Sonicated DNA Samples. DNA extracts were sonicated in a bath for 0, 5, 10, 20, 40 and 80 minutes. Electrophoresis was run at 60 V for 5 minutes followed by 100 V for 60 minutes. The displayed standard (STD) is a 50 bp hyperladder.

4. Discussion

The aim of this degradation study was to create a set of progressively degraded DNA samples that are representative to those encountered in casework. These DNA samples are later used in evaluating the robustness of the novel DNA markers for mtDNA quantification and amplification. Two degradation methods were trialed, specifically heat treatment and sonication. DNA integrity was assessed using agarose gel electrophoresis and a real time assay for quantification of DNA fragments. Agarose gel electrophoresis is a commonly applied method to visualise DNA quality and quantity. Electrophoresis separates DNA based on size, conformation and charge, with smaller DNA fragments migrating furthest. Intact DNA molecules are observed as discrete and clear bands, while degradation weakens band intensity and increases migration rate into the lower molecular weight regions.

The structure of DNA is composed of a nitrogenous base bonded to a deoxyribose sugar and phosphate backbone [80]. Exposure to high temperatures has been reported to degrade nucleic acids into fragments due to the depurination of DNA involving the hydrolysis of DNA glycosidic bonds, followed by phosphodiester cleavage [15]. For this reason, the recovery of DNA becomes limited in hot temperate environments [14, 15, 17]. It should be noted that degradation is different to denaturation, which is the rendering of double stranded DNA strands

to single strands using heat (typically at 95°C). As shown in Figure 8, the discrete band of heat-treated extracts gradually weakens into a series of smaller fragments, indicating the degradation of DNA. The presence of smear fragments becomes more evident for samples heated at higher temperatures and for longer periods (Figure 8). At 99°C, degradation smears were fairly consistent between treatments, while at 125°C smear fragments became progressively more degraded with faster migration into lower sized regions (Figure 8). It is proposed that at 99°C, DNA was still relative stable as previous studies have found DNA to be stable below 100°C [16]. In agreement with these studies, the findings confirm that temperature and treatment time are variables, which affect the extent of thermal degradation in nucleic acids.

While previous studies did not observe DNA degradation until 130°C, degradation was observed in this study for samples heated at 125°C [16]. Albeit, the temperature was 5°C lower than previous methods, it must be considered that the current samples were heated for longer time periods and thereby may account for the observed difference [16]. Indeed when heated for longer periods of time, DNA smears were shown to become gradually lighter in band intensities whilst also migrating further into lower sized fragments (Figure 9). Accordingly, the findings of this study conclude heating time as a significant variable in the progressive degradation of heated DNA and support that of previous studies [15].

The determination of the range of degraded fragment sizes from standards was shown to be variable and arbitrary at times. This was largely due to its estimation by visual means. For this reason, other methods of estimating fragment sizes were used including densitometry and the Agilent 2100 Bioanalyser (Agilent Technologies, CA, USA). However, both methods displayed poor sensitivity to fragment analysis (data not shown), requiring further alternatives methods. Accordingly, the extent of degraded fragments were assessed using a qPCR assay that quantifies a small autosomal and large autosomal target, and reports degradation as relative index. Evidently displayed in Figure 10, the equal concentrations (ng/µL) of small and large autosomal targets in untreated samples at 0 minutes, indicates an expected quality of intact double stranded DNA. In contrast, the heating of DNA samples reports an increase of small autosomal fragments and a decreasing of large autosomal fragments (Figure 10). This marked response was significant ($p < 0.05$) and reveals fragmentation of DNA fragments into smaller fragments.

Additionally, degradation index displayed as a relative ratio of the two fragment quantities showed no to very little degradation in unheated controls (Figure 11). Degradation indexes were shown to deviate to less than five degradation units for treatment times from 0 to 60 minutes, while treatments times greater than 60 minutes reported degradation indexes above

five units with more variable S.E.M, indicating putative breakdown of DNA molecules. As heating progressed, degradation indexes were shown to increase with significance ($p < 0.05$) recorded at 120 and 240 minutes, showing a continuum of reduced sample quality. At a maximum index of 7.31 for 120 minutes and 21.40 for 240 minutes, these samples would be classified as degraded and highly degraded for forensic applications such as STR profiling [78].

Note that the statistical values reported here are based on the minimum sample number for validation studies, being 5 biological replicates [81]. Operational laboratories frequently use 50 replicates to represent a more precise estimate of variation. Thereby it might be beneficial for additional samples to be included in future degradation studies as a means to reflect the precise variation of the procedure. A further limitation is noted in that the real time assay used to report quantities and DIs are based on genomic DNA [76]. As a result, assessment of degradation on non-autosomal targets such as mtDNA which is circular and more stable is left unaccounted. To gain a reliable assessment of sample quality, it would be beneficial to use an equivalent counterpart assay for non-autosomal targets such as mtDNA. Such an assay is developed and tested in later studies. Despite these noted limitations, the results of this study propose that above unstable temperatures, the quality of DNA begins to exponentially degrade with heating time. In support of previous studies, it is further concluded that temperature and time are variables that can significantly affect the integrity of DNA [14, 15, 17].

In contrast to heat treatment, sonication is the application of ultrasonic high frequency waves to molecules. When applied to nucleic acids, cavitation forces are generated, which breaks the phosphodiester bonds and results in a range of DNA fragment sizes. For this reason, sonication is commonly used to produce appropriate sized libraries for MPS (300 to 500 bp) and accordingly was considered as a degradation method [82].

The electrophoresis results for DNA samples placed in a sonication bath are displayed in Figure 12. Bright and discrete nucleic bands are seen between controls and treated samples indicating no degradation had occurred (Figure 12). This was unexpected as previous studies have reported sonication to efficiently prepare a set of degraded human DNA samples to defined fragment lengths [18]. However, experiments were repeated numerous times and at variable time frames but still had no observable signs of degradation. The difference may be accounted for by the mode of sonication used – being a sonication bath that releases frequencies throughout the bath as opposed to a sonication probe. It is plausible, that the original frequency waves generated in the sonication bath had gradually dispersed to the extent cavitation forces were unable to be produced within DNA solutions. For future studies, it is recommended that a probe based sonication method is adapted, such as Covaris, as

cavitation forces are directly introduced into the DNA solution. Accordingly, whilst sonication may be commonly used to fragment DNA in libraries for MPS, the sonication bath method used here was unsuitable to create a set of degraded fragments as opposed to the thermal method.

There are few studies which have looked at thermal degradation as a potential method to preparing sequencing libraries. Most commonly, DNase, sonication (Covaris) or PCR amplification are used to produce fragment libraries. Yet the thermal results presented here show a potential preparation of DNA libraries using heat-treatment. In theory, if the relationship between thermal degradation and DNA fragment size can be modelled, then it may be possible to prepare MPS libraries using heat-treatment.

In conclusion, the two methods of heat treatment and sonication have been discussed and evaluated on its suitability to degrade DNA. Of the two methods tested, heat treatment was chosen as it was a method that showed nucleic acids to degrade into fragments with treatment time and was also a variable that could be optimally controlled. Furthermore, temperature is inevitably an abiotic factor of all environments where DNA may be deposited and thereby selection would relate to samples recovered from forensic cases. Overall, the degradation study created a set of progressively degraded samples by heating equal sample volumes at 125°C for 0, 30, 60, 120 and 240 minutes. It is proposed that these samples are reflective of casework qualities of various categories: non-degraded, mildly degraded, degraded and severely degraded. In extension to the study, it is expected that these prepared DNA samples will be usable for downstream forensic processes such as genomic DNA profiling and mtDNA testing using MPS technologies, in later studies.

Chapter 3: Quantification of Mitochondrial DNA

This chapter titled ‘Quantification of Mitochondrial DNA’ reports on the development and preliminary testing of a real-time assay. The real-time assay quantifies the copy number of mitochondrial DNA in reference and heat-treated DNA samples. It is prepared as a manuscript for submission to journals. It is noted that the assay was developed latter to the rest of the studies and is discussed retrospectively.

Development of the MitoQ assay as a real-time quantification of mitochondrial DNA in degraded samples

Ka Tak Wai[1] · Peter Gunn[1] · Mark Barash[1]

[1]University of Technology Sydney, Centre for Forensic Science, Ultimo NSW 2007, Australia

Ka Tak Wai
Ka.T.Wai@student.uts.edu.au
ORCHID: 0000-0003-0109-4048

Abstract Mitochondrial DNA is a reliable genetic material for identifying maternally related haplogroups and ancestries. Study of the human mitochondrial genome is particularly useful when nuclear DNA is degraded or limited as the copy number of mitochondrial DNA is far greater than the copy number of nuclear DNA. Normal mitochondrial DNA copy number has been estimated to 100 copies per buccal epithelial cell, 4000 copies in skeletal cells and 7000 copies in myocardial cells. It is essential to reduce this variability and accurately quantify the exact copy number of mitochondrial DNA, especially in compromised samples of a forensic or ancient nature. While useful, the testing of mitochondrial DNA is often long, costly and comes with limited success. The accurate quantification of mitochondrial DNA using quantitative PCR assays can be used to better inform the success of amplifications and quantities required for downstream testing. As a result, this study develops a real-time assay for the quantification of mitochondrial DNA copy number and assesses its performance on a set of degraded DNA samples. This assay has been shown to be highly specific for the human mitochondrial genome with no amplification of nuclear pseudogenes being observed. Additionally, a high sensitivity was measured to 280 copies of mitochondrial DNA. Minimal variation was observed between cycles of replicates, indicating the assay to be robust and reproducible. Overall, this study presents a real-time assay that is sensitive and robust to quantifying mitochondrial DNA copy number in degraded samples. Furthermore, there is potential to incorporate the assay as an additional target in current qPCR assays which use a 6-dye chemistry and provide a complete overview of a sample's quality and quantity.

Keywords: DNA quantification; Mitochondrial DNA; Forensic assay; Copy number; Degraded DNA; Real-time PCR

Acknowledgements This research is supported by an Australian Government Training Program Scholarship.

Introduction

The accurate and reliable quantification of nucleic acids underpins the success of many PCR-based applications including forensic DNA profiling and sequencing. Under- amplification can lead to allele drop outs, heterozygote imbalances and low sequence coverages. Conversely, over- amplification can increase stutter peaks, introduce non-specific amplification and reduce overall coverage representation of sample sequences [1]. Numerous methods exist quantifying nucleic acids, including absorbance, fluorescent dyes and qPCR [2]. While each method has its limitations and advantages, qPCR usually provides quantifications, which are highly sensitive, specific and reproducible. There are commercial kits which are readily available to quantitate nucleic acids, though most only target the nuclear genome. As a result, the quantification of non-autosomal DNA has been mostly relative to estimations of nuclear DNA [3]. For mitochondrial DNA (mtDNA), these estimations have been shown to be unreliable at times and require a more specific method of detection [4, 5]. In response, there have been a number of *in-house* qPCR assays, which have been developed for the quantification of mtDNA [4, 6-8]. However, in many cases, it was found that these assays were also targeting pseudogenes existing as nuclear inserts of mitochondrial DNA (NUMTs) in the non-coding regions of nuclear DNA [9]. While these assays are human specific, a high sequence homology with nuclear DNA may also allow the co-amplification of DNA from nuclear and mitochondrial genomes and lead to an overestimation of mtDNA quantities.

As a result, an assay that is genome- and species- specific is required for the accurate quantification of human mtDNA. Limited studies have yet to develop such an assay for forensic or degraded DNA samples [10, 11]. One study has shown it is able to amplify and sequence the complete human mitochondrial genome and avoid the co-amplification of mitochondrial-like sequences in the nucleus [11]. This assay uses long-range PCR primers to amplify intact mitochondrial genomes of approximately 2000 bp overlapping DNA fragments. In most cases, the long-range amplification of DNA regions can be readily achieved in reference or pristine DNA samples. However, amplification would be variable in forensic and ancient DNA samples due to its fragmented or degraded nature. Accordingly an assay targeting a reduced amplicon of mtDNA is required for the quantification of forensic DNA samples.

In response, this study designs MitoQ as a real-time assay for the quantification of human mitochondrial genomes in artificially degraded samples. The early testing of this assay shows it features a high sensitivity and specificity in estimating mtDNA copy number, which is comparable to the performance of other real-time assays.

Materials and methods

DNA samples

DNA samples from five individuals were quantified in this study. The DNA samples were collected from the buccal epithelial cells of saliva, by rubbing the oral mucosa of participants with cotton swabs (Classiq Swabs™, Copan Diagnostics, CA, USA). Samples were collected with informed consent, under ethical procedures approved by the Human Ethics Committee for the University of Technology Sydney (Approval Number: 2015000296). DNA was extracted using the QIAamp® DNA Mini Kit (Qiagen, Hilden, Germany), following the manufacturer's protocol for buccal swabs [12]. An equal volume of each biological extract was heat-treated at 125°C for 30, 60, 120, 240 minutes. Further details of heat-treatment have been reported [13]. A reference sample for each individual was left untreated as a control for artificial degradation (total *n*=25). The quality of each DNA sample was assessed by agarose gel electrophoresis and the Quantifiler™ Trio DNA Quantification Kit (Life Technologies, CA, USA) on the QuantStudio™ 6 Flex Real-Time PCR System (Applied Biosystems, CA, USA) following the manufacturer protocols [14]. Additional controls included human sperm samples differentially extracted as nuclear DNA controls using Chelex-100 (Bio-Rad Laboratories, CA, USA) [11] and a no template control (NTC). The quality of sperm extracts as a control for nuclear DNA was confirmed with STR profiling, using the SureID® Compass Human DNA Identification Kit, following kit recommendations (Health Gene Technologies, Ningbo, China). Standard de-salted primers (Sigma-Aldrich, Sydney, Australia) were evaluated in a PCR run with positive-mitochondrial samples and negative controls. The HotStarTaq® *Plus* Master Mix Kit (Qiagen, Hilden, Germany) was used for PCR amplification, following manufacturer's instructions [15]. Amplification was carried out for 5 minutes at 95°C followed by 35 cycles of 30 seconds at 94°C, 30 seconds at 55°C, 1 minute at 72°C and a final extension of 10 minutes at 72°C on the Veriti® 96-Well Thermal Cycler (Applied Biosystems, CA, USA). PCR products were run on a 2% agarose gel for 5 minutes at 60V and 40 minutes at 100V, with molecular marker, HyperLadder™ 50bp (Bioline, London, UK). Products were visualised on the InGenius^3 system (Syngene, Cambridge, UK).

DNA standards

Synthetic DNA standards were used to calibrate concentrations of the qPCR assay. Two complementary oligonucleotides were synthesised and purified via reverse-phase cartridge purification (Sigma-Aldrich, Sydney, Australia). The DNA sequences correspond to positions 10,128 to 10,197 of the revised Cambridge Reference Sequence (rCRS) for human mitochondrial genomes [16, 17] (Table 1).

Table 1 DNA sequences of qPCR standards, primers and probe. The target sequence corresponds to a 70 bp region of the mitochondrial genome located between 10,128 to 10,197 rCRS

qPCR standard (forward strand)
5'-CTA CCA CAA CTC AAC GGC TAC ATA GAA AAA TCC ACC CCT TAC GAG TGC GGC TTC GAC CCT ATA TCC CCC G-3'
qPCR standard (reverse strand)
5'-CGG GGG ATA TAG GGT CGA AGC CGC ACT CGT AAG GGG TGG ATT TTT CTA TGT AGC CGT TGA GTT GTG GTA G-3'
Forward primer
CTA CCA CAA CTC AAC GGC TA
Reverse primer
CGG GGG ATA TAG GGT CGA A
Probe
Cyt 5.5 CCA CCC CTT ACG AGT GCG GC Iowa Black® RQ-Sp

Forward and reverse template strands were separately reconstituted in TE buffer, up to 100 µM. The primary stock solution was made up of 100 µM of dsDNA and was generated by combining equimolar volumes of the respective forward and reverse strands. The quantity of the stock solution was confirmed using the Qubit™ dsDNA HS Assay Kit (Life Technologies, CA, USA) on the Qubit® 2.0 Fluorometer, following the manufacturer guidelines (Invitrogen, CA, USA) [18]. Secondary stocks were generated by TE dilution to 2.8E+08 copies/µL. From the secondary stock, a working dilution series (2.8E+08 to 2.8E+02 copies/µL) was used for the quantification of standards. Aliquots of the primary and secondary stocks were stored at -20°C. The dilution series was prepared monthly and stored at 4°C.

Real-time qPCR

The MitoQ assay was run in a PrimePCR™ format with the SsoAdvanced™ Universal Probes Supermix (Bio-Rad Laboratories, CA, USA) [19]. In 20 µL reaction volumes, the PrimePCR assay (20X) was combined with SsoAdvanced Probe Supermix and 2 µL template DNA or control. The target probe was Cyt 5.5 and reactions were run in triplicate for standards and duplicate for samples on the CFX 96 Touch™ Real-Time PCR Detection System (Bio-Rad Laboratories, CA, USA). Amplification was performed for 2 minutes at 95°C followed by 40 cycles of 5 seconds at 95°C and 30 seconds at 60°C.

Concordance assays

The performance of the MitoQ assay was compared to genomic DNA quantities and its respective mtDNA estimations. The genomic DNA was quantified using the Quantifiler™ Trio DNA Quantification Kit (Life Technologies, CA, USA) on the QuantStudio™ 6 Flex Real-Time PCR System (Applied Biosystems, CA, USA), following the manufacturer protocols [14].

A concordant *in-house* assay for quantification of mtDNA was also run following the published procedures [20]. The primers, probes and DNA standards were obtained as synthetic oligonucleotides (Integrated DNA Technologies, IA, USA) and the two targets were run as separate assays. FAM probe was substituted in both of the published assays. The assay was

run in a PrimePCR™ format, using the SsoAdvanced™ Universal Probes Supermix (Bio-Rad Laboratories, CA, USA) on the CFX 96 Touch™ Real-Time PCR Detection System (Bio-Rad Laboratories, CA, USA).

Data analysis

The amplification data was analysed using the CFX Manager™ Software, Version 3.1 (Bio-Rad Laboratories, CA, USA). For the standard plot analysis settings, cycle baselines were automatically subtracted to curve fit and quantification cycles were determined in an independent regression model [21]. The genomic assay was detected using the QuantStudio™ Real-Time PCR Software (v1.2) (Applied Biosystems, CA, USA), following the manufacturer's protocols [14].

Results and discussion

Sensitivity and reproducibility

The assay demonstrates a high sensitivity to quantifying low copies of synthetic DNA standards. Among three replicates, 280 copies of mtDNA were quantified at 24.10 cycles [21] (Fig. 1).

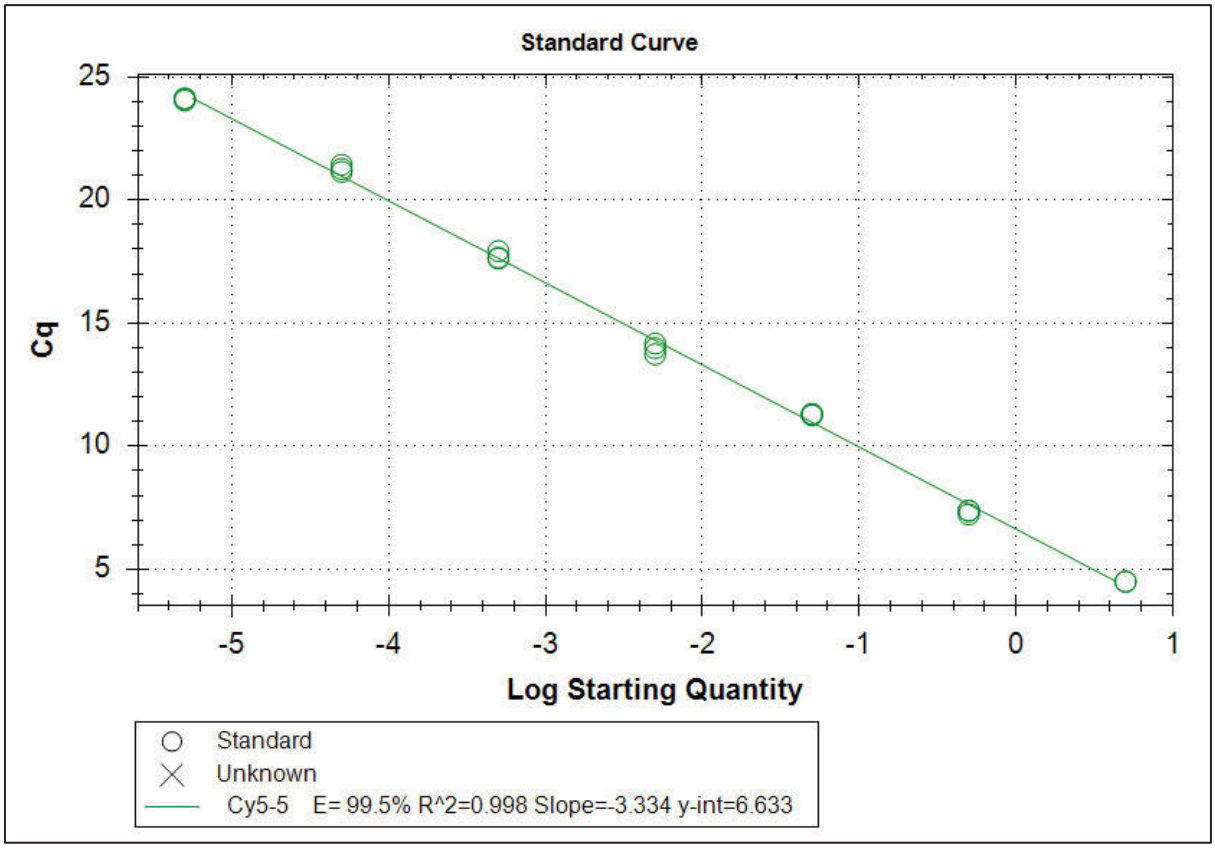


Fig. 1 Real-time quantification of mitochondrial DNA standards using the developed MitoQ assay. The quantification of the seven ten-fold dilution series (*n*=3) from 2.8E+08 to 2.8E+02 copies/µL (presented in log scale). The quantification cycle indicates the PCR cycle at which the target product is amplified to a detectable fluorescent signal [21]

This was observed in a seven ten-fold dilution series (2.8E+08 to 2.8E+02 copies) that showed a high amplification efficiency (99.5%) and increased on an average of 3.27 cycles for each ten-fold dilution. The assay's sensitivity was limited to 280 copies of mtDNA as lower concentrations (28 copies) were amplified at non-optimal efficiencies (>110%). Additionally, lower concentrations also quantified extraneous amounts of background DNA at times (data not shown). The quantification of background DNA in controls is not optimal as it limits the use of quality controls to assess an assays performance [22]. However, the quantification reflects the ubiquitous nature of mtDNA as well as the assay's high sensitivity [3]. The sensitivity of this assay is comparable to regions of mtDNA which were targeted in other qPCR assays [6, 20]. Commonly mtDNA testing is used in cases where nuclear genotyping is limited or unsuccessful. In these cases, the DNA samples are often compromised and degraded in a range of small fragments. As a result, the smaller amplicon target of the MitoQ assay at 70 bp may be more readily recovered in fragments found in degraded and ancient DNA samples [6, 11, 23].

Furthermore the assay was highly reproducible for all sample replicates. All synthetic DNA standards showed low variability (*n*=3), though slight increases in standard deviation were seen in low copy number standards (Table 2).

Table 2 Standard deviation between replicates of DNA standards, quantified using the MitoQ mitochondrial DNA qPCR assay. The standard deviation is reported as a cycle of quantification and represents the variability between replicates of DNA standards (*n*=3)

Standard copy number	Standard deviation (*n*=3)
2.8E+02 copies	0.045 cycles
2.8E+03 copies	0.143 cycles
2.8E+04 copies	0.166 cycles
2.8E+05 copies	0.217 cycles
2.8E+06 copies	0.030 cycles
2.8E+07 copies	0.085 cycles
2.8E+08 copies	0.004 cycles

The variation in low copy standards may arise as a stochastic effect of pipetting low volumes of DNA as higher volumes have been reported with increased assay reproducibility [24]. Nonetheless standard deviations between replicate samples (range, 0.004 to 0.217 cycles) remain below variabilities for quantification standards (SD < 0.5 C_q) and indicates a highly reproducible qPCR assay.

Specificity

The assay amplifies an encoding subunit (10,128 to 10,197 rCRS) of the NADH Dehydrogenase (ND3) genes (NC 012920.1 Region: 10,059 to 10,404). This DNA region is part of a larger amplicon, which was amplified with primers validated to prevent the co-amplification of nuclear pseudogenes [11]. A Basic Local Alignment Search Tool (BLAST) of the target in this study showed high specificity to human mitochondrial genomes. A low homology was also shown to nDNA (expectation value, E-value >1) (Fig. 2).

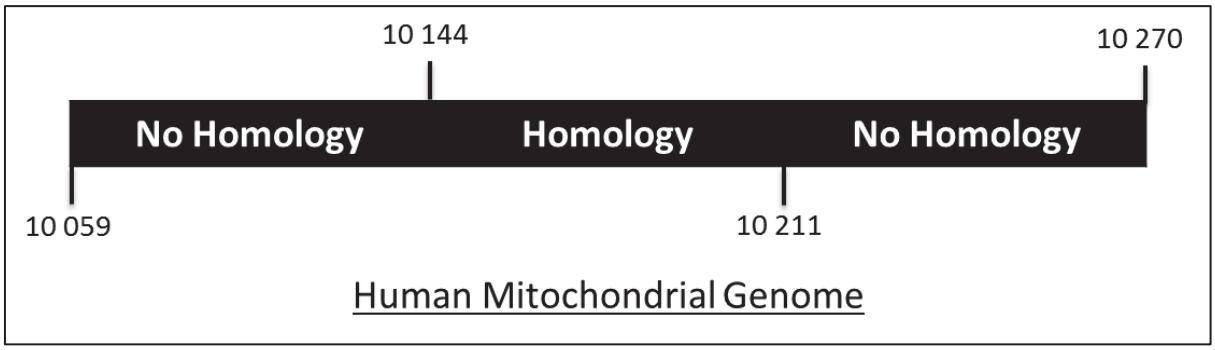


Fig. 2 Evaluation of target mitochondrial DNA region for homology [25]. A region within mitochondrial DNA (10,059 to 10,270 rCRS) was evaluated for homology based on E-value. Low homology (E-value > 1) was seen between regions 10,059 to 10,144 rCRS and 10,211 to 10,270 rCRS. Homology (E-value < 1) was identified between regions 10,144 to 10,211 rCRS

The E-value infers the sequence similarity in known genomes. Using a statistical model, a high similarity is reported when an E-value of less than 1 is observed [26]. The only potential homology shown in BLAST was on the reverse primer (10,197 to 10,179 rCRS) to unintended templates, chromosome 17 primary (NC_000017.11) and alternate assemblies (NC_018928.2). However, gel electrophoresis confirmed correct amplification of the intended 70 bp product and no amplification in negative controls (sperm extracts) (Online Resource 1).

Previous research has found an absence of mtDNA in sperm and has led to its use in mtDNA-free studies [11, 27-33]. This study confirmed the presence of nDNA in sperm extracts with STR profiling (Online Resource 2). The lack of amplification in these controls shows the MitoQ is highly specific to the mitochondrial genome and indicates the absence of mtDNA in sperm extracts. However, in order to indicate the assay's amplification to be human-specific, further studies need to be carried out on a range of different species samples.

Estimation of copy number

The MitoQ assay was able to quantify mtDNA copy number in a range of pristine and compromised DNA samples (Fig. 3). The compromised samples were composed of DNA extracts which were heat-treated at progressive times.

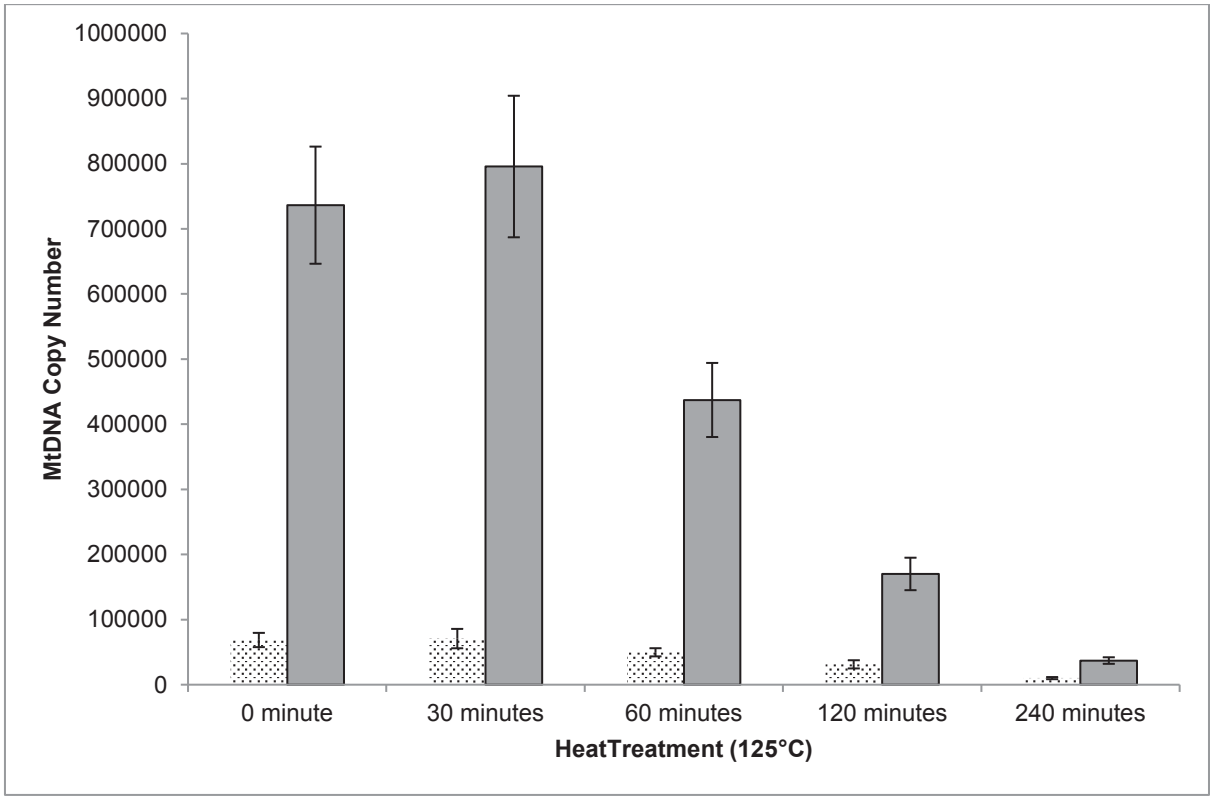


Fig. 3 Real-time estimation of mitochondrial DNA copy number in artificially degraded samples. Mean (*n*=5) estimation of mitochondrial copy number using the developed MitoQ assay, compared to another real-time qPCR assay [20]. The dotted column represents the mtDNA copy number for the MitoQ ND3 target. The filled column represents the mtDNA copy number for the other HV1 (A1/A1R) target. The quantified samples are heat-treated at 125°C for 0, 30, 60, 120 and 240 minutes. Error bars represent S.E.M. between biological replicates

Multiple copies of mtDNA (68,000 to 10,000 copies) were detected in all samples for the 70 bp ND3 target. In comparison to the other previously published HV1 (A1/A1R) mitochondrial targets [20], a lower copy number was reported. However, the amplification of one of the HV1 targets (16,112-16,071) was variable with little to no amplification being recorded at times (Online Resource 3). Most likely this was due to the original sequence of the forward primer (L16131) that was designed based on individuals of known mtDNA sequence [20]. As a result, variants (16,124- and 16,128-) relative to the human mtDNA genome (NC_000017.11) may have caused mispriming and low amplification. A re-design of the forward primer, which accounted for the two deletions in **bold** (5'- GCC ACC ATG AAT AT**T** GTA **C**GG T-3') showed improved priming and amplification efficiency (Online Resource 4). However, the amplification of the DNA standards remained below optimal efficiencies (84.5%) and accordingly mtDNA quantities, which were amplified with the forward L16131 (2F) primers, were excluded for reporting (data not shown).

The difference in copy number between the assays can be better resolved by quantifying synthetic standards that contain the DNA sequence of all three targets. A pure nuclear DNA sample such as sperm extracts should be included to monitor any potential amplification of

nuclear pseudogenes. While there was difference between the reported quantities, it is proposed that all targets would be suitable for the quantification of mtDNA depending on the downstream application. The MitoQ assay would be optimal for evaluating DNA samples for downstream mitochondrial processes such as whole mitochondrial genome sequencing [13]. This was observed with the assay's quantification of buccal cells that provided a 1:147 ± 54 (mean ± SD, *n*=5), genomic to mtDNA copy number estimation, which is consistent with the reported values of 1:100-1000 [13, 34, 35]. Most likely the MitoQ assay can be used for other biological sources as other longer mtDNA targets have been quantified in ancient DNA samples, skeletal remains and, telogenic and catagenic hair shafts with variable success rates [36]. Similarly, the assay's performance can be expanded to evaluating these additional sample types. Additionally, the concordance of the assay can also be tested on other quantification technologies such as digital droplet PCR.

MtDNA copy number was shown to overall decrease for prolonged heat-treated samples (Fig. 3). The treatment of DNA samples at high temperatures has been reported to result in a breakage of bonds in DNA strands, in particular breakage of phosphodiester and glycosidic bonds. As a result, the loss of intact DNA molecules reduces the amplification of target sequences between forward and reverse primers. This most likely explains the reduced copies of heated samples as the frequency of strand-breaks is increased with prolonged treatment. The quantification of copy numbers in the longest heat-treated DNA samples (240 minutes) shows the assay is able to detect DNA fragments, which are highly degraded. This information is valuable as it provides an accurate estimation of the whole mtDNA copy number and optimise the overall success of downstream DNA processes. It is recognised that the degradation of DNA samples is not the only determining factor of amplification success. Other factors such as PCR inhibitors and environmental (non-human DNA) contaminants may influence the efficiency of amplification. As a result, future studies can focus on the assay's performance with known inhibitors as well as evaluate its potential cross reactivity to other non-human species.

Conclusions

The preliminary evaluation of the MitoQ assay has shown it is able to quantify mtDNA copy number in pristine and compromised DNA samples. This is achieved by the amplification and detection of the mt-ND3 fragment. This assay is specific to the mitochondrial genome as shown by the lack of NUMTs in the design and amplification. Its short amplicon (70bp) was highly sensitive to amplifying degraded DNA samples of saliva extracts. The amplification of DNA standards was highly efficient and showed minimal deviation between replicates, indicating the assay to be robust and reproducible. While the assay was developed as a singleplex, the detection of the MitoQ target can be potentially incorporated into other quantification kits by adding the Cyt 5.5 dye to its existing chemistry. Such an assay will conserve the consumption of samples and provide an overall assessment of the DNA sample for downstream uses, which becomes time and cost effective to the end-user.

Compliance with ethical standards

Conflict of interest The authors have declared no conflict of interest.

Ethical Approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

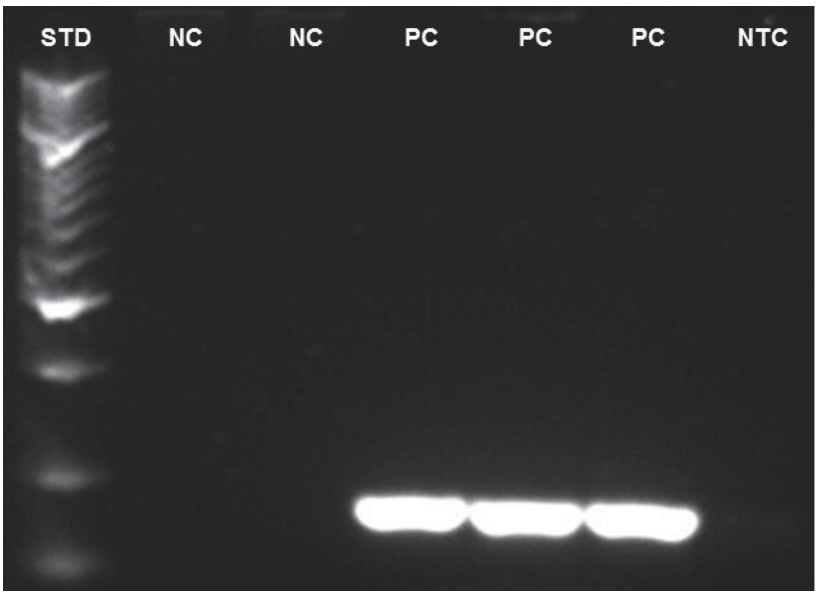
Informed consent Informed consent was obtained from all individual participants included in the study.

References

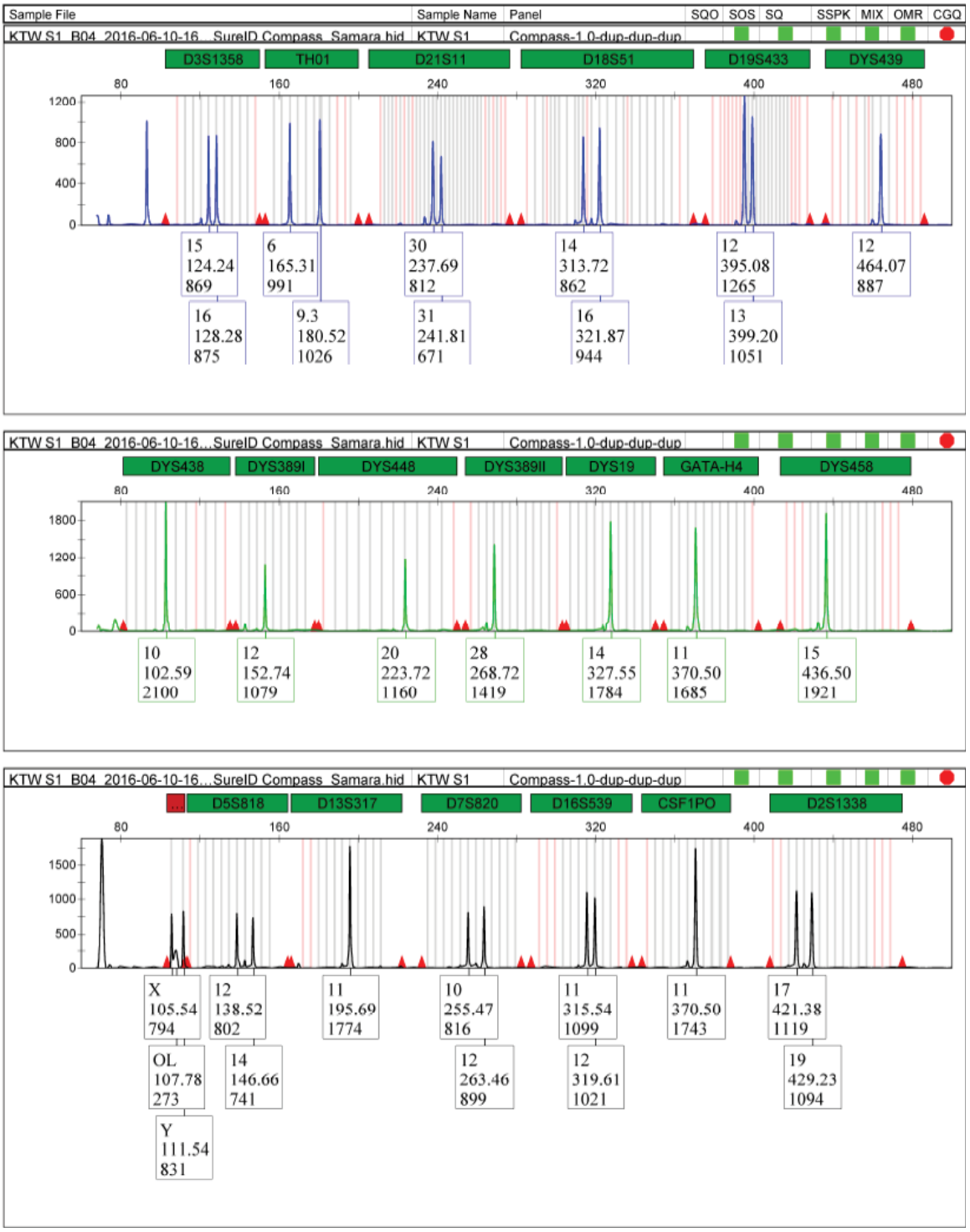
1. Robin JD, Ludlow AT, LaRanger R, Wright WE, Shay JW (2016) Comparison of DNA quantification methods for next generation sequencing. Scientific reports 6:24067
2. Wieczorek P (2017) To NanoDrop®or Not to NanoDrop®: Choosing the Most Appropriate Method for Nucleic Acid Quantitation. Promega Corporation, Wisconsin
3. Satoh M, Kuroiwa T (1991) Organization of multiple nucleoids and DNA molecules in mitochondria of a human cell. Experimental cell research 196 (1):137-140
4. Tobe SS, Linacre AM (2008) A technique for the quantification of human and non-human mammalian mitochondrial DNA copy number in forensic and other mixtures. Forensic Science International: Genetics 2 (4):249-256
5. Malik AN, Shahni R, Rodriguez-de-Ledesma A, Laftah A, Cunningham P (2011) Mitochondrial DNA as a non-invasive biomarker: accurate quantification using real time quantitative PCR without co-amplification of pseudogenes and dilution bias. Biochemical and biophysical research communications 412 (1):1-7
6. Kavlick MF, Lawrence HS, Merritt RT, Fisher C, Isenberg A, Robertson JM, Budowle B (2011) Quantification of human mitochondrial DNA using synthesized DNA standards. Journal of forensic sciences 56 (6):1457-1463
7. Andréasson H, Nilsson M, Budowle B, Lundberg H, Allen M (2006) Nuclear and mitochondrial DNA quantification of various forensic materials. Forensic science international 164 (1):56-64
8. Sprouse ML, Phillips NR, Kavlick MF, Roby RK (2014) Internal Validation of Human Mitochondrial DNA Quantification Using Real-Time PCR. Journal of forensic sciences 59 (4):1049-1056
9. Lopez JV, Yuhki N, Masuda R, Modi W, O'Brien SJ (1994) Numt, a recent transfer and tandem amplification of mitochondrial DNA to the nuclear genome of the domestic cat. Journal of Molecular Evolution 39 (2):174-190
10. Walker JA, Hedges DJ, Perodeau BP, Landry KE, Stoilova N, Laborde ME, Shewale J, Sinha SK, Batzer MA (2005) Multiplex polymerase chain reaction for simultaneous quantitation of human nuclear, mitochondrial, and male Y-chromosome DNA: application in human identification. Analytical biochemistry 337 (1):89-97
11. Ramos A, Santos C, Alvarez L, Nogués R, Aluja MP (2009) Human mitochondrial DNA complete amplification and sequencing: A new validated primer set that prevents nuclear DNA sequences of mitochondrial origin co-amplification. Electrophoresis 30 (9):1587-1593
12. Qiagen (2016) QIAamp® DNA Mini and Blood Mini Handbook. Qiagen, Hilden
13. Wai KT, Barash M, Gunn P (2018) Performance of the Early Access AmpliSeq™ Mitochondrial Panel with degraded DNA samples using the Ion Torrent™ platform. Electrophoresis
14. Thermo Fisher Scientific (2017) Quantifiler™ HP and Trio DNA Quantification Kits. Thermo Fisher Scientific, California
15. Qiagen (2010) HotStarTaq® Plus PCR Handbook. Qiagen, Hilden
16. Anderson S, Bankier A, de Bruijn M, Coulson AR, Drouin J, Eperon I, Nierlich D, Roe B, Sanger F, Schreier P, Smith A, Staden R, Young I (1981) Sequence and organization of the human mitochondrial genome. Nature 290 (5806):457-465
17. Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N (1999) Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. Nature genetics 23 (2):147-147
18. Life Technologies (2015) Qubit® dsDNA HS Assay Kits. Life Technologies, California
19. Bio-Rad Laboratories (2014) PrimePCR™ Assays, Panels, and Controls for Real-Time PCR. Bio-Rad Laboratories, California
20. Alonso A, Martí P, Albarrán C, Garcí P, Garcí O, de Simón LF, Garcí J, Sancho M, de la Rúa C, Fernández-Piqueras J (2004) Real-time PCR designs to estimate nuclear and mitochondrial DNA copy number in forensic and ancient DNA studies. Forensic Science International 139 (2):141-149

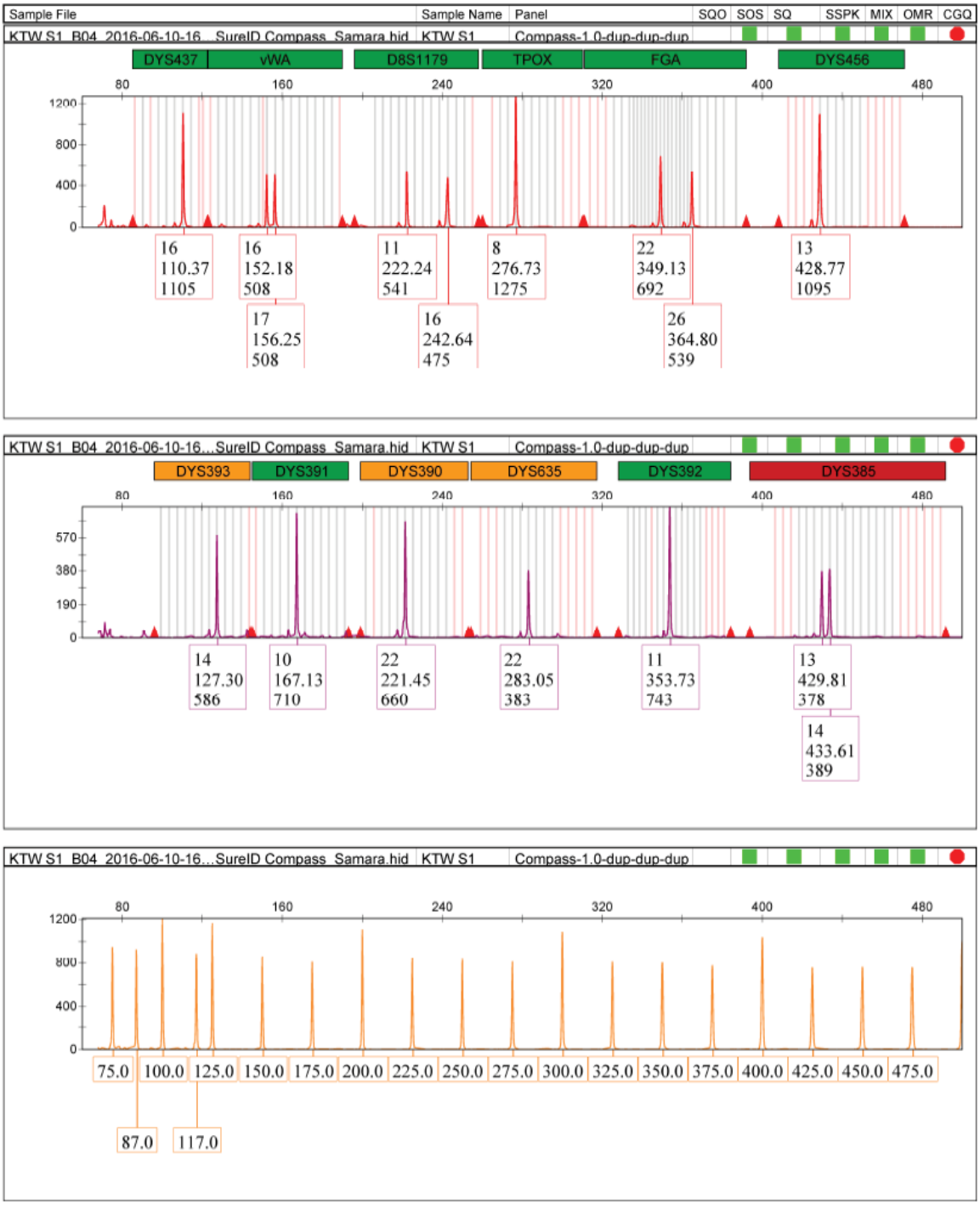
21. Ellison SL, English CA, Burns MJ, Keer JT (2006) Routes to improving the reliability of low level DNA analysis using real-time PCR. BMC biotechnology 6 (1):33
22. Wilson MR, Stoneking M, Holland MM, DiZinno JA, Budowle B (1993) Guidelines for the use of mitochondrial DNA sequencing in forensic science. Crime Lab Digest 20 (4):68-77
23. Goodwin C, Higgins D, Tobe SS, Austin J, Wotherspoon A, Gahan ME, McNevin D (2018) Singleplex quantitative real-time PCR for the assessment of human mitochondrial DNA quantity and quality. Forensic Science, Medicine and Pathology:1-6
24. Peccoud J, Jacob C (1996) Theoretical uncertainty of measurements using quantitative polymerase chain reaction. Biophysical journal 71 (1):101-108
25. Consortium GP (2015) A global reference for human genetic variation. Nature 526 (7571):68-74
26. Pearson WR (2013) An introduction to sequence similarity ("homology") searching. Current protocols in bioinformatics:3.1. 1-3.1. 8
27. Sharpley MS, Marciniak C, Eckel-Mahan K, McManus M, Crimi M, Waymire K, Lin CS, Masubuchi S, Friend N, Koike M (2012) Heteroplasmy of mouse mtDNA is genetically unstable and results in altered behavior and cognition. Cell 151 (2):333-343
28. Lane N (2012) The problem with mixing mitochondria. Cell 151 (2):246-248
29. Luo S-M, Ge Z-J, Wang Z-W, Jiang Z-Z, Wang Z-B, Ouyang Y-C, Hou Y, Schatten H, Sun Q-Y (2013) Unique insights into maternal mitochondrial inheritance in mice. PNAS 110 (32):13038-13043
30. Schwartz M, Vissing J (2002) Paternal inheritance of mitochondrial DNA. New England Journal of Medicine 347 (8):576-580
31. Sato M, Sato K (2013) Maternal inheritance of mitochondrial DNA by diverse mechanisms to eliminate paternal mitochondrial DNA. Biochimica et Biophysica Acta 1833:1979-1984
32. Luo S-M, Schatten H, Sun Q-Y (2013) Sperm Mitochondria in Reproduction: Good or Bad and Where Do They Go? Journal of Genetics and Genomics 40 (1):549-556
33. Chan DC, Schon EA (2012) Eliminating Mitochondrial DNA from Sperm. Developmental Cell 22 (3):469-470
34. Wai T, Ao A, Zhang X, Cyr D, Dufort D, Shoubridge EA (2010) The role of mitochondrial DNA copy number in mammalian fertility. Biology of reproduction 83 (1):52-62
35. Shin J, Kim KC, Lee DC, Lee HR, Shim JY (2017) Association between Salivary Mitochondrial DNA Copy Number and Chronic Fatigue according to Combined Symptoms in Korean Adults. Korean journal of family medicine 38 (4):206-212
36. Alonso A, Albarrán C, Martín P, Garcıa P, Garcıa O, de la Rúa C, Alzualde A, de Simón LF, Sancho M, Piqueras JF Multiplex–PCR of short amplicons for mtDNA sequencing from ancient DNA. In: International Congress Series, 2003. Elsevier, pp 585-588

Online Resources

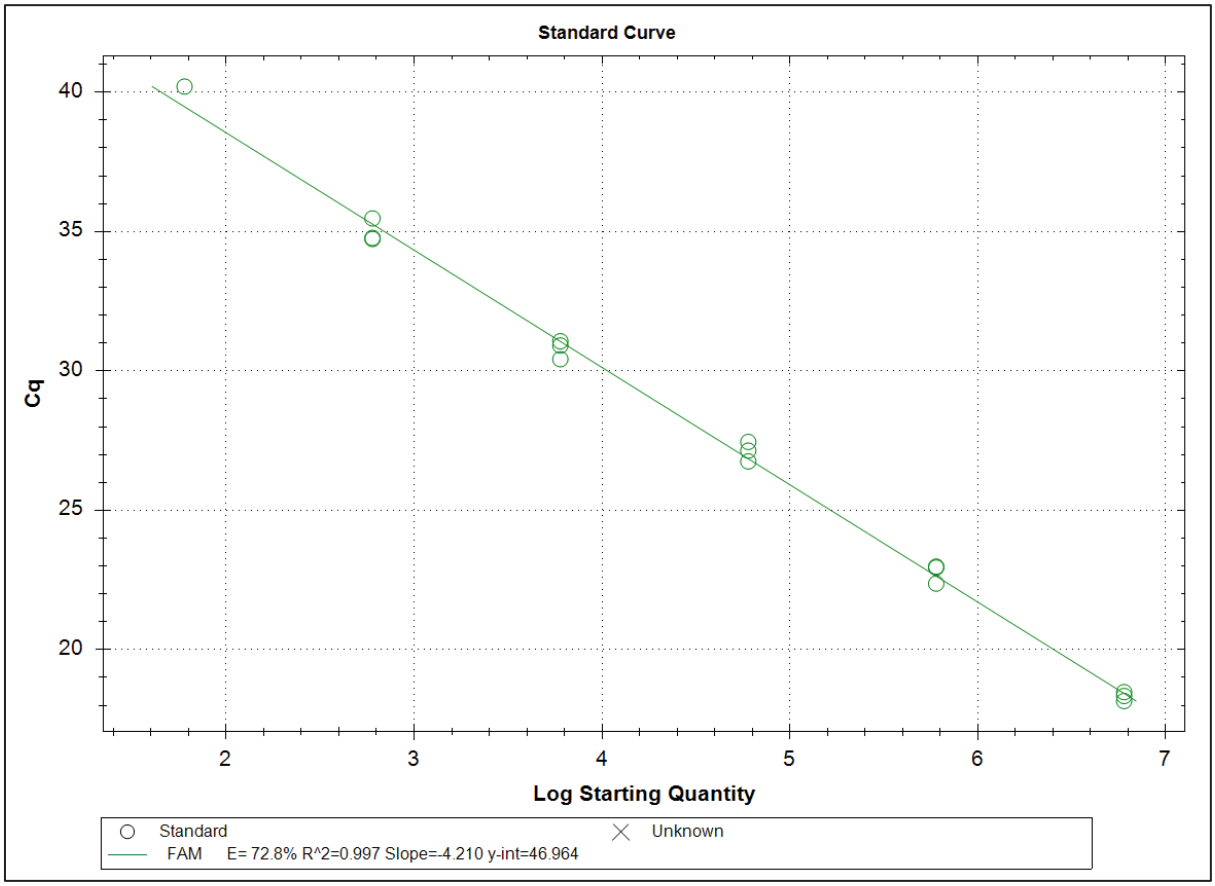


Online Resource 1 Electrophoresis of PCR products amplified with the MitoQ mitochondrial DNA assay. No amplification observed in negative controls (NC), sperm extracts. Positive amplification of 70 bp product observed in positive controls (PC), saliva extracts. No amplification present in no template control (NTC). Loaded standard (STD) is HyperLadder™ 50bp (Bioline, London, UK)

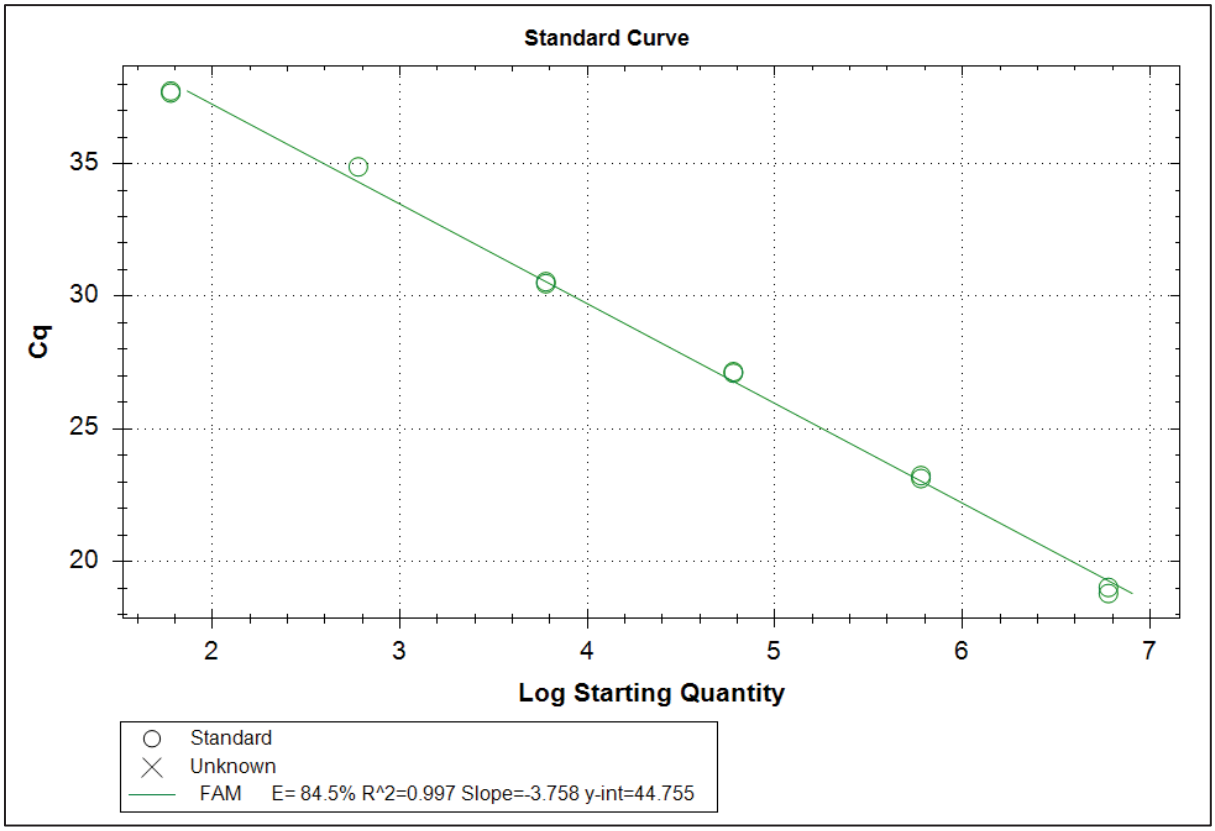

Sample File
Sample Name
Panel
SQO
SOS
SQ
SSPK
MIX
OMR
CGQ
KTW S1_B04_2016-06-10-16...SureID Compass_Samara.hid
KTW S1
Compass-1.0-dup-dup-dup
D3S1358
TH01
D21S11
D18S51
D19S433
DYS439
15 124.24 869
16 128.28 875
6 165.31 991
9.3 180.52 1026
30 237.69 812
31 241.81 671
14 313.72 862
16 321.87 944
12 395.08 1265
13 399.20 1051
12 464.07 887
DYS438
DYS389I
DYS448
DYS389II
DYS19
GATA-H4
DYS458
10 102.59 2100
12 152.74 1079
20 223.72 1160
28 268.72 1419
14 327.55 1784
11 370.50 1685
15 436.50 1921
D5S818
D13S317
D7S820
D16S539
CSF1PO
D2S1338
X 105.54 794
OL 107.78 273
Y 111.54 831
12 138.52 802
14 146.66 741
11 195.69 1774
10 255.47 816
12 263.46 899
11 315.54 1099
12 319.61 1021
11 370.50 1743
17 421.38 1119
19 429.23 1094



Online Resource 2 STR profile of sperm extracts. Sperm extracts were used as a control for human nuclear DNA. Amplification of STR loci, part of the SureID® Compass Human DNA Identification Kit (Health Gene Technologies, Ningbo, China) confirms the presence of nuclear DNA in controls



Online Resource 3 Real-time quantification of synthetic mitochondrial DNA standards with previously published L16131 (2F) and H16379 (4R) primers [20]. Quantification of six ten-fold dilution series (*n*=3) from 6.0E+06 to 60 copies (presented in log scale). The quantification cycle determines the PCR cycle at which the target product is amplified to detectable thresholds. An amplification efficiency of 72.8% was calculated for standards



Online Resource 4 Real-time quantification of synthetic mitochondrial DNA standards with modified L16131 (2F) and H16379 (4R) primers [20]. Quantification of six ten-fold dilution series (*n*=3) from 6.0E+06 to 60 copies (presented in log scale). Quantification cycle determines PCR cycle at which target product is amplified to detectable thresholds. An amplification efficiency of 84.5% was calculated for standards

Chapter 4: Profiling of Genomic DNA

1. Introduction

For over 30 years, forensic DNA profiling has been identifying and reporting DNA evidence in criminal and civil proceedings. DNA profiling targets tandemly repeated DNA sequences across the human genome, where the length of these short tandem repeats (STRs) is highly variable and discriminating between individuals (with the exception of genetically identical twins) [73]. The majority of STRs are non-coding and provide no additional information on a person's disease or phenotype status.

With more than 20,000 STRs identified in the human genome, the most discriminating STR loci have been selected and are routinely reported for forensic identifications [83]. Loci sets include the FBI's Combined DNA Index System (CODIS) and European Standard Set (ESS) [74, 84, 85]. The core STR loci are part of a number of forensic kits, which are validated for human identifications. Examples include the PowerPlex® 21 System (Promega Corporation, Wisconsin, USA), GlobalFiler® (Applied Biosystems, California, USA) and SureID® Compass (Health Gene Technologies, Ningbo, China) [3, 86]. Genomic DNA profiling involves using these multiplex kits for PCR amplification, followed by CE separation and fluorescent detection [87].

In cases where DNA profiling is unsuitable due to degradation or low yields of DNA, analysis of mtDNA can be considered as an alternate means of identification. Although not as discriminating as genomic DNA, mtDNA sequencing can still be of investigative value by inferring a person's ancestry and confirm their maternal relationships in certain cases. Thus, the testing of mtDNA is reserved for cases of unreportable profiles, such as in ancient remains, the re-opening of cold cases, compromised samples or where there is a lack of appropriate reference samples [20].

To determine the suitability of mtDNA analysis using MPS technologies for samples degraded in a previous study (Chapter 2), genomic DNA was profiled and assessed. Profiling involved amplification of STR loci with the PowerPlex® 21 System (Promega Corporation, Wisconsin, USA), followed by separation via CE on the 3500 Genetic Analyser® (Applied Biosystems, California, USA).

The results showed a significant number of allele peaks with reduced heights as well as allele dropouts in degraded samples. In addition, PCR artifacts of stutter and heterozygote imbalance were pronounced in treated samples. Overall, the study identifies a reduction in allele calling for the most degraded samples.

2. Method

2.1 DNA Amplification

A set of artificially degraded DNA samples (see Chapter 2) were amplified using the PowerPlex® 21 System (Promega Corporation, Wisconsin, USA). The PowerPlex® 21 System amplifies 21 loci (20 autosomal STR loci and Amelogenin) that are highly discriminating between individuals and have been adopted by Australian operational forensic laboratories [88]. The 20 autosomal loci are: D1S1656, D2S1338, D3S1358, D5S818, D6S1043, D7S820, D8S1179, D12S391, D13S317, D16S539, D18S51, D19S433, D21S11, CSF1PO, FGA, Penta D, Penta E, TH01, TPOX, and vWA [3]. Following manufacturer guidelines, in a 25 µL reaction mix, 0.5 ng sample DNA was combined with PowerPlex® 21 5X MasterMix and PowerPlex® 21 5X Primer Pair Mix. PCR reactions were amplified for 1 min at 96°C followed by 30 cycles of 10 seconds at 94°C, 1 minute at 59°C, 30 seconds at 72°C and a final extension of 10 minutes at 60°C before holding at 4°C on the Veriti® 96-Well Thermal Cycler (Applied Biosystems, California, USA).

2.2 Capillary Electrophoresis

Amplified PCR product (1 µL) was combined with 0.5 µL of WEN Internal Lane Standard (ILS) 500 and 9.5 µL of Hi-Di™ formamide. The reaction was denatured at 95°C, followed by a snap cooling. All samples were loaded onto the 3500 Genetic Analyser® (Applied Biosystems, California, USA). Channels were calibrated and set to detect loci labelled with fluorescein (blue), JOE (green), TMR-ET (yellow), CXR-ET (red) or CC5 ILS 500 (orange) probes. Amplified samples were injected at 1.2 kV for 24 seconds, followed by separation at 1.2 kV for 24 minutes at a run temperature of 60°C, as recommended by the manufacturer [3]. For quality control and allele sizing, allelic ladders from the PowerPlex® 21 Allelic Ladder Mix were included in each column of the plate run.

2.3 STR Profiling Analysis

Fragment sizing and allele detection of the PowerPlex® 21 System (Promega Corporation, Wisconsin, USA) loci was performed using the software package GeneMapper® ID-X 1.4 (Applied Biosystems, California, USA). Analysis involved peak assignment to evaluate the number of alleles called at each sample. Table 9 lists the thresholds used for allele calling and determination of PCR anomalies such as stutter and heterozygote balance. Thresholds were arbitrarily set to values suggested by the manufacturer in an import file [3] (Table 9). The PowerPlex® 21 Panels, Bins and Stutter values were provided by the manufacturer as text

files and, can be found at [89]: https://www.promega.com.au/resources/software-firmware/genemapper-id-software-panels-and-bin-sets/

Table 9: Allele Calling thresholds used for analysis of peaks produced by the PowerPlex® 21 System (Promega Corporation, Wisconsin, USA).

Threshold	Cut-off value
Peak detection (by default)	50 RFU
Heterozygote peaks	175 RFU
Homozygote peaks	700 RFU
Stutter peaks	0.2 PHR
Heterozygote imbalance peaks	0.7 PHR

*Cut-off values are set in accordance to those validated by the kit manufacturer [3]. PHR (Peak height ratio). RFU (Relative fluorescence unit).

By default, the software is able to detect peaks that exhibit a fluorescence of at least 50 relative fluorescence units (RFU). A 175 RFU calling threshold was applied for heterozygote peaks. For homozygote peaks, a 700 RFU calling threshold was used. These thresholds were applied in reference to the kits' default settings [3] and reflect values used in an operational Australian laboratory (Forensic and Analytical Science Service, Sydney, Australia).

Peak height ratio (PHR) was calculated using the formula proposed by Gill *et al.,* [90]. Specifically, PHR at any given locus (x) was calculated using the following formula:

$$Peak\ height\ ratio\ (PHR) = \frac{Height\ of\ smallest\ peak\ (\emptyset \min)}{Height\ of\ largest\ peak\ (\emptyset \max)}$$

Stutter was defined when two peaks at the same locus were separated by four bases (n ± 4) for tetranucleotide repeats or five bases (n ± 5) for pentanucleotide repeats (±0.5 bases). A cut-off value for stutter was applied at 0.2 PHR, in consideration of the allele and, locus size. A cut-off of 0.7 PHR was set for heterozygote balance where a heterozygote imbalance was reported for peaks between 0.2 to 0.7 PHR. Again, the heterozygote balance was considered in respect to the locus size as well as the distance between the heterozygote pair.

3. Results

The suitability of degraded DNA samples for mitochondrial sequencing was assessed based on evaluation of sample genomic DNA profiles for peak height, allele drop-ins and allele drop-outs.

To determine the effect of heat treatment on allele peak height with respect to amplicon size, a degradation plot (peak height vs locus size) was constructed for each time point (Figure 13). The degradation plot records the peak height (RFU) across all STR fragments (locus dropouts were excluded as it was undetected). The degradation slope (ΔPeak Height/ΔLocus Size) reports the change in peak height of increasing STR fragment size. Accordingly, the degradation slope (*n*=5) was calculated to be -1.58 for 0 minutes, -21.61 for 30 minutes, -21.71 for 60 minutes, -25.84 for 120 minutes and -24.98 for 240 minutes (Figure 13). These slopes indicates an overall peak height reduction as STR fragments increase in size.

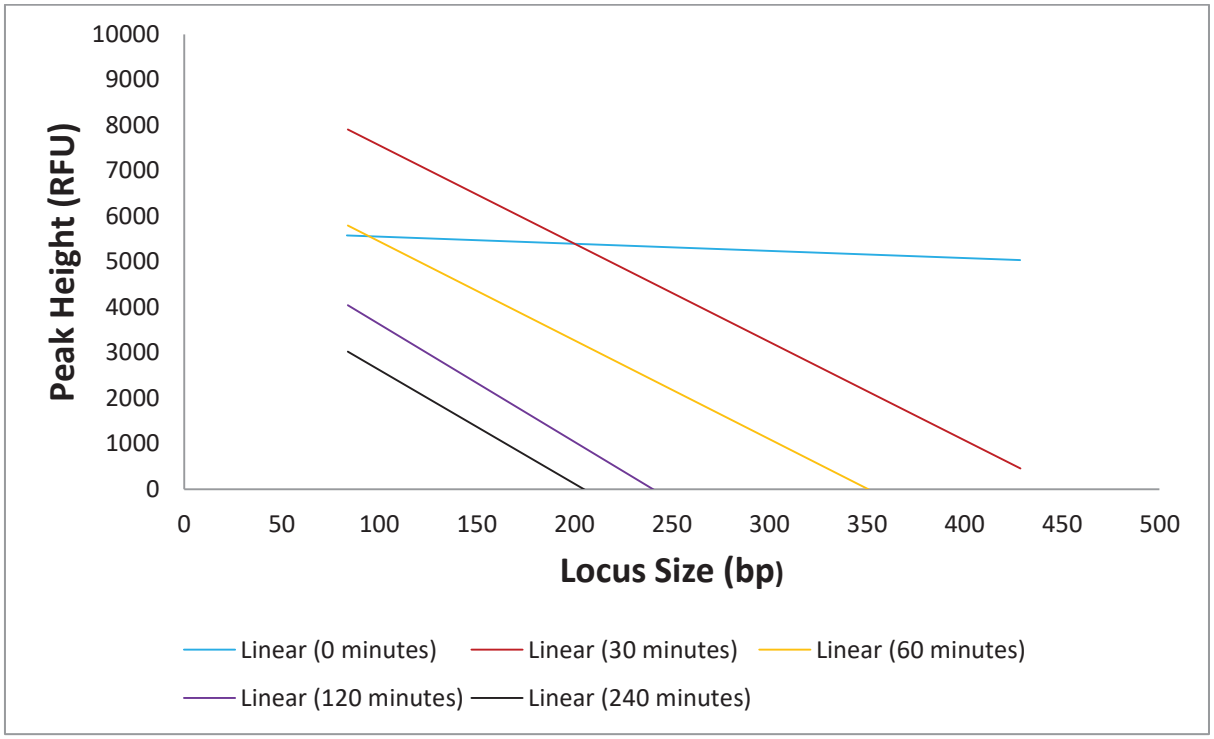


Figure 13: Degradation Plot. Peak height values recorded across loci of increasing fragment size in artificially degraded DNA samples. Five DNA samples were degraded by heating at 125°C for five time points (*t*=0, 30, 60, 120, 240 minutes) each (*n*=25).

Alleles were reported for treated samples, however were at lower peak heights than untreated samples. Figure 14 shows overall peak heights and percentage of alleles called for each treatment group. Peak height (mean ± S.E.M) was shown to decrease for prolonged treated samples with 4,665 ± 1,589 RFU for 0 minutes, 4,331 ± 1,445 for 30 minutes, 2,812 ± 527 RFU for 60 minutes, 1,931 ± 134 RFU for 120 minutes and 1,643 ± 189 RFU for 240 minutes. The reduction in RFU reads reflects a decrease in the percentage of alleles called in profiles. All alleles were called at 0 minutes and in some instances at 30 minutes. Allele drop-outs became notable for treatment groups at 60, 120 and 240 minutes with the percentage of alleles (mean ± S.E.M) called at 80 ± 8%, 57 ± 3% and 40 ± 4%, respectively.

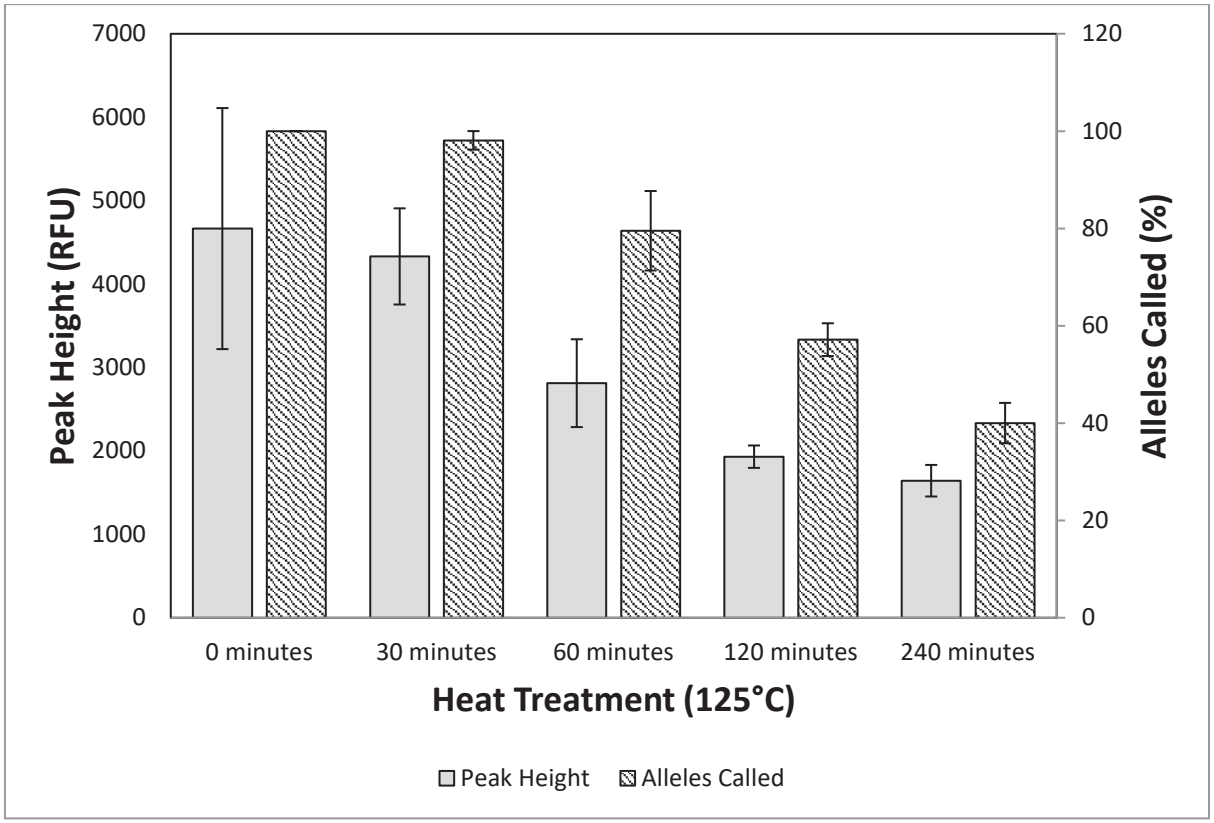


Figure 14: Degradation Study. Amplification of the PowerPlex® 21 System (Promega Corporation, Wisconsin, USA) of degraded DNA at 125°C for 0, 30, 60, 120 and 240 minutes, respectively. Average peak height (shaded column) and average alleles called (diagonal ruled column) are displayed with a 50 RFU detection threshold. Allele calling was set to 175 RFU for heterozygotes and 700 RFU for homozygotes. Error bars represent the S.E.M between five biological replicates heated at five time points each (*n*=25).

Figure 15 shows majority of allele drop-outs to occur at loci with sizes starting at 128 bp (vWA). As displayed, Amelogenin was reportable for all treatment groups and allowed sex determination in all cases (Figure 15). This was expected given it is from a small locus (84 bp) and routinely recovered in compromised samples [4]. Other loci which were reported for all treatments were TH01 (71 bp), D8S1179 (77 bp), D16S539 (78 bp) and D3S1358 (98 bp).

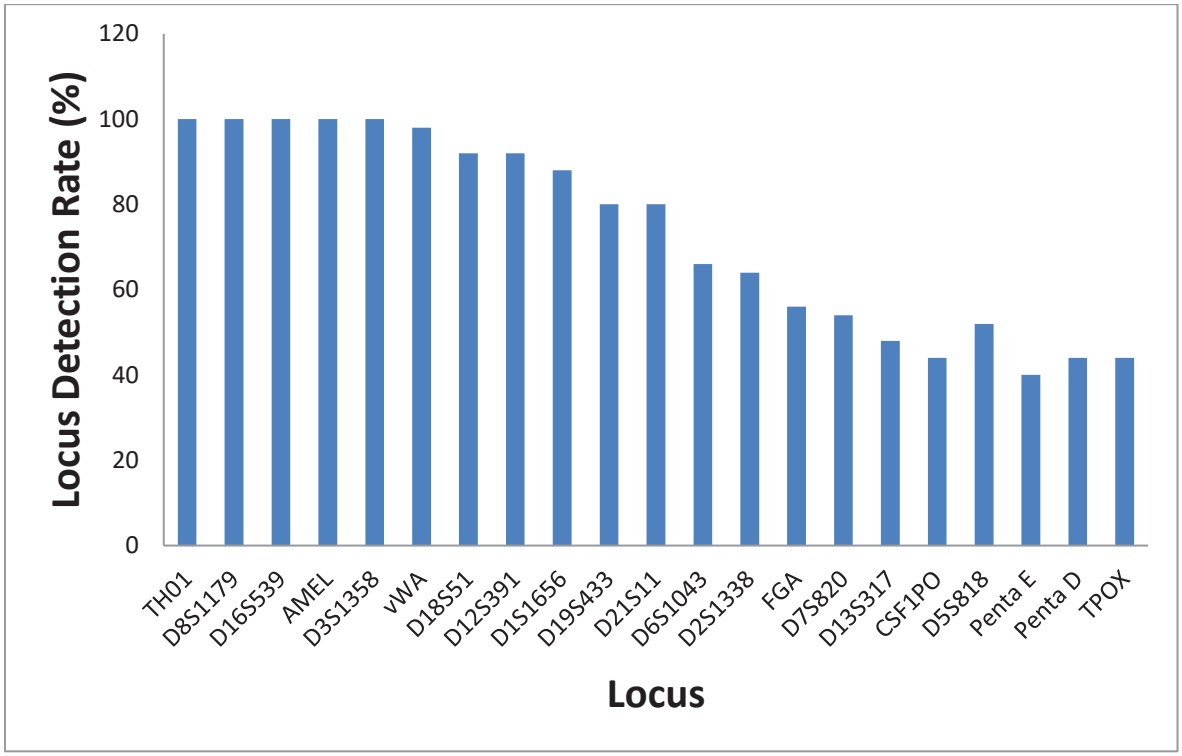
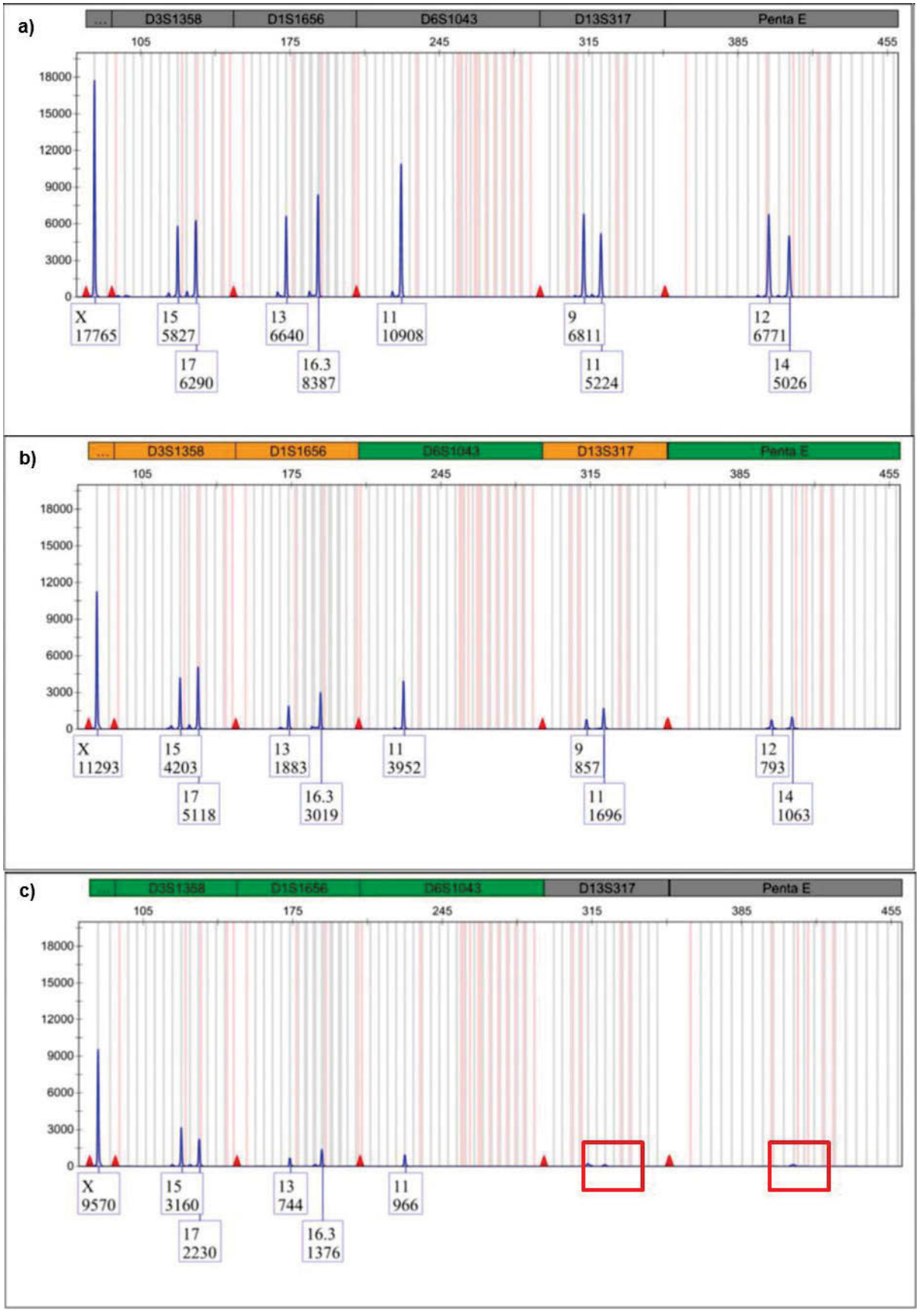


Figure 15: Individual Locus Detection. The individual detection rate for loci included in the PowerPlex 21® System (Promega Corporation, Wisconsin, USA) is reported for 25 artificially degraded DNA samples. Loci are arranged in order of increasing fragment size (bp). Locus Detection Rate is calculated by the count of alleles at respective sites (homozygote alleles, counted twice) (*n*=50).

An example of allele drop-outs is displayed in Figure 16 for one of the biological replicates. For this individual, alleles for Amelogenin, D3S1358, D1S1656, D6S1043, D13S317 and Penta E were reportable up to 30 minutes. At 60 minutes of heat treatment, four alleles had dropped-out at the D13S317 and Penta E STR locus. At 120 minutes of heat treatment, two additional alleles had dropped out at the D6S1043 locus. By 240 minutes treatment, the D1S1656 locus had also completely dropped out.


a)
...
D3S1358
D1S1656
D6S1043
D13S317
Penta E
105
175
245
315
385
455
18000
15000
12000
9000
6000
3000
0
X
17765
15
5827
17
6290
13
6640
16.3
8387
11
10908
9
6811
11
5224
12
6771
14
5026
b)
...
D3S1358
D1S1656
D6S1043
D13S317
Penta E
105
175
245
315
385
455
18000
15000
12000
9000
6000
3000
0
X
11293
15
4203
17
5118
13
1883
16.3
3019
11
3952
9
857
11
1696
12
793
14
1063
c)
...
D3S1358
D1S1656
D6S1043
D13S317
Penta E
105
175
245
315
385
455
18000
15000
12000
9000
6000
3000
0
X
9570
15
3160
17
2230
13
744
16.3
1376
11
966

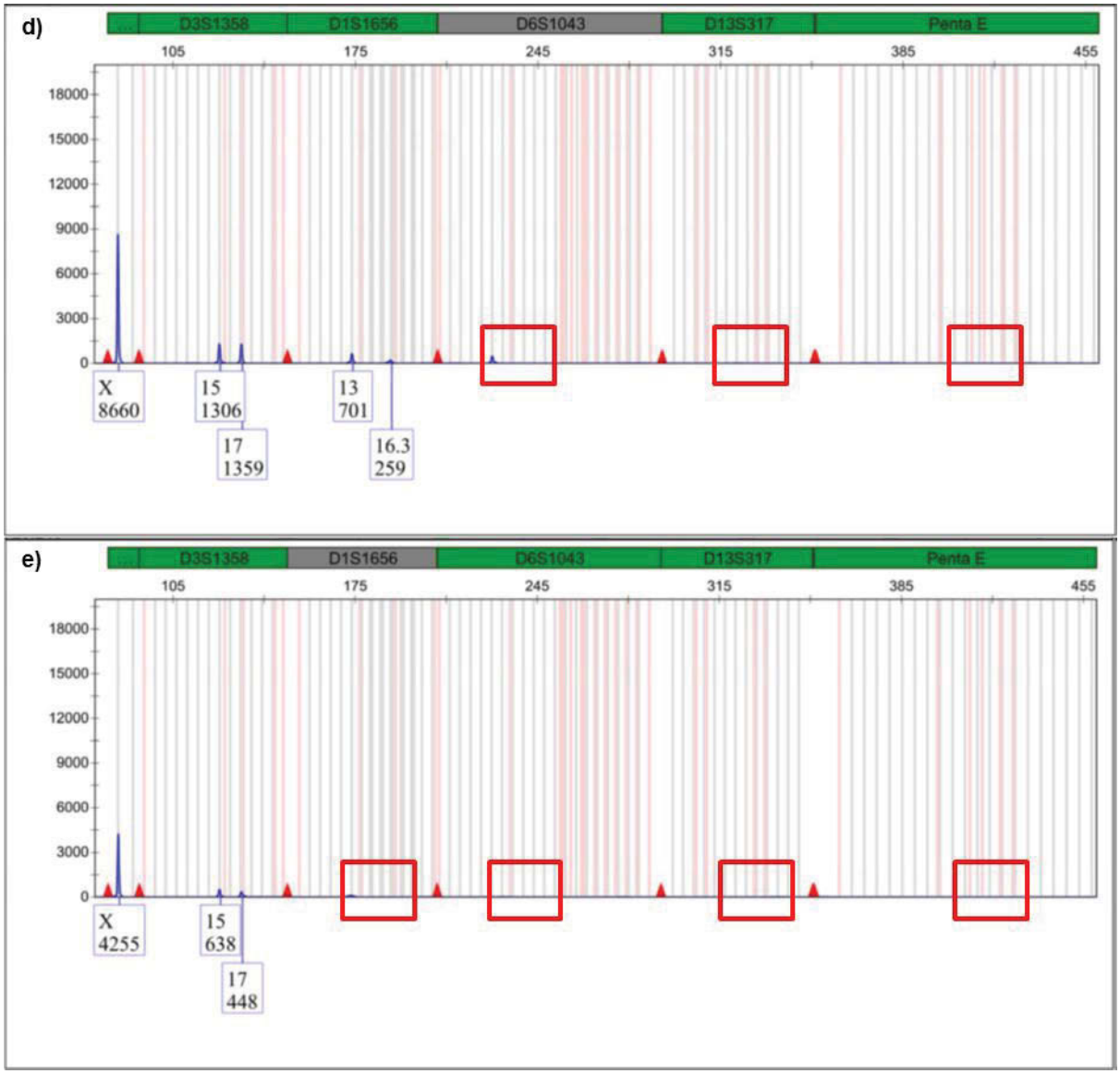


Figure 16: Comparison of allele drop-outs across variable heat treatment times. Samples are heated for **a)** 0 minute, **b)** 30 minutes, **c)** 60 minutes, **d)** 120 minutes and **e)** 240 minutes at 125°C and compared for allele drop-outs (highlighted in red). For simplicity, the DNA profiles display the five loci (amelogenin, D3S1358, D1S1656, D6S1043, D13S317 and Penta E) which are labelled with the fluorescein (blue) dye.

In one instance, a drop-out of two alleles (9, 12) were called with two contaminating alleles (8, 13) at the D7S820 locus (Figure 17). This highlights the potential for miscalling alleles in degraded samples. For another sample at 60 minutes of heat-treatment, a single allele drop-in was reported at the FGA locus. Overall, these instances of allele drop-outs were frequently seen at 60, 120 and 240 minutes treatment.

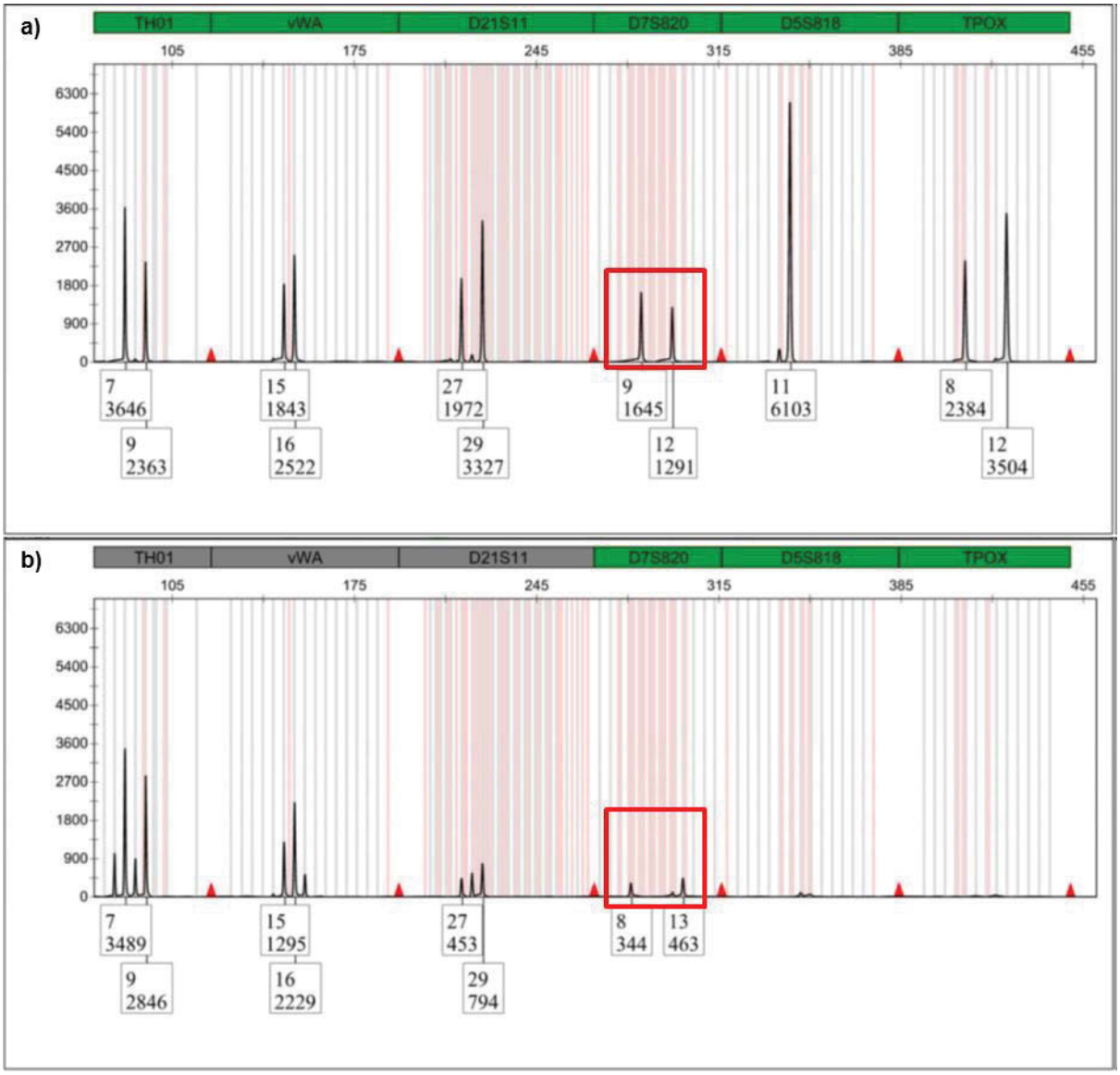


Figure 17: Calling of allele drop-ins. Calling of allele drop-ins (8, 13) in place of alleles (9, 12) at the D7S820 locus (highlighted in red). Profiles are of DNA extracts heated at 125°C for **a)** 0 minutes and **b)** 120 minutes. The DNA profiles only display the five loci (TH01, vWA, D21S11, D7S820, D5S818 and TPOX) which are labelled with the TMR-ET (yellow) dye.

4. Discussion

DNA profiling provides a degree of discrimination that is capable of identifying individuals in criminal and civil matters. However, in cases where genomic DNA fails to produce a reliable profile, alternative non-autosomal markers such as mtDNA can be examined for potential identifications [91]. Yet before pursing this analysis, genomic DNA profiles should be assessed and deemed unusable as a means to direct sample processing. In this study, the quality of genomic DNA profiles for samples previously degraded by a thermal treatment (Chapter 2) is reported and assessed for potential mitochondrial sequencing. The parameters used to

evaluate the profile quality and its suitability for forensic genotyping were peak height and the percentage of alleles called. There was also an assessment of allele drop-outs and drop-ins. The DNA samples were amplified at loci included in the PowerPlex® 21 System (Promega Corporation, Wisconsin, USA). The products were resolved on the 3500 Genetic Analyser® (Applied Biosystems, California, USA) and the profiles were analysed as peaks on electropherograms [92, 93].

The height of a peak is measured in relative fluorescence units (RFU) and represents the relative amount of fluorescent dyes, which are covalently attached to a DNA fragment. The height of a peak has been shown to be approximately proportional to the input amount of template DNA as well as its integrity and size [94]. This relationship has found peak heights to be relatively constant across all loci of increasing molecular weights for DNA of pristine quality [87] [95]. Though there is a preferential amplification of smaller alleles. The quality of DNA can be assessed as a degradation slope where the relationship between peak height and molecular weight has been found to best follow an exponential or 'ski' slope model [88].

The degradation slope for control samples at 0 minutes was shown to be -1.58 RFU/bp, indicating a relative reduction in peak height as allele size increases [87]. The slight slope decrease is considered reasonable as incomplete PCR efficiencies tends to preferentially amplify lower molecular weight markers [90, 96]. By extension it has been established that the relative peak height difference between markers of low and high molecular weight becomes overstated in fragmented and degraded samples [90, 96]. These results are supported by the lower degradation slopes recorded for heat treated samples (-21.61 for 30 minutes, -21.71 for 60 minutes, -25.84 for 120 minutes and -24.98 for 240 minutes).

As well as a low degradation slope, treated samples recorded lower peak heights when compared to controls. Notably, samples treated at 60, 120 and 240 minutes displayed an overall peak height of 3,340 RFU or lower. At this height, allele calling was reduced as some peaks were below thresholds for allele assignment. Though the calling threshold set was arbitrary (175 RFU for heterozygote, 700 RFU for homozygote), it was set in view of the values recommended by the kit manufacturer as well as the values used by one of the Australian DNA laboratories [3]. At this threshold, the percentage of alleles that were reported was shown to progressively decrease with prolonged treatments. Samples degraded at 240 minutes recorded the most allele drop-outs, with only 40 ± 4% alleles called in these cases. With profiles limited to a few alleles, it is expected that these DNA samples would benefit from mtDNA sequencing as additional information such as ancestry and matrilineal identifications can be inferred.

The majority of allele drop-outs occurred in high molecular weight loci (typically greater than 128 bp). In all samples, only four of 21 (19%) loci were reportable at THO1 (71 bp), D8S1179 (77 bp), D16S539 (78 bp), Amelogenin (84 bp) and D3S1358 (98 bp). The recovery of the four STR loci in all samples is consistent with profiles reported in compromised samples [4]. More so, the gradual loss of peak height in large alleles is common in fragmented DNA samples. This reflects the strand breaks in DNA that reduce the efficiency of PCR and a limitation of kits to amplify STR loci that are larger than the fragments of template DNA [95]. Hence, a reduction in the number of alleles called is reported for the compromised samples.

Given that thermal degradation breaks DNA into a number of random fragments, the addition of more template DNA may increase peaks to a threshold, which can be detected. This largely depends on the sensitivity of the kit [95]. However, it is often the case that forensic samples encountered in an operational setting are already compromised [70]. Alternatively, if DNA template is limited, considerations can be made to increase the number of amplification cycles (> 30 cycles) which may improve the overall yield for detection. This is referred to as low copy number DNA profiling, which can result in the accumulation of artifacts. Another possible solution would be to modify the design of the amplification panel to type smaller amplicons that have a higher chance to remain intact in degraded samples [97].

A reduced number of alleles were observed with an increase of amplification artifacts. These artifacts were characteristic of degradation and include heterozygote imbalance and allele drop-ins and drop-outs. Heterozygote balance is defined as the ratio between two peak heights of heterozygote alleles. A heterozygote balance of 1 indicates a complete balance between the heterozygote alleles. Heterozygote imbalance is commonly caused by the sampling of non-equal amounts of starting template. Through PCR, one allele is preferentially amplified to the other resulting in an imbalance in peak height. At times, the imbalance led to the calling of low peaks as allele drop-outs, and erroneously designated a single allele at a heterozygote site. This was resolved by a manual inspection and assignment to the reference profile. The factors shown to increase heterozygote imbalance include low template input, PCR inhibition and DNA fragmentation, which would be most applicable in this case given the prior treatment of profiled samples [95, 98]. The observance of heterozygote imbalance especially in prolonged treated samples, presented a significant challenge to interpreting profiles. In turn, a great deal of sample analysis was required to assign true allele peaks and affected the overall quality of the profiles.

Furthermore, a number of additional DNA peaks were profiled in the heat-treated samples. These peaks can be seen at the D7S820 locus where two heterozygote peaks (8, 13) were called in place of the reference alleles (9, 12) (Figure 17). It is possible that these additional

peaks are typed from an external contaminating DNA source. This is a common phenomenon especially when analysing DNA samples of a trace or degraded nature. This is because the original degraded DNA sample is outcompeted by the contaminating DNA and becomes amplified. Hence, the contaminating alleles become pronounced while the original alleles are reduced or have dropped-out.

It would be useful to test if the observed heterozygote balance as well as peak height and allele drop-out effects continue for samples degraded beyond those tested in this study. To date, limited studies have examined the degradation performance of the PowerPlex® 21 System (Promega Corporation, Wisconsin, USA) [88]. A model predicting genomic DNA profile recovery from degradation, with respect to the kit, would be highly useful to forensic practitioners. While kit robustness has been validated for reaction volumes, mixtures, direct amplification, reproducibility and concordance, degradation studies will offer further insights into kit capabilities and constraints. For this, it is recommended that a similar study is repeated with a larger set and range of degraded samples and thereby build on the preliminary results [78]. However, this remains outside the scope of the current project.

In summary, forensic genotyping has revealed important insights into the recovery of genomic profiles for thermally degraded DNA samples. Firstly, it reaffirms preferential amplification of lower molecular weight markers that leads to the observance of greater peak heights. Secondly, peak height was correlated to the observance heterozygote imbalance, allele drop-ins and allele drop-outs, especially in degraded samples. Thirdly, samples treated beyond 30 minutes failed to call all alleles found in reference samples. It is expected that these treated samples would benefit most from mtDNA typing. Hence, mitochondrial genomes were sequenced on an MPS platform and reported in Chapter 5.

Chapter 5: Sequencing of Mitochondrial DNA

This chapter titled ‘Sequencing of Mitochondrial DNA’ provides an evaluation of the performance of the Early Access AmpliSeq™ Mitochondrial Panel for amplification of reference and heat-treated DNA samples for sequencing on MPS platforms. It has been accepted for publication (elps.201700371.R2) in a special issue of Electrophoresis, dedicated to *Massively Parallel Sequencing*.

Performance of the Early Access AmpliSeq™ Mitochondrial Panel with degraded DNA samples using the Ion Torrent™ platform

Ka Tak Wai[1] · **Mark Barash**[1] · **Peter Gunn**[1]

[1]University of Technology Sydney, Centre for Forensic Science, Ultimo NSW 2007, Australia

Mr. Ka Tak Wai

Centre for Forensic Science
University of Technology Sydney
P.O. Box 123
Ultimo NSW 2007 Australia

Ka.T.Wai@student.uts.edu.au
ORCHID: 0000-0003-0109-4048

Abbreviations: DI, degradation index; **HVR**, hypervariable region; **INDEL**, insertion and deletion; **MPS**, massively parallel sequencing; **mtDNA**, mitochondrial DNA; **rCRS**, revised Cambridge Reference Sequence

Keywords: Degraded DNA / Forensic / Ion Torrent PGM™ / Massively parallel sequencing / Mitochondrial DNA

Word count: 5,571

The Early Access AmpliSeq™ Mitochondrial Panel amplifies whole mitochondrial genomes for phylogenetic and kinship identifications, using Ion Torrent™ technology. There is currently limited information on its performance with degraded DNA, a common occurrence in forensic samples. This study evaluated the performance of the Panel with DNA samples degraded *in vitro*, to mimic conditions commonly found in forensic investigations. Purified DNA from five individuals was heat-treated at five time points each (125°C for 0, 30, 60, 120 and 240 minutes; total *n*=25). The quality of DNA was assessed via a real-time DNA assay of genomic DNA and prepared for massively parallel sequencing on the Ion Torrent™ platform. Mitochondrial sequences were obtained for all samples and had an amplicon coverage averaging between 66X to 2,803X. Most amplicons (157/162) displayed high coverages (452 ± 333X), while reads with less than 100X coverage were recorded in 5 amplicons only (90 ± 5X). Amplicon coverage was decreased with prolonged heating. At 72% strand balance, reads were well balanced between forward and reverse strands. Using a coverage threshold of 10 reads per SNP, complete sequences were recovered in all samples and resolved kinship and, haplogroup relations. Additionally, the HV1 and HV2 regions of the reference and 240 minute heat-treated samples (*n*=10) were Sanger-sequenced for concordance. Overall, this study demonstrates the efficacy of a novel forensic Panel that recovers high quality mitochondrial sequences from degraded DNA samples.

1 Introduction

There is value in using mitochondrial DNA (mtDNA) for ancestry and kinship analysis. This is true in cases of unsuccessful nuclear DNA typing or when there is lack of reference profiles [1]. Traditionally, mtDNA sequencing has been restricted to the hypervariable regions (HVR) that exhibit the most polymorphisms [2]. However, the control region covers only 7% of the human mitochondrial genome and limits the resolution of matrilineal and ancestry inferences [3-5]. Whole genome sequencing offers additional sequences from which mitochondrial haplogroups can be determined to the highest resolution [6, 7].

For the last forty years, chain-terminating technology has been a consensus method of DNA sequencing [8]. However, using this technology to sequence whole genomes can be highly labour intensive and costly. Alternatively, massively parallel sequencing (MPS) can retrieve billions of ssDNA molecules, amplified from panels containing targeted or whole genome markers [9]. Both targeted and whole genome panels have been used as *in-house* methods to amplify mitochondrial genomes for sequencing [7, 10-12]. However, based on panel design, the performance of *in-house* panels can be highly variable in accuracy, coverage and strand balance. This is not optimal for forensic laboratories that require a standard panel of optimised performance. Therefore a standard panel is required to normalise mitochondrial MPS testing across forensic DNA laboratories.

The Early Access AmpliSeq™ Mitochondrial Panel (Applied Biosystems, CA, USA) offers a solution to standardising MPS testing of mitochondrial genomes. The Panel uses 162 primer pairs to amplify whole mitochondrial genomes for MPS uses. To date, only earlier versions of the Panel have been evaluated using degraded DNA samples [13]. This study assessed the performance of the MPS panel to amplify mitochondrial genomes in reference and degraded DNA samples. Coverage, strand balance and accuracy of variant calls for familial and phylogenetic relations were evaluated and the potential of the MPS Panel to supplement the CE workflow of operational forensic laboratories is discussed.

2 Materials and methods

2.1 Samples

Saliva samples (Classiq Swabs™, Copan Diagnostics, CA, USA) were collected from one unrelated individual of Swiss ancestral background and four related Fijian-Indian individuals. The ancestry information was self-reported and based on the biogeographical ancestry of all five grandparents. DNA samples were collected with informed consent and handled according to ethical procedures approved by the Human Ethics Committee for the University of Technology Sydney (Approval Number: 2015000296). DNA was extracted using the QIAamp® DNA Mini Kit (Qiagen, Hilden, Germany), following the manufacturer's protocol for buccal swabs [14]. Equal volumes of each biological extract were heat-treated at 125°C for 30, 60, 120, 240 minutes. An untreated sample for each individual was used as a reference to artificial degradation (total *n*=25). The quantity and degradation of DNA was measured using the Quantifiler™ Trio DNA Quantification Kit (Life Technologies, CA, USA), on the QuantStudio™ 6 Flex Real-Time PCR System (Applied Biosystems, CA, USA), following manufacturer's protocol [15]. DNA concentrations were normalised (4 ng/µL) and stored at -20°C.

2.2 Library preparation

Each DNA template was amplified for 162 mitochondrial amplicons using two primer pools included in the Early Access AmpliSeq™ Mitochondrial Panel (Applied Biosystems, CA, USA) in two separate reactions. Following the manufacturer's 2-in-1 method for low copy number samples, 0.1 ng gDNA of each sample was combined with 5X Ion AmpliSeq™ HiFi Mix and 10X primer pool, included in the Precision ID Library Kit (Applied Biosystems, CA, USA) [16]. Amplification included 2 minutes at 99°C followed by 21 cycles of 15 seconds at 99°C, 4 minutes at 60°C and overnight hold at 10°C on the Veriti® 96-Well Thermal Cycler (Applied Biosystems, CA, USA). Products from two primer pools were combined for each respective sample and post-PCR primers were digested with FuPa reagent included in the kit.

Amplicons were ligated to Ion P1 Adapter and Ion Xpress™ Barcode adapters and, purified using AMPure™ XP reagent (Beckman Coulter, CA, USA). Size (bp) of library fragments were

assessed using the Agilent High Sensitivity DNA Kit on the Agilent 2100 Bioanalyser (Agilent Technologies, CA, USA), following standard protocols [17]. Quantity of libraries was determined using the KAPA SYBR® FAST ABI Prism qPCR Kit (Kapa Biosystems, MA, USA) on the QuantStudio™ 6 Flex Real-Time PCR System (Applied Biosystems, CA, USA) [18]. Libraries were diluted (8 pM) and equal volumes pooled for template preparation.

Pooled libraries (25 µL) were fixed onto template beads and amplified using the Ion PGM™ Template OT2 200 Kit (Applied Biosystems, CA, USA) [19]. Amplification was carried out on the Ion OneTouch™ 2 System (Applied Biosystems, CA, USA). Amplified templates were enriched for monoclonal Ion Sphere™ Particles (ISPs) on the Ion OneTouch™ Enrichment System (Applied Biosystems, CA, USA). Quality of ISPs were evaluated with the Ion Sphere™ Assay (Life Technologies, CA, USA) on Qubit® 2.0 Fluorometer (Invitrogen, CA, USA) [20].

2.3 DNA sequencing

25 mtDNA template libraries were loaded onto a Ion 316™ Chip using the Ion PGM™ Hi-Q™ Sequencing Kit (Applied Biosystems, CA, USA), following the manufacturer's recommendations [21]. Libraries were sequenced on the Ion Torrent PGM™ System (Applied Biosystems, CA, USA) using the protocol for 500 nucleotide flows.

2.4 DNA sequence analysis

DNA sequences were reconstructed by pooling all barcoded libraries to respective samples and trimming adapter sequences 20 bases from the 3' and 5' end, using Torrent Suite™ software (Applied Biosystems, CA, USA). Sequences were formatted to the human mitochondrial genome by alignment to the revised Cambridge Reference Sequence (rCRS) [22, 23]. Sequence variants, SNPs, insertions and deletions (INDELs) were reported using the Ion PGM™ System: Torrent Variant Caller (Applied Biosystems, CA, USA) plugin as variant caller files. Binary alignment map files of aligned sequences and variants were inspected using Integrative Genomics Viewer (Broad Institute, MA, USA) [24, 25]. Variants were imported into MitoTool, a third party online software designed with PhyloTree Build 17, and used for

assignment of mitochondrial haplogroups [26, 27]. HaploGrep 2 (v2.1.0) was used as a secondary confirmation of haplogroups [28]. A minimum arbitrary threshold of 10X coverage reads was used to call mitochondrial variants and a threshold of 0.05 was set for point heteroplasmy detection. In line with forensic convention, length heteroplasmy was reported to the most dominant variant of all detected sequences [29].

2.5 Statistical analysis

Coverages of mitochondrial amplicons were pooled and presented as overall reads for each treatment (n=5). Strand balance was calculated as a percentage of lower reads/higher reads between complementary amplicon strands [30]. ANOVA (one-way) was used to measure strand balance differences across treatments. The statistical tests (α=0.05) were performed using Microsoft Excel 2010 (Microsoft Corporation, WA, USA).

2.6 CE concordant sequence

HV1 (15,971–16,410 rCRS) and HV2 regions (15–389 rCRS) of reference and 240 minute heat-treated samples (n=10) were sequenced by CE systems for concordant data. PCR amplification was prepared with the HotStarTaq Plus Master Mix Kit (Qiagen, Hilden, Germany) using 4 ng genomic DNA and 0.5 µM HV1 and HV2 M13-tailed primers [31, 32]. The Veriti® 96-Well Thermal Cycler (Applied Biosystems, CA, USA) was used for amplification with parameters set to 95°C for 5 minutes followed by 35 cycles of 94°C for 30 seconds, 62°C for 45 seconds, 72°C for 1 minute and final extension of 72°C for 10 minutes [33]. Products were sequenced using the BigDye™ Terminator v3.1 Cycle Sequencing Kit on the 3730XL DNA Analyser (Applied Biosystems, CA, USA) [33]. Raw DNA sequences were analysed and aligned using DNA Sequencing Analysis Software (Applied Biosystems, California, USA).

3 Results and discussion

3.1 DNA quality

The quality of DNA was decreased with heat-treatment, marked by the higher degradation indexes (DI) (Fig. 1) [34].

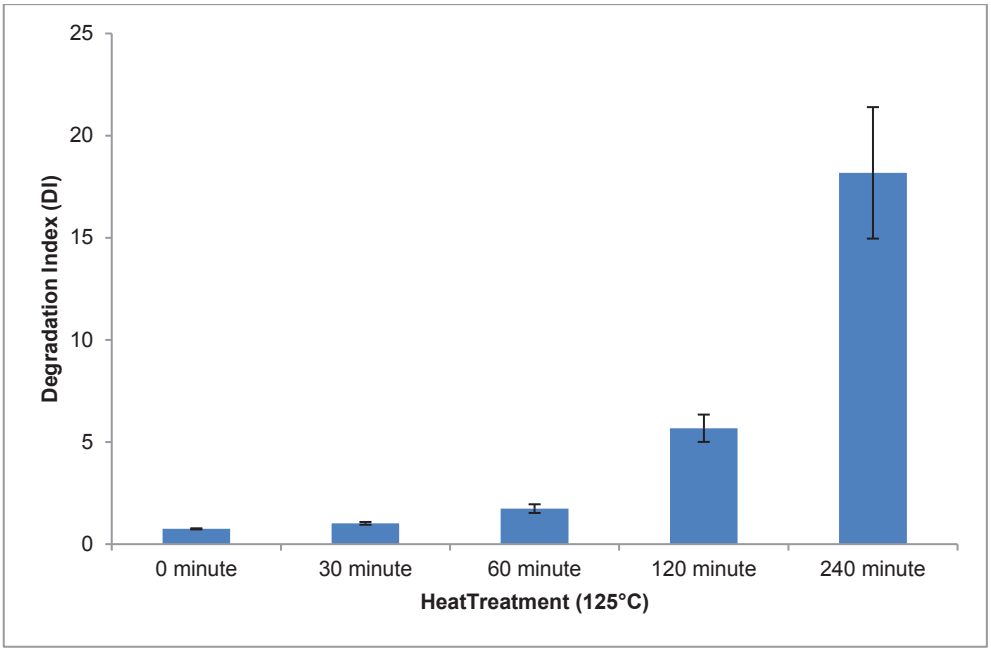


Figure 1. Quality of genomic DNA in heat-treated DNA extracts. Mean (*n*=5) degradation index (DI) of heat-treated DNA samples determined using the Quantifiler™ Trio DNA Quantification Kit (Life Technologies, CA, USA). DI calculated as ratio, small autosomal target/large autosomal target [15]. Error bars represent S.E.M. between replicates.

The DNA quantities (ng/μL) used to derive DI's of each treatment group is provided in Supporting Information Figure 1. Little to no degradation was observed in reference samples (4 ng/μL), with a DI 0.75 ± 0.03 (mean ± S.E.M.). Mild degradation was seen in the heat-treated DNA samples at 30 minutes (DI 1.02 ± 0.07, 3.52 ng/μL) and 60 minutes (DI 1.74 ± 0.21, 2.46 ng/μL). Prolonged heat-treated samples at 120 minutes (DI 5.64 ± 0.67, 1.41 ng/μL) and 240 minutes (DI 18.18 ± 3.22, 0.44 ng/μL) were highly degraded, indicating occurrence of putative DNA fragmentation.

Note in this study, degradation refers to the fragment quality of DNA. While heat-treatment was used for *in vitro* degradation, it is recognised that the method is limited as it does not truly mimic forensic DNA samples. Other factors such as the quantity of DNA and presence of co-inhibitors in extracts can also be encountered in compromised samples. Consequently it is strongly recommended that these variables can be assessed in separate studies.

Degradation of DNA has been previously shown to start at 100°C and become completely degraded with longer exposures [35]. In these cases, heat was found to degrade nucleic acids

into fragments due to DNA depurination, and a break of glycosidic and phosphodiester bonds [36]. Therefore these findings support the progressive degradation of heat-treated DNA samples that were determined from a real-time assay. However, there is a limitation to infer mtDNA quality from genomic real-time assays as mtDNA has been shown to be more stable to degradation [37, 38]. Hence, a real-time duplex assay for mtDNA targets of which few are available, would offer a more accurate assessment of DNA quality [39, 40].

3.2 Sequencing metrics

A total of 173 megabases of MPS data was generated for 25 samples using 500 flow runs. An average live ISP density of 47% (1,503,921 reads) was obtained and 131 million bases were correctly aligned to the rCRS. Mean coverage depth for alignment quality was reported to 7,875X at mean read lengths of 99 bp fragments.

3.3 Coverage

The Panel shows high amplicon coverage for reference mtDNA samples, though coverage reads are non-uniformly distributed, with some reads as low as 66X (amplicon 3,473-3,596 rCRS) and other reads as high as 2,803X (amplicon 10,482-10,577 rCRS) (Fig. 2).

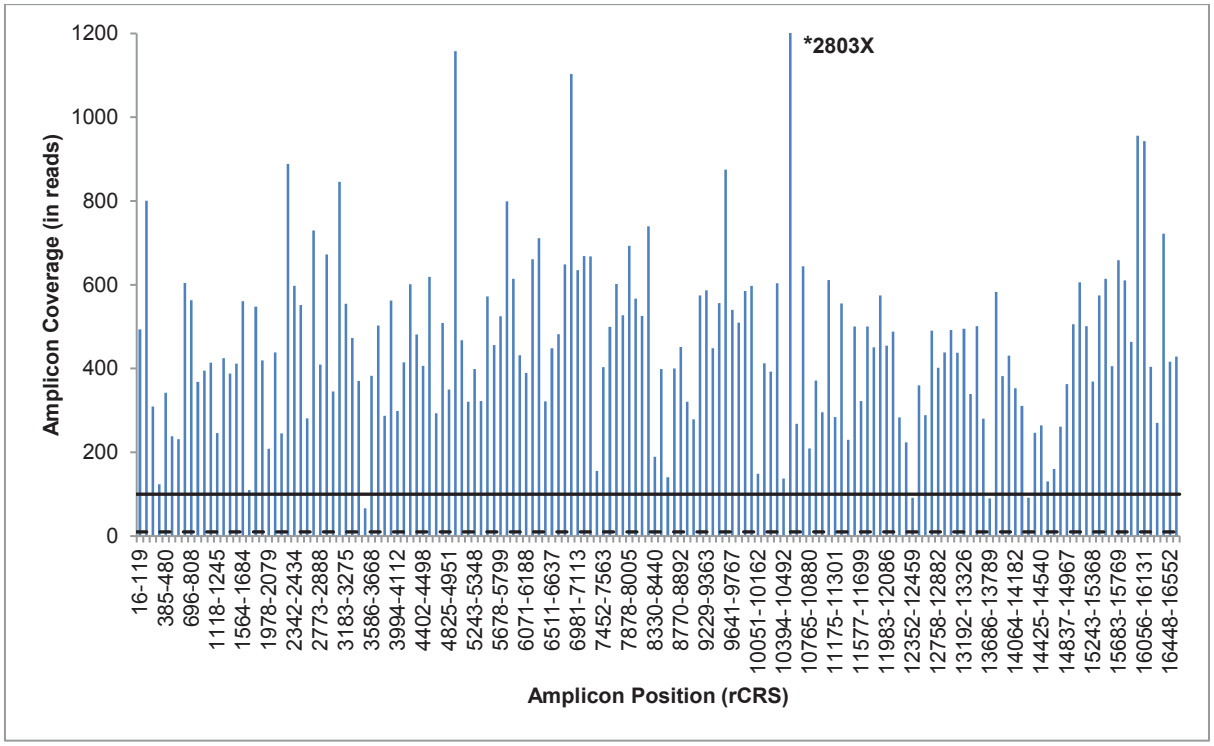


Figure 2. Amplicon Coverage. Average coverage (in reads) of 162 amplicons spanning the mitochondrial genome in reference samples (*n*=5). Amplicons part of the Early Access AmpliSeq™ Mitochondrial Panel (Applied Biosystems, CA, USA). Thresholds for variant calling were arbitrary set at 10X minimum (- - - - -) and 100X maximum (———).

Overall amplicon coverage (*n*=5) was 468 ± 21 reads (mean ± S.E.M.) for time 0 samples, 348 ± 8 reads for 30 minute samples, 395 ± 12 reads for 60 minute samples, 339 ± 12 reads for 120 minute samples and 274 ± 11 reads for 240 minute samples (Fig. 3).

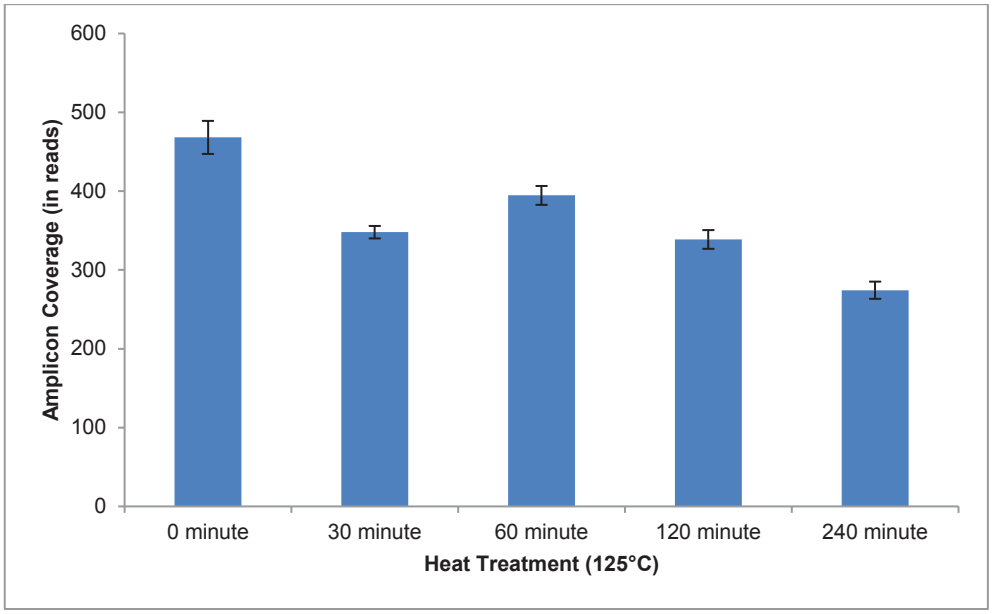


Figure 3. Amplicon Coverage for Degraded DNA. Mitochondrial sequencing amplicon coverage (in reads) of DNA samples (*n*=5) heated at 125°C for 0, 30, 60, 120 and 240 minutes. Amplicons (*n*=162) part of the Early Access AmpliSeq™ Mitochondrial Panel (Applied Biosystems, CA, USA). Error bars represent S.E.M.

A total of 97% of amplicons (157/162) demonstrated reads greater than 100X on both strands. Only five amplicons (3,473-3,596 rCRS, 10,394-10,492 rCRS, 12,352-12,459 rCRS, 13,686-13,789 rCRS and 14,276-14,367 rCRS) were under-reported with reads below 100X. While there was a marginal decrease in reads for prolonged heating times, the decrease did not affect the detection of variants. This trend agrees with a previous study that recovered full mitochondrial genomes for DNase-treated samples (at 50 reads) [13]. In comparison, at a coverage threshold of 50 reads, 92% (23/25) of samples in this study displayed complete genomes and reported all variants. When lowered to 10 reads, all samples were sequenced for complete genomes (Supporting Information Table 1).

The difference in recovery of genomes may be explained by sequencing samples of varying qualities on a single chip. The multiplexing of different quality samples has been considered to reduce read quality as there is a preferential amplification towards high quality DNA samples [41]. Consequently separating samples of pristine and compromised quality may improve sequence recovery. Alternatively reducing the number of samples may also improve coverage and sequence quality as more reads can be assigned to each sample [42, 43]. Though chip density was not maximised (ISP, 47%) which indicates manual library preparation and chip

loading could be further optimised or automated. The highest number of samples multiplexed in a sequence run has been previously reported to be 15 samples on a single Ion 316™ Chip, albeit using a different custom MPS panel [44]. Here it is demonstrated that chip capacity can be extended to at least 25 samples, using this whole genome Panel. Therefore it is suggested that the pooling of samples of similar qualities as well as reducing the number of samples in sequence runs, may result in overall higher coverages.

The low reads reported for two samples may indicate potential nucleotide sites (14,766 and 14,783 rCRS) which are difficult to sequence. Both sites follow a 4-mer poly-A-stretch (AAAA) region. Studies have suggested that it is the homopolymer stretch that lowers the efficiency of reads in this region [7, 44]. Low coverage in homopolymer regions of the Ion PGM™ or other MPS platforms has been described, as have high coverages in amplicons with few homopolymeric repeats [45, 46]. Bragg et al. [47] suggest the most likely explanation to be an inaccurate flow-call of the proton-based system. As well as homopolymer stretches, G/C content and sample quality are also known to affect the accuracy of flow-calls [45]. The over- and under- calling of nucleotides would also explain the non-uniform distribution of amplicon reads that were observed in homopolymeric regions as it seems that amplicon reads are mostly affected by sequence composition over other variables such as base size. The two samples which failed to reproduce complete genome sequences were both heat-treated at 125°C for 240 minutes. Naturally, DNA samples of poor or degraded qualities are likely to record low to no reads, because of the fragmented state of the DNA [12, 13, 48]. Therefore the observed lower coverages most likely arise from a complication in the assigned flow-call algorithm and DNA quality as opposed to the design of the Panel.

Point heteroplasmy was detected in all samples (ranges, 1-16,569 rCRS) at a threshold of 0.05. Three point heteroplasmic variants were each detected in the Swiss individual (214R, 750R, 16221Y) and the Fijian-Indian father (750R, 10586R, 15043R), while the same point heteroplasmy was shared between the Fijian-Indian mother, son and daughter (750R, 15043R, 15258Y). Manual inspection of mitochondrial variants showed INDELs leading to

length heteroplasmy. All reference samples were found to contain an uninterrupted C stretch at HV2 (303-315 rCRS) as 310C, 315.1C and 315.2C. Additionally one heat-treated sample showed deletions, 523- at 212 reads and 524- at 214 reads (Supporting Information Figure 2).

3.4 Strand balance

Amplicons of the Panel were well balanced between forward and reverse strands, though balance was non-uniform across the mitochondrial genome. For five reference samples, strand bias (<50%) was observed in 19/162 (12%) amplicons. In particular, three amplicons (299-411 rCRS, 13,686-13,789 rCRS and 14,276-14,367 rCRS) showed extreme strand bias (<10%) (Fig. 4).

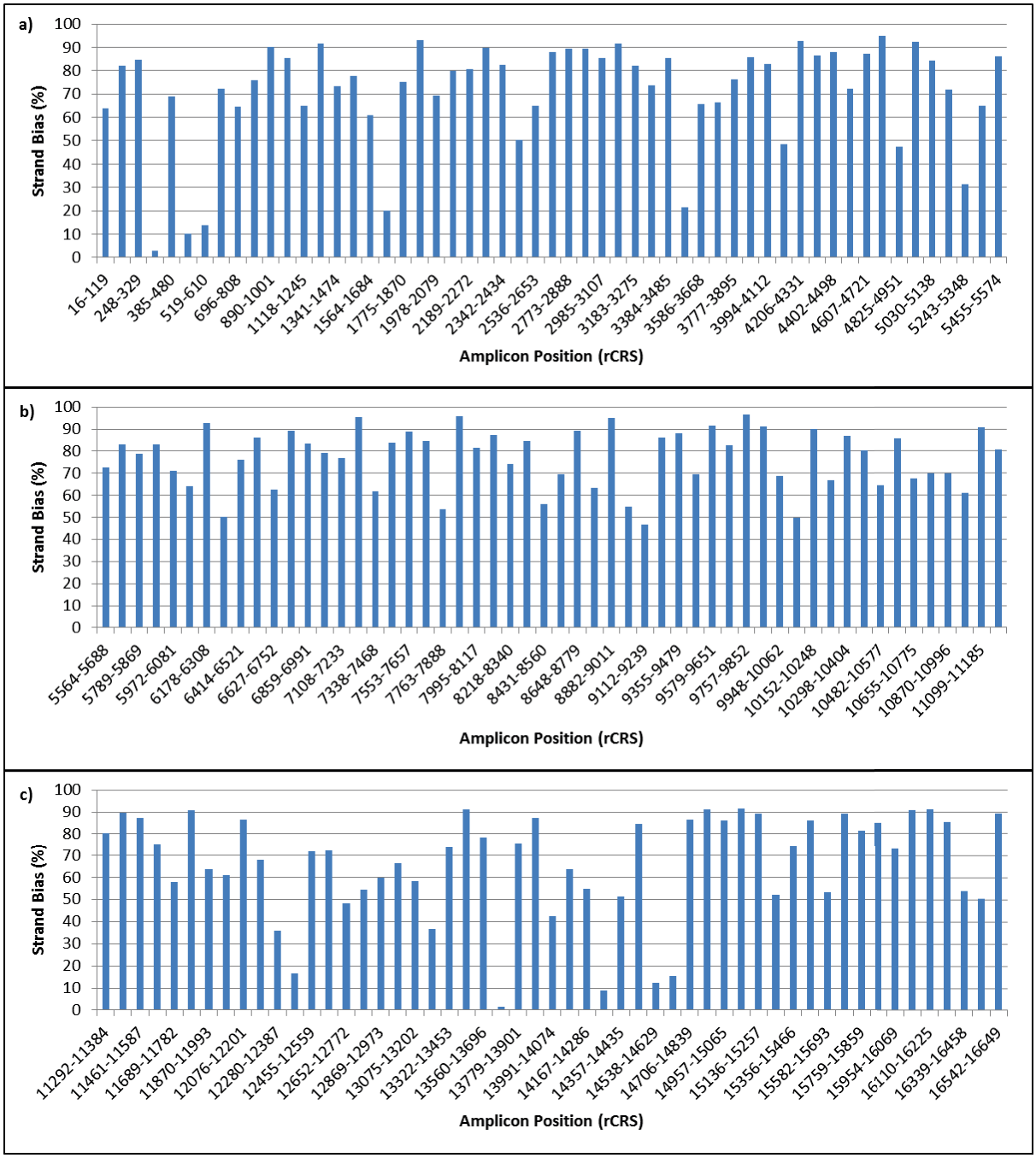


Figure 4. Overall amplicon strand bias for five reference samples. Average strand bias (*n*=5) of mitochondrial amplicons part of the Early Access AmpliSeq™ Mitochondrial Panel (Applied Biosystems, CA, USA). Strand bias (%) between forward strand and reverse strands was calculated as lower reads/higher reads. **a)** amplicons 16-119 rCRS to 5,455-5,574 rCRS, **b)** amplicons 5,564-5,688 rCRS to 11,175-11,301 rCRS, **c)** amplicons 11,292-11,384 rCRS to 16,542-166,496 rCRS.

While most strands were balanced in this study, extreme strand bias has led to the erroneous designation of SNPs in other MPS panels [49]. Other studies using different *in-house* mitochondrial panels have also observed imbalance at similar positions suggesting that strand bias in these regions is inherent in the sequence of DNA [44]. Since strands are equal in length

and complementary, strand bias can quite possibly be sequence and platform dependent [50, 51]. Previously, strand bias has been attributed to the multiple continuous stretches of homopolymers in the DNA region [44]. This is plausible as homopolymers have been shown to reduce reads more so in strands with poly-C-stretches than any other repeats [30].

Overall strand bias was 72% ± 14% (mean ± S.E.M.) for reference samples and 60% ± 3% for heat-treated samples at 30, 60, 120 and 240 minutes (*n*=5) (data not shown). Strand bias was significantly increased in degraded samples (*P*=0.000, one-way ANOVA). Originally it was hypothesised that strand bias would remain unaffected by the nature of the sample (degraded or non-degraded) as forward and reverse strands undergo the same treatment. However, random breakage of covalent bonds in strands is increased with heat-treatment and thereby may account for the imbalanced reads [35, 52].

3.5 Variant detection

As expected more variants were detected in the whole mitochondrial genome than the HV1 (15,971-16,410 rCRS) and HV2 (15-389 rCRS) regions. In non-treated samples, the number of detected HVR variants compared to the whole mitochondrial genome were 7/15 for sample 1, 8/33 for sample 2, 8/34 for sample 3, 8/33 for sample 4 and 8/33 for sample 5 (Table 1).

Table 1. Haplogroup and phylogenetic summary of individual reference samples sequenced for whole mitochondrial genomes (1-16,569 rCRS)

Sample	Variants[a] (rCRS)	Private Mutations (rCRS)	HV1 and HV2 Haplogroup	Whole Mitochondrial Haplogroup (Ancestry)
1: Swiss Individual	**214R**, **263G**, **310C**, **315.1C**, **315.2C**, 750R, 1438G, 4769G, 8860G, 13830C, 14470A, 15326G, **16093C**, **16221Y**, 16519C	-	H10	H10e1 (Europe/UK) [57]
2: Fijian-Indian Mother	**73G**, **195A**, **263G**, **310C**, **315.1C**, **315.2C**, 489C, 524-, 750R, 1438G, 1598A, 2706G, 4769G, 7028T, 8701G, 8860G, 9540C, 10398G, 10400T, 10873C, 11719A, 12007A, 12705T, 14766T, 14783C, 15043R, 15259Y, 15301A, 15326G, 15431A, **16179-**, **16223T**, 16519C	16179-	M30	M30d1 (South/SE Asia) [58]
3: Fijian-Indian Father	**73G**, **152C**, **263G**, **310C**, **315.1C**, **315.2C**, 489C, 709A, 750R, 1438G, 1888A, 2706G, 3921T, 4769G, 7028T, 8701G, 8860G, 9540C, 10398G, 10400T, 10586R, 10873C, 11719A, 12477C, 12705T, 14323A, 14766T, 14783C, 15043R, 15301A, 15326G, **16129A**, **16223T**, 16519C	152C	M5a	M5a (India/South Africa) [58]

4: Fijian-Indian Son	**73G**, **195A**, **263G**, **310C**, **315.1C**, **315.2C**, 489C, 524-, 750R, 1438G, 1598A, 2706G, 4769G, 7028T, 8701G, 8860G, 9540C, 10398G, 10400T, 10873C, 11719A, 12007A, 12705T, 14766T, 14783C, 15043R, 15259Y, 15301A, 15326G, 15431A, **16179-**, **16223T**, 16519C	16179-	M30	M30d1 (South/SE Asia) [58]
5: Fijian-Indian Daughter	**73G**, **195A**, **263G**, **310C**, **315.1C**, **315.2C**, 489C, 524-, 750R, 1438G, 1598A, 2706G, 4769G, 7028T, 8701G, 8860G, 9540C, 10398G, 10400T, 10873C, 11719A, 12007A, 12705T, 14766T, 14783C, 15043R, 15259Y, 15301A, 15326G, 15431A, **16179-**, **16223T**, 16519C	16179-	M30	M30d1(South/SE Asia) [58]

[a]HV1 (15,971-16,410 rCRS) and HV2 (15-389 rCRS) variants are indicated in **bold**. Each sample is identified with the individual's declared ancestry. Based on detected variants (SNPs, INDELs) within the HV1 and HV2 region or whole mitochondrial genome, a most probable haplogroup is assigned with the biogeographical ancestry. Private mutations (rCRS) list variants in samples, which are absent in predicted haplogroup.

These variants were mostly concordant to the HV1 and HV2 CE sequences of reference and 240 minute heat-treated samples. Compared to the typing of single DNA molecules in CE systems, MPS platforms have a greater sensitivity to detect intra-individual sequences as it types a multitude of DNA fragments [53]. This greater sensitivity accounts for the differences in base calls that were mostly found in heteroplasmic positions, such as 214R and 16043R. Other sequence differences were due to unreported bases by the CE system, mostly likely arising as artifacts of sequencing.

Using whole genome sequencing, a remaining 109 variants were called in addition to HV1 and HV2 regions. While mitochondrial haplogroups were inferable from HV1 and HV2 sequences, the additional variants were able to increase the resolution of haplogroups to a phylogenetic relevance (Table 1). This highlights the benefits of a whole genome panel over HV1 and HV2 sequencing, which is not recommended in forensic casework [29]. Variant assignment of mitochondrial haplogroups was consistent and accurate between reference and degraded samples of the same individual. A few variants not previously observed in haplogroups were seen (Table 1). This was expected as it is common for individuals to acquire private mutations that differ to the overall variants of historical haplogroups because of the high mutation rate of mtDNA [30, 54-56]. Nonetheless, it has been shown that a majority of detected variants will allow haplogroups to be assigned with a sufficient reliability. Specifically, mtDNA sequences of high scoring quality (>90% quality, HaploGrep) have been shown to correctly assign haplogroups [30]. Herein, the quality scores of variants were >90% and thereby indicates a reliable and, accurate alignment of variants to haplogroups.

3.5.1 Kinship analysis

Kinship relations were correctly assigned to individuals of the same maternal lineage. Within the five individuals, five unique variants were detected for Individual 1 and nine unique variants for Individual 2 which could be used for kinship exclusion. All heat-treated samples were used for kinship analysis at 10X minimum coverage. The results demonstrated that even the prolonged-treated samples could be used for exclusion of the Swiss individual and Fijian-Indian father as maternal lineages of the Fijian-Indian mother, son and daughter, who all shared the same haplogroup (M30d1) (Table 1) [57, 58].

3.5.2 Phylogenetic analysis

Haplogroup phylogenies were accurately identified to declared ancestries (Table 1). Although haplogroups for four related Fijian-Indian individuals were distributed across South/SE Asia and South Africa, the admixed ancestry has been previously observed [58]. Most likely the

admixture results from a migration of the M haplogroup from Asia into the pacific islands, started by the African expansion and continued from British colonisation [58]. Nonetheless, the Panel shows diverse coverage of haplogroups of different ancestries, in this case, of European and Fijian-Indian ancestries. Previously the earlier version of this Panel has inferred genetic ancestries in South Africa, Russia, Israel, New Guinea, Algeria, China, Italy, China, Australia, Thailand and the Netherlands [13]. The additional inference of samples from Swiss and Fijian-Indian heritage in this study shows the Panel can be further applied to worldwide haplogroups.

3.6 Workflow

The Panel workflow was completed in 5 days (from library preparation to sequencing) for two Ion 316™ chips. Manual preparation of the Panel required significant hands-on time and has been shown to cause variability in chip loading and densities which are difficult to reproduce [59, 60]. However, the protocol can be automated onto the Ion Chef™ (Life Technologies, CA, USA) and favour the standard workflow of forensic laboratories. An automated protocol facilitates reproducibility and has also been identified to increase through-put capacities and reduce turnaround times and processing costs compared to Sanger sequencing workflows [59]. For forensic use, issues such as contamination, in particular the carry-over of 'ghost' barcodes as well as the unprecedented volume of MPS data and associated bioinformatics expertise required for analysis relative to CE platforms, have been raised [51]. More-so the logistics of storing MPS files in secured formats are considered. To address contamination, laboratories are considering the preparation of forensic and reference samples in separate sequence runs (private communication). While this may reduce sample contamination, it also results in inter-run variability between different chip uses [41]. Considering this, it is best to multiplex with barcode adapters and pool samples based on chip capacity and samples of similar qualities. This will be most cost-effective and achieve even coverages, as supported by this study. All these challenges have been raised by European laboratories and seconded by Australian laboratories [61]. Australian forensic DNA laboratories are yet to implement MPS

workflows, however ongoing laboratory validations will likely lead to the introduction of MPS and associated panels in the foreseeable future. It is likely that MPS of the HV1 and HV2 regions will first be introduced into the workflow of mtDNA analyses, as this is readily compatible with existing techniques such as dideoxynucleotide sequencing.

4 Concluding remarks

The Early Access AmpliSeq™ Mitochondrial Panel (Applied Biosystems, CA, USA) is able to target and amplify whole mitochondrial genomes using the Ion Torrent™ technology. The Panel amplifies genomes at coverages that can reliably call variants for haplogroups, even in highly degraded samples. The Panel is amendable to the forensic identification of kinship and phylogenetic relations. In view of this performance, it is concluded that the Panel can potentially translate as a commercial and standard workflow into operational forensic laboratories that consider introducing MPS.

5 References

[1] Wilson, M. R., Stoneking, M., Holland, M. M., DiZinno, J. A., Budowle, B., *Crime Lab Digest* 1993, *20*, 68-77.

[2] Horai, S., Hayasaka, K., *American journal of human genetics* 1990, *46*, 828.

[3] Schlebusch, C. M., Naidoo, T., Soodyall, H., *Electrophoresis* 2009, *30*, 3657-3664.

[4] van Oven, M., Vermeulen, M., Kayser, M., *Investigative genetics* 2011, *2*, 6.

[5] Ballantyne, K. N., van Oven, M., Ralf, A., Stoneking, M., Mitchell, R. J., van Oorschot, R. A., Kayser, M., *Forensic Science International: Genetics* 2012, *6*, 425-436.

[6] Coble, M. D., Just, R. S., O'Callaghan, J. E., Letmanyi, I. H., Peterson, C. T., Irwin, J. A., Parsons, T. J., *International journal of legal medicine* 2004, *118*, 137-146.

[7] Park, S., Cho, S., Seo, H. J., Lee, J. H., Kim, M.-Y., Lee, S. D., *Journal of Korean medical science* 2017, *32*, 587-592.

[8] Sanger, F., Nicklen, S., Coulson, A. R., *Proceedings of the National Academy of Sciences* 1977, *74*, 5463-5467.

[9] Yang, Y., Xie, B., Yan, J., *Genomics, proteomics & bioinformatics* 2014, *12*, 190-197.

[10] Ramos, A., Santos, C., Alvarez, L., Nogués, R., Aluja, M. P., *Electrophoresis* 2009, *30*, 1587-1593.

[11] Gabriel, M. N., Huffine, E. F., Ryan, J. H., Holland, M. M., Parsons, T. J., *Journal of Forensic Science* 2001, *46*, 247-253.

[12] Parson, W., Huber, G., Moreno, L., Madel, M.-B., Brandhagen, M. D., Nagl, S., Xavier, C., Eduardoff, M., Callaghan, T. C., Irwin, J. A., *Forensic Science International: Genetics* 2015, *15*, 8-15.

[13] Chaitanya, L., Ralf, A., Oven, M., Kupiec, T., Chang, J., Lagacé, R., Kayser, M., *Human mutation* 2015, *36*, 1236-1247.

[14] Qiagen, *QIAamp® DNA Mini and Blood Mini Handbook*, Qiagen, Hilden 2016.

[15] Thermo Fisher Scientific, *Quantifiler™ HP and Trio DNA Quantification Kits*, Thermo Fisher Scientific, California 2017.

[16] Thermo Fisher Scientific, *Precision ID Panels with Ion PGM™ System*, Thermo Fisher Scientific, California 2016.

[17] Agilent Technologies, *Agilent High Sensitivity DNA Kit Guide*, Agilent Technologies, Waldbronn 2013.

[18] Kapa Biosystems, *Kapa Probe Fast qPCR Master Mix (2X) Kit*, Kapa Biosystems, Massachusetts 2016.

[19] Thermo Fisher Scientific, *Ion PGM™ Template OT2 200 Kit*, Thermo Fisher Scientific, California 2015.

[20] Thermo Fisher Scientific, *Ion Sphere™ Assay on the Qubit™ 2.0 Fluorometer*, Thermo Fisher Scientific, California 2017.

[21] Thermo Fisher Scientific, *Ion PGM™ Hi-Q™ Sequencing Kit*, Thermo Fisher Scientific, California 2017.

[22] Anderson, S., Bankier, A., de Bruijn, M., Coulson, A. R., Drouin, J., Eperon, I., Nierlich, D., Roe, B., Sanger, F., Schreier, P., Smith, A., Staden, R., Young, I., *Nature* 1981, *290*, 457-465.

[23] Andrews, R. M., Kubacka, I., Chinnery, P. F., Lightowlers, R. N., Turnbull, D. M., Howell, N., *Nature genetics* 1999, *23*, 147-147.

[24] Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., Mesirov, J. P., *Nature biotechnology* 2011, *29*, 24-26.

[25] Thorvaldsdóttir, H., Robinson, J. T., Mesirov, J. P., *Briefings in bioinformatics* 2013, *14*, 178-192.

[26] Fan, L., Yao, Y.-G., *Mitochondrion* 2011, *11*, 351-356.

[27] van Oven, M., *Forensic Science International: Genetics Supplement Series* 2015, *5*, e392-e394.

[28] Weissensteiner, H., Pacher, D., Kloss-Brandstätter, A., Forer, L., Specht, G., Bandelt, H.-J., Kronenberg, F., Salas, A., Schönherr, S., *Nucleic acids research* 2016, *44*, W58-W63.

[29] Parson, W., Gusmao, L., Hares, D., Irwin, J., Mayr, W., Morling, N., Pokorak, E., Prinz, M., Salas, A., Schneider, P., *Forensic Science International: Genetics* 2014, *13*, 134-142.

[30] Seo, S. B., Zeng, X., King, J. L., Larue, B. L., Assidi, M., Al-Qahtani, M. H., Sajantila, A., Budowle, B., *BMC genomics* 2015, *16*, S4.

[31] Grisedale, K. S., 2014.

[32] Qiagen, *HotStarTaq® Plus PCR Handbook*, Qiagen, Hilden 2010.

[33] Thermo Fisher Scientific, *BigDye™ Terminator v3.1 Cycle Sequencing Kit*, Thermo Fisher Scientific, California 2016.

[34] Vernarecci, S., Ottaviani, E., Agostino, A., Mei, E., Calandro, L., Montagna, P., *Forensic Science International: Genetics* 2015, *16*, 77-85.

[35] Karni, M., Zidon, D., Polak, P., Zalevsky, Z., Shefi, O., *DNA and cell biology* 2013, *32*, 298-301.

[36] Zhang, L., Wu, Q., *Journal of biosciences* 2005, *30*, 599-604.

[37] Holt, A., Wootton, S. C., Mulero, J. J., Brzoska, P. M., Langit, E., Green, R. L., *Forensic Science International: Genetics* 2016, *21*, 145-157.

[38] Foran, D. R., *Journal of forensic sciences* 2006, *51*, 766-770.

[39] Alonso, A., Martí, P., Albarrán, C., Garcí, P., Garcí, O., de Simón, L. F., Garcí, J., Sancho, M., de la Rúa, C., Fernández-Piqueras, J., *Forensic Science International* 2004, *139*, 141-149.

[40] Kavlick, M. F., Lawrence, H. S., Merritt, R. T., Fisher, C., Isenberg, A., Robertson, J. M., Budowle, B., *Journal of forensic sciences* 2011, *56*, 1457-1463.

[41] Sims, D., Sudbery, I., Ilott, N. E., Heger, A., Ponting, C. P., *Nature Reviews Genetics* 2014, *15*, 121-132.

[42] Fumagalli, M., *PLoS One* 2013, *8*, e79667.

[43] Shearer, A. E., Hildebrand, M. S., Ravi, H., Joshi, S., Guiffre, A. C., Novak, B., Happe, S., LeProust, E. M., Smith, R. J., *BMC genomics* 2012, *13*, 618.

[44] Zhou, Y., Guo, F., Yu, J., Liu, F., Zhao, J., Shen, H., Zhao, B., Jia, F., Sun, Z., Song, H., *Forensic Science International: Genetics* 2016, *22*, 11-21.

[45] Ross, M. G., Russ, C., Costello, M., Hollinger, A., Lennon, N. J., Hegarty, R., Nusbaum, C., Jaffe, D. B., *Genome biology* 2013, *14*, R51.

[46] Churchill, J. D., King, J. L., Chakraborty, R., Budowle, B., *International journal of legal medicine* 2016, *130*, 1169-1180.

[47] Bragg, L. M., Stone, G., Butler, M. K., Hugenholtz, P., Tyson, G. W., *PLoS Comput Biol* 2013, *9*, e1003031.

[48] Parson, W., Strobl, C., Huber, G., Zimmermann, B., Gomes, S. M., Souto, L., Fendt, L., Delport, R., Langit, R., Wootton, S., *Forensic Science International: Genetics* 2013, *7*, 543-549.

[49] Guo, Y., Li, J., Li, C.-I., Long, J., Samuels, D. C., Shyr, Y., *BMC genomics* 2012, *13*, 666.

[50] Quail, M. A., Smith, M., Coupland, P., Otto, T. D., Harris, S. R., Connor, T. R., Bertoni, A., Swerdlow, H. P., Gu, Y., *BMC genomics* 2012, *13*, 1.

[51] Børsting, C., Morling, N., *Forensic Science International: Genetics* 2015, *18*, 78-89.

[52] Poptsova, M. S., Il'icheva, I. A., Nechipurenko, D. Y., Panchenko, L. A., Khodikov, M. V., Oparina, N. Y., Polozov, R. V., Nechipurenko, Y. D., Grokhovsky, S. L., *Sci Rep* 2014, *4*, 1697-1712.

[53] Just, R. S., Irwin, J. A., Parson, W., *Forensic Science International: Genetics* 2015, *18*, 131-139.

[54] Schneider, S., Excoffier, L., *Genetics* 1999, *152*, 1079-1089.

[55] Levin, L., Zhidkov, I., Gurman, Y., Hawlena, H., Mishmar, D., *Genome biology and evolution* 2013, *5*, 876-890.

[56] King, T. E., Fortes, G. G., Balaresque, P., Thomas, M. G., Balding, D., Delser, P. M., Neumann, R., Parson, W., Knapp, M., Walsh, S., Tonasso, L., Holt, J., Kayser, M., Appleby, J., Forster, P., Ekserdjian, D., Hofreiter, M., Schurer, K., *Nature Communications* 2014, *5*, 5631.

[57] Brotherton, P., Haak, W., Templeton, J., Brandt, G., Soubrier, J., Adler, C. J., Richards, S. M., Der Sarkissian, C., Ganslmeier, R., Friederich, S., *Nature communications* 2013, *4*, 1764.

[58] Kayser, M., Brauer, S., Cordaux, R., Casto, A., Lao, O., Zhivotovsky, L. A., Moyse-Faurie, C., Rutledge, R. B., Schiefenhoevel, W., Gil, D., *Molecular biology and evolution* 2006, *23*, 2234-2244.

[59] Mogensen, H. S., Børsting, C., Morling, N., *Forensic Science International: Genetics Supplement Series* 2015, *5*, e368-e369.

[60] Duke, J., Lind, C., Mackiewicz, K., Ferriola, D., Papazoglou, A., Derbeneva, O., Wallace, D., Monos, D., *International journal of immunogenetics* 2015, *42*, 346-358.

[61] Alonso, A., Müller, P., Roewer, L., Willuweit, S., Budowle, B., Parson, W., *Forensic Science International: Genetics* 2017.

Acknowledgements *The provision of the Early Access Panel by Thermo Fisher Scientific is greatly appreciated. The technical support provided by Dr. Beatrice Goh and Dr. Lucy Dagostino of Thermo Fisher Scientific is greatly appreciated. This research is supported by an Australian Government Research Training Program Scholarship.*

Conflict of interest *The authors have declared no conflict of interest.*

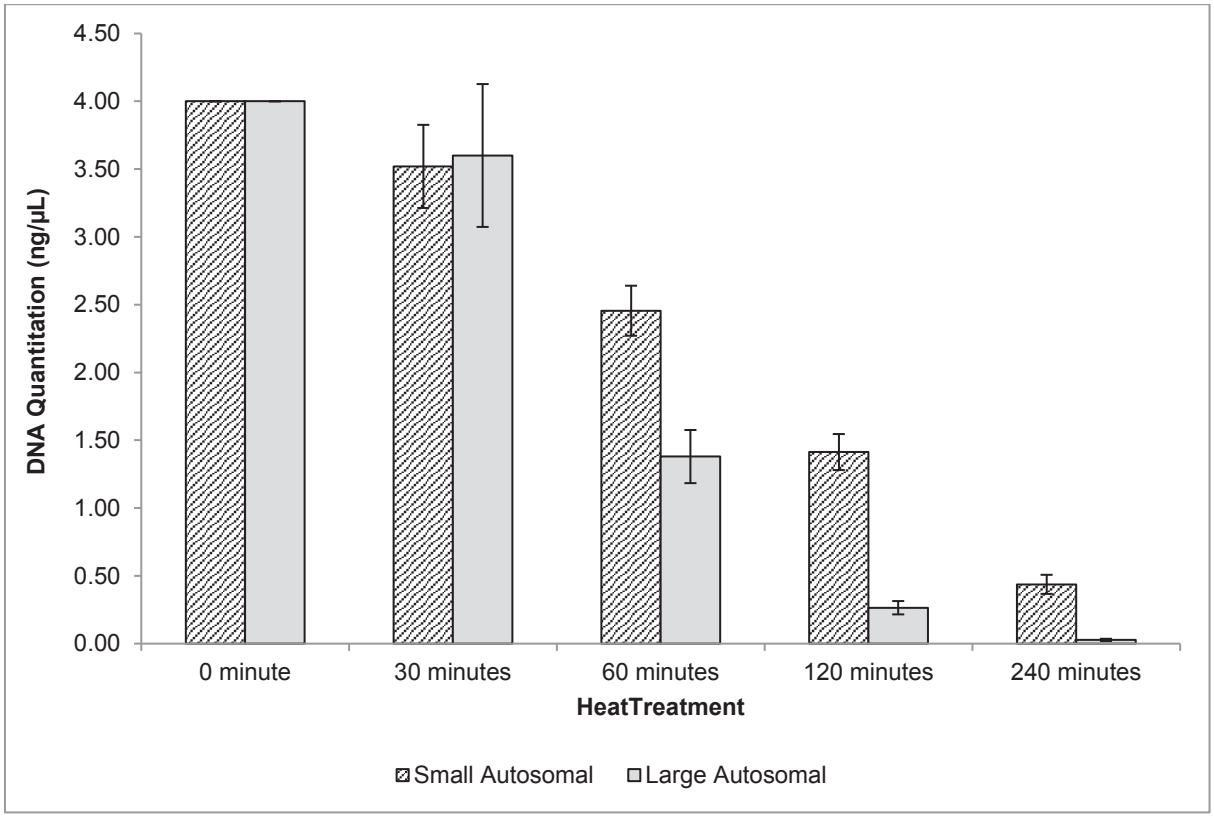


Figure S1. Determination of DNA Concentration in Heat Treated DNA Samples. DNA concentrations (ng/µL) of small autosomal and large autosomal targets for each respective treatment group were combined and plotted as mean ± S.E.M (n= 5). Small autosomal target quantities presented as diagonal ruled column. Large autosomal target quantities presented as shaded column. Relative to the control treatment group (0 minute), quantities of small and large autosomal targets were reduced ($p < 0.05$) in heat-treatment groups at 30, 60, 120 and 240 minutes.

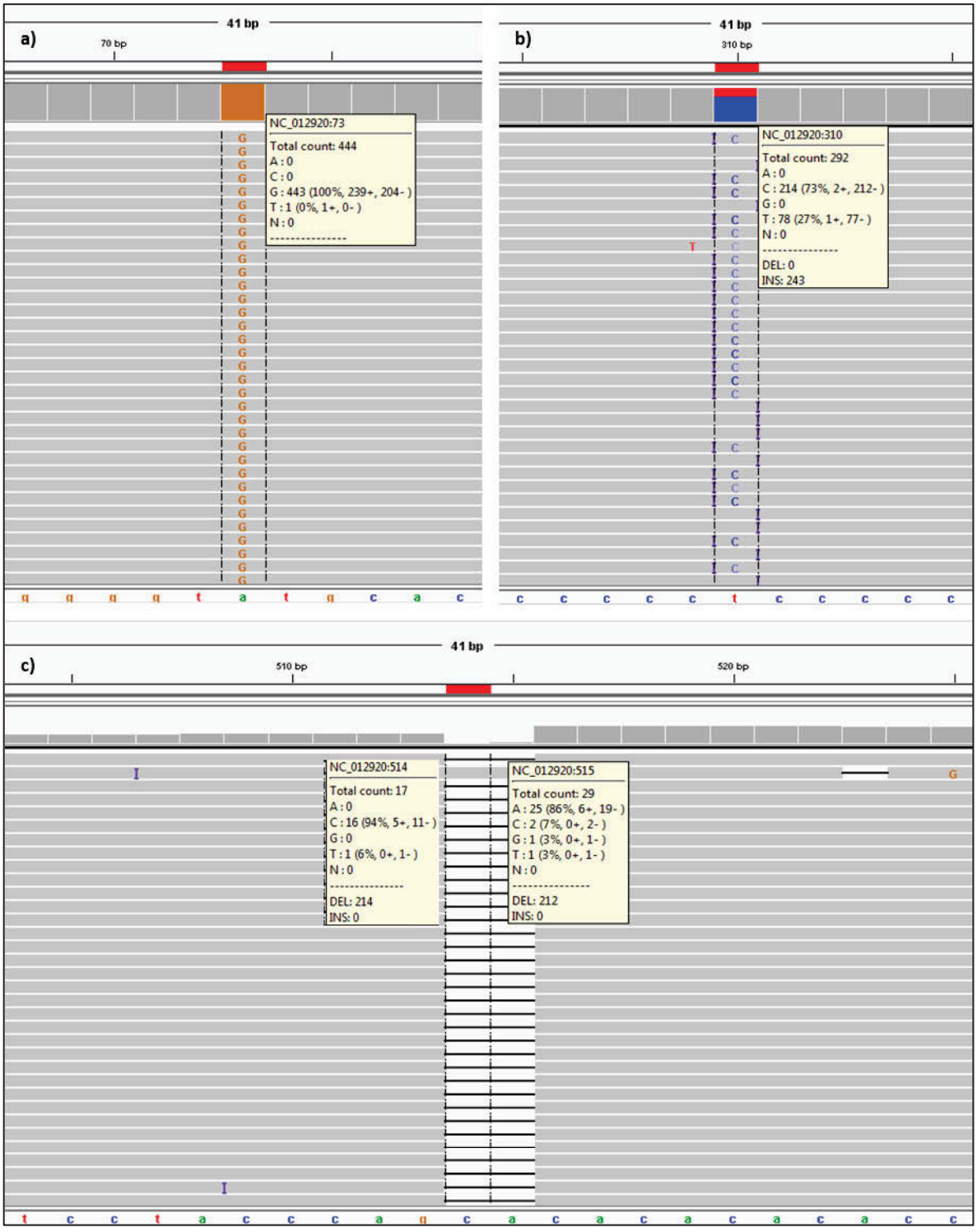


Figure S2. Variant Inspection. Display of variants on Integrative Genomics Viewer (Broad Institute, MA, USA). **a)** SNP variant (73G), **b)** length heteroplasmy (310C, 315.1C, 315.2C) and **c)** deletion variants (523-, 524-). Bottom panel shows rCRS for human mitochondrial genomes (NC_012920).

Table S1. Reported Variants, SNPs, insertions and deletions (INDELs) at coverages 10X, 50X and 100X for whole mitochondrial genomes amplified with the Early Access AmpliSeq™ Mitochondrial Panel (Applied Biosystems, CA, USA)

	Coverage		
Sample (*n*=5)	*10X*	*50X*	*100X*
0 minutes	149	147	139
30 minutes	147	145	138
60 minutes	149	149	141
120 minutes	148	148	140
240 minutes	150	145	132
Total Variants	743	734	690

Reported variants are inclusive of length heteroplasmy. Samples (*n*=5) represent DNA extracts of five biological replicates artificially degraded at 125°C for 0, 30, 60, 120 and 240 minutes

Chapter 6: Final Discussion and Conclusions

1. Introduction

MtDNA analysis is a common method of choice for analysing DNA samples that have failed to produce a complete STR profile. Usually this is seen in compromised DNA samples that are reduced in quality and/or limited in quantities of genomic DNA. Due to the technology, time and costs associated with mtDNA analysis, traditional sequencing of mitochondrial genomes has been restricted to the HV1 and HV2 regions [99]. This was extended to the entire control region and accounted for the phantom mutations and artificial recombination that were seen in single strand sequencing artefacts [100, 101]. The emergence of MPS technologies in human DNA testing is expanding the analysis of mtDNA to the complete genome. This has provided an increased resolution of mitochondrial haplogroups and ancestries.

In order to transition to the mtDNA-MPS testing of the complete mitochondrial genome, it is necessary to use primer pairs that cover the entire 16,569 bp genome [6, 102]. Sometimes these primers are designed *in-house*. While these primers may have been internally validated, they have often not been tested by the wider forensic community and lack the required standardisation necessary for forensic applications. Hence, there has been need to introduce a standard panel, which minimises this variability and provides a quality sequence data that is compatible across laboratories. As a result, commercial providers have developed and released mtDNA panels for performance testing. The main aim of this study was to evaluate the performance of one of these panels, the Early Access AmpliSeq™ Mitochondrial Panel (Applied Biosystems, CA, USA) with DNA samples similar to those found in casework.

The initial studies involved creating a set of DNA samples which were degraded *in vitro* to qualities analogous to those found in forensic scenarios. The methods used to assess the quality and quantity of these DNA samples were real-time qPCR assays and gel electrophoresis. Furthermore, a real-time qPCR assay was developed for the precise quantification of mtDNA. Currently available methods for quantifying mtDNA have been limited in specificity and as a result, the development of this assay addresses such a limitation [103].

The overall evaluation of the Panel involved the sequencing of degraded DNA samples for whole mitochondrial genomes, using the Ion Torrent™ technology (Applied Biosystems, CA, USA). Bioinformatics and sequence analysis software tools were used to assess the DNA sequences for coverage, variant calling (including point and length heteroplasmy) and resolution of mitochondrial haplogroups.

2. Mitochondrial DNA Quantification

The DNA extracts of buccal epithelial cells were quantified using a custom developed real-time quantification assay for mtDNA. This assay provided a precise quantification of mtDNA copy number by amplifying a 70 bp encoding region of the mt-ND3 gene. In comparison to other assays for mtDNA, this sequence does not exist as a NUMT or pseudogene in nuclear DNA and thereby corrects for possible overestimations [103]. The assay has been shown to be highly sensitive to a minimum of 280 copies of standard mtDNA as well as being reproducible between replicates.

The results from this study show a reduced number of mtDNA molecules between reference and heat-treated samples. The loss of intact DNA molecules was caused by a heat breakage of phosphodiester and glyosidic bonds that effectively prevented the binding of primer sequences and amplification of template sequences. As a result, the greater reduction of mtDNA copy number in prolonged heated DNA samples indicates a progressive breakdown of DNA molecules. This degraded quality of DNA was confirmed with other real time assays, agarose electrophoresis gels and STR typing of nuclear DNA.

The reduction of mtDNA copy number with treatment was concordant with quantifications reported in other *in-house* qPCR assays that amplify the HV1 mitochondrial region [103]. In comparison, the quantities of mtDNA provided by the assay amplifying the mt-ND3 region was conservative compared to assays that target other sequences. While the conservative quantifications may indicate potential improvements for amplification efficiency, the difference between assays can be better resolved by a simultaneous quantification of the mt-ND3 and HV1 sequences. Further work can look at optimising assay parameters such as cycling temperatures and times. Other technologies such as digital droplet PCR can also be considered to improve assay performance.

Though no non-specific amplification was detected, cross reactivity studies could be extended to mammalian and non-mammalian species as well as bacterial and fungal populations. This would be especially important for domestic animals that are likely to be environmental contaminants such as dogs, cats, rabbits and fish [69].

The assay is able to quantify mtDNA copy number in epithelial buccal cells with progressive levels of DNA degradation. To indicate its full range, testing of the assay can include other types of samples of different qualities, which are common in mitochondrial casework such as skeletal remains, hair fibers and ancient remains. This can extend to the testing of inhibitors commonly extracted with biological samples and its potential effect on the assay performance.

Overall, these results indicate that the developed *in-house* assay can provide quantifications of mtDNA, which are specific, sensitive and reproducible. In heat-treated samples, it reflects there is a degradation of DNA and the concomitant reduction in copy number. Applying such an assay to biological samples of a compromised nature will indicate the extent of degradation and inform the appropriate downstream processes for sample use including the sequencing of samples for mtDNA.

3. Sequencing of Mitochondrial DNA

Compromised DNA samples containing 0.1 ng of genomic DNA were amplified for whole mtDNA genomes using the Early Access AmpliSeq™ Mitochondrial Panel (Applied Biosystems, CA, USA). The quantities of genomic DNA provided by the Quantifiler™ Trio DNA Quantification Kit (Life Technologies, CA, USA) were estimated to contain 2,900 copies of mtDNA from 0.1 ng of genomic DNA. Sequencing of amplified fragments was performed on the Ion Torrent PGM™ platform using an Ion 316™ Chip (Applied Biosystems, CA, USA). Output sequences were evaluated focusing on coverage (amplicon and strand balance), variant calls and the alignment of mitochondrial haplogroups and ancestries. A consensus of mitochondrial haplotypes was confirmed when there was concordance to reference sequences and sequences of the HV1 and HV2 locations typed by dideoxynucleotide sequencing and CE.

The results from this study indicate that the inherent sequences of DNA at certain locations can affect the quality of MPS data. While coverage was overall high, it was evident that sequence complexity causes an unequal distribution of reads between amplicons. In particular, the repeating structure of adenosine bases at positions 14,766 and 14,783 rCRS were reported at lower coverages to the other mitochondrial variants. This extended to point and length heteroplasmy, especially in the C-stretches of HV2, 303-315 rCRS, which recorded the highest coverage reducing effect [104]. The C-tracts of these regions are often excluded from comparisons between reference and casework samples as often the calling of these sequences is ambiguous [30]. Coverage variation between amplicons was not confined to the performance of the Panel amplification as other panels report similar readings at such sites [105-107]. As a result, there are limited strategies for improving coverages affected by the inherent sequence of mtDNA and thereby to the design of the Panel. Rather methods for improving coverage should focus on processes involving the library set-up of DNA samples. These include indexing samples of similar qualities and multiplexing of sample numbers to a chip's sequencing output capacity. Such set-ups minimise preferential amplification of reference samples over samples of a compromised quality and ensures an even representation of sample genomes in the MPS data.

Depending on sample quality, library preparation of mitochondrial genomes can follow three different Panel amplification strategies. These include the conservative method for non-degraded or reference samples, 2-in-1 method for low copy number samples and the full method for very low copy number samples [64]. In this study, the 2-in-1 method provided sequence data for all samples without maximising the chip output capacity. Further studies should focus on comparing the sequence and coverage output of samples, using different methods. Reaction volumes should also be optimised for each method, which can possibly accommodate to smaller sample volumes. This is especially important for the analysis of mtDNA samples, which are frequently limited and degraded. This study controls degradation through an *in vitro* method of heating DNA samples. Other sample factors are also known to interfere with mtDNA testing, including inhibitors and environmental degradation. Such variables should be the focus of follow-up studies to validate the full potential of the Panel.

The calling of mitochondrial variants was accurate when compared to the reference sequences of each sample. HV1 and HV2 regions, which were additionally sequenced with chain-terminating technologies, were mostly concordant to MPS sequences called for reference and 240 minute heat-treated samples. In comparison to CE systems, the parallel sequencing of DNA molecules at high coverage depths improves the sensitivity of MPS platforms to detect intra-individual sequences [108]. This greater sensitivity accounts for the differences of base calling in heteroplasmic positions, 214R and 16093Y. Other sequence differences were due to unreported bases by the CE system, mostly likely arising as artifacts of sequencing.

The increased resolution seen in the whole mitochondrial genome panel using MPS technologies was useful in identifying mitochondrial haplogroups. Sequencing of whole mtDNA gives resolutions that is improved to the sequencing of HV1 and HV2 regions alone. The latter was shown to provide only broad haplogroup assignments. Haplotype assignments were correct, given the known ethnic background and migration history of the DNA donors. Overall, the concordance of mtDNA sequences between the CE and MPS technology as well as the additional data provided by the whole genome Panel confirms the strong performance of the Panel in forensic mtDNA testing.

Prior to the use of the MPS Panel in forensic mtDNA testing, there are some workflow changes, which should be considered from the current dideoxynucleotide/CE protocols. The strict set-up practices which are used in CE workflows to prevent extraneous contamination of mtDNA should be maintained in the protocols of MPS testing [34]. It is recommended that an automated set-up is introduced for the library and template preparation stages with the Ion Chef™ (Life Technologies, CA, USA) platform. In comparison to manual methods, this has

been reported to provide greater efficiency, reproducibility and faster turn-around times in forensic laboratories [109].

Overall, the results of this study indicate the Panel is capable of sequencing whole mitochondrial genomes even when DNA has been degraded, on the Ion Torrent™ platform. While the quality of sequence data does reduce as degradation progresses, its effect on data acquisition and analysis is minimal. Even in most compromised samples, the mitochondrial haplogroups and ancestries of individuals were still inferable. The introduction of this standard Panel is leading to an improved consistency of mtDNA testing with MPS technology, which is beneficial to its forensic acceptance and use. As MPS continues to translate into forensic genetics, additional Panels for testing of mtDNA, STR and phenotypic SNPs will become readily available. In which case, the evaluations of the Panel performed in this study will become a useful indicator of performance and quality.

4. Summary

The original aims of this research were to review the use of mtDNA and improve its testing in forensic genetics. The work of these studies has achieved these aims. The development of an *in-house* qPCR assay for the amplification of mt-ND3 has provided a method of mtDNA quantification that is specific, sensitive and robust. The evaluation of the Early Access AmpliSeq™ Mitochondrial Panel (Applied Biosystems, CA, USA) has indicated its ability to sequence mitochondrial genomes in compromised samples and contribute to the overall use of MPS panels in human DNA testing.

In conclusion, these studies have highlighted the progression of mtDNA testing and its identification value in cases of limited STR profiles. Developing an *in-house* qPCR assay has improved the assessment of mtDNA samples that were previously limited to the quantity and quality of the nuclear genome. Simply using a target specific sequence of mtDNA in the many *in-house* qPCR assays can overcome the under- and over- estimation of genomes and lead to improved decision-making for downstream applications. The additional data produced from a whole genome Panel compared to a HV1 and HV2 amplification has extended the reporting of mitochondrial haplogroups to the maximum phylogenetic resolution.

The sequencing of complete mitochondrial genomes in compromised samples has become more feasible with the development of MPS platforms. The early use of this technology in USA and European laboratories has shown its conformity to the ‘gold standard’ of CE systems. In Australia, MPS technologies are undergoing research and evaluations studies for operational casework. This has focused on the workflow and set-up of the technology as well as the

bioinformatics associated with sequence analysis. There is also the reviewing of the extra information of MPS sequences and its treatment. While the additional sequences are valuable for identification, it is also possible to infer non-forensic traits such as inherited medical conditions, which may raise ethical considerations.

The past use of mtDNA sequencing has helped identify many individuals from the 2004 Indonesian Tsunami to missing persons cases to the historical remains of Edward ‘Ned’ Kelly [2, 54, 110]. However, there are many past samples, which remain unidentified from its continually degrading state. As the feasibility of complete mtDNA analysis continues to develop with MPS technologies and its associated panels, the possibility of identifying these past samples is becoming realised in DNA identification programs. This includes establishing an identification program for Australia’s missing persons as well as the revival of cold cases [111]. Additionally, it extends to the identification of the many unrecovered causalities of war and the repatriation of indigenous remains to Country [52, 112].

Appendices

Appendices

Appendix I: Sample Consent Form

School of Mathematical and Physical Sciences
Centre for Forensic Science

UTS
UNIVERSITY OF TECHNOLOGY SYDNEY

CONSENT FORM

I _____ hereby consent to providing:

- ☐ buccal swab (saliva) sample/s
- ☐ hair sample/s
- ☐ ancestry details

I understand that these sample(s) will be used for research purposes only in the project:

Evaluation of novel forensic DNA typing kits, ultilising a next generation sequencing (NGS) technology

I understand that the research project has been approved by the UTS Human Ethics Committee (Approval Number: 2015000296)

I understand that these sample(s) and all products derived from them:

Will be securely stored under all the responsibility of the researcher, Ka Tak Wai

Will remain unidentifiable to personnel not involved in the research project

Will be used to obtain a DNA profile and **will not** be used to obtain personal or medical information about myself (except ancestry details)

Maternal Grandmother Ancestry: _____

Maternal Grandfather Ancestry: _____

Paternal Grandmother Ancestry: _____

Paternal Grandfather Ancestry: _____

_____ ____/____/____
Signed by (participant) Date

_____ ____/____/____
Signed by (witness) Date

1

Appendix II: Genotype Table

Biological Group 1										
	0 minutes		**30 minutes**		**60 minutes**		**120 minutes**		**240 minutes**	
locus	**Allele 1**	**Allele 2**	**Allele 1**	**Allele 2**	**Allele 1**	**Allele 2**	**Allele 1**	**Allele 2**	**Allele 1**	**Allele 2**
AMEL	X	Y	X	Y	X	Y	X	Y	X	Y
D3S1358	16	18	16	18	16	18	16	18	16	18
D1S1656	17.3		17.3		17.3		17.3			
D6S1043	11	18	11	18	11	18	11	18		
D13S317	11		11							
Penta E	13	18	13	18						
D16S539	9	11	9	11	9	11	9	11	9	11
D18S51	12		12		12		12			
D2S1338	17	23	17	23						
CSF1PO	11	12	11	12						
Penta D	9	11	9	11						
TH01	7	9	7	9	7	9	7	9	7	9
vWA	15	16	15	16	15	16	15	16	15	16
D21S11	27	29	27	29	27	29	27	29		
D7S820	9	12	9	12			8	13		
D5S818	11		11							
TPOX	8	12	8	12						
D8S1179	10	16	10	16	10	16	10	16	10	16
D12S391	21	27	21	27	21	27	21	27	21	27
D19S433	14	16.2	14	16.2	14	16.2	14	16.2		
FGA	20	23	20	23						

Appendix II: Genotype Table

Biological Group 2										
	0 minutes		**30 minutes**		**60 minutes**		**120 minutes**		**240 minutes**	
locus	**Allele 1**	**Allele 2**	**Allele 1**	**Allele 2**	**Allele 1**	**Allele 2**	**Allele 1**	**Allele 2**	**Allele 1**	**Allele 2**
AMEL	X		X		X		X		X	
D3S1358	15	17	15	17	15	17	15	17	15	17
D1S1656	16	16.3	16	16.3	16	16.3	16	16.3	16	16.3
D6S1043	11	14	11	14	11	14	11	14		
D13S317	11		11		11					
Penta E	11	14	11	14	11	14				
D16S539	9	12	9	12	9	12	9	12	9	12
D18S51	14	17	14	17	14	17	14	17	14	17
D2S1338	19	20	19	20	19	20	19	20	19	20
CSF1PO	12	14	12	14	12	14				
Penta D	10	11	10	11	10	11				
TH01	7	9	7	9	7	9	7	9	7	9
vWA	16	17	16	17	16	17	16	17	16	17
D21S11	30	31.2	30	31.2	30	31.2	30	31.2		
D7S820	11		11		11					
D5S818	11	12	11	12	11	12				
TPOX	8		8		8					
D8S1179	10	13	10	13	10	13	10	13	10	13
D12S391	23	24	23	24	23	24	23	24	23	24
D19S433	14	14.2	14	14.2	14	14.2	14	14.2	14	14.2
FGA	20	25	20	25	20	25	20	25		

Appendix II: Genotype Table

Biological Group 3										
	0 minutes		**30 minutes**		**60 minutes**		**120 minutes**		**240 minutes**	
locus	**Allele 1**	**Allele 2**	**Allele 1**	**Allele 2**	**Allele 1**	**Allele 2**	**Allele 1**	**Allele 2**	**Allele 1**	**Allele 2**
AMEL	X	Y	X	Y	X	Y	X	Y	X	Y
D3S1358	14	17	14	17	14	17	14	17	14	17
D1S1656	13	16	13	16	13	16	13	16	13	16
D6S1043	11		11		11					
D13S317	9	13	9	13	9	13				
Penta E	11	12	11	12						
D16S539	9	12	9	12	9	12	9	12	9	12
D18S51	13	15	13	15	13	15	13	15	13	15
D2S1338	18	23	18	23	18	23			18	23
CSF1PO	11	12	11	12	11	12				
Penta D	9	11	9	11	9	11				
TH01	9	10	9	10	9	10	9	10	9	10
vWA	18	19	18	19	18	19	18	19	18	19
D21S11	30	32.2	30	32.2	30	32.2	30	32.2		
D7S820	10		10		10					
D5S818	11	12	11	12	11	12				
TPOX	8	11	8	11	8	11				
D8S1179	11	15	11	15	11	15	11	15	11	15
D12S391	22	23	22	23	22	23	22	23	22	23
D19S433	14.2	15.2	14.2	15.2	14.2	15.2	14.2	15.2		
FGA	20		20		20		20			

Appendix II: Genotype Table

Biological Group 4										
	0 minutes		**30 minutes**		**60 minutes**		**120 minutes**		**240 minutes**	
locus	**Allele 1**	**Allele 2**	**Allele 1**	**Allele 2**	**Allele 1**	**Allele 2**	**Allele 1**	**Allele 2**	**Allele 1**	**Allele 2**
AMEL	X	Y	X	Y	X	Y	X	Y	X	Y
D3S1358	15	17	15	17	15	17	15	17	15	17
D1S1656	13	16	13	16	13	16	13	16		
D6S1043	11	14	11	14	11	14				
D13S317	9	11	9	11	9	11				
Penta E	12	14	12	14						
D16S539	9	12	9	12	9	12	9	12	9	12
D18S51	13	17	13	17	13	17	13	17		
D2S1338	20	23	20	23	20	23				
CSF1PO	11	14	11	14						
Penta D	10	11	10	11						
TH01	9	10	9	10	9	10	9	10	9	10
vWA	16	19	16	19	16	19	16	19	16	19
D21S11	30	32.2	30	32.2	30	32.2	30	32.2		
D7S820	10	11	10	11	10	11				
D5S818	11		11		11					
TPOX	8		8							
D8S1179	13	15	13	15	13	15	13	15	13	15
D12S391	23		23		23		23		23	
D19S433	14	15.2	14	15.2	14	15.2				
FGA	20	25	20	25	20	25				

Appendix II: Genotype Table

Biological Group 5										
	0 minutes		**30 minutes**		**60 minutes**		**120 minutes**		**240 minutes**	
locus	**Allele 1**	**Allele 2**	**Allele 1**	**Allele 2**	**Allele 1**	**Allele 2**	**Allele 1**	**Allele 2**	**Allele 1**	**Allele 2**
AMEL	X		X		X		X		X	
D3S1358	15	17	15	17	15	17	15	17	15	17
D1S1656	13	16.3	13	16.3	13	16.3	13	16.3		
D6S1043	11		11		11					
D13S317	9	11	9	11						
Penta E	12	14	12	14						
D16S539	12		12		12		12		12	
D18S51	13	17	13	17	13	17	13	17	13	17
D2S1338	19	23	19	23	19	23				
CSF1PO	12	14	12	14						
Penta D	9	11	9	11						
TH01	9	10	9	10	9	10	9	10	9	10
vWA	16	18	16	18	16	18	16	18	16	18
D21S11	30	31.2	30	31.2	30	31.2	30	31.2		
D7S820	10	11	10	11	10	11				
D5S818	12		12		12					
TPOX	8	11	8	11						
D8S1179	13	15	13	15	13	15	13	15	13	15
D12S391	22	23	22	23	22	23	22	23		
D19S433	14	15.2	14	15.2	14	15.2	14	15.2		
FGA	20		20		20	25				

Appendix III: Summary Table of Variants

Sample 1

Position	Reference	0 minute	30 minute	60 minute	120 minute	240 minute
214	A	G	G	G	G	G
263	A	G	G	G	G	G
303	-	C/CC	C/CC	C/CC	C/CC	C/CC
311	-	C	C	C	C	C
750	A	G	G	G	G	G
1438	A	G	G	G	G	G
4769	A	G	G	G	G	G
8495	A	A	A	A	A	Artefact
8860	A	G	G	G	G	G
13830	T	C	C	C	C	C
14470	T	A	A	A	A	A
15326	A	G	G	G	G	G
16093	T	C	C	C	C	C
16221	C	T	T	T	T	T
16519	T	C	C	C	C	C

*Key: Variants highlighted in green are heteroplasmy variants. Variants highlighted in red were at low coverages (<100X).

Sample 2

Position	Reference	0 minute	30 minute	60 minute	120 minute	240 minute
73	A	G	G	G	G	G
195	T	A	A	A	A	A
263	A	G	G	G	G	G
303	-	C	-	C/CC	C	C
311	-	C	C	C	C	C
489	T	C	C	C	C	C
514	CA	-	-	-	-	-
750	A	G	G	G	G	G
1438	A	G	G	G	G	G
1598	G	A	A	A	A	A
2706	A	G	G	G	G	G
4769	A	G	G	G	G	G
7028	C	T	T	T	T	T
8495	A	A	A	A	Artefact	A
8701	A	G	G	G	G	G
8860	A	G	G	G	G	G
9540	T	C	C	C	C	C
10398	A	G	G	G	G	G
10400	C	T	T	T	T	T
10873	T	C	C	C	C	C
11719	G	A	A	A	A	A
12007	G	A	A	A	A	A
12705	C	T	T	T	T	T
14766	C	T	T	T	T	T
14783	T	C	C	C	C	C
15043	G	A	A	A	A	A
15259	C	T	T	T	T	T
15301	G	A	A	A	A	A
15326	A	G	G	G	G	G
15431	G	A	A	A	A	A
16179	C	-	-	-	-	-
16223	C	T	T	T	T	T
16519	T	C	C	C	C	C
16642	A	G	G	G	G	G

*Key: Variants highlighted in red were at low coverages (<100X)

Sample 3

Position	Reference	0 minute	30 minute	60 minute	120 minute	240 minute
73	A	G	G	G	G	G
152	T	C	C	C	C	C
263	A	G	G	G	G	G
303	-	C	C	C		C
311	-	C	C	C	C	C
489	T	C	C	C	C	C
709	G	A	A	A	A	A
750	A	G	G	G	G	G
1438	A	G	G	G	G	G
1888	G	A	A	A	A	A
2706	A	G	G	G	G	G
3921	C	T	T	T	T	T
4769	A	G	G	G	G	G
7028	C	T	T	T	T	T
8701	A	G	G	G	G	G
8860	A	G	G	G	G	G
9540	T	C	C	C	C	C
10398	A	G	G	G	G	G
10400	C	T	T	T	T	T
10586	G	A	A	A	A	A
10873	T	C	C	C	C	C
11719	G	A	A	A	A	A
12477	T	C	C	C	C	C
12705	C	T	T	T	T	T
14323	G	A	A	A	A	A
14766	C	T	T	T	T	T
14783	T	C	C	C	C	C
15043	G	A	A	A	A	A
15301	G	A	A	A	A	A
15326	A	G	G	G	G	G
16129	G	A	A	A	A	A
16223	C	T	T	T	T	T
16519	T	C	C	C	C	C
16642	A	G	G	G	G	G

*Key: Variants highlighted in red were at low coverages (<100X).

Sample 4

Position	Reference	0 minute	30 minute	60 minute	120 minute	240 minute
73	A	G	G	G	G	G
195	T	A	A	A	A	A
263	A	G	G	G	G	G
303	-	C	C	C	C	C/CC
311	-	C	C	C	C	C
489	T	C	C	C	C	C
514	CA	-	-	-	-	-
750	A	G	G	G	G	G
1438	A	G	G	G	G	G
1598	G	A	A	A	A	A
2706	A	G	G	G	G	G
4769	A	G	G	G	G	G
7028	C	T	T	T	T	T
8495	A	Artefact	A	A	A	A
8701	A	G	G	G	G	G
8860	A	G	G	G	G	G
9540	T	C	C	C	C	C
10398	A	G	G	G	G	G
10400	C	T	T	T	T	T
10873	T	C	C	C	C	C
11719	G	A	A	A	A	A
12007	G	A	A	A	A	A
12705	C	T	T	T	T	T
14766	C	T	T	T	T	T
14783	T	C	C	C	C	C
15043	G	A	A	A	A	A
15259	C	T	T	T	T	T
15301	G	A	A	A	A	A
15326	A	G	G	G	G	G
15431	G	A	A	A	A	A
16179	C	-	-	-	-	-
16223	C	T	T	T	T	T
16519	T	C	C	C	C	C
16642	A	G	G	G	G	G

*Key: Variants highlighted in red were at low coverages (<100X).

Sample 5

Position	Reference	0 minute	30 minute	60 minute	120 minute	240 minute
73	A	G	G	G	G	G
195	T	A	A	A	A	A
263	A	G	G	G	G	G
303	-	C	C	C	C	C
311	-	C	C	C	C	C
489	T	C	C	C	C	C
514	CA	-	-	-	-	-
750	A	G	G	G	G	G
1438	A	G	G	G	G	G
1598	G	A	A	A	A	A
2706	A	G	G	G	G	G
4769	A	G	G	G	G	G
7028	C	T	T	T	T	T
8701	A	G	G	G	G	G
8860	A	G	G	G	G	G
9540	T	C	C	C	C	C
10398	A	G	G	G	G	G
10400	C	T	T	T	T	T
10873	T	C	C	C	C	C
11719	G	A	A	A	A	A
12007	G	A	A	A	A	A
12705	C	T	T	T	T	T
14766	C	T	T	T	T	T
14783	T	C	C	C	C	C
15043	G	A	A	A	A	A
15259	C	T	T	T	T	T
15301	G	A	A	A	A	A
15326	A	G	G	G	G	G
15431	G	A	A	A	A	A
16179	C	-	-	-	-	-
16223	C	T	T	T	T	T
16519	T	C	C	C	C	C
16642	A	G	G	G	G	G

*Key: Variants highlighted in red were at low coverages (<100X).

Appendix IV: Kinship Matching of Variants

Position	Reference	Sample 1	Sample 2	Sample 3	Sample 4	Sample 5
73	A	A	G	G	G	G
152	T	T	T	C	T	T
195	T	T	A	T	A	A
214	A	G	A	A	A	A
263	A	G	G	G	G	G
303	-	C/CC	C	C	C	C
311	-	C	C	C	C	C
489	T	T	C	C	C	C
514	CA	CA	del or G	CA	G	G
709	G	G	G	A	G	G
750	A	G	G	G	G	G
1438	A	G	G	G	G	G
1598	G	G	A	G	A	A
1888	G	G	G	A	G	G
2706	A	A	G	G	G	G
3921	C	C	C	T	C	C
4769	A	G	G	G	G	G
7028	C	C	T	T	T	T
8701	A	A	G	G	G	G
8860	A	G	G	G	G	G
9540	T	T	C	C	C	C
10398	A	A	G	G	G	G
10400	C	C	T	T	T	T
10586	G	G	G	A	G	G
10873	T	T	C	C	C	C
11719	G	G	A	A	A	A
12007	G	G	A	G	A	A
12477	T	T	T	C	T	T
12705	C	C	T	T	T	T
13830	T	C	T	T	T	T
14470	T	A	T	T	T	T
14766	C	C	T	T	T	T
14783	T	T	C	C	C	C

Appendix IV: Kinship Matching of Variants

15043	G	G	A	A	A	A
15259	C	C	T	C	T	T
15301	G	G	A	A	A	A
15326	A	G	G	G	G	G
15431	G	G	A	G	A	A
16093	T	C	T	T	T	T
16129	G	G	G	A	G	G
16179	C	C	del or T	C	T	T
16221	C	T	C	C	C	C
16223	C	C	T	T	T	T
16519	T	C	C	C	C	C
16642	A	A	G	G	G	G

*Key: Highlighted in yellow are the SNP variants detected to be exclusive for identification, with respect to the samples sequenced.

Appendix V: Low Coverage Amplicons

Nucleotide Position (rCRS)	Sequence	Reported Polymorphisms*
3,473 – 3,596	tcaccaaagagcccctaaaacccgccacatctacca tcaccctctacatcaccgccccgaccttagctctca ccatcgctcttctactatgaaccccccctccccatac ccaaccccctggtcaa	68
10,394 – 10,492	ctgaaccgaattggtatatagtttaaacaaaacgaa tgatttcgactcattaaattatgataatcatattta ccaaatgcccctcatttacataaatat	38
12,352 – 12,459	actataaccaccctaaccctgacttccctaattccc cccatccttaccaccctcgttaaccctaacaaaaaa aactcataccccccattatgtaaaatccattgtcgca	64
13,686 – 13,789	actaaaccccattaaacgcctggcagccggaagcc tattcgcaggatttctcattactaacaacatttccc ccgcatcccccttccaaacaacaatcccccctct	63
14,276 – 14,367	cctgacccctctccttcataaattattcagcttcct acactattaaagtttaccacaaccaccaccccatca tactctttcacccacagcac	48

***Reported Polymorphisms detected from [113].**

Appendix VI: Poster Presentation at ISFG

Evaluation of the Early Access AmpliSeq™ Mitochondrial Panel utilising Massively Parallel Sequencing

170031

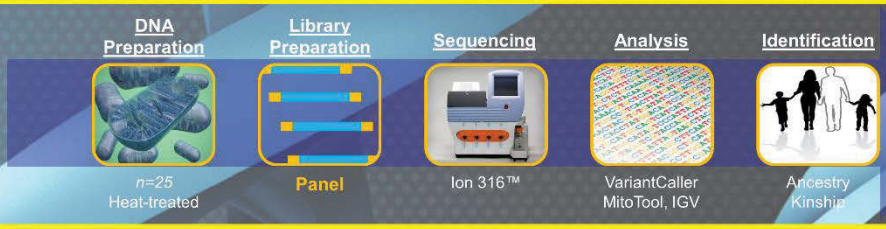
Ka Tak Wai[1*], Mark Barash[1] and Peter Gunn[1]

[1] University of Technology Sydney, Centre for Forensic Science, Ultimo NSW 2007, Australia

Background

- There is value in using mitochondrial DNA for ancestry and kinship analysis
- The AmpliSeq™ Panel (Applied Biosystems, CA, USA) offers a commercial solution to standardising mitochondrial DNA sequencing
- The Panel contains 162 primer pairs that cover the entire mitochondrial genome for sequencing applications
- **This study evaluates the performance of the Panel for coverage, strand balance and accuracy of variant calls in reference and artificially-degraded DNA samples (125°C for 30, 60, 120 and 240 mins)**

Method



Results

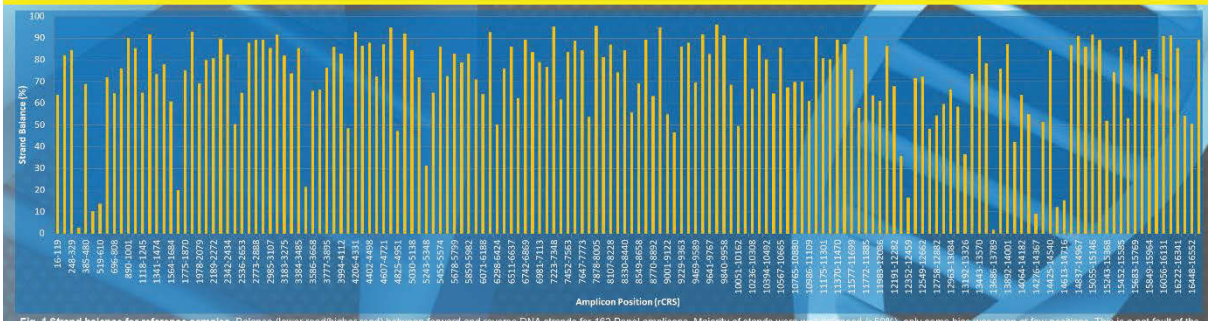


Fig. 1 Strand balance for reference samples. Balance (lower read/higher read) between forward and reverse DNA strands for 162 Panel amplicons. Majority of stands were well balanced (>50%), only some bias was seen at few positions. This is a not fault of the Panel, but inherent in regions of repeating sequences, that are also difficult to read with other panels

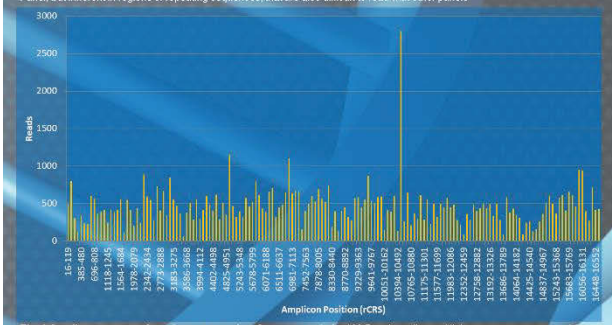


Fig. 2 Amplicon coverage for reference samples. Coverage reads for 162 Panel amplicons. High coverage was reported for all DNA amplicons, though coverage was not uniform at positons

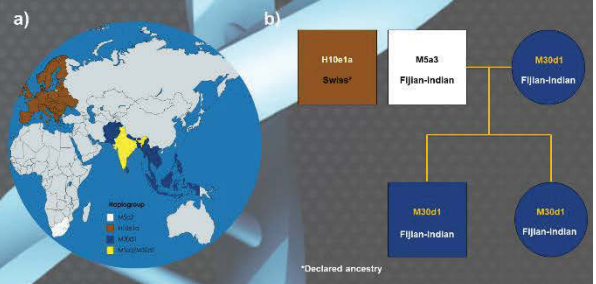


Fig. 3 Haplogroup assignment for all samples. Assignment of mitochondrial haplogroups for **a)** ancestry and **b)** kinship identification at 10X coverage. All mitochondrial haplogroups were accurately placed with known ancestry and kinship relations. Only heat-were heat-treated samples failed to call variants (at reliable coverages) for mitochondrial haplogroups

Conclusions

- Strands were well balanced at 72%
- OVER FOUR HUNDRED coverage reads were reported on average across the 162 amplicons
- Haplogroups were inferred for most HEAT-TREATED samples
- RESOULTION of kinship and ancestry relations were consistent with reference samples and known family histories

Acknowledgements

- The supply of the Panel for testing by Thermo Fisher Scientific is greatly appreciated
- This research is supported by an Australian Government Research Training Program Scholarship

*Ka.T.Wai@student.uts.edu.au

Centre for Forensic Science

UTS: THINK.CHANGE.DO

References

References

1. Wright, K., Mundorff, A., Chaseling, J., Maguire, C., and Crane, D.I., *An Evaluation of the Thai Tsunami Victim Identification DNA Operation.* Forensic Science Policy & Management: An International Journal, 2015. **6**(3-4): p. 69-78.
2. Hartman, D., Benton, L., Spiden, M., and Stock, A., *The Victorian missing persons DNA database–two interesting case studies.* Australian Journal of Forensic Sciences, 2015. **47**(2): p. 161-172.
3. Ensenberger, M.G., Hill, C.R., McLaren, R.S., Sprecher, C.J., and Storts, D.R., *Developmental validation of the PowerPlex® 21 System.* Forensic Science International: Genetics, 2014. **9**: p. 169-178.
4. Whitaker, J., Clayton, T., Urquhart, A., Millican, E., Downes, T., Kimpton, C., and Gill, P., *Short tandem repeat typing of bodies from a mass disaster: high success rate and characteristic amplification patterns in highly degraded samples.* Biotechniques, 1995. **18**(4): p. 670-677.
5. Parson, W., Strobl, C., Huber, G., Zimmermann, B., Gomes, S.M., Souto, L., Fendt, L., Delport, R., Langit, R., and Wootton, S., *Evaluation of next generation mtGenome sequencing using the Ion Torrent Personal Genome Machine (PGM).* Forensic Science International: Genetics, 2013. **7**(5): p. 543-549.
6. Parson, W., Huber, G., Moreno, L., Madel, M.-B., Brandhagen, M.D., Nagl, S., Xavier, C., Eduardoff, M., Callaghan, T.C., and Irwin, J.A., *Massively parallel sequencing of complete mitochondrial genomes from hair shaft samples.* Forensic Science International: Genetics, 2015. **15**: p. 8-15.
7. Chaitanya, L., Ralf, A., Oven, M., Kupiec, T., Chang, J., Lagacé, R., and Kayser, M., *Simultaneous whole mitochondrial genome sequencing with short overlapping amplicons suitable for degraded DNA using the ion torrent personal genome machine.* Human mutation, 2015. **36**(12): p. 1236-1247.
8. Tobe, S.S. and Linacre, A.M., *A technique for the quantification of human and non-human mammalian mitochondrial DNA copy number in forensic and other mixtures.* Forensic Science International: Genetics, 2008. **2**(4): p. 249-256.
9. Margulis, L., *Origin of eukaryotic cells: evidence and research implications for a theory of the origin and evolution of microbial, plant, and animal cells on the Precambrian earth.* 1970.
10. Gray, M.W., *Mitochondrial Evolution.* Cold Spring Harbor Perspectives in Biology, 2012. **4**(9): p. a011403.
11. Ettema, T.J., *Evolution: Mitochondria in the second act.* Nature, 2016.

12. Gray, M.W., *The pre-endosymbiont hypothesis: a new perspective on the origin and evolution of mitochondria.* Cold Spring Harbor perspectives in biology, 2014. **6**(3): p. a016097.
13. Malik, A.N., Shahni, R., Rodriguez-de-Ledesma, A., Laftah, A., and Cunningham, P., *Mitochondrial DNA as a non-invasive biomarker: accurate quantification using real time quantitative PCR without co-amplification of pseudogenes and dilution bias.* Biochemical and biophysical research communications, 2011. **412**(1): p. 1-7.
14. Foran, D.R., Gehring, M.E., and Stallworth, S.E., *The recovery and analysis of mitochondrial DNA from exploded pipe bombs.* Journal of forensic sciences, 2009. **54**(1): p. 90-94.
15. Zhang, L. and Wu, Q., *Single gene retrieval from thermally degraded DNA.* Journal of biosciences, 2005. **30**(5): p. 599-604.
16. Karni, M., Zidon, D., Polak, P., Zalevsky, Z., and Shefi, O., *Thermal degradation of DNA.* DNA and cell biology, 2013. **32**(6): p. 298-301.
17. Buś, M.M., Nilsson, M., and Allen, M., *Analysis of Mitochondrial DNA from a Burned, Ninhydrin-Treated Paper Towel.* Journal of forensic sciences, 2016.
18. Bender, K., Farfán, J., and Schneider, P.M., *Preparation of degraded human DNA under controlled conditions.* Forensic science international, 2004. **139**(2): p. 135-140.
19. Wachsmuth, M., Huebner, A., Li, M., Madea, B., and Stoneking, M., *Age-related and heteroplasmy-related variation in human mtDNA copy number.* PLoS Genet, 2016. **12**(3): p. e1005939.
20. Budowle, B., Bieber, F.R., and Eisenberg, A.J., *Forensic aspects of mass disasters: strategic considerations for DNA-based human identification.* Legal Medicine, 2005. **7**(4): p. 230-243.
21. Sharpley, M.S., Marciniak, C., Eckel-Mahan, K., McManus, M., Crimi, M., Waymire, K., Lin, C.S., Masubuchi, S., Friend, N., and Koike, M., *Heteroplasmy of mouse mtDNA is genetically unstable and results in altered behavior and cognition.* Cell, 2012. **151**(2): p. 333-343.
22. Lane, N., *The problem with mixing mitochondria.* Cell, 2012. **151**(2): p. 246-248.
23. Luo, S.-M., Ge, Z.-J., Wang, Z.-W., Jiang, Z.-Z., Wang, Z.-B., Ouyang, Y.-C., Hou, Y., Schatten, H., and Sun, Q.-Y., *Unique insights into maternal mitochondrial inheritance in mice.* PNAS, 2013. **110**(32): p. 13038-13043.
24. Schwartz, M. and Vissing, J., *Paternal inheritance of mitochondrial DNA.* New England Journal of Medicine, 2002. **347**(8): p. 576-580.
25. Sato, M. and Sato, K., *Maternal inheritance of mitochondrial DNA by diverse mechanisms to eliminate paternal mitochondrial DNA.* Biochimica et Biophysica Acta, 2013. **1833**: p. 1979-1984.

26. Luo, S.-M., Schatten, H., and Sun, Q.-Y., *Sperm Mitochondria in Reproduction: Good or Bad and Where Do They Go?* Journal of Genetics and Genomics, 2013. **40**(1): p. 549-556.
27. Chan, D.C. and Schon, E.A., *Eliminating Mitochondrial DNA from Sperm.* Developmental Cell, 2012. **22**(3): p. 469-470.
28. Stewart, J.E., Fisher, C.L., Aagaard, P.J., Wilson, M.R., Isenberg, A.R., Polanskey, D., Pokorak, E., DiZinno, J.A., and Budowle, B., *Length variation in HV2 of the human mitochondrial DNA control region.* Journal of Forensic Science, 2001. **46**(4): p. 862-870.
29. Walsh, S., Chaitanya, L., Clarisse, L., Wirken, L., Draus-Barini, J., Kovatsi, L., Maeda, H., Ishikawa, T., Sijen, T., and de Knijff, P., *Developmental validation of the HIrisPlex system: DNA-based eye and hair colour prediction for forensic and anthropological usage.* Forensic Science International: Genetics, 2014. **9**: p. 150-161.
30. Budowle, B., DiZinno, J.A., and Wilson, M.R. *Interpretation guidelines for mitochondrial DNA sequencing*. in *Tenth International Symposium on Human Identification*. 1999.
31. Jamieson, A. and Bader, S., *A Guide to Forensic DNA Profiling*. 2016: Wiley.
32. Andrews, R.M., Kubacka, I., Chinnery, P.F., Lightowlers, R.N., Turnbull, D.M., and Howell, N., *Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA.* Nature genetics, 1999. **23**(2): p. 147-147.
33. Behar, D.M., van Oven, M., Rosset, S., Metspalu, M., Loogväli, E.-L., Silva, N.M., Kivisild, T., Torroni, A., and Villems, R., *A "Copernican" reassessment of the human mitochondrial DNA tree from its root.* The American Journal of Human Genetics, 2012. **90**(4): p. 675-684.
34. Parson, W., Gusmao, L., Hares, D., Irwin, J., Mayr, W., Morling, N., Pokorak, E., Prinz, M., Salas, A., and Schneider, P., *DNA Commission of the International Society for Forensic Genetics: revised and extended guidelines for mitochondrial DNA typing.* Forensic Science International: Genetics, 2014. **13**: p. 134-142.
35. den Dunnen, J.T., Dalgleish, R., Maglott, D.R., Hart, R.K., Greenblatt, M.S., McGowan-Jordan, J., Roux, A.F., Smith, T., Antonarakis, S.E., and Taschner, P.E., *HGVS recommendations for the description of sequence variants: 2016 update.* Human mutation, 2016.
36. Sanger, F. and Coulson, A.R., *A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase.* Journal of molecular biology, 1975. **94**(3): p. 441-448.
37. Sanger, F., Nicklen, S., and Coulson, A.R., *DNA sequencing with chain-terminating inhibitors.* Proceedings of the National Academy of Sciences, 1977. **74**(12): p. 5463-5467.

38. Anderson, S., Bankier, A., de Bruijn, M., Coulson, A.R., Drouin, J., Eperon, I., Nierlich, D., Roe, B., Sanger, F., Schreier, P., Smith, A., Staden, R., and Young, I., *Sequence and organization of the human mitochondrial genome.* Nature, 1981. **290**(5806): p. 457-465.
39. Alvarez-Cubero, M.J., Saiz, M., Martínez-García, B., Sayalero, S.M., Entrala, C., Lorente, J.A., and Martinez-Gonzalez, L.J., *Next generation sequencing: an application in forensic sciences?* Annals of human biology, 2017. **44**(7): p. 581-592.
40. Behjati, S. and Tarpey, P.S., *What is next generation sequencing?* Archives of disease in childhood-Education & practice edition, 2013. **98**(6): p. 236-238.
41. Caratti, S., Turrina, S., Ferrian, M., Cosentino, E., and De Leo, D., *MiSeq FGx sequencing system: A new platform for forensic genetics.* Forensic Science International: Genetics Supplement Series, 2015. **5**: p. e98-e100.
42. Quail, M.A., Smith, M., Coupland, P., Otto, T.D., Harris, S.R., Connor, T.R., Bertoni, A., Swerdlow, H.P., and Gu, Y., *A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers.* BMC genomics, 2012. **13**(1): p. 1.
43. Børsting, C. and Morling, N., *Next generation sequencing and its applications in forensic genetics.* Forensic Science International: Genetics, 2015. **18**: p. 78-89.
44. Li, J.Z., Absher, D.M., Tang, H., Southwick, A.M., Casto, A.M., Ramachandran, S., Cann, H.M., Barsh, G.S., Feldman, M., and Cavalli-Sforza, L.L., *Worldwide human relationships inferred from genome-wide patterns of variation.* science, 2008. **319**(5866): p. 1100-1104.
45. Prestes, P.R., Mitchell, R.J., Daniel, R., Sanchez, J.J., and van Oorschot, R.A., *Predicting biogeographical ancestry in admixed individuals–values and limitations of using uniparental and autosomal markers.* Australian Journal of Forensic Sciences, 2016. **48**(1): p. 10-23.
46. Haak, W., Balanovsky, O., Sanchez, J.J., Koshel, S., Zaporozhchenko, V., Adler, C.J., Der Sarkissian, C.S., Brandt, G., Schwarz, C., and Nicklisch, N., *Ancient DNA from European early neolithic farmers reveals their near eastern affinities.* PLoS Biol, 2010. **8**(11): p. e1000536.
47. Sykes, B., *The seven daughters of Eve: The science that reveals our genetic ancestry*. 2010: WW Norton & Company.
48. Van Oven, M. and Kayser, M., *Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation.* Human mutation, 2009. **30**(2): p. E386-E394.
49. Quintáns, B., Alvarez-Iglesias, V., Salas, A., Phillips, C., Lareu, M., and Carracedo, A., *Typing of mitochondrial DNA coding region SNPs of forensic and anthropological*

interest using SNaPshot minisequencing. Forensic science international, 2004. **140**(2): p. 251-257.

50. King, T.E., Fortes, G.G., Balaresque, P., Thomas, M.G., Balding, D., Delser, P.M., Neumann, R., Parson, W., Knapp, M., Walsh, S., Tonasso, L., Holt, J., Kayser, M., Appleby, J., Forster, P., Ekserdjian, D., Hofreiter, M., and Schurer, K., *Identification of the remains of King Richard III.* Nature Communications, 2014. **5**: p. 5631.
51. Coble, M.D., Loreille, O.M., Wadhams, M.J., Edson, S.M., Maynard, K., Meyer, C.E., Niederstatter, H., Berger, C., Berger, B., Falsetti, A.B., Gill, P., Parson, W., and Finelli, L.N., *Mystery Solved: The Identification of the Two Missing Romanov Children Using DNA Analysis.* PLoS ONE, 2009. **4**(3): p. e4838.
52. Army, A. *The Fromelles Project - Australian Army*. 2016 [cited 2016 17th August]; Available from: http://www.army.gov.au/our-work/unrecovered-war-casualties-army/fromelles/the-fromelles-project.
53. Gill, P., Ivanov, P.L., Kimpton, C., Piercy, R., Benson, N., Tully, G., Evett, I., Hagelberg, E., and Sullivan, K., *Identification of the remains of the Romanov family by DNA analysis.* Nature genetics, 1994. **6**(2): p. 130-135.
54. Deng, Y.-J., Li, Y.-Z., Yu, X.-G., Li, L., Wu, D.-Y., Zhou, J., Man, T.-Y., Yang, G., Yan, J.-W., and Cai, D.-Q., *Preliminary DNA identification for the tsunami victims in Thailand.* Genomics, proteomics & bioinformatics, 2005. **3**(3): p. 143-157.
55. Ellison, S.L., English, C.A., Burns, M.J., and Keer, J.T., *Routes to improving the reliability of low level DNA analysis using real-time PCR.* BMC biotechnology, 2006. **6**(1): p. 33.
56. Consortium, G.P., *A global reference for human genetic variation.* Nature, 2015. **526**(7571): p. 68-74.
57. Parson, W. and Dür, A., *EMPOP—a forensic mtDNA database.* Forensic Science International: Genetics, 2007. **1**(2): p. 88-92.
58. Murakami, H., Ota, A., Simojo, H., Okada, M., Ajisaka, R., and Kuno, S., *Polymorphisms in control region of mtDNA relates to individual differences in endurance capacity or trainability.* The Japanese journal of physiology, 2002. **52**(3): p. 247-256.
59. Taylor, R.W. and Turnbull, D.M., *Mitochondrial DNA mutations in human disease.* Nature Reviews Genetics, 2005. **6**(5): p. 389-402.
60. Vasta, V., Ng, S.B., Turner, E.H., Shendure, J., and Hahn, S.H., *Next generation sequence analysis for mitochondrial disorders.* Genome Medicine, 2009. **1**(10): p. 1.
61. Iacono, W.G., *Forensic "Lie Detection" Procedures Without Scientific Basis.* Journal of Forensic Psychology Practice, 2000. **1**(1): p. 75-86.

62. Smith, M. and Mann, M., *Recent developments in DNA evidence.* Trends & Issues in Crime and Criminal Justice, 2015(506): p. 1.
63. Purcell, A., *Could DNA phenotyping construct a likeness of the Gold Coast rapist?* , in *The Sydney Morning Herald*. 2016, Fairfax Media: Online.
64. Thermo Fisher Scientific, *Precision ID Panels with Ion PGM™ System*. 2016, California: Thermo Fisher Scientific.
65. Eduardoff, M., Xavier, C., Strobl, C., Casas-Vargas, A., and Parson, W., *Optimized mtDNA Control Region Primer Extension Capture Analysis for Forensically Relevant Samples and Highly Compromised mtDNA of Different Age and Origin.* Genes, 2017. **8**(10): p. 237.
66. Chaitanya, L., van Oven, M., Weiler, N., Harteveld, J., Wirken, L., Sijen, T., de Knijff, P., and Kayser, M., *Developmental validation of mitochondrial DNA genotyping assays for adept matrilineal inference of biogeographic ancestry at a continental level.* Forensic Science International: Genetics, 2014. **11**: p. 39-51.
67. Robin, J.D., Ludlow, A.T., LaRanger, R., Wright, W.E., and Shay, J.W., *Comparison of DNA quantification methods for next generation sequencing.* Scientific reports, 2016. **6**: p. 24067.
68. Satoh, M. and Kuroiwa, T., *Organization of multiple nucleoids and DNA molecules in mitochondria of a human cell.* Experimental cell research, 1991. **196**(1): p. 137-140.
69. Kavlick, M.F., Lawrence, H.S., Merritt, R.T., Fisher, C., Isenberg, A., Robertson, J.M., and Budowle, B., *Quantification of human mitochondrial DNA using synthesized DNA standards.* Journal of forensic sciences, 2011. **56**(6): p. 1457-1463.
70. Andréasson, H., Nilsson, M., Budowle, B., Lundberg, H., and Allen, M., *Nuclear and mitochondrial DNA quantification of various forensic materials.* Forensic science international, 2006. **164**(1): p. 56-64.
71. Goodwin, C., Higgins, D., Tobe, S.S., Austin, J., Wotherspoon, A., Gahan, M.E., and McNevin, D., *Singleplex quantitative real-time PCR for the assessment of human mitochondrial DNA quantity and quality.* Forensic Science, Medicine and Pathology, 2018: p. 1-6.
72. Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G., and Erlich, H. *Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction*. in *Cold Spring Harbor symposia on quantitative biology*. 1986. Cold Spring Harbor Laboratory Press.
73. Gill, P., Jeffreys, A.J., and Werrett, D.J., *Forensic application of DNA 'fingerprints'.* Nature, 1985. **318**(6046): p. 577-579.
74. Yang, Y., Xie, B., and Yan, J., *Application of next-generation sequencing technology in forensic science.* Genomics, proteomics & bioinformatics, 2014. **12**(5): p. 190-197.
75. Qiagen, *QIAamp® DNA Mini and Blood Mini Handbook*. 2016, Hilden: Qiagen.

76. Holt, A., Wootton, S.C., Mulero, J.J., Brzoska, P.M., Langit, E., and Green, R.L., *Developmental validation of the Quantifiler® HP and Trio Kits for human DNA quantification in forensic samples.* Forensic Science International: Genetics, 2016. **21**: p. 145-157.
77. Thermo Fisher Scientific, *Quantifiler™ HP and Trio DNA Quantification Kits*. 2017, California: Thermo Fisher Scientific.
78. Vernarecci, S., Ottaviani, E., Agostino, A., Mei, E., Calandro, L., and Montagna, P., *Quantifiler® Trio Kit and forensic samples management: a matter of degradation.* Forensic Science International: Genetics, 2015. **16**: p. 77-85.
79. Agilent Technologies, *Agilent High Sensitivity DNA Kit Guide*. 2013, Waldbronn: Agilent Technologies.
80. Watson, J.D. and Crick, F.H. *The structure of DNA*. in *Cold Spring Harbor symposia on quantitative biology*. 1953. Cold Spring Harbor Laboratory Press.
81. Butler, J.M., Tomsey, C.S., and Kline, M.C. *Can the validation process in forensic DNA typing be standardized?* in *Proceedings of the 15th International Symposium on Human Identification*. 2004. Citeseer.
82. Sambrook, J. and Russell, D.W., *Fragmentation of DNA by sonication.* Cold Spring Harbor Protocols, 2006. **2006**(4): p. pdb. prot4538.
83. Collins, J.R., Stephens, R.M., Gold, B., Long, B., Dean, M., and Burt, S.K., *An exhaustive DNA micro-satellite map of the human genome using high performance computing.* Genomics, 2003. **82**(1): p. 10-19.
84. Hares, D.R., *Addendum to expanding the CODIS core loci in the United States.* Forensic Science International: Genetics, 2012. **6**(5): p. e135.
85. Gill, P., Fereday, L., Morling, N., and Schneider, P.M., *New multiplexes for Europe—amendments and clarification of strategic development.* Forensic science international, 2006. **163**(1): p. 155-157.
86. Wang, D.Y., Gopinath, S., Lagacé, R.E., Norona, W., Hennessy, L.K., Short, M.L., and Mulero, J.J., *Developmental validation of the GlobalFiler® Express PCR Amplification Kit: a 6-dye multiplex assay for the direct amplification of reference samples.* Forensic Science International: Genetics, 2015. **19**: p. 148-155.
87. Edwards, A., Civitello, A., Hammond, H.A., and Caskey, C.T., *DNA typing and genetic mapping with trimeric and tetrameric tandem repeats.* American journal of human genetics, 1991. **49**(4): p. 746.
88. Bright, J.-A., Taylor, D., Curran, J.M., and Buckleton, J.S., *Degradation of forensic DNA profiles.* Australian Journal of Forensic Sciences, 2013. **45**(4): p. 445-449.
89. Promega. *GeneMapper® ID and ID-X Files for PowerPlex® Systems*. [cited 2017 1/05/2017]; Available from:

https://www.promega.com.au/resources/software-firmware/genemapper-id-software-panels-and-bin-sets/.

90. Gill, P., Sparkes, R., and Kimpton, C., *Development of guidelines to designate alleles using an STR multiplex system.* Forensic Science International, 1997. **89**(3): p. 185-197.

91. Coble, M., Andersen, M.M., Buckleton, J.S., Taylor, D., and Bright, J.-A., *Non-Autosomal Forensic Markers*, in *Forensic Dna Evidence Interpretation*. 2016, CRC Press LLC.

92. Buel, E., Schwartz, M.B., and LaFountain, M.J., *Capillary electrophoresis STR analysis: comparison to gel-based systems.* Journal of Forensic Science, 1998. **43**(1): p. 164-170.

93. Lazaruk, K., Walsh, P.S., Oaks, F., Gilbert, D., Rosenblum, B.B., Menchen, S., Scheibler, D., Wenz, H.M., Holt, C., and Wallin, J., *Genotyping of forensic short tandem repeat (STR) systems based on sizing precision in a capillary electrophoresis instrument.* Electrophoresis, 1998. **19**(1): p. 86-93.

94. Moore, D. and Thomson, J., *How degraded is our DNA? A review of single source live case work samples with optimal DNA inputs processed with the PowerPlex® ESI17 Fast kit.* Forensic Science International: Genetics Supplement Series, 2015. **5**: p. e467-e469.

95. Chung, D.T., Drábek, J., Butler, J.M., McCord, J.M., and Opel, K.L., *A study on the effects of degradation and template concentration on the amplification efficiency of the STR Miniplex primer sets.* Journal of Forensic Science, 2004. **49**(4): p. 1-8.

96. Tully, G., Sullivan, K.M., and Gill, P., *Analysis of 6 VNTR loci by 'multiplex'PCR and automated fluorescent detection.* Human genetics, 1993. **92**(6): p. 554-562.

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

98. Kloosterman, A.D. and Kersbergen, P., *Efficacy and limits of genotyping low copy number (LCN) DNA samples by multiplex PCR of STR loci.* Journal de la Société de Biologie, 2002. **197**(4): p. 351-359.

99. Bär, W., Brinkmann, B., Budowle, B., Carracedo, A., Gill, P., Holland, M., Lincoln, P.J., Mayr, W., Morling, N., and Olaisen, B., *DNA Commission of the International Society for Forensic Genetics: guidelines for mitochondrial DNA typing.* International journal of legal medicine, 2000. **113**(4): p. 193-196.

100. Brandstätter, A., Niederstätter, H., Pavlic, M., Grubwieser, P., and Parson, W., *Generating population data for the EMPOP database—an overview of the mtDNA*

sequencing and data evaluation processes considering 273 Austrian control region sequences as example. Forensic science international, 2007. **166**(2): p. 164-175.

101. Parson, W. and Bandelt, H.-J., *Extended guidelines for mtDNA typing of population data in forensic science.* Forensic Science International: Genetics, 2007. **1**(1): p. 13-19.
102. Fendt, L., Zimmermann, B., Daniaux, M., and Parson, W., *Sequencing strategy for the whole mitochondrial genome resulting in high quality sequences.* BMC genomics, 2009. **10**(1): p. 139.
103. Alonso, A., Martí, P., Albarrán, C., Garcí, P., Garcí, O., de Simón, L.F., Garcí, J., Sancho, M., de la Rúa, C., and Fernández-Piqueras, J., *Real-time PCR designs to estimate nuclear and mitochondrial DNA copy number in forensic and ancient DNA studies.* Forensic Science International, 2004. **139**(2): p. 141-149.
104. Seo, S.B., Zeng, X., King, J.L., Larue, B.L., Assidi, M., Al-Qahtani, M.H., Sajantila, A., and Budowle, B., *Underlying data for sequencing the mitochondrial genome with the massively parallel sequencing platform Ion Torrent™ PGM™.* BMC genomics, 2015. **16**(1): p. S4.
105. Zhou, Y., Guo, F., Yu, J., Liu, F., Zhao, J., Shen, H., Zhao, B., Jia, F., Sun, Z., and Song, H., *Strategies for complete mitochondrial genome sequencing on Ion Torrent PGM™ platform in forensic sciences.* Forensic Science International: Genetics, 2016. **22**: p. 11-21.
106. Ross, M.G., Russ, C., Costello, M., Hollinger, A., Lennon, N.J., Hegarty, R., Nusbaum, C., and Jaffe, D.B., *Characterizing and measuring bias in sequence data.* Genome biology, 2013. **14**(5): p. R51.
107. Churchill, J.D., King, J.L., Chakraborty, R., and Budowle, B., *Effects of the Ion PGM™ Hi-Q™ sequencing chemistry on sequence data quality.* International journal of legal medicine, 2016. **130**(5): p. 1169-1180.
108. Just, R.S., Irwin, J.A., and Parson, W., *Mitochondrial DNA heteroplasmy in the emerging field of massively parallel sequencing.* Forensic Science International: Genetics, 2015. **18**: p. 131-139.
109. Mogensen, H.S., Børsting, C., and Morling, N., *Template preparation of AmpliSeq™ libraries using the Ion Chef™.* Forensic Science International: Genetics Supplement Series, 2015. **5**: p. e368-e369.
110. Blau, S., Catelli, L., Garrone, F., Hartman, D., Romanini, C., Romero, M., and Vullo, C., *The contributions of anthropology and mitochondrial DNA analysis to the identification of the human skeletal remains of the Australian outlaw Edward 'Ned'Kelly.* Forensic science international, 2014. **240**: p. e11-e21.

111. Ward, J., *Best practice recommendations for the establishment of a national DNA identification program for missing persons: A global perspective.* Forensic Science International: Genetics Supplement Series, 2017.
112. Nagle, N., Van Oven, M., Wilcox, S., van Holst Pellekaan, S., Tyler-Smith, C., Xue, Y., Ballantyne, K.N., Wilcox, L., Papac, L., and Cooke, K., *Aboriginal Australian mitochondrial genome variation–an increased understanding of population antiquity and diversity.* Scientific reports, 2017. **7**.
113. Lott, M.T., Leipzig, J.N., Derbeneva, O., Xie, H.M., Chalkia, D., Sarmady, M., Procaccio, V., and Wallace, D.C., *mtDNA variation and analysis using MITOMAP and MITOMASTER.* Current protocols in bioinformatics, 2013: p. 1.23. 1-1.23. 26.