Alternative DNA technologies for obtaining DNA profiles from cartridge cases

Teigan Jae Tarapi

A thesis submitted in fulfilment of the requirements for the degree of Master of Science in Forensic Science, the University of Auckland, 2024.

ABSTRACT

Obtaining a DNA profile from unfired cartridges and fired cartridge cases is desirable in a forensic investigation as it can link an individual to an offence. Conventional nuclear DNA (nDNA) profiling technologies have proven largely unsuccessful for this evidence type, as any DNA recovered is often low in quality and quantity. The advent of massively parallel sequencing (MPS) has allowed for the simultaneous analysis of many short DNA amplicons with a greater sensitivity for highly degraded samples. This research aimed to determine whether two different MPS technologies, the Precision ID Whole mtDNA Genome Panel and the ForenSeq™ DNA Signature Prep Kit, could successfully generate DNA profiles from unfired cartridges and fired cartridge cases.

Previously developed custom MPS workflows were trialled on touch DNA samples recovered from unfired .223 Remington cartridges. Mitochondrial DNA (mtDNA) variants concordant with the corresponding reference haplotype were generated from 82.1% of the unfired samples, with 25% generating fully concordant mtDNA haplotypes. STR and iiSNP alleles could be recovered from unfired cartridges, with typed iiSNPs more likely to be typed concordant to a reference profile. The addition of a second purification, post-library preparation, minimised the impact of adapter-dimers on MiSeq FGx™ sequencing performance and was included into the final mtDNA workflow. Less DNA was recovered from a fired cartridge case, resulting in the smaller concentrations of mitochondrial and ForenSeq™ libraries. mtDNA haplotypes were generated in 78.6% of the fired samples, with 17.9% generating fully concordant mtDNA haplotypes. The presence of non-concordant minor mtDNA variants suggests that mtDNA haplotypes should be called using only major variants, ignoring any low-level heteroplasmy. STR and iiSNP alleles were also recovered from fired cartridge cases; however, allele dropouts were common, indicating that only partial loci profiles can be used in a forensic context.

These methods could successfully obtain DNA profiles from cartridges and cartridge cases. Ultimately, this research demonstrates that MPS technologies should be strongly considered for further optimisation, validation, and implementation into the analysis of firearm evidence recovered from crime scenes.

ACKNOWLEDGEMENTS

Firstly, thank you to the University of Auckland and the Institute of Environmental Science and Research (ESR Ltd) for the various sources of funding that made this research possible. Thank you to ESR Ltd for providing access to all the facilities and resources used during this research.

To my supervisors, I am beyond grateful for all the time you have dedicated to my weekly meetings, laboratory training and endlessly revising my many drafts. SallyAnn Harbison, thank you for always pushing me to my full potential. Bethany Forsythe, thank you for your commitment to this research and your continuous support from the other side of the world. Ryan England, thank you for agreeing to become one of my supervisors last minute, carrying out a part of the lab work for the fired samples and for your constant patience. It has been truly motivating and inspiring to work alongside you all, I really could not have done this without all of your support.

To Leah Tottey, thank you so much for overseeing the collection of all samples and teaching me everything I now know about firearms. Thank you to all the volunteers who participated in this research. I really appreciate that you took the time to donate a large amount of samples. To the R&D, Forensic Biology and P.E. teams, thank you for answering my endless questions, providing a never-ending supply of swabs and envelopes, and always being willing to help. Thank you to all the current and past students; it has been amazing to learn alongside a group of people so passionate about the same field of study. Our lunchtime chats have been certainly memorable, and I would not have survived this past year without our twice a day coffee trips to PNF!

To my Beau whānau, thank you for the much-needed distraction every Saturday from 10:30 to close. To my friends and family, thank you for your love, encouragement and for being so understanding when it took me 1-2 working days to reply to your messages. Finally, to Mum, Dad, and Ana - thank you for always believing in me, all the "quick" phone calls and your never-ending love and support.

TABLE OF CONTENTS

ABBREVIATIONS

.22 LR	.22 Long Rifle
.223 Rem	.223 Remington
ANEG	Amplification negative control
ANOVA	Analysis of Variance
APOS	Amplification positive control
a-STR	autosomal STR
bp	Base pairs
CE	Capillary electrophoresis
CR	Control region
DNA	Deoxyribonucleic acid
DNL	Denatured, normalised library
ENEG	Extraction negative control
EPOS	Extraction positive control
ESR	Institute of Environmental Science and Research Ltd.
gDNA	genomic DNA
GM-HTS	GeneMarker HTS
iiSNPs	Identity informative SNPs
LHP	Length heteroplasmy
MPS	Massively parallel sequencing
mtDNA	Mitochondrial DNA
nDNA	Nuclear DNA
NTC	No template control
PHI	Personal health information sites
PHP	Point length heteroplasmy
PNL	Pooled, normalised library
rCRS	revised Cambridge Reference Sequence
SNP	Single nucleotide polymorphism
SOP	Standard operating procedure
STR	Short tandem repeats
STS	Sanger type sequencing
x-STR	X chromosome STR
y-STR	Y chromosome STR

1 INTRODUCTION

Firearm-associated crime in Aotearoa, New Zealand has increased over recent years. In 2018, there were 901 recorded cases of firearm-related offences by the New Zealand Police (1). In 2022, this number had increased to 1,442. Firearm-related offences are categorised depending on the committed crime, which can vary from volume offences such as intimidation and threats, to more serious offences such as manslaughter, homicide, and grievous assaults (1). This is reflected in local media, with reports on recent firearm-related offences ranging from multiple aggravated robberies where the offender armed himself with a small pistol (2), to a man being left in a serious condition due to a gunshot injury, resulting in a manhunt for the offender (3).

Serious firearm-related offences lead to a forensic investigation where common types of forensic evidence recovered from the crime scene include unfired cartridges, fired bullets and cartridge casings (4). As it is likely an offender touched a cartridge while loading a firearm, forensic investigators try to obtain DNA profiles from such evidence using conventional DNA profiling techniques; where the targeted amplification of short tandem repeats (STRs) combined with capillary electrophoresis (CE) sees the generation of a DNA profile (5). Unfortunately, the success rate of using these DNA profiling methods has been extremely limited (6–8) primarily due to the low quantity and quality of touch DNA present on cartridges and cartridge cases.

Massively parallel sequencing (MPS), also commonly referred to as next generation sequencing (NGS), is an alternative method for DNA typing in forensic science. One main advantage of MPS is that it can use shorter amplicons to target genetic markers, which are currently not routinely utilised as a means of identification in conventional forensic DNA typing (9). These alternative genetic markers include the whole mitochondrial genome and single nucleotide polymorphisms (SNPs), which are extremely sensitive and more likely to be recovered from heavily degraded, low DNA quantity forensic samples. So far, limited research efforts have investigated whether MPS technologies can obtain DNA profiles from firearm evidence, with the targeting of mitochondrial DNA appearing promising (4). This research project aims to determine if targeting such alternative genetic markers via short amplicon MPS technologies is a viable DNA typing workflow for successfully obtaining DNA profiles from such evidence.

1.1 FIREARMS AND FIREARM EVIDENCE BACKGROUND

1.1.1 Firearms

In New Zealand, the Arms Act 1983 promotes the safe possession and use of firearms through strict controls and regulations (10). In this act, a firearm is defined as *"anything from which any shot, bullet, missile or other*

projectile can be discharged by force of explosive". Traditionally, firearms are categorised into two main groups: handguns or long arms (11), with the latter the type used in this thesis. Typical handguns include revolvers and pistols, whereas long arms include shotguns or rifles. Rifles can be further differentiated depending on their reloading mechanism (11). One type of mechanism is 'single shot', where the ammunition is manually loaded by an individual into a single-shot firearm each time the firearm is to be fired. Separately, there is also a repeating mechanism, where numerous ammunitions are manually loaded into a magazine, allowing multiple shots to be fired in one round. Repeating firearms can then be further categorised depending on the mechanical motion undertaken to load and fire the ammunition repeatedly (11). These motions include lever, bolt, pump, or semiautomatic actions. Regardless of the type of firearm used, ammunition is often a common and crucial piece of evidence found at firearm-related crime scenes (12).

1.1.2 Cartridges

Cartridges are often referred to as a single unit of ammunition. Ammunition consists of a cartridge casing that encloses the propellant (typically a black powder), primer, and projectile (bullet or pellets). Cartridge casings are mainly constructed of either metal or plastic. Plastic cartridges are mostly used as shotgun ammunition and are referred to as shotgun shells or shotshells, where metallic cartridges are regularly used in rifles (13). Metals and alloys used to make cartridge cases include nickel, aluminium, stainless steel, and brass (12). Different types of cartridges are further categorised into groups depending on the location of the primer. The primer can be located centrally or in the rim, resulting in centrefire and rimfire cartridges, respectively.

Standard cartridge nomenclature consists of the calibre followed by the manufacturer's name or the case length. Calibre is a numerical value and is expressed as either millimetres (e.g., 9mm) or in hundredths of an inch (e.g., .223), depending on geographical origin. The value itself approximately represents the diameter of the barrel of the firearm that the cartridge is designed to be fired from (11,13).

The most popular cartridge used in New Zealand and found at firearm-related crime scenes is a .**22 Long Rifle (.22 LR)**. Favoured for small game and vermin hunting, .22 LR cartridges are the most common type of ammunition not only in NZ but worldwide (13). Often also found at crime scenes in New Zealand is the **.223 Remington (.223 Rem)** cartridge case. Developed in 1957 as a military cartridge, this the type of ammunition used by the New Zealand Police in their rifles (Personal communication with ESR Ltd. Physical Firearms Specialist, 2024).

1.1.2.1 From cartridge-to-cartridge case; the firing mechanism of a firearm

To fire a firearm, the trigger must be pulled. This causes the release of the firing pin, which strikes the primer, igniting black powder contained within the cartridge case. Gases released by the combustion of the powder causes the release of the bullet from the cartridge case (11). The bullet projectiles out of the barrel of the firearm towards a target, and the cartridge case is pushed in the opposite direction, against the breechface. Unfired ammunition represents a cartridge with the components still contained within, whereas fired ammunition represents an empty cartridge case and a bullet.

1.1.3 Current forensic examinations of firearm-related evidence

In New Zealand, the analysis of cartridges and cartridge cases recovered from crime scenes is case-dependent (Personal communication with ESR Ltd. Firearms Specialist, 2023). Case managers use their experience to determine what would be the most effective for this evidence type. Typically, if the type of firearm crime is volume crime, retrieval of cartridges and cartridge cases is performed by Scene of Crime Officers from the New Zealand Police, who then propose what analysis methods are undertaken. If the crime is serious and forensic investigators are invited to the crime scene, a lead firearms expert will make the decision regarding the analysis of firearm evidence obtained from the scene. The main provider of forensic services in New Zealand is the Institute of Environmental Science and Research (ESR Ltd). There are three main types of forensic analyses and examinations that can be performed on unfired cartridges and/or fired cartridge cases retrieved from a crime scene. These are fingerprint analyses, physical examinations (microscopic analysis and comparisons to submitted firearms) or biological examinations (deoxyribonucleic acid (DNA) profiling). This research specifically focused on improving the biological examination of cartridges and fired cartridge cases.

1.2 FORENSIC DNA ANALYSIS

Since the 1980s, DNA has been utilised by forensic investigators as a means of identification**.** Visible biological material or items thought to contain biological traces collected from a crime scene undergo DNA analysis in hopes of obtaining a DNA profile. DNA profiles have high discriminatory power, allowing for the differentiation between, or exclusion of individuals. They can also provide crucial links between an individual to an object or a location (14). In general, a typical forensic DNA typing workflow consists of sampling, DNA extraction, quantification, the targeted amplification of STRs, capillary electrophoresis, and finally, the analysis of STR profiles.

1.2.1 STRs and Forensic Application

A majority of the human DNA genome is similar between individuals with minor variations occurring at specific regions. The location of some variable regions in the human genome is known, allowing them to act as a genetic marker. Conventionally, genetic markers targeted in forensic investigations consist of **short tandem repeats (STRs)**. Also referred to as microsatellites, STRs consist of DNA sequences, typically 2 - 6 base pairs (bps) in length, that are repeated in tandem (15,16). The total length of an STR depends on how many repeats are present. STR length is highly variable between individuals and globally utilised in forensic identifications (16). This variation in the length of STRs arises due to a high mutation rate, which is predominantly caused by DNA polymerase slippage during DNA replication (17). Commercial companies (e.g. Promega Corporation, Applied Biosystems) have created STR multiplex kits that allow for the targeted amplification of STR loci in one reaction (5). In New Zealand, several different autosomal STR multiplex kits; GlobalFiler™, SGMPlus™, MiniFiler™ and Identifiler™ Plus, are used to create a DNA profile that may be uploaded to the Databank.

DNA profiling is highly sensitive, and able to obtain profiles from minute and degraded biological samples such as saliva on cigarette butts, a single human hair, the charred remains of a disaster victim and items that have been simply touched by a person (18,19). As a result, this has expanded the different types of evidence that are subjected to DNA analysis.

1.2.2 Touch DNA

Locard's exchange principle states that when an object contacts another object, material is exchanged (11,20). This theory can be extended to the deposition of DNA through touch; when an individual touches a surface or item, their DNA will be deposited onto that surface or item. In 1997 it was discovered that DNA typing could generate DNA profiles from touched sources (19) and since then has been utilised in forensic investigations.

The terms touch DNA and trace DNA are commonly used to describe DNA and DNA profiles obtained from such deposition circumstances. In the forensic literature, these terms are used seemingly interchangeably. In this thesis, **trace DNA** refers to any minute levels of DNA obtained from a sample that is unable to be attributed to a specific body fluid (21). A review on DNA transfer by van Oorschot et al. (22) stated that if the action of transfer and/or the originating source of DNA is known, then using a term that indicates this is acceptable. Therefore, **touch DNA** in this thesis will be used to refer to DNA that has been deposited through the action of an individual touching or handling an object or item.

1.2.2.1 Composition of touch DNA

In the absence of biological fluids, touch DNA provides an opportunity for the generation of a DNA profile (23). The cellular composition of touch DNA is currently largely unknown, and for a considerable time, touch DNA was believed to be composed of shed outermost skin cells. These outermost skin cells are dead keratinocytes, scientifically known as corneocytes. Corneocytes are anucleate (24); therefore, it is commonly presumed that they do not contain nuclear DNA, complicating this logic. Recently, Burrill et al. (25) specifically addressed this apparent paradox and found that with successful lysing of corneocytes, DNA can be recovered. DNA from corneocytes is highly fragmented, and it was suggested that SNP profiling or the use of short amplicons might improve the recovery and subsequent value of touch DNA (25).

It has also been proposed that touch DNA can arise from many more different forms of genetic material. These include nucleated epithelial cells, fragmented cells and nuclei, and cell-free DNA (21,24,26). These possible sources of touch DNA are all capable of arising exogenously from hand contact with other body parts such as hair, face, and mouth or endogenously from within the hands. Endogenously, it has been suggested that small amounts of fragmented DNA arise from cornified skin layers and become present in sweat on an individual's fingers (27). Sweat contains cell-free nucleic acids that also can contribute to the composition of touch DNA (24).

1.2.2.2 Factors that impact touch DNA transfer, deposition, and recovery

Multiple studies have shown that it is possible to generate a DNA profile from a touched item (23,28,29), and a review by Burrill et al. (21) found that the measurable quantity of touch DNA recovered from various touched surfaces ranged from 0 to 169 ng. There are several factors that contribute to the quantity of touch DNA transferred and, therefore, its ability to be recovered and used for DNA typing.

Firstly, there is variation between individuals and their ability to deposit touch DNA. The term 'shedder status' reflects an individual's tendency to deposit DNA on physical contact (30). Individuals can be categorised as either good or poor shedders, depending on the ability to generate their DNA profile from a touched item (28). Where intra-variation in a person's ability to shed DNA can be seen (31), there is larger variation in DNA shedding between individuals (32). There is no standardised approach to assigning shedder status, with different studies defining what constitutes a good, intermediate, or poor shedder differently, and this approach is likely overly simplistic with inter-variation in shedder status unable to accurately represent the general population (30).

Another factor that is known to impact touch DNA transfer is the washing of hands. Generally, the amount of transferrable DNA on an individual's hands increases the longer they go without washing them (31). How long contact is made with a substrate and whether pressure with friction is applied also impacts the amount of touch DNA transferred (22). It has also been shown that most touch DNA is transferred during the initial contact (19); however, recently, it has been suggested that a layering phenomenon can occur with multiple instances of contact accumulating the amount of recoverable touch DNA (12). Touch DNA can be transferred directly from an individual's hands to an item (28), or it can be transferred indirectly, where an individual's DNA is present on a surface, but no contact was made between the individual and that surface (22). Consequently, non-self-DNA can be detected in a touch DNA deposit made by an individual 15 minutes after shaking hands with another person (33). One study that did not give their participants any restrictions regarding activities before deposition commonly found non-self-DNA in their touch DNA deposits, resulting in a majority of the typed DNA profiles being a two-person mixture, with the depositor mainly being the major contributor (32). Due to these factors, the amount of touch DNA deposited is immensely variable.

1.2.3 Touch DNA and firearm evidence

Recovering sufficient amounts of quality DNA from cartridges and cartridge cases for successful DNA typing is extremely challenging. This is due to low amounts of touch DNA transferred to the cartridge upon handling, inhibitory chemical interactions between nucleic acids and metallic cartridge surfaces and DNA degradation through exposure to high temperatures during the firing process (4,8,26,34). Each factor is discussed in further detail in the following subsections.

1.2.3.1 The transfer of touch DNA to metal

The chemical and physical properties of the touched substrate can impact the transfer and recovery of touch DNA (22,35,36). A study looking at the transfer and recovery of touch DNA on different substrates found that porous surfaces (e.g., paper, cardboard) yielded the largest number of profiles compared to nonporous surfaces (e.g., glass, metal) (36). Different types of nonporous surfaces can then impact the amount of DNA able to be recovered, with metal surfaces proven to be the most challenging (37) as less DNA is primarily transferred through touch to a metal substrate (38).

A key contributing factor is the amount of pressure applied during contact. An increase in finger pressure when touching an item increases the amount of DNA deposited (39). When loading a cartridge into a firearm (if no gloves are worn), direct contact between an individual's fingers and the cartridge is made. Cartridges are either manually loaded directly into the firearm's chamber or can be placed within a magazine. Moore et al. (40) suggested that loading cartridges into a magazine requires additional pressure that would lead to larger DNA deposits. Similarly, Dieltjes et al. (41) suggested that the last cartridge to be loaded into a magazine

(and therefore the first to be fired) requires more force to load and is the most suitable for touch DNA recovery. In contrary, Jansson et al. (7) found no clear correlation between the loading order of cartridges and the amount of DNA recovered.

1.2.3.2 The impact of metal on DNA

Cartridges used as ammunition for rifles are typically made from metal. Metal ions are commonly co-extracted alongside DNA from ammunition impacting downstream DNA analysis methods by inhibiting quantitation and amplification processes (42). Different types of metal and metal ions are known to have different impacts on DNA degradation and overall DNA recovery rates. Holland et al. (4) determined that mtDNA degradation was not only higher in samples from copper and brass cartridges but also occurred faster in comparison to aluminium cartridges. As a result, mtDNA recovery yields were significantly higher from aluminium samples compared to brass and copper samples. Similarly, it has been shown that more nuclear DNA can be recovered from nickel cartridge cases in comparison to brass cartridges (43–45).

Both brass and copper cartridges contain copper ions (Cu^{2+}). It has been suggested that copper ions mainly have a negative effect by causing damage to the DNA template rather than through PCR inhibition (46)**.** Copper is known to have a high binding affinity to DNA, specifically at base residues, leading to destabilisation of the double stranded helix and inducing damage (47,48). Copper also promotes the oxidation of DNA through acting as a catalyst for hydrogen peroxide (a reactive oxidative species) oxidation remarkably faster than other divalent cations (47). *In vivo* DNA damage by copper ions is avoided through active repair processes, however, these processes are unavailable *in vitro* resulting in the degradation of DNA (4). This has a negative impact on conventional DNA genetic markers targeted for DNA typing (4). Zinc ions also recovered from brass surfaces may further contribute synergistically to DNA damage through binding to alternative sites to copper ions within a DNA template (46).

1.2.3.3 The impact of firing on DNA

Cartridge cases and any touch DNA are exposed to high temperatures and pressures when a firearm is fired. To determine whether this harsh environment contributes to the limited success rate of obtaining DNA profiles from ammunition, Polley et al. (8) compared the amount of DNA recovered from unfired cartridges and fired cartridges. It was suggested that considerable loss of DNA occurs as a cartridge is loaded and fired from a firearm, with two samples (one unfired cartridge and one fired cartridge) meeting the study's minimum input threshold for amplification. The unfired sample gave a partial, degraded DNA profile, and the fired sample gave a single peak at the Amelogenin locus. This indicated that where enough DNA may be recovered from ammunition to warrant amplification, the quality of this DNA may be too degraded for STR amplification (8).

One study investigated the maximum temperatures that external cartridge surfaces reach during the firing process and whether this can impact the quantity and quality of DNA deposited on a cartridge (49). An infrared thermal camera determined that the maximum temperature attained on a brass cartridge was 62.85 ^{o}C and 98.85 ^{o}C for a copper cartridge. It was argued that the temperatures reached during the firing of a brass cartridge are less than the temperatures typically used in PCR amplification (49). Furthermore, PCR amplifications take about three hours (5) whereas once a firearm trigger is pulled, it only takes 1.2 ms for the cartridge case to leave the firearm (49). No significant difference between the amount of touch DNA recovered from unfired and fired cartridges has been observed, suggesting that firing does not impact recovery of DNA quantity (6). Conversely, Montpetit et al. (50) saw a similar trend to Polley et al., with less DNA and fewer interpretable profiles obtained from fired cartridge cases in comparison to unfired cartridges within their research.

It is likely that a combination of the factors discussed throughout Section 1.2.3 accumulate to result in any DNA recovered from cartridges and cartridge cases being low in quality and quantity.

1.3 DNA PROFILING AND FIREARM EVIDENCE

A recent review by Montpetit (34) summarised the different sampling, extraction and amplification methods used to perform DNA typing on ammunition, providing an overview of the attempts made in forensic research to obtain DNA profiles. This review concluded that DNA typing from both unfired and fired cartridges can be achieved through multiple different DNA sampling, extraction, and STR amplification methods.

1.3.1 DNA recovery and sampling

Optimal recovery of DNA is influenced by the surface from which it is being recovered (51), with different sampling methods more successful for different substrates. For touch DNA samples it is vital to minimise any loss of DNA by selecting an appropriate sampling method. A focus in forensic research has been on optimising DNA recovery from cartridge cases resulting in efforts to determine the best sampling method and the creation of new sampling methods specific for this evidence type (41,44,45,50,52).

A comparison between (single) swabbing, vacuum filtration, and tape-lifting methods on different types of ammunition found that both swabbing and tape-lifting outperformed vacuum filtration (44). Overall success was limited, with the best DNA recovery amounting to less than 25% of the original known amount of deposited DNA (44). In a separate study, a traditional single swabbing method was compared to a soaking method, where the cartridges/cartridge cases were soaked in a lysis buffer (10 mM Tris–HCL, 10 mM EDTA,

50 mM NaCl, 2% SDS) and proteinase K solution (50). It was found that this soaking method obtained more interpretable DNA profiles than swabbing (50). There were concerns that using a soaking method could lead to oxidisation of the cartridge case and cause the partial dissolving of the cartridge surface over an extended period (41), thus impacting any physical examinations that rely on the identification of markings and striations made during the firing process. It was found that if soaking is restricted to 20 minutes there is no effect on microscopic striation detail quality (53).

It is hard to directly compare between different studies as the method of DNA deposition, the type of DNA deposited, the type of ammunition used, and the method of sampling all vary. Additional method variation is introduced through different downstream amplification methods, and variation also naturally arises from varying amounts of touch DNA deposition by individuals, as described in Section 1.2.2.2. As Montpetit (34) concluded, the only inference able to be drawn from previous studies about DNA recovery and sampling methods is that soaking, tape lifting, rinsing and both double and single swabbing can all provide DNA typing results.

1.3.1.1 Success of DNA typing

The success rate for conventional DNA typing through CE-STR analysis has varied from reports of no success to 27.2% of samples having profiles obtained (8,50). One study was unable to detect any touch DNA on unfired cartridge samples (8). Only one fired cartridge case recovered 288 pg of DNA, which when amplified by CE-STR, resulted only in the typing of the Amelogenin locus. It has been reported that larger STR targets fail to amplify from samples recovered from both fired cartridge cases and unfired cartridges (7). Consequently, amplification kits that target shorter amplicons have greater DNA typing success. Horsman-Hall et al. (6) found that samples recovered from cartridges held for 30 seconds had a greater percentage of loci typed using the AmpFlSTR™ MiniFiler™ PCR Amplification Kit (Thermo Fisher Scientific), compared to the PowerPlex 16® BIO System or Identifiler™ kits. Out of these three different amplification kits, MiniFiler™ targets the shortest amplicons. Specifically, MiniFiler™ is a mini-STR assay whereby shorter amplicons are used to target larger STR loci, that typically tend to be the first to dropout (54). The use of shorter amplicons results in a higher possibility that DNA profiles can be obtained from highly degraded forensic samples that contain a fragmented DNA template (54). Montpetit et al. (50) used MiniFiler™ on cartridge and cartridge case samples that had less than 50 pg of DNA recovered and were able to obtain interpretable DNA profiles in 19.25 % of these samples.

In reality, the success rate of obtaining reliable and usable DNA profiles from cartridges and or cartridge cases recovered from a crime scene is even lower. Where Dieltjes et al. (41) found that 163 out of 616 (26.5%) criminal cases were able to result in at least one reportable DNA profile, only 6.9% of the individual samples

from these cases produced a reliable DNA profile. Specifically, this saw 4,085 cartridges, bullets and cartridge casings tested, with only 283 items producing a STR profile (41). This is unsurprising as tightly controlled experimental conditions are not seen in the live casework, with any firearm evidence exposed to harsh environmental insults, further degrading any DNA present. An expectation of low success rates may impact the decision of whether to even process such samples to begin with (Personal communication with ESR Ltd. Forensic Biology Senior Science Leader, 2024). Based on the yield of low-level data obtained in previous research and casework attempts, it can be concluded that conventional DNA typing (CE-STR) has had limited success in obtaining DNA profiles from cartridges and cartridge cases.

1.3.2 The current problem

Since mid-2017, 144 samples recovered from such evidence types have been submitted for DNA analysis at ESR Ltd (Personal communication with ESR Ltd. Firearms Specialist, 2024). Only seven of the submitted samples have provided successful results, of which six generated a sufficient number of alleles to be uploaded onto the Databank and one sample was restricted to a limited comparison with a reference sample. The increasing number of firearm crimes and the minimal number of these crimes able to be successfully solved using conventional DNA profiling technologies highlights the need for research into an alternative method for obtaining DNA profiles from evidence recovered from firearm-associated crime scenes.

1.4 Massively Parallel Sequencing

Massively parallel sequencing (MPS) is a DNA typing method that sees the rapid, high-throughput sequencing of entire genomes or targeted regions of DNA. Since its implementation into the forensic field, it has become apparent that MPS has several advantages over conventional DNA typing methods.

Firstly, MPS allows for the analysis of more genetic markers in one single run compared to conventional CE-STR methods (9,55–58). Different CE-STR assays can be combined into one single multiplex MPS assay, offering a more time-efficient alternative to CE-STR, which requires individual assays for the targeting of different genetic markers such as autosomal STRs, X-chromosome STRs and Y-chromosome STRs (59). Several analyses with different CE-STR assays inevitably uses more forensic DNA sample, which in casework is already valuable and limited (60). A multiplex MPS assay offers an efficient approach for gaining the greatest amount of genetic information from the smallest amount of sample input. Secondly, MPS identifies target DNA fragments based on their nucleotide sequence and fragment length (55,57). This expands the range of genetic markers that can be targeted for forensic use as MPS allows for the targeting of both short amplicons and amplicons of the same length (55,61), previously limited with conventional CE-STR methods. The use of short

amplicons improves the likelihood of obtaining a DNA profile from highly degraded samples (59) commonly seen in forensic casework.

An alternative type of genetic marker that can be targeted for forensic investigations using MPS are **single nucleotide polymorphisms (SNPs)**. SNPs are a variation at a single nucleotide base at a unique location in the human DNA genome (62). SNPs have a low mutation rate, are highly abundant, and can be analysed from short amplicons, making them extremely useful markers for degraded forensic samples (9,62,63). Furthermore, SNPs make up 85% of genetic variation in humans and provide additional information about identity, lineage, ancestry, and phenotypic traits (9,63). One type of SNP, identity informative SNPs (iiSNPs), helps with human identification in forensic applications. iiSNPs are often bi-allelic and less informative than STRs, with at least 50-60 iiSNPs required to obtain the same discrimination power as current STR forensic genetic markers (63–65). As a result, STRs have remained the 'gold standard' DNA marker in forensic science. The advent of MPS now allows many SNPs to be targeted alongside STR markers in a single MPS multiplex assay, providing forensic scientists an opportunity to effectively capitalise on the advantages of using SNPs as a genetic marker.

1.4.1 MPS platforms for forensic use

A typical MPS workflow sees the inclusion of a library preparation step. After DNA extraction and quantification, PCR enrichment sees the targeted amplification of genetic markers before adaptors are added to uniquely identify each sample, creating a library. Libraries are then purified, normalised, and pooled together before sequencing. There are currently two benchtop MPS sequencers commonly employed for forensic use, the Ion Torrent PGM™ and/or Ion S5™ (both by ThermoFisher Scientific) and the MiSeq FGx™ benchtop sequencing instrument (Illumina)(66). Both platforms use a sequencing by synthesis approach but employ different sequencing technologies to do so.

The Ion Torrent PGM™ and Ion S5™ Systems use a semi-conductor technology for sequencing. This technology sees the detection of hydrogen ions produced during template driven DNA synthesis: as a base is added to the synthesised DNA strand, protons are naturally released as by-products (67). In comparison, the MiSeq FGx™ sequencing platform uses a reversible dye terminator technology (68). The sequencing by synthesis approach undertaken is similar to Sanger sequencing. All nucleotide bases and a polymerase are washed over the flow cell. Bases are incorporated through complementary binding to clonal clusters of library fragments (which contain the target amplicon), that have previously hybridised to the same flow cell surface during cluster generation (68). Each of the four nucleotide bases then emit a light signal at a different wavelength once incorporated, and this light signal (fluorescence) is recorded by a camera in real-time (69). A MiSeq FGx™ benchtop sequencer is the MPS sequencing platform currently employed at ESR Ltd.

1.4.2 MiSeq FGx™ Forensic Genomic Solution

The MiSeq FGx™ benchtop sequencing instrument is a part of the MiSeq FGx™ Forensic Genomic Solution. Developed in 2015, the MiSeq FGx Solution was the first commercial workflow to utilise MPS technologies for forensic genomics and the DNA analysis of forensic samples (70). This workflow has successfully undergone evaluations (61,71), and a developmental validation in accordance with SWGDAM guidelines (57).

The MiSeq FGx Solution workflow consists of four different components:

1) ForenSeq™ DNA Signature Prep Kit
2) MiSeq FGx™ Sequencing Reagent Kit
3) MiSeq FGx™ benchtop sequencing platform
4) ForenSeq™ Universal Analysis Software

1.4.2.1 ForenSeq™ DNA Signature Prep Kit

The ForenSeq™ DNA Signature Prep Kit (ForenSeq™ Kit) is used for the targeted amplification of up to 231 forensically relevant loci and the generation of libraries for sequencing (70). Two different primer mixes are included: DNA Primer Mix A (DPMA) and DNA Primer Mix B (DPMB). DPMA sees the amplification of Amelogenin, 27 autosomal STRs (a-STR), 24 y-STRs, seven X-STRs and 94 identity informative SNPs (iiSNPs), whereas DPMB sees the amplification of all DPMA targets, 24 phenotypic informative SNPs (piSNPs) and 56 biogeographical ancestry SNPs (aiSNPs) (57). The ForenSeq™ Kit uses two independent PCR reactions to generate sequencing libraries (56). The first PCR reaction amplifies, and tags targeted gDNA regions, and the second reaction sees the addition of indexed adaptors to enable the simultaneous sequencing of all libraries (56,57).

Both the ForenSeq™ Kit and the MiSeq FGx™ platform have previously undergone internal validation studies at ESR (72). These studies looked at sensitivity, repeatability, operator reproducibility, and application of the manufacturer's workflow to case-work samples. Overall, it was concluded that the ForenSeq™ Kit and MiSeq FGx™ platform are fit for purpose and use in the New Zealand justice sector. The compatibility of the 59 ForenSeq™ STR loci with STR loci currently targeted for DNA typing by forensic investigators becomes important for implementing the ForenSeq™ Kit into casework practices, especially in jurisdictions with well-established DNA databases. Table 1.1 lists the 24 a-STR loci targeted by various amplification kits currently used for DNA profiling in New Zealand.

Table 1.1 Compatibility of ForenSeq™ DNA Signature Prep Kit with loci targeted by STR multiplexes used at ESR Ltd. Green indicates that loci is seen in that multiplex.

Locus	GlobalFiler™	Identifiler™ Plus	SGM Plus™	MiniFiler™	ForenSeq™
D3S1358					
vWA					
D16S539					
CSF1PO					
TPOX					
Y-indel					
Amelogenin					
D8S1179					
D21S11					
D18S51					
DYS391					
D2S441					
D19S433					
TH01					
FGA					
D22S1045					
D5S818					
D13S317					
D7S820					
SE33					
D10S1248					
D1S1656					
D12S391					
D2S1338					

As seen in Table 1.1, the ForenSeq™ Kit contains 22 of the same STRs currently used in New Zealand. Several studies have compared the concordance of genotypes obtained from various orthogonal CE-STR kits with genotypes obtained by the ForenSeq™ Kit (9,58,60). Fully concordant genotypes between the ForenSeq™ Kit and CE-STR technologies were generated, with MPS allowing for the detection of further genetic information (58,60,73). This proves the ForenSeq™ Kit is a viable alternative option for DNA profiling and can provide information that is consistent with STR profiles that have been uploaded to the New Zealand Databank.

1.4.2.2 ForenSeq™ Kit performance with degraded samples

The performance of the ForenSeq™ Kit has been tested on casework-type DNA samples (9) and DNA samples that have been manually degraded (55,57). The ForenSeq™ Kit was able to detect equal or more STR alleles from aged swabs and bone samples, which had previously produced poor results with CE-STR (9). Jäger et al.

(57) manually fragmented DNA samples into nonspecific fragments. It was found that both short and medium length ForenSeq™ STRs were able to provide more information compared to long-length STRs, increasing the possibility of recovering genetic information from degraded samples. Similarly, Sharma et al. (55) found that loci that dropped out or gave low read coverage in serially degraded DNA samples were typically long-length amplicons such as PentaE and DXS8378, which are 362-481 and 434-458 base pairs respectively (74). In general, SNP loci are more robust and ≥ 15% more likely to be typed compared to a-STRs (57). However, it has also been found that iiSNPs are overall the most sensitive to degradation (55) with rs1736442, rs2920816, rs1031825 and rs7041158 common loci that dropout, due to low read counts, even in high-quantity DNA samples (61).

Sharma et al. (55) also compared the performance of the ForenSeq™ Kit to a CE-STR assay, the PowerPlex® Fusion System (Promega), The 20 loci in common have target amplicons that are 85 to 306 bp in length in the ForenSeq™ Kit, shorter than the amplicons in the PowerPlex® System (72 to 464 bp). Perhaps, consequently, the MPS data typed 90% of the loci while CE-STR data typed only 35% in severely degraded samples (55).

1.4.3 Conclusion

Several studies have shown that the ForenSeq™ DNA Signature Prep Kit not only provides concordant results with genotypes produced from conventional DNA typing methods but also has an increased sensitivity and success rate for obtaining profiles from degraded forensic samples (9,55,57,73). Furthermore, touch DNA has been able to be successfully amplified with the ForenSeq™ Kit and saw more a-STR loci typed using MPS compared to using CE-STR (75). This indicates that using MPS for DNA typing is highly valuable for limited forensic samples, such as touch DNA (75). Based on this knowledge, it is hypothesised that the use of shorter STR amplicons and SNPs in the ForenSeq™ DNA Signature Prep Kit will allow for successful DNA typing from cartridges and cartridge cases.

1.5 MTDNA, AN ALTERNATIVE GENETIC MARKER

In situations where nuclear DNA is too degraded or in extremely low quantities and analysis is not possible, **mitochondrial DNA (mtDNA)** is targeted as a genetic marker. mtDNA typing is a robust sequencing technique used to generate DNA profiles from small, highly degraded forensic samples (76,77).

1.5.1 mtDNA background

Mitochondria are double-membraned organelles located within the cytoplasm of eukaryote cells (78). In 1963, it was discovered that mitochondria contain their own DNA - mtDNA (79). The human mtDNA genome (mitogenome) is 16,569 bp long and organised in a circular structure. Like nDNA, mtDNA is composed of two

strands of nucleotides. These strands differ in their base composition, resulting in a heavy, purine-rich strand and a light pyrimidine-rich strand (78,80).

The mitogenome is divided into a coding and a control region. The coding region contains a majority of the mitogenome and encodes 37 genes that are responsible for the production of RNAs and enzymes necessary for mitochondrial function (78,81). The control region (also referred to as the D-loop region or non-coding region) is approximately 1125 bp in length and contains the heavy strands origin of DNA replication as well as promoters for the transcription of genes on both strands (78). The control region is noncoding and there are fewer selection pressures for this region (82). As a result, there are specific sections in the control region that are considered 'hot spots' for mutations (in comparison to the coding region) (78,83). These sections are referred to as hypervariable region one (HV1), hypervariable region two (HV2) and hypervariable region three (HV3) (81,84). As the names suggest, these regions are highly variable between individuals and, therefore, highly discriminating (77). Due to this, hypervariable regions were initially the main target for forensic mtDNA typing (78).

1.5.2 mtDNA and forensic application

Utilising mtDNA for forensic investigations has two main typing advantages over nuclear DNA (78,80). Firstly, mitochondria have a high copy number per cell. There are hundreds of mitochondria within a single human cell, with a single mitochondrion containing multiple copies of mtDNA (85). In comparison, nDNA is only present in two copies per cell (77,86). A high copy number makes it more likely that a copy of mtDNA can survive in highly degrading conditions, increasing its sensitivity as a profiling technique (16,77,78,82). Additionally, mitochondria have a double-membrane system offering further protection from degradation. This increases the likelihood of recovering mtDNA from highly degraded forensic samples such as bones, fingernails, telogen hairs and ancient remains (77,82). Secondly, mtDNA is maternally inherited and does not undergo recombination (87). Excluding mutations, this means that individuals of the same maternal line will have the same mtDNA genome (82). Whilst this means that mtDNA has a lower discriminatory ability than nDNA analysis, targeting mtDNA is extremely useful for disaster victim identification (DVI), and missing person cases as maternally related individuals can provide reference samples for direct comparison (78,82).

A mtDNA profile is referred to as a haplotype, in which any nucleotide base differing to the revised Cambridge Reference Sequence (rCRS) is reported. Haplotypes are treated as a single locus in forensic analysis. As a result, identification occurs between different maternal lineages rather than between individuals. Haplotypes can be further categorised into groups of similar haplotypes, forming a haplogroup. Haplogroups can be used to determine the distribution of a haplotype globally (88).

1.5.3 Conventional mtDNA typing

Traditionally, forensic mtDNA typing targeted the hypervariable regions within the noncoding control region (CR) using capillary electrophoresis-based Sanger-type sequencing (CE-STS) (82). Where this method has been used for over three decades in forensic science (88), sequencing only the CR restricts the power of discrimination and can lead to ambiguous haplogroup assignments (89). Furthermore, many variants required for defining different haplogroups are located in the coding region (90). This means haplotype and haplogroup assignments can be improved by expanding from sequencing only the CR to sequencing the entire mitogenome (89). However, using STS to target the whole mitochondrial genome becomes labour-intensive, time-consuming, and expensive (82,88,89,91). This makes it not practical for routine whole mitogenome analysis of forensic samples (92).

1.5.4 mtDNA and MPS

The advent of MPS has allowed for the accessible targeting of the whole mtDNA genome as a genetic marker overcoming limitations that may arise with CE-STS (90). MPS is more cost-effective per sample and is becoming increasingly accessible for forensic laboratories through the recent development of commercial kits that target the whole mitogenome (88). MPS is also more sensitive, and through the generation of sequencing information, allows for the detection of low-level heteroplasmy, resolving true minor variants from background noise and deconvoluting mixtures (88). Specifically, MPS can detect heteroplasmy at threshold levels of 1-2%, whereas CE-STS can detect heteroplasmy down to only 10-20% (91,93).

Since implementation into forensic analysis, mtDNA haplotypes have always been reported based on a comparison to the rCRS (94–96). This enables an easy transition for forensic laboratories from conventional CE-STS typing to MPS (94). Furthermore, mtDNA genomes typed by MPS are concordant with those generated by STS (92). Commercial kits have been developed for MPS of only the mtDNA CR, such as ForenSeq™ mtDNA Control Region (Verogen) and PowerSeq™ CRM Nested System (Promega). However, targeting the whole mitogenome generates more information and improves maternal lineage identification. Specifically, in the New Zealand population, the discriminatory power of mtDNA as a forensic marker greatly increases when the entire mitogenome is targeted (97). There are several approaches for targeting and enriching the whole mtDNA genome for MPS. These are:

1.5.4.1 Large amplicon approach

This approach sees the long-range PCR enrichment of large amplicons. Specifically, two large overlapping amplicons that are approximately 8,500 bp in length (89,91) and span the entire mitogenome are used. The use of large amplicons is applicable for the analysis of high-quality DNA samples (89), which in forensics tend to be samples, such as buccal swabs, used for the generation of reference mtDNA profiles. Utilising these

large amplicon targets is not practical for casework samples, where the DNA template is likely highly fragmented and degraded.

1.5.4.2 Small amplicon approach

For highly degraded samples, a small amplicon approach is desirable. This approach has been referred to as 'mito-mini' and was first developed for conventional mtDNA profiling of the CR (98). It has more recently been adapted to MPS technologies, allowing the enrichment of the entire mtDNA genome through two primer pools of short overlapping fragments. Chaitanya et al. (90) designed 161 amplicons ranging from 144 bp to 230 bp in length, which successfully amplified the whole mitogenome in two primer pool multiplex reactions. This approach has since been commercialised into panels including:

- The **Precision ID mtDNA Whole Genome Panel**, which is an amplification panel from Thermo Fisher Scientific (Waltham, MA, USA). Specifically developed for forensic samples, this kit utilises a two-pool multiplex assay with a small overlapping amplicon design to allow for the sequencing of the entire mitochondrial genome (99). There are 81 primer pairs contained within each pool, and the resulting PCR amplicons are approximately 163 bps in size.

- The **ForenSeq™ mtDNA Whole Genome Kit** (Verogen, Inc., San Diego, CA, USA), which sees the amplification of the whole mtDNA genome through 245 amplicons (with an average size of 131 bps). This kit is part of the ForenSeq™ Whole Genome Solution, a streamlined workflow that sees the preparation of libraries specifically for sequencing on the MiSeq FGx™ platform (100).

- The **PowerSeq® Whole Mito System** (Promega, Madison, WI, USA), which sees the amplification of the entire mitochondrial genome through 161 small amplicons ranging from 92 – 254 bp in size (101). Released in 2023, the PowerSeq® System includes a library preparation workflow compatible with MiSeq and MiSeq FGx™ platforms.

1.5.4.3 Other approaches

Where the above approaches are examples of targeted amplification, a DNA capture approach could be performed instead. One example of this is hybridisation capture, whereby target DNA sequences (in the form of a library) are hybridised to either DNA or RNA probes (102,103). The probes are then bound to magnetic beads and any DNA that was not captured by the beads is removed. Hybridisation capture has been successfully coupled with MPS, and the enriched libraries can be sequenced on both Ion and Illumina sequencing platforms (102). The target mtDNA fragments for hybridisation capture are < 100 bp in length,

allowing for mtDNA profiling of extremely challenging and fragmented samples (103). The main limitation with this approach is that it is more labour intensive than PCR enrichment (102,104).

1.5.5 Detection of true mtDNA variation

The increased sensitivity of MPS generated data has resulted in some challenges, the main one being the detection of true low-level mtDNA variants. Heteroplasmy, a type of true mtDNA variation, needs to be correctly discerned from artefacts such as nuclear mitochondrial pseudogenes (NUMTs), DNA damage, low levels of contamination, background noise, and stochastic, sequencing and PCR errors (93).

1.5.5.1 mtDNA heteroplasmy

Heteroplasmy is a phenomenon that occurs within mtDNA. This is when there are two or more mtDNA subpopulations within one individual, tissue, single cell, or mitochondrion (78,93). As multiple copies of the mitogenome exist in a cell and replicate autonomously, a mutation can arise in one mitogenome and become present in only a portion of mitogenomes, resulting in heteroplasmy. There are two types of heteroplasmy: length heteroplasmy (LHP) and point length heteroplasmy (PHP). LHP is when an individual has mtDNA genomes that differ in their length, and PHP is when individuals have mtDNA genomes that differ at a single nuclear position (93).

mtDNA mutations can arise at the somatic level and randomly occur throughout an individual's life. mtDNA replication has a higher mutation rate compared to nuclear DNA (78). This mutation rate is estimated to be approximately 5 to 15 times greater and is due to both the mitochondrial genome's cellular location and its susceptibility to mutations (105). Within a mitochondrion, the mtDNA is located within close proximity to oxidative free radicals released as byproducts from catabolic reactions in the electron transport system (105). mtDNA also lacks the presence of histone proteins (106). This restricts mtDNA protection, which, coupled with a decreased proofreading function of mtDNA polymerase compared to nuclear DNA polymerase, means that any errors made in mtDNA replication are unlikely to be repaired (81,105,106). The cellular location and frequency of each mutation influences whether they are harmful or not to an individual. The mutations that have no impact are detected as heteroplasmy.

Other mutations can be pathogenic and lead to mtDNA disorders. Extending the analysis of mtDNA from the CR to the entire mitogenome can reveal personal genetic information about the donor (107). There have been 92 pathogenic variants at 92 unique sites in the mitogenome confirmed by Marshall et al. (107). Strict privacy measures at ESR Ltd. require that such sites are not visible to the analyst during any part of mtDNA data

analysis. GeneMarker® HTS, a bioinformatic software that provides an analysis workflow for MPS generated mtDNA data, allows the analyst to mask such personal health information (PHI) sites.

Mutations can also occur at the germline level and are passed to offspring through maternal inheritance. A bottleneck effect sees a reduction of the number of maternal mtDNA molecules passed into the developing oocyte (106). If a mtDNA molecule with a mutation passes through this bottleneck, it could be present in a higher proportion in the oocyte. The mutation/variant can become fixed in the offspring through genetic drift (106), resulting in two versions of the mitogenome. During foetal development different mtDNA molecules, including low-level variants, can segregate throughout embryo leading to these variants being present in different frequencies in different tissues. This means that the rate of heteroplasmy can vary between various forensic sample types. It has been determined that buccal cells are a better reference sample source when investigating evidentiary hairs, compared to blood cells (108). This is thought to be because both buccal cells and hair follicles arise from the ectoderm germ layer during foetal development (108). So far, no study has investigated the heteroplasmy rates of touch DNA samples. Since the epidermis (the outermost skin layer) also derives from the ectoderm, this suggests that buccal cells could also be used to obtain comparative reference samples for touch DNA samples.

Heteroplasmy can complicate forensic analysis and interpretation of mtDNA (93). Where MPS allows for the detection of lower levels of heteroplasmy, which can see the differentiation of maternal relatives (109), it can also introduce additional sources of false positive error. It is essential that only true mtDNA variation is correctly called as heteroplasmy. There are several ways this can be achieved. Bioinformatic analysis tools, such as GeneMarker HTS and mitoSAVE (an Excel workbook), allow the user to set a minor variant frequency (110). This minor variant frequency threshold can help distinguish true PHP from possible false positives, such as sequencing errors and background noise. True heteroplasmy can also be determined through an understanding of heteroplasmy prevalence in the human population. At least one PHP is seen in around 25% of individuals (93). Furthermore, an individual is expected to have up to a maximum of three PHPs (94), with more indicating the presence of a mixture or low-level contamination. To determine true mtDNA variation within an individual, forensic analysts need to be aware of and able to identify all of the different artefacts that can pose as heteroplasmy.

LHP is largely ignored in CE-STS analysis as it complicates the interpretation of data (111). Using MPS allows for the analysis of LHP, however, employing different sequencing methodologies and technologies introduces variation in how LHP presents and is then interpreted (111). One solution that is recommended for population database samples is to call LHP only when it is present in high frequencies and classified as a major variant (112).

1.5.5.2 NUMTs

Nuclear mitochondrial pseudogene (NUMT) sequences are fragments of mtDNA that have translocated into the nDNA genome. Initially entering the nuclear genome through insertion events, the duplication of NUMTs has resulted in their continual evolution (94). Over half of the NUMTs discovered by Dayama et al. (113) were present in low frequencies, indicating that they had integrated recently into the nuclear genome. This means that some NUMTs are phylogenetically young (94) and mtDNA translocation into the nuclear genome is an ongoing process (113).

The detection of NUMTs has increased due to the sensitivity of MPS and development of small-amplicon sequencing assays that target the entire mitogenome (94). MPS generated mtDNA data shows that true mitochondrial reads outnumber NUMT sequence reads (114). As nDNA is present in only two copies per cell, NUMTs are present in low frequencies (115) and can be mistaken for true low-level PHP variants. Highly degraded forensic samples subjected to mtDNA analysis often result in NUMT signals that are indistinguishable to background noise (94). In comparison, NUMTs are more frequently seen in high-quality samples with high concentrations of DNA, e.g., reference samples, that have been amplified using short amplicon assays. This is because most NUMTS are 100 to 500 bp in length (94). As a small amplicon approach for whole mtDNA genome enrichment targets mtDNA fragments in this same size range, co-amplification of NUMTs can occur through non-specific primer binding (90,115). NUMT interference in reference samples can be mitigated through using a long amplicon assay. In cases where a short amplicon assay is used, one option is to dilute the samples prior to amplification to limit the nDNA template present in the sample while ensuring multiple copies of mtDNA remain (94).

To further avoid NUMT interference, Woerner et al. (116) created an application called 'Remove the Numts!' (RtN!), which filters out NUMT and noise artifacts from MPS generated data through the mapping of data to human haplotype and NUMT sequence databases. A similar approach is used by the Universal Analysis Software (UAS) bioinformatic analysis platform whereby the automatic alignment of sequenced mtDNA data to a reference human NUMT compilation occurs with any aligned reads removed before analysis. Woerner et al. (116) saw 57% of low-frequency reads removed; however, warned that there was still the risk of true mtDNA variant removal through an automatic approach. Alternatively, Cihlar et al. (114) has created a workflow to help analysts distinguish between mtDNA and NUMT sequences through manually assessing different molecular and bioinformatic traits.

1.5.5.3 Sequencing errors and DNA damage

Errors that arise throughout sequencing can also complicate heteroplasmy detection. The MiSeq sequencing system is known to have a low substitution error rate, with the misincorporation of C nucleotides the most prevalent error type (117). Homopolymer sequences, which are difficult for the MiSeq sequencing chemistry (89), and the quality of the sample being sequenced can influence this error rate. Touch DNA recovered from ammunition has a higher rate of sequencing error compared to buccal samples, reflecting the low DNA quality and quantity of this sample (117). Specifically for this sample type, G nucleotide substitutions are commonly seen with some error sites reported to exceed a 2% minor variant threshold. It is recommended that adjustments of this threshold for samples with damaged DNA may be required to avoid false calls of heteroplasmy. Sequencing error can also be identified through strand imbalance; when a variant is seen in either the forward or reverse sequencing read but not the other (94). True heteroplasmy, in contrast, is detected at the same nucleotide position on both sequencing strands.

Forensic samples recovered from crime scenes are often degraded and damaged through harsh environmental exposures. Such damage can include oxidation, deamination and depurination of the template DNA (117). Background noise levels have been reported to be elevated in DNA recovered from metallic cartridges and can marginally impact heteroplasmy detection (4). In addition, forensic samples are often of low quantity and result in low template amplification, which can also produce stochastic errors where the wrong base is incorporated during PCR (118). These types of errors are not reproducible; therefore, using replicates of the same sample can help mitigate and distinguish DNA damage from true PHPs (4,94).

1.5.6 Commercial whole mtDNA MPS workflows for use in NZ

Previous research at ESR Ltd. investigated which commercial whole mtDNA MPS panel should be implemented into casework in New Zealand (97). These short amplicon assays were the Precision ID mtDNA Whole Genome Panel (Precision ID panel) an alpha version of the PowerSeq™ Whole Mito System, and the ForenSeq™ mtDNA Whole Genome Solution. Each panel was evaluated in terms of how well they generated concordant haplotypes, obtained full mitogenome coverage, and accurately detected PHP and LHP variants. As mentioned previously, ESR Ltd. has a MiSeq FGx™ sequencing instrument; how each of these amplification panels performed in a workflow suitable for this platform was also assessed. It was determined that the Precision ID panel clearly outperformed both the ForenSeq™ and PowerSeq® panels. Adapted so that it could be sequenced on the MiSeq FGx™ platform, the final custom workflow was able to produce full mitochondrial haplotypes and accurately detect LHP and low-level PHP variants (97). It was concluded that the Precision ID MiSeq FGx™ workflow was the most suitable for further validation in New Zealand.

1.5.6.1 Precision ID mtDNA Whole Genome Panel

The Precision ID panel has undergone several evaluations (119–122) as well as an internal validation for use on the Ion S5™ sequencer (123). Many of these evaluation studies investigated the performance of the Precision ID panel and MPS workflow on forensic casework-like samples in comparison to conventional CE-STS workflows. One study analysed 15 forensically relevant samples (e.g., hair shafts, ancient remains) that formerly had only resulted in partial haplotypes using CE-STS and were able to generate six partial and nine full mitogenome sequences using the Precision ID panel and MPS workflow (121). These results were not only concordant with the data previously generated from CE-STS but were also able to provide additional sequence information and increase discriminatory power. The Precision ID panel has been used alongside CE-STS to help identify remains located in a Viking Age burial site in Sweden (124). While CE-STS was able to provide a broad assignment of a haplogroup to most individuals, the Precision ID panel provided additional coding region information, improving the precision of haplogroup assignment (124). Where this panel was designed for use on the Ion S5™ sequencer, it has been found that consistent and concordant haplotypes are still able to be produced on the MiSeq™ FGx sequencer (120). It has been repeatedly concluded that this panel produces consistent, reliable sequencing results and provides an increased discrimination power compared to traditional control region only analysis**,** making it useful for forensic application (120–122).

1.5.7 Conclusion

The emergence of MPS in forensic DNA typing has led to the use of the whole mitogenome as a forensic genetic marker. Specifically, using a short amplicon approach has proven successful for obtaining genetic information from highly degraded forensic samples (121,124). The Precision ID panel is extremely sensitive and has been successfully utilised in a high performing custom sequencing workflow at ESR that has demonstrated applicability to both reference and casework type samples (e.g., telogen hairs) (97). The limit of this workflow currently remains unknown, creating an opportunity for research into determining its performance with different low DNA quantity forensic samples with previously limited nDNA typing success, such as touch DNA recovered from firearm ammunition. So far, partial control region (CR) haplotypes have been able to be recovered from touch DNA deposited on unfired copper, brass and aluminium cartridges using MPS technologies (4). It is hypothesised that using the Precision ID panel will also allow for successful mtDNA profiling from both cartridges and cartridge cases.

1.6 RESEARCH AIMS AND OBJECTIVES

The overall aim of this research was to determine if alternative mtDNA and DNA MPS technologies can successfully generate DNA profiles from firearm evidence types. Research objectives were set out in a tiered approach to investigate this.

The first set of objectives were:

- To determine if mtDNA haplotypes could be successfully recovered from unfired cartridges using the commercially available Precision ID Whole mtDNA Genome Panel kit (Thermo Fisher Scientific).
- To determine if nuclear DNA (nDNA) STR and iiSNP profiles could also be recovered from unfired cartridges using the commercially available ForenSeq™ DNA Signature Prep kit (Verogen).

These initial objectives determine if DNA profiles can be recovered from unfired cartridges using sequencing workflows previously evaluated and developed at ESR, Ltd (72,97).

The second objective was to optimise the mtDNA sequencing workflow to determine if the quality of mtDNA haplotypes could be improved.

This led on to the third objectives which were:

- To determine if mtDNA haplotypes could be recovered from fired cartridge cases using the Precision ID Whole mtDNA Genome Panel kit (Thermo Fisher Scientific) and the optimised mtDNA sequencing workflow.
- To determine if nDNA STR and iiSNP profiles could be recovered from fired cartridge cases using the ForenSeq™ DNA Signature Prep kit (Verogen) and the DNA sequencing workflow.

Finally, a comparison on the two approaches determined whether either or both are a suitable alternative for obtaining DNA profiles from fired cartridge cases.

2 METHODOLOGY

2.1 INTRODUCTION

This chapter outlines the experimental methods used in this research. A general description of the final two methods used to obtain mtDNA and nDNA profiles is provided, with any developments and/or optimisations discussed in subsequent chapters. The workflow for both sequencing methods starts using the same experimental steps before diverging into separate workflows.

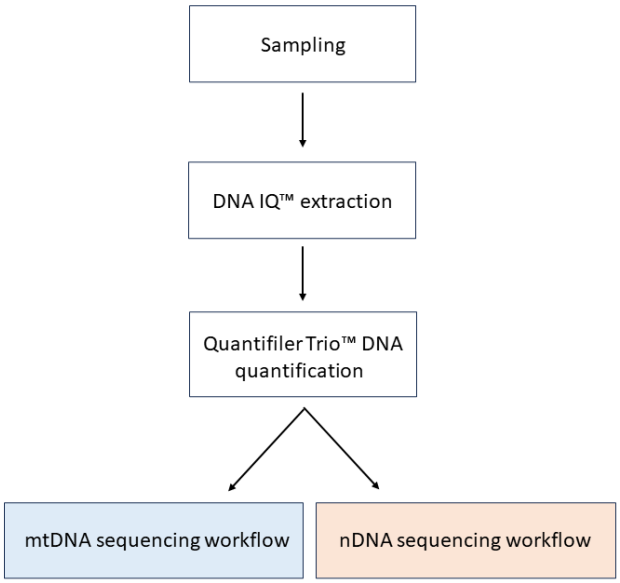

Figure 2.1 Summary of experimental steps undertaken by all samples before diverging into separate mtDNA and nDNA sequencing workflows

2.1.1 Ethical approval

This research project was approved by the Auckland Health Research Ethics Committee (AHREC) on the 27th of April 2023 (reference number: AH25804) for three years. This approval allowed for the collection of buccal swabs and touch DNA from seven volunteers. An amendment to increase the wait time between handling individual cartridges from 5 minutes to 15 minutes was approved on the 16th of June 2023.

2.1.2 Participant recruitment

Participants were recruited from Mount Albert Science Centre (MASC), ESR Ltd through a site-wide email. This contained a participant information sheet and a consent form (Appendix 7.1).

Participants confirmed their participation by handing a signed consent form to an independent third-party representative. Two sampling collection periods were carried out. During the first collection period, each participant received an individual sample deposition pack which included two sterile swabs for buccal swab collection and eight .223 Rem cartridges for touch DNA collection. During the second collection period, each participant received an individual sample deposition pack which included four .223 Rem cartridges for touch DNA collection.

2.1.3 Sample deposition pack creation

The creation of all sample deposition packs was performed in a Biological Safety Cabinet (Class 2, type A2) to minimise contamination. All .223 Rem cartridges (Figure 2.2) were supplied by ESR Ltd. .223 Rem was the ammunition used as it is commonly recovered as evidence from crime scenes in New Zealand and has a large external surface, maximising the area for touch DNA deposition. Each cartridge was cleaned with 70% ethanol before being placed in an envelope. This was to remove any background contamination such as the manufacturers DNA or any previous handlers' DNA. Swab casings were cut with sterile scissors to allow the buccal swabs to dry once collected.

Figure 2.2 Unfired .223 Remmington cartridge

2.2 SAMPLE COLLECTION

2.2.1 Buccal swab samples

Buccal swabs were taken at least 30 minutes after the participant last consumed food or a beverage. Participants undertook their own buccal swab collection by swabbing the inside of their cheek for approximately one minute before placing the used swab back into its original casing. Each participant provided two buccal swabs so that both their mtDNA and DNA profiles could be obtained. After collection, buccal swabs were dried and stored at room temperature until sampling.

2.2.2 Touch DNA samples

To determine the best method of touch DNA deposition for this research, a review into the experimental design of previous studies was conducted. Since factors such as handwashing, secondary transfer through contact with other people and an individual's own ability to deposit touch DNA all impact the quantity of touch DNA deposited (22) and introduce variation into an experiment, it was important to determine how previous studies conducted experiments to generate DNA profiles from touch DNA while controlling such variables.

Martin et al. (29) had participants wash their hands and wait 15 minutes before touching an item for 15 seconds. During this 15-minute period, participants were instructed to continue with their normal everyday activities excluding wearing gloves or rewashing their hands. This method was based on the success of Phipps et al. (31) and Templeton et al. (125) who both obtained DNA profiles from items that were touched 15 minutes after handwashing. Phipps et al. (31) enforced tighter restrictions during the 15-minute wait period and asked their participants not to eat, touch other people or wear gloves. Phipps et al. (31) chose their 15-minute wait period based on the results of Lowe's (28) study, which investigated the tendency of different individuals to deposit DNA through touch. In this study, participants washed their hands and waited 15 minutes, 2 hours or 6 hours before handing a tube. Again, participants were instructed to continue with normal activities but avoid wearing gloves. It was found that the 15-minute wait period was the best time interval for the largest inter-variation between individuals and their ability to deposit touch DNA with shedder status (good or bad) able to be defined (28).

Since the aim of this research was to determine whether alternative MPS technologies can successfully profile touch DNA recovered from cartridges and cartridge cases, waiting a longer period (~ 6 hours) after handwashing would lead to higher percentage of a DNA profile being recovered (28), and maximise the chances of a successful result. This was not practical as sampling would take significantly more time and increase variability in a participant's activities within the given timeframe. As touch DNA profiles have been obtained 15 minutes after handwashing the method described below was used.

2.2.2.1 Unfired cartridge samples

Prior to deposition of touch DNA, participants washed their hands. 15 minutes elapsed before they handled a .223 Rem cartridge for 15 seconds, making sure to apply a medium pressure with their fingers. Each touched cartridge was then placed in a separate white paper envelope. This whole process was repeated seven more times so that each participant handled eight cartridges in total. Touched cartridges were stored in a white paper envelope at room temperature until sampling.

During the 15-minute waiting period, participants were allowed to continue using their personal/work electronic devices and read/write. Participants were instructed not to touch other people, wear gloves and to avoid eating.

2.2.2.2 Fired cartridge case samples

The same method of touch DNA deposition outlined in Section 2.2.2.1 was carried out for the fired cartridge case samples with the same restrictions in place. Instead of placing the handled cartridge into a white envelope, the participants loaded the cartridge directly into the chamber of a bolt-action rifle (Figure 2.3A). The bolt action rifle was contained within a remote firing cart at all times (Figure 2.3C).

Figure 2.3 Photographs of the firearm and remote firing cart used in this research. ***A)*** *Remmington Model Seven bolt-action .223 Rem calibre rifle* ***B)*** *The remote firing cart used to fire the bolt-action rifle.* ***C)*** *The bolt-action rifle contained within the remote firing cart.*

Once the cartridge was loaded, the participant left the room, and the contained rifle was fired by a trained firearm expert. The cartridge case was expelled out of the rifle onto the ground which was covered in a piece of clean white paper. The cartridge case was recovered using a DNA-free, EtO-sterilised Forensic Swab (Sarstedt, Nümbrecht, Germany) by the firearm expert, and the swab was placed into an unused white paper envelope (Appendix 7.2). After firing, the chamber of the firearm and a blank cartridge were cleaned with

70% ethanol. The blank cartridge was fired, ensuring that no residual ethanol remained. This whole process was repeated three more times so that four fired cartridge cases were collected per participant.

2.3 ANTI-CONTAMINATION LABORATORY PROCEDURES

Separate laboratories for pre- and post-amplification experiments were used to prevent contamination in this research. In the pre-amplification laboratory, sampling (Section 2.4), extraction (Section 2.5), quantification (Section 2.6), and the preparation of samples for amplification was conducted (Sections 2.7.1.1 and 2.8.1.1). Following amplification, the PCR products were transferred into a post-amplification laboratory and the experiments outlined in Section 2.7.2 or Section 2.8.1.2 onwards were conducted. Reference samples with high concentrations of DNA were processed separately (in separate batches and on different days) in the pre-amplification laboratory to the expected low DNA copy number cartridge/cartridge case samples to avoid any cross-contamination.

A weekly cleaning schedule is in place at ESR Ltd for the research laboratories. All surfaces and equipment inside a Biological Safety Cabinet (BSC, Class 2, type A2) are cleaned with Virkon (1% solution), 70% ethanol and irradiated with UV. All laboratory benches are also cleaned with Virkon (1% solution) and 70% ethanol.

Additionally, all BSC and bench spaces were cleaned both prior to and after use. Until samples had been uniquely identified using sequencing indexes (Section 2.7.3.2 and 2.8.1.2), Virkon and 70% ethanol were used for cleaning. After the addition of adapters, only 70% ethanol was used.

2.4 SAMPLING

The sampling and DNA extraction methods used for this research project emulated the methods currently in use for casework samples at ESR Ltd.

2.4.1 Buccal swabs

A sterile scalpel was used to cut approximately a quarter of the buccal swab head. Once cut, this portion was transferred into an appropriately labelled 1.5 mL tube (Eppendorf, Hamburg, Germany). The remaining swab head was transferred back into the original swab casing and stored at room temperature. The sampling of buccal swabs was performed on a laboratory bench.

2.4.2 Cartridge and cartridge case swabs

Touch DNA was sampled from .223 Rem cartridges and cartridge cases using the double swab method (126). This method is used to collected trace DNA samples from non-porous surfaces and items at ESR Ltd. Each individual cartridge and cartridge case had one 'wet sample' and one 'dry sample' obtained from it.

2.4.2.1 Standard ESR double swab procedure

Unfired cartridges were held by the bullet with a double gloved hand. Fired cartridge cases were held by inserting sterile tweezers into the empty cartridge case. Sterile double ended swabs (Medical Wire & Equipment, Corsham, Wiltshire, England) were used to swab each unfired cartridge and fired cartridge case. The swab head was moistened with sterile water and then rubbed over the entire surface of the .223 Rem cartridge/cartridge case before being placed back into its original casing creating a 'wet sample'. A second dry swab was then rubbed over the same surface area to absorb any excess liquid left on the cartridge/cartridge case and then placed back into its original casing creating a 'dry sample'. Pressure was applied during both the sampling processes to ensure that any DNA present would transfer from the cartridge/cartridge case to the swab. The end of the swab casings was removed allowing the used swabs to air dry while being stored at room temperature before further processing.

2.4.2.2 Processing of swabs from double swabbing procedure

A sterile scalpel was used to cut the entire swab head from the swab shaft of the 'wet sample' before being transferred into an appropriately labelled 1.5 mL tube (Eppendorf). Using the same scalpel, the entire head of the 'dry sample' swab was cut from the shaft and transferred into the same 1.5 mL tube (Eppendorf). Samples were then taken directly through a DNA IQ™ extraction or were stored overnight at 4 °C.

2.5 DNA EXTRACTION

The DNA IQ™ System (Promega, Madison, Wisconsin, USA) was used for DNA extraction. This extraction kit utilises paramagnetic resin to efficiently bind, isolate and purify all genomic DNA (gDNA) from a forensic sample.

The DNA extraction method was performed as follows:

1. Lysis buffer was prepared by adding 1 M DTT to a stock solution in volumes calculated by using the formulas in Table 2.1
2. The 1.5 mL tubes (Eppendorf) containing portions of, or entire swab heads had a predetermined amount of the prepared Lysis buffer added (Table 2.1)

3. All tubes were vortexed to mix and then incubated at 70^{o}C for 30 minutes.
4. Samples and all liquid were transferred into new tubes containing Spin Baskets. The old tubes were retained.
5. The new tubes were centrifuged at maximum speed for two minutes.
6. The Spin Baskets were removed using sterile tweezers and transferred back into the original 1.5 mL tube for storage.
7. The resin bottle was vortexed until the resin was suspended. 7 μL of resin was added to each tube of eluted sample and vortexed for three seconds.
8. Samples were then incubated at room temperature for five minutes. During this incubation period, the samples were vortexed for three seconds every minute. This allowed the DNA to bind to the resin.
9. Samples were vortexed for two seconds, centrifuged and placed on a magnetic stand.
10. The lysis retentate solution was carefully removed and discarded without disturbing the resin pellet.
11. 100 μL of the prepared Lysis buffer was added to each resin pellet.
12. Each tube was removed from the magnetic stand and vortexed for two seconds before being placed back on the magnetic stand.
13. All lysis solution was carefully removed and discarded without disturbing the resin pellet.
14. 100 μL of the 1 x Wash buffer was added to each pellet.
15. Samples were removed from the magnetic stand, vortexed for two seconds and then placed back on the magnetic stand.
16. All wash solution was removed carefully not to disturb the resin pellet and discarded.
17. Steps 14 to 16 were repeated twice more, seeing a total of three washes. All the solution was removed after the last wash.
18. All samples were air-dried for five minutes while remaining in the magnetic stand.
19. 25 μL of Elution buffer was then added to each sample.
20. Samples were then vortexed and incubated at 65 ^{o}C for five minutes.
21. Samples were then vortexed for two seconds and immediately placed on the magnetic stand while they were still hot.
22. The eluted solution was entirely transferred to a newly labelled 1.5 mL tube (Eppendorf).
23. Steps 19 to 21 were repeated with the eluted solution transferred to the same tube in step 22 to ensure that the final extracted volume was greater than 45 μL for all samples.

Table 2.1 Volumes of lysis buffer and 1 M DTT required for DNA IQ™ Extraction. Final volumes are dependent on sample type and amount.

Number of samples	Lysis Buffer (µL)	DTT (µL)
Controls = A	(150 + 100) x A = D	(G/100) x 2.5 µL = _____
Portion of swab size = B	(150 + 100) x B = E	
2 swab size = C	(400 + 100) x C = F	
	Total = D + E + F + 100 = G	

2.6 DNA QUANTIFICATION

Quantification of all extracted samples was carried out using the Quantifiler™ Trio DNA Quantification Kit, hereinafter referred to as Quantifiler™ Trio (Thermo Fisher Scientific, Waltham, MA, USA). Quantifiler™ Trio is designed to quantify both the total amount of human autosomal DNA and human male nuclear DNA, within a sample, in real-time. There are two different targets for the quantification of human autosomal DNA: a small autosomal marker (80 bp) and a large autosomal marker (214 bp). The use of these two markers allows for an estimation of the amount of degradation within a sample (127). Separately, human male DNA is quantified through a single Y chromosome marker. Quantifiler™ Trio also includes an Internal Positive Control (IPC) assay. The IPC helps determine if the quantity of DNA in each sample is too high and will lead to downstream analysis complications, whether the instrument failed to operate as expected, or if PCR inhibitors are present (127).

All Quantifiler™ Trio reactions were set up in a MicroAmp® Optical 96-Well plate (Thermo Fisher Scientific) as per the manufacturer's user guide (127) to have a total volume of 20 µL. Each reaction consisted of:

- 2 µL of either standard, NTC, DNA extract sample or control
- 8 µL of Quantifiler™ Trio Primer Mix
- 10 µL Quantifiler™ THP PCR Reaction Mix

All standards and the NTC were prepared as duplicates and the samples and controls were prepared once.

The Applied Biosystems® 7500 Real-Time PCR System was used to carry out all quantifications using the thermal cycling conditions provided by the manufacturer. Data was collected in real-time then analysed using the HID Real-Time PCR Analysis Software v1.2 (Thermo Fisher Scientific). A quality check on the standard curve was performed, with the values for specific parameters having to meet certain requirements for the quantitation results to be accepted. These quality requirements were that the standard cure slope had to be between -3.0 to -3.6, the correlation coefficient (R^2) had to be above 0.98 and the amplification efficiency value had to be between 90-110%.

2.7 MTDNA SEQUENCING USING A CUSTOM WORKFLOW

The following method was specifically developed for mtDNA sequencing on the MiSeq FGx™ Forensic Genomics System, hereafter referred to as the MiSeq FGx™ (97).

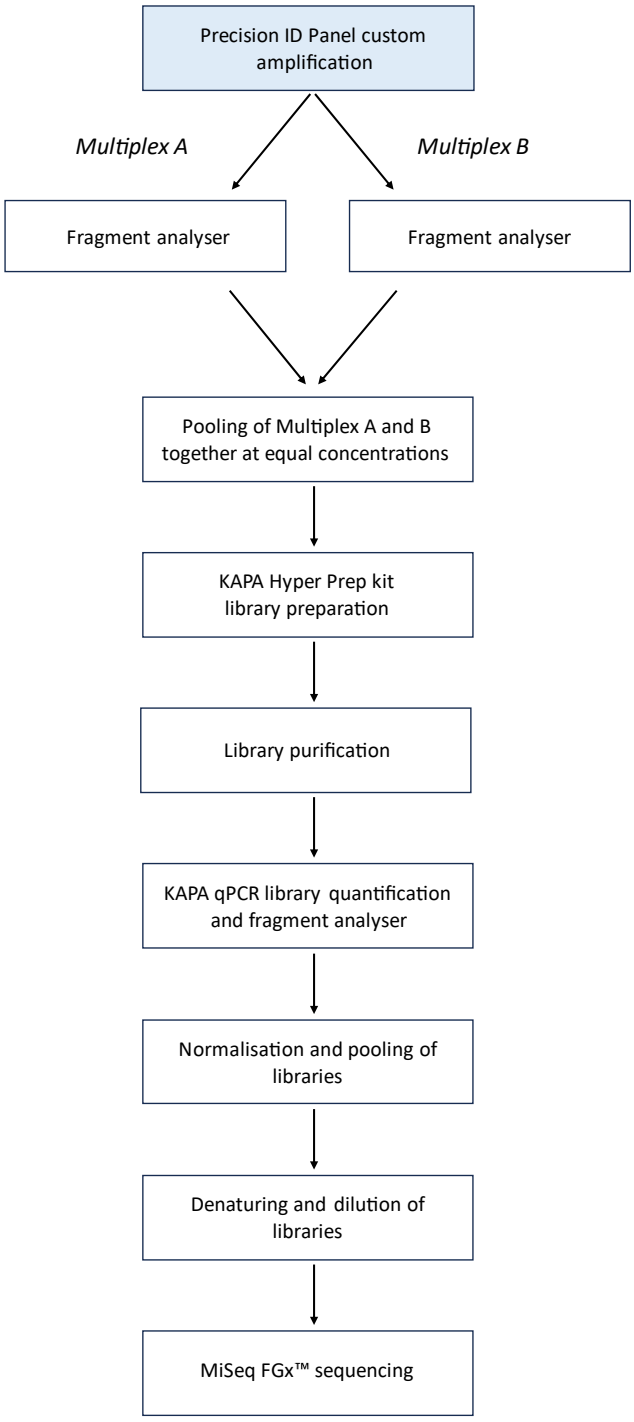

Figure 2.4 Custom mtDNA MiSeq FGx™ sequencing workflow, adapted from Forsythe (97) *and herein after referred to as the mtDNA sequencing workflow.*

2.7.1 Targeted amplification with Precision ID mtDNA Whole Genome Panel

As the Precision ID panel is specifically designed for use with Ion Torrent sequencing platforms (Thermo Fisher Scientific), the reagents used, and manufacturer's guidelines (128) are not suitable for the MiSeq FGx™ platform. To overcome this, a custom amplification method was used (97). This method sees primers from the Precision ID panel combined with the 2x QIAGEN Multiplex PCR Master Mix from the QIAGEN® Multiplex PCR Kit (QIAGEN, Hilden, Germany) for the targeted amplification of the entire mtDNA genome. As there are two multiplexes of primers in the Precision ID panel, two amplification reactions (A and B) were created for each sample. To save on both reagents and samples, half volumes of the reaction volumes recommend by the QIAGEN® manufacturers were used (97,129).

2.7.1.1 Custom Precision ID and QIAGEN® Multiplex PCR Kit amplification method

As 0.1 ng of gDNA is recommended as input for the Precision ID panel as per manufacturers guidelines (128). The optimal input volume was calculated using the small autosomal marker genomic concentration (ng/µL) determined during Quantifiler™ Trio quantification (Section 2.6) and the following equation:

$$DNA\ input\ vol\ (\mu L) = \frac{0.1\ ng}{Genomic\ conc\ (ng/\mu L)\ \times DF}$$

The final DNA extract input volume was between 2 – 7.5 µL. A dilution factor (*DF*) was used to ensure that none of the samples had an input volume lower than 2 µL. Dilutions were completed using nuclease-free water. Any sample whose genomic concentration resulted in less than 0.1 ng of DNA was left undiluted and 7.5 µL of that sample was used. A total input volume of 7.5 µL was required. If the DNA input volume was below 7.5 µL, nuclease-free water was used to ensure that the total input volume was 7.5 µL.

The custom mtDNA amplification method was then performed as follows:

1. Extracted samples were diluted using nuclease free water if necessary, using the calculated Dilution Factor (*DF*)
2. Two Master Mixes were made in two Lo-bind 1.5 mL tubes (Eppendorf) using the reagents and volumes in Table 2.2.

The following steps were then preformed in duplicate (once for each multiplex):

3. In a new 96-well plate, the pre-calculated amount of nuclease free water was added if required.
4. To the same 96-well plate, extracted DNA samples were added so that the total volume in each well was 7.5 µL.
5. 17.5 µL of Master Mix A or B was added, making the total volume in each well 25 µL.

6. The plate was then sealed, vortexed and centrifuged to ensure all the components were mixed and sitting at the bottom of the plate wells.
7. Thermal cycling was then preformed using a Proflex™ Thermal Cycler (Thermo Fisher Scientific) (Table 2.3).

Table 2.2 Volumes and reagents required for creating the custom amplification master mix for each multiplex.

Reagent	Multiplex A - Volume (µL)	Multiplex B - Volume (µL)
2 X QIAGEN Multiplex PCR Master Mix	12.5 x number of samples	12.5 x number of samples
Precision ID Panel A	5 x number of samples	-
Precision ID Panel B	-	5 x number of samples

The thermal cycling conditions (Temperature/Time) were carried out in accordance with both Thermo Fisher Scientific and QIAGEN manufacturers protocols (99,129) as determined by Forsythe (97). In Forsythe's work, 26 cycles of denaturation, annealing and extension was used. This was increased to 30 cycles for this research to account for the expected low copy number of any template mtDNA obtained from cartridges and cartridge cases while still minimising the risk of stochastic amplification of any DNA damage (130).

Table 2.3 Proflex™ thermal cycling conditions for custom amplification using the Precision ID mtDNA whole genome panel.

Temperature	Time	Number of Cycles	PCR Step
95°C	15 minutes	-	Polymerase activation
94 °C	30 seconds	30	Denaturation
60 °C	4 minutes		Annealing and extension
72 °C	10 minutes	-	Final extension
10 °C	∞	-	Hold

2.7.2 Amplicon quantitation and pooling

The resulting amplicons were quantified to firstly determine the concentration (ng/µL) of the amplicons in each multiplex so they could be pooled together in equal amounts for each sample, and secondly determine the amount of each pooled amplicon sample required for library preparation.

2.7.2.1 Amplicon quantification using Fragment Analyzer™

Amplicon quantification was completed using the High Sensitivity NGS Analysis Kit and the 5200 Fragment Analyzer™ System (Agilent©, Santa Clara, CA, USA). The High Sensitivity NGS Analysis Kit provides information on amplicons from 1 to 6,000 bp in length and the 5200 Fragment Analyzer™ System is an automated parallel capillary electrophoresis (CE) platform that separates amplicons based on their bp length.

2 µL of sample from each multiplex was added to 22 µL of HS NGS Fragment Diluent Marker Solution. Reference samples, cartridge samples and positive controls were diluted 1 in 100 using nuclease-free water. Negative controls were left undiluted. The HS NGS Fragment Diluent Marker Solution contains a lower marker (1 bp) and an upper marker (6,000 bp). A HS NGS Fragment DNA ladder was used in each CE run to allow for the size of the amplicons to be determined (Figure 2.5A).

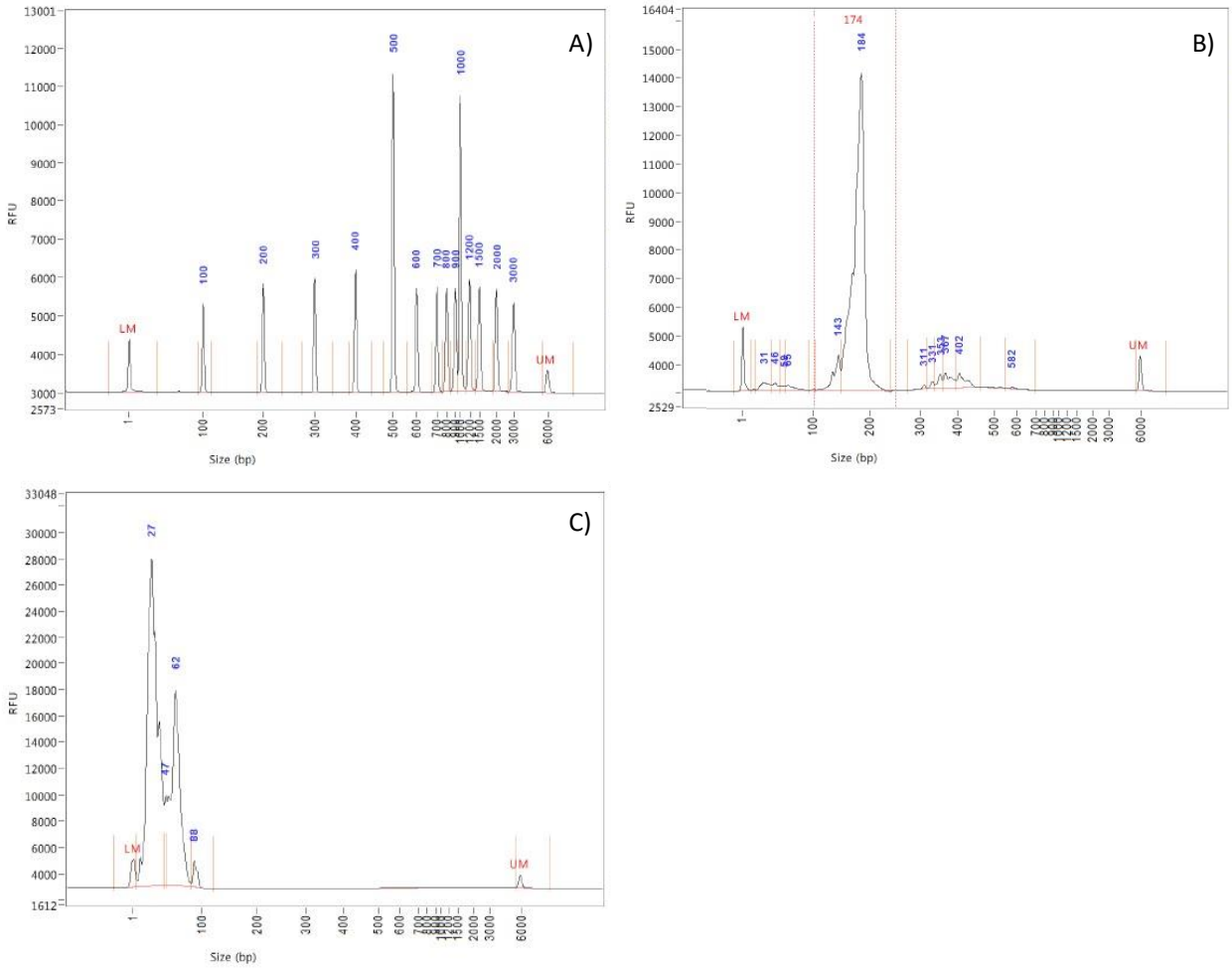

Figure 2.5 Examples of Fragment Analyzer electropherograms. A lower marker (LM) and upper marker (UM) is seen in each graph. ***A)*** *Example of the HS NGS Fragment DNA ladder.* ***B)*** *Example of a 1:100 diluted reference buccal sample. Precision ID amplicon peaks are seen between 100 to 250 bp (as indicated by the two large red lines), with a median length of 184 bp.* ***C)*** *Example of a negative control with unbound primers and primer-dimer peaks present from 0 to 100 bp.*

2.7.2.2 Amplicon pooling

The concentration (ng/µL) of the amplicon peaks seen in Fragment Analyzer electropherograms (EPGs) from 100 to 250 bp (as shown in Figure 2.5B), were used to determine the volume of each multiplex to be pooled together. Only peaks in this range were used as the average amplicon length of the Precision ID panel is 163 bp (99), therefore, the inclusion of any primer-dimers or off target DNA fragments was avoided (97).

Firstly, the ratio of the target amplicon concentrations (*R*) had to be calculated. This was done for each sample, by dividing the multiplex sample with the highest target amplicon concentration by multiplex sample with the lowest target amplicon concentration:

$$R = \frac{Highest\ amplicon\ concentration}{Lowest\ amplicon\ concentration}$$

As a total volume of 25 µL is required for downstream analyses, the volume of the target amplicon with the highest concentration (*V1)* was calculated by dividing 25 µL by *R* plus 1:

$$V1 = \frac{25\ \mu L}{(1+R)}$$

The volume of the target amplicon with the lowest concentration was then calculated by subtracting *V1* from 25 µL:

$$V2 = 25\ \mu L - V1$$

The V1 and V2 calculated volumes (µL) were then used to pool together multiplex A and multiplex B sample*s* in equal concentrations.

2.7.3 Library preparation using KAPA HyperPrep Kit

The KAPA HyperPrep Kit from KAPA Biosystems Inc, (Wilmington, MA, USA) was used to prepare libraries compatible with the MiSeq FGx™ platform (Illumina). The library prep workflow is made up of three steps: end repair and A-tailing; adapter ligation; and a post ligation clean-up. Following Forsythe (97), all reactions were prepared in half the volume stated in the manufacturer's protocol (131) to save on reagents and sample.

A maximum input amount of 125 ng in a volume of 25 µL was required for the library preparation of each sample. If required, 10 mM Tris HCl (pH 8.5) was added to each pooled amplicon product to ensure that this amount and total reaction volume was obtained.

2.7.3.1 End repair and A-tailing

End repair and A-tailing of the amplicons occurred in a single enzymatic reaction. End repair produced blunt-ended 5' phosphorylated dsDNA fragments and A-tailing added deoxyadenosine monophosphate (dAMP) to the 3'-ends of the fragments (132) in preparation for adapter ligation. A master mix was created by combining 3.5 µL of End Repair & A-Tailing Buffer with 1.5 µL of End Repair & A-Tailing Enzyme Mix for each sample. This

5 µL of master mix was then added to 25 µL of input DNA. All prepared reactions were incubated in a Proflex™ thermal cycler (Table 2.4).

Table 2.4 Proflex™ thermal cycling conditions for end repair and A-tailing incubation, as per (97)

Step	Temperature	Time
Pre-cooling	4°C	∞
End repair and A-tailing	30°C	20 minutes
	65 °C	30 minutes
Hold	4 °C	∞

2.7.3.2 Adapter ligation

For adapter ligation, the KAPA Dual-Indexed Adapter Kit (KAPA UDIs, KAPA Biosystems Inc.) was used. This Kit has been designed to construct libraries with a unique combination of dual indices that are able to be sequenced on an Illumina® instrument (133). There are 96 adapters included within this kit, all of which are full length and contain both P5 and P7 flow cell oligo sequences to allow for complementary annealing to an Illumina® flow cell.

A master mix was created by combining 15 µL of Ligation buffer and 5 µL of DNA Ligase for each sample. Reactions were set out in the proportions described in Table 2.5.

Table 2.5 KAPA HyperPrep adapter ligation enzymatic reaction components and their required volumes

Reaction Component	Volume (µL)
End repair and A-tailing Product	30
KAPA Dual-Indexed Adapter Stock (15 µM)	5
Master Mix	20
Total Reaction Volume	**55**

All prepared adapter ligation reactions were incubated on a Proflex™ thermal cycler as per the manufacturer's recommendations (131).

2.7.3.3 Post-Ligation clean-up/Library purification

Any un-ligated adapters and adapter-dimer molecules need to be removed prior to cluster generation on the MiSeq FGx™ sequencer. The Agencourt® AMPure XP (Beckman Coulter®, Brea, CA, USA) purification system was utilised for this, as it has been validated for clean-up use with the KAPA HyperPrep kit manufacturers protocol (131). This purification system uses solid-phase reversible immobilisation paramagnetic bead technology to selectively bind DNA fragments. Excess un-ligated adapters and any adapter-dimers or primer-dimers are able to be removed, leaving behind a more purified library (134).

The KAPA HyperPrep Kit manufacturer's protocol was followed using 0.8 x AMPure beads for each total volume of adapter ligation reaction product. This saw 44 µL of AMPure beads added to 55 µL of adapter ligated reaction product. Each sample was then washed twice with freshly made 80% Ethanol (80% absolute ethanol:20% MilliQ water), with all ethanol completely removed after the second wash. 55 µL of 10 mL Tris-HCl (pH 8.5) was then immediately added to maximise library yield, and after a five-minute incubation, 50 µL of the final eluted solution (now containing the purified libraries) was transferred to a new plate and used for subsequent experimental steps.

2.7.4 Preparation of libraries for normalisation

Before the pooling of libraries, each library must be individually normalised to the same molarity. This is so each library is present at equal amounts within the sequencing pool, ensuring that an equal number of sequencing clusters for each library are generated (69). Normalisation requires two pieces of information: the average fragment length of all libraries within a sequencing run (bp) and the average molarity (pM) of each library.

2.7.4.1 Sizing of libraries

To determine the average fragment length of each library (bp), a second analysis using the High Sensitivity NGS Analysis Kit, and the 5200 Fragment Analyzer™ System (Agilent©) was performed following the same protocol as outlined in Section 2.7.2.1 with an input volume of 2 µL for each undiluted library. The smear analysis function on the ProSize software (v4.0.2.7) was used to determine the presence of any primer-dimers from 1 to 100 bp, any non-converted amplicons where the adapters did not ligate from approximately 100 to 200 bp, any adapter-dimers from approximately 200 to 280 bp (average of 230 bp), and finally, successfully converted libraries from 280 to 700 bp. As any fragment that contains both adapters will be sequenced, regardless of if they are adapter-dimers or have successfully ligated to an amplicon, all fragments between 200 to 700 bp were measured to determine the **average fragment length of each library *(AFL)***.

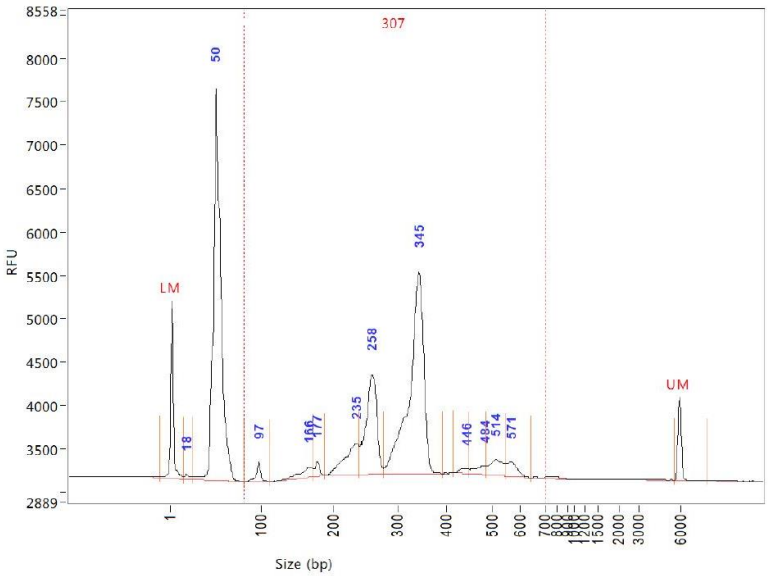

Figure 2.6 Example of a Fragment Analyzer library electropherogram following library preparation with the KAPA HyperPrep Kit. The first and last peak are the lower marker (LM) and upper marker (UM) respectively. The red dotted lines show the peaks of DNA fragments that are from 200 bp to 700 bp in length.

2.7.4.2 Library quantification

All samples were quantified to determine the molarity of each library. Quantification was carried out using KAPA Library Quantification Kit for Illumina Platforms from KAPA Biosystems Inc. This kit contains KAPA SYBR FAST qPCR Master Mix (2X), Library Quantification Primer Premix (10X) and six Library Quantification DNA Standards ranging from 0.0002 to 20 pM. The Library Quantification Primer Premix contains specific primers that target Illumina P5 and P7 flow cell oligo sequences (135). Each sample was serially diluted using QIAGEN Elution Buffer (QIAGEN) to give a final dilution factor of 2000; ensuring that the prepared libraries had a molarity between 0.0002 to 20 pM.

All the KAPA Library Quantification Kit reactions were set up in duplicate on a MicroAmp® Optical 96-Well plate (Thermo Fisher Scientific) as per the manufacturer's user guide (135) to have a total volume of 20 µL. Each reaction consisted of:

- 4 µL of either standard, NTC (QIAGEN elution buffer), or library sample
- 4 µL of nuclease free water
- 12 µL of KAPA qPCR Master Mix

The 7500 Real-Time PCR System (Thermo Fisher Scientific) was used following thermal cycling conditions provided by the manufacturer's user guide (135). The same quality requirements discussed in Section 2.6 were used to perform a quality check on the standard curve generated from the KAPA Library Quantification Kit. If the standard curve passed this quality check, the average quantity of each sample was recorded. This

was the molarity of each diluted library, and hereinafter referred to as the **diluted qPCR molarity *(Dil.qPCR.Mol)***.

2.7.5 Library normalisation and pooling

The diluted qPCR molarity *(Dil.qPCR.Mol)* obtained for each library was size adjusted using the size of the KAPA Library Quantification DNA Standards (452 bp), divided by the average fragment length *(AFL)* of that library. The dilution factor *(DF)* of 2000 used in the previous step (Section 2.7.4.2) was then used to multiple the product, giving the molarity of the undiluted prepped libraries *(Mol.unDil)*.

$$Mol.unDil\ (pM) = \left(Dil.qPCR.Mol\ (pM)\ \times \frac{452\ bp}{AFL} \right) \times DF$$

The *Mol.unDil* was converted from pM to nM and multiplied by the normalised library total volume (10 µL) before dividing by the targeted normalised concentration (4 nM) as follows:

$$Final\ library\ volume\ (\mu L) = \frac{Library\ total\ volume\ (\mu L) \times Mol.unDil\ (nM)}{Normalised\ concentration\ (nM)}$$

This gave the final library volume (µL) for each sample to be normalised to 4 nM. Nuclease-free water was added to ensure that each well had a final volume of 10 µL. An input volume of at least 2 µL for each library sample was required.

5 µL of each normalised library was then combined into the same 1.5 mL Lo-bind tube (Eppendorf) creating a **pooled, normalised library (PNL)**. Up to 33 libraries were pooled together into the same tube. The molarity of the PNL was then determined using a Qubit™ fluorometer (Thermo Fisher Scientific) and the Qubit™ dsDNA HS assay kit (Thermo Fisher Scientific) according to the manufacturers protocol (136). The Qubit value (ng/µL) was divided by the total average fragment length (total AFL) of all the libraries in the PNL to give the molarity of the PNL in nM. This is summarised in the following equation:

$$PNL\ molarity\ (nM) = \frac{Qubit\ Value\ (ng/\mu L) \times\ 1{,}000{,}000}{total\ AFL\ (bp)\ \times 660}$$

This was done to ensure that the PNL had a final targeted molarity of 4 nM.

2.7.6 Dilution and denaturing of a pooled, normalised library

The dilution and denaturing of libraries, from dsDNA to ssDNA, is required for both binding and clustering of the ssDNA library fragments to the flow cell (69). Forsythe (97) determined the optimal final loading concentration of the mtDNA libraries to be 6 pM. At this concentration over clustering or under clustering on the flow cell was avoided. Therefore, for this research, the 4 nM PNL were diluted and denatured to give a final loading concentration of 6 pM.

0.1 NaOH was freshly prepared by combining 2 µL of HP3 (Verogen, Inc) and 38 µL of nuclease free water. 5 µL of this 0.1 NaOH and 5 µL of the PNL was combined in a new 1.5 mL Lo-bind tube (Eppendorf) to create a **denatured, normalised library (DNL).** The DNL was vortexed, centrifuged and then left to incubate for 5 minutes, allowing the NaOH to denature the double-stranded mtDNA library fragments into single-stranded mtDNA library fragments.

10 µL of 200 nM Tris-HCl and 980 µL of HT1 was added to the DNL to give a concentration of 20 pM. From this, 180 µL of the 20 pM DNL, 30 µL of 20 pM PhiX Control v3 Library (Illumina, Inc) and 390 µL of HT1 was combined in a new 1.5 mL Lo-bind tube (Eppendorf) to give a final DNL concentration of 6 pM and a final total volume of 600 µL.

The 20 pM PhiX control was freshly made using 2 µL of 10 nM PhiX control library and 3 µL 10 mM Tris-HCL. PhiX Control v3 Library (Illumina) is derived from the bacteriophage genome PhiX and is an Illumina library that can be used as a quality control within each sequencing run. An input volume of 30 µL equates to a 5% PhiX spike in; therefore, 5% of all the generated sequencing reads should align to the PhiX genome (97). PhiX also provides nucleotide diversity, ensuring that there is balanced fluorescent signals within a sequencing run and improving overall quality (137).

The 6 pM DNL was then incubated at 96 ^{o}C for two minutes to ensure no double-stranded mtDNA fragments remained, before being immediately covered by ice until loading into the sequencing reagent cartridge.

2.7.7 Sequencing set-up

The MiSeq FGx^{TM} sequencing instrument and the 600 cycle MiSeq Reagent Kit v3 (Illumina, Inc.) were used for sequencing. The 6 pM DNL was removed from ice, and 600 µL (the total volume) was loaded into the MiSeq reagent cartridge.

All mtDNA sequencing runs were performed using the 'Research Use Only Run' mode on the MiSeq FGx Control Software. A sample sheet was created for each sequencing run. Each sample sheet contained

technical details about the sequencing run, the ID of each sample in the final DNL, the specific KAPA Unique Dual Indexed (UDI) Adapter that was added to each individual sample, and the corresponding P7 Index sequence and the P5 Index sequence with that adapter. During the sequencing set-up, this sheet was uploaded to the MiSeq FGx™ internal computer.

2.7.8 Sequencing analysis

The MiSeq FGx™ instrument's internal computer processed all the sequencing data generated during a sequencing run. Sequencing clusters were converted into nucleotide base calls with poor quality clusters filtered out by a chastity filter. The quality score of each base call was recorded in one .bcl file for each sequencing run cycle. The recorded KAPA UDI adapter sequences in the sample sheet were used to demultiplex each cluster on the flow cell, into individual sample libraries. Base calls within each demultiplexed cluster were compiled together to form sequencing reads and are recorded into two FastQ files (one for read 1 and one for read 2) for each individual library (69). Each FastQ file contained all the nucleotide base calls and the associated quality score for each base.

At the end of each sequencing run, three run quality metrics were provided on the internal computer's interface. These gave an indication of the sequencing run performance, and each value was recorded at the end of each sequencing run.

1. **Cluster density (K/mm2)**. Cluster density is directly related to the loading concentration of the library pool. This value (reads per square millimetre of flow cell) should be between 1200 – 1400 K/mm2 for well-balanced libraries using the 600 cycle MiSeq Reagent Kit v3 (138).
2. **Cluster passing filter (%)**, shows the percentage of clusters that pass the Illumina chastity filter. Only clusters that pass this filter are converted into base calls.
3. **Q-score all cycles (%)**. A Q30 quality score is measured throughout the entire sequencing run. Quality scores (Q-score) indicate the probability of an incorrectly called base during sequencing, with Q30 predicting the probability of one in 1000 base calls being incorrect. The Q-score all cycles value shows the percentage of bases with a Q-score greater than 30. In the 600 cycle MiSeq Reagent Kit v3, it is expected that > 70% of bases in the entire run will have a Q-Score greater than 30 (138).

2.7.9 Bioinformatic analysis

GeneMarker® HTS v2.6.0 (GM-HTS, SoftGenetics, Inc., State College, PA, USA) was the software application used for the filtering, trimming, aligning and analysis of all mtDNA sequencing data. GM-HTS offers a streamlined analysis workflow for MPS generated mtDNA data and has been previously evaluated for forensic application (97,139). GM-HTS has a unique motif alignment technology that is suitable for both Illumina and Ion Torrent generated data (140). This software employs a Burrows-Wheeler hash alignment algorithm to align sequencing reads (139,141). A motif alignment can be performed where a motif file is applied to the sequencing reads. This ensures phylogenetic alignment where the samples sequence is aligned to phylogenetically correct sequence motifs, and variant calls are made based on current phylogeny knowledge (112,139,142).

After the completion of each mtDNA sequencing run, the generated FastQ files were transferred from the MiSeq FGx™ internal computer into the GM-HTS software.

The sequencing results obtained from each mtDNA sample underwent two alignments in GM-HTS. GM-HTS has two separate setting sections: Alignment Settings and Mito Variant Filter Settings. Both sections contain different parameters that can be left at default GM-HTS values or set to custom values by the analyst.

Alignment One

The first alignment saw the use of a **strict sort paired-end, motif alignment** and allowed for the calling of each samples haplotype and any minor (PHP) variants.

Alignment Setting parameters were set as follows:

- Sequencing reads were aligned to the default rCRS.
- Soft clipping based on quality scores (Q) set to ≤ 25.
- Soft clipping of the 3' ends of reads if bases are mismatched at the end of an alignment (140).
- Match proportion set to ≥ 90%.
- Identity score set to ≥ 90%.
- Sequencer set to Illumina.
- The default GM-HTS motif file was used for motif alignment.
- A BED file with all the regions of interest/Precision ID panel targeted amplicons was used with a strict sort paired-end read alignment. This required amplicons to be sequenced in both the forward and reverse direction to be included in the alignment.
- 92 PHI nucleotide positions were masked using a custom file by Forsythe (97), which was based on recommendations by Marshall et al. (107) (Appendix 7.3).

Mito Variant Filter Setting parameters were set as follows:

- Allele score difference set to ≤ 10.
- Balance ratio set to ≤ 2.5 for SNPs and ≤ 5 for indels.
- Minimum total coverage of 100 reads.
- Minimum variant allele coverage of 40 reads.
- Minor variant frequency ≥ 10%.

Alignment Two

The second alignment was used to resolve any dropouts in the mitogenome that occurred during the first alignment, providing full mitogenome coverage and allow for analysis of the HV2 cytosine stretch (from bases 303 to 315)(97). The same values for the Mito Variant Filter Setting parameters and Alignment Setting parameters as the first alignment were used except **the strict sort paired-end read alignment was removed.**

GM-HTS automatically creates a project for each alignment performed on this software. Numerous text files are generated for each sample and additional analyses on this data was performed using Haplogrep3 (v3.2.1), RStudio (v2023.09.1+494), R (v4.3.1) and Microsoft Excel 2016. All processed sequencing data was visualised on the GM-HTS software (v2.6.0).

2.7.9.1 mtDNA sequence nomenclature

GM-HTS follows recommended international forensic guidelines (96,112) for sequence nomenclature and alignment. The generated sequence is aligned to the rCRS with any differences between the two sequences recorded as a polymorphism. These polymorphisms are hereby referred to as variants and are recorded based on the variants numerical position and base difference, relative to the rCRS. For example, if at nucleotide position 31, an adenine base (A) is observed in a sample, but at the same nucleotide position in the rCRS, a thymine base (T) is observed, the variant is denoted as *31A*. The base difference to the rCRS is always noted as a suffix. A nomenclature system by the International Union of Pure and Applied Chemistry (IUPAC) is used to help with analysis of any confirmed positions of ambiguity. The IUPAC nomenclature system and additional examples of sequence nomenclature are outlined in SWGDAM guidelines (96) and Forsythe's work (97).

2.8 nDNA sequencing using the ForenSeq™ DNA Signature Prep Kit

The following sequencing workflow saw the generation of nDNA profiles. Samples were sampled (Section 2.4), extracted (Section 2.5) and quantified (Section 2.6) prior to entering this workflow.

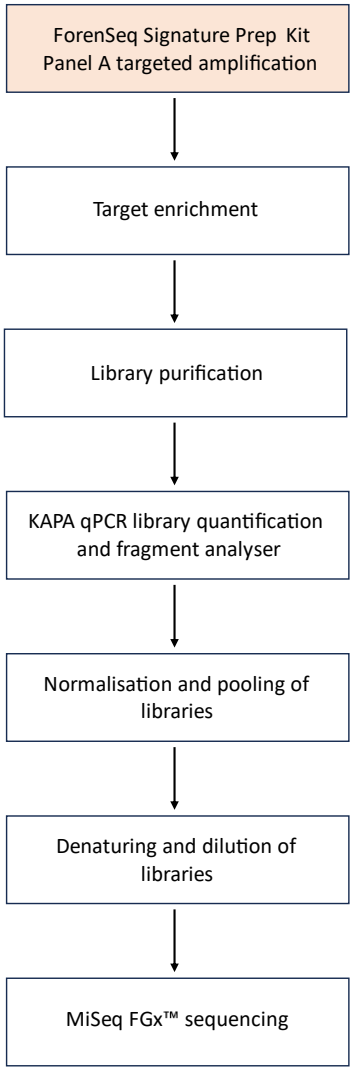

Figure 2.7 ForenSeq™ DNA Signature Prep Kit sequencing workflow (herein referred to as the ForenSeq™ sequencing workflow)

2.8.1 Library preparation using the ForenSeq™ DNA Signature Prep Kit

The ForenSeq™ Prep Kit library preparation contains two PCR steps: target amplification and target enrichment (74). During target amplification, DPMA was added to the input DNA samples, seeing DPMA primers bind to specific regions upstream of either their STR or SNP target. A PCR reaction (PCR1) then saw the amplification of these target STRs and SNPs. During target enrichment, Index 1 (i7) and Index 2 (i5) adapters were added to the PCR1 amplicons in a second PCR reaction (PCR2). Addition of i7 and i5 adapters

saw the tagging of DNA markers with unique barcodes. These barcodes identify each sample, ensuring they are ready for analysis (74).

2.8.1.1 PCR1 – Targeted amplification

As per the manufacturer's recommendations, 1 ng of gDNA in 5 µL (a concentration of 0.2 ng/µL) was used for PCR1 (74). After DNA quantification (Section 2.6), the small autosomal marker genomic concentration (ng/µL) was used to determine the DNA concentration of each sample. Any sample that had a concentration higher than 0.2 ng/µL was diluted with Ultrapure water. Any sample that had a concentration lower than 0.2 ng/µL was left undiluted and added at the maximum input volume.

A master mix was created (in a DNA free hood) by combining 4.7 µL of PCR1, 0.3 µL of FEM and 5 µL of DPMA for each sample. For PCR1, 10 µL of master mix was added to 5 µL input DNA. All reactions were incubated using a Proflex™ as per the manufacturer's recommended cycling conditions (74).

2.8.1.2 PCR2 – Enrichment of targets

PCR2 sees the addition of i7 and i5 adapters to each sample. These adapter indices are added in different combinations, uniquely identifying each sample. The ForenSeq™ Prep Kit comes with eight i5 adapters indices and twelve i7 adapters indices allowing for up to 96 samples to be processed simultaneously.

PCR2 reaction mix, and the i7 and i5 adapters were added to each PCR1 product following the manufacturer's recommendations (74). All reactions were then incubated using a Proflex™ also following the manufacturer's recommendations (74).

2.8.1.3 Library purification

Post PCR2, library purification is required to remove any adapter-dimers, unamplified DNA fragments, leftover reaction components (e.g. dNTPs) and un-ligated primers and/or adapters from the PCR2 product (69). The ForenSeq™ DNA Signature Prep Kit contains Sample Purification Beads (SPB); magnetic beads that utilise a solid phase reversible immobilisation technology to purify libraries.

The method for library purification outlined in the manufacturers reference guide was followed (74). This saw 45 µL of magnetic SPBs combined with 45 µL of the PCR2 product in a new 96 well plate. Once loaded, the plate was placed on a plate shaker for 30 seconds at 1800 rpm before being centrifuged for two minutes. All the supernatant was removed and 200 µL of freshly made 80% ethanol (80% absolute ethanol:20% MilliQ water) was added. This was then repeated, so that each individual sample underwent two ethanol washes.

52.5 µL of Resuspension Buffer was added to each sample and then incubated for two minutes at room temperature. Following this, if the supernatant was completely clear, 50 µL of the resuspended purified library (the final purified PCR2 product) was transferred to a new 96 well plate.

2.8.2 Preparation of libraries for normalisation

The ForenSeq™ Prep Kit uses a bead-based normalisation whereby Library Normalisation Beads are added to each purified library and bind DNA fragments of the same molarity (69). Where this approach is time efficient allowing normalisation to occur collectively for each individual sample, it has proven less reliable for libraries generated from samples with low amounts of DNA (69). As bead-based normalisation sees the binding of all DNA fragments of the same molarity, this includes non-amplifiable molecules which can contribute to the molarity of that sample and overestimate true library concentration (71). One potential solution is to have varying volumes of beads depending on the DNA concentration of each sample (143), however, this would be more time-consuming. Furthermore, bead-based normalisation does not quantitatively measure libraries (143), therefore a quality check cannot be performed, and the analyst is unaware if non target DNA fragments (e.g., adapter-dimers) are present.

An alternative approach is to use quantitative PCR (qPCR). This method is better suited for low copy number DNA samples (69), as it can determine the total amount of adapter bound amplified product, such as target DNA fragments and adapter-dimers. Similarly, an electrophoresis-based method can be used to quantify libraries, determine the size distribution of DNA fragments, and show the presence of any adapter-dimers/primer-dimers/unused primers. It has been determined that using a combination of qPCR and electrophoresis-based methods for library normalisation provides better sequencing results, compared to bead-based normalisation (71,72,143). Therefore, library normalisation was carried out based on data obtained from Fragment Analyzer (Agilent©) and library quantification using the KAPA Library Quantification Kit for Illumina Platforms (KAPA Biosystems Inc).

2.8.2.1 Fragment Analyzer

The High Sensitivity NGS Analysis Kit and the 5200 Fragment Analyzer™ System (Agilent©) was used to determine the **average fragment length *(AFL)*** of each library, following the same method outlined in Section 2.7.2.1. 2 µL of each PCR2 product was left undiluted and added to 22 µL of HS NGS Fragment Diluent Marker Solution. The resulting electropherograms (EPGs) from the Fragment Analyzer were then used to check the quality of each library through the identification of any unused primers, adapter-dimers, and successful adapter-ligated libraries.

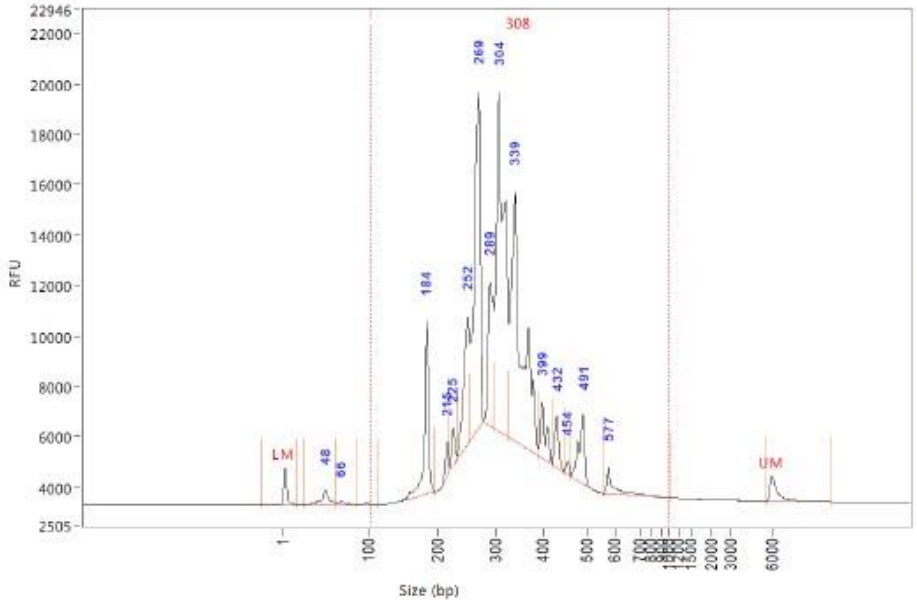

Figure 2.8 Example of a Fragment Analyzer EPG for ForenSeq™ libraries. Small peaks around 48 bp are unbound primers. The larger peak around 180 bp represents adapter-dimers. Successfully converted DNA libraries are seen from 200 to 600 bp. Red vertical lines from 100 to 1000 bp show all DNA fragments that will be sequenced.

The smear analysis function on the Fragment Analyzer data analysis software (ProSize, v4.0.2.7) gave the average fragment length of all DNA fragments between 100 to 1,000 bp in each library. This range encompasses all fragments that contain both adapters and therefore will be sequenced. Peaks from 160 to 180 bp indicate the presence of adapter-dimers and peaks from 200 to 600 bp indicate the presence of adapter ligated library fragments (69,72). Fragments that are under 100 bp indicate the presence of un-ligated primers. These do not have both adapters attached and will not be sequenced.

2.8.2.2 Library quantification

The KAPA Library Quantification Kit (KAPA Biosystems Inc) was used to quantify the molarity of each library using qPCR. 2 µL of each PCR2 product was serially diluted using EB buffer (QIAGEN) to give a final dilution factor *(DF)* of 10,000; ensuring that each PCR2 product had a molarity between 0.0002 – 20 pM. The rest of the method for was conducted the same as described in Section 2.7.4.2 and gave the diluted qPCR molarity for each library *(Dil.qPCR.Mol).* Please refer to Section 2.7.4.2 for further details.

2.8.3 Library normalisation and pooling

Having determined the *AFL* (bp) and molarity of each library (pM), the molarity of the undiluted libraries was calculated. The diluted qPCR molarity *(Dil.qPCR.Mol)* obtained for each library was size adjusted. This was completed using the size of the KAPA Library Quantification DNA Standards (452 bp), divided by the *AFL.* The

dilution factor *(DF)* of 10,000 used in library quantification (Section 2.8.2.2) was then used to multiple the product, giving the molarity of the undiluted prepped libraries *(Mol.unDil)*.

$$Mol.unDil\ (pM) = \left(Dil.qPCR.Mol\ (pM)\ \times \frac{452\ bp}{AFL} \right) \times DF$$

The molarity of the undiluted prepped libraries was then converted from pM to nM. This was then multiplied by the final dilution total volume (10 µL) before dividing by the targeted normalised concentration (1.2 nM). This is summarised in the following equation:

$$Final\ library\ volume\ (\mu L) = \frac{Library\ volume\ (\mu L) \times Mol.unDil\ (nM)}{Normalised\ concentration\ (nM)}$$

This gave the final library volume (µL) of each sample to be added into a new 96-well plate. Nuclease-free water was added to ensure that each well had a final volume of 10 µL. An input volume of at least 2 µL for each library sample was required. 5 µL of each normalised library was then pooled together into the same 1.5 mL lo-bind tube (Eppendorf), creating a **Pooled Normalised Library (PNL)**. Up to 32 libraries were pooled together into one PNL.

The concentration of the PNL (ng/µL) was determined using a Qubit™ fluorometer (Thermo Fisher Scientific) and the Qubit™ dsDNA HS Assay Kit (Thermo Fisher Scientific) according to the manufacturer's protocol (136). A PNL molarity within 1.0 to 1.4 nM is required for subsequent successful sequencing steps, with 1.2 nM being the most optimal target. The Qubit™ concentration (ng/µL) was divided by the total average fragment length (total AFL) of all the libraries in the PNL to give the molarity of the PNL in nM. This is summarised in the following equation:

$$PNL\ molarity\ (nM) = \frac{Qubit\ Value\ (ng/\mu L) \times\ 1{,}000{,}000}{total\ AFL\ (bp)\ \times 660}$$

2.8.4 Library dilution and denaturing

It has been determined that a final loading concentration between 10 – 14 pM provides the most optimal sequencing results for the ForenSeq™ Signature Prep Kit amplified libraries, with 12 pM being the ideal target (72). The 1.2 nM PNL was denatured and diluted to reach this optimal loading range.

6 µL of the PNL was combined with 6 µL of 0.1 N NaOH in a 1.5 mL lo-bind tube (Eppendorf) to create a **Denatured Normalised Library (DNL)**. The presence of NaOH sees the denaturation of ds-DNA library fragments into ss-DNA library fragments, necessary for the binding of library fragments to the sequencing flow cell. If 1.2 nM was not obtained for the PNL, the input volume of PNL was adjusted as described by England (72).

6 µL of 20 pM PhiX (1% spike in) and 2 µL of the Human Sequencing Control (HSC) was then added to the DNL. The addition of HSC to the DNL is used to help determine the performance of each sequencing run and acts as a positive sequencing control for the MiSeq FGx™ platform (57). Finally, 580 µL of the hybridisation buffer, HT1, was added to the DNL to give a total volume of 600 µL. HT1 dilutes the DNL to ensure that it reaches the optimal loading concentration.

The DNL was incubated for 2 minutes at 96°C before being inverted twice and then immediately covered by ice for 5 minutes to ensure the libraries were completely denatured.

2.8.5 Sequencing set-up

The MiSeq FGx™ sequencing instrument and MiSeq FGx™ Reagent Kit (Illumina) was used for the sequencing of all ForenSeq™ amplified samples. All the ForenSeq™ sequencing runs were performed using the 'Forensic Genomics" mode on the MiSeq FGx™ Control Software. The 10 to 14 pM DNL was removed from ice, and 600 µL (the full volume) was loaded into a MiSeq FGx™ reagent cartridge.

All ForenSeq™ sequencing runs were performed using the 'Forensic Genomics' mode on the MiSeq FGx Control Software. For ease of processing, a sample sheet (text file) was remotely uploaded to the UAS server. This sheet contained information about the name of each sample, type of sample and which combination of adapters (i5 and i7) were added to each sample. The sample sheet was then used to demultiplex each cluster in the final library pool based on the pre-recorded adapter sequences attached to each individual sample.

2.8.6 Sequencing analysis

At the end of each sequencing run, ForenSeq™ UAS run quality metrics are displayed on both the MiSeq FGx™ internal computer interface and the UAS server. These quality metrics provided an indication of the performance of each sequencing run, and their values were recorded at the end of each sequencing run. The run quality metrics displayed for all 'Forensic Genomic' mode runs were Cluster density, Clusters passing filter, Phasing and Prephasing. These metrics and their target values are defined in (144).

2.8.7 Bioinformatic analysis

The ForenSeq™ Universal Analysis Software (UAS) (v1.3.6897) bioinformatically processes the generated raw sequencing data. The recorded adapter combinations in the sample sheet are used to demultiplex the sequenced clusters and organise them into their respective libraries. The base calls for each cluster within a library are combined into sequence reads within a FastQ file. For each library, two FastQ files are generated: one for read 1 and one for read 2. The UAS automatically undertakes further processing of FastQ files, seeing the alignment of the sequenced regions to target regions within a human reference genome (69). Variant calling of target STR and iiSNP loci was completed using the UAS default analysis threshold settings: a 1.5% Analytical Threshold (AT), 4.5% Interpretable Threshold (IT) and STR Stutter Thresholds (144). The AT represents the lower limit of detection and requires > 1.5% of total sequencing reads to align to a locus, with a minimum coverage of 11 reads, in order to be called (69). Any allele not reaching this level of coverage is not included within the UAS reports. The IT is then used as a conservative threshold for the calling of alleles. For most loci this is set to 4.5%, requiring > 4.5% of total sequencing reads to align to a locus with a minimum coverage of 31 reads (69) for classification as an allele. If both alleles fail to reach this threshold and are not called, this results in locus dropout.

The UAS Web Module (v1.3.6887) was used to visualise processed sequencing data and the Sample Details Report (.xlsx) was downloaded for the further analysis of each sample. FastQC (v.0.12.1) was used to generate quality control reports of FastQ files. Additional analyses were undertaken using RStudio (v2023.09.1+494), R (v4.3.1) and Microsoft Excel 2016.

3 PHASE ONE: GENERATION OF REFERENCE PROFILES

3.1 INTRODUCTION

The sequencing methodologies outlined in Chapter Two were utilised to investigate if alternative MPS technologies could generate DNA profiles from cartridge cases. To evaluate the performance of these methods, four phases of experimental work were conducted. The first phase saw the application of the mtDNA and ForenSeq™ sequencing workflows to buccal swabs to generate reference profiles. All participants provided two buccal swabs. This allowed for the generation of two reference profiles for each participant: their mitochondrial haplotype and their STR/SNP profile.

3.1.1 Control samples

In accordance with the SWGDAM guidelines (96), two sets of controls used consistently during this research: an extraction control set and an amplification control set.

The extraction positive sample (EPOS) was blood from a volunteer that had previously had both their mtDNA haplotype and DNA profile sequenced. Their blood was obtained by using a sterile lancet to prick their finger, and then collected in a 1.5 mL tube (Eppendorf) before being divided out into 5 µL aliquots in fresh 1.5 mL tubes. The extraction negative sample (ENEG) was used as a reagent blank and was taken through the entire experimental process (from extraction to sequencing).

The second set of controls, APOS and ANEG, were used as amplification controls. Different amplification positive controls (APOS) were used for the two independent sequencing workflows. For the mtDNA sequencing workflow, the APOS was AmpFℓSTR™ 9947A (Thermo Fisher Scientific), a known DNA template from a human lymphoid cell line, GM09947A (145). GM09947A, herein referred to as 9947A, is a part of the National Institute of Standards and Technology mtDNA standard reference material 2392, and its haplotype has previously been successfully generated using different MPS platforms (146). For the ForenSeq™ sequencing workflow, 2800M control DNA was used as the APOS control. 2800M was provided as part of the ForenSeq™ Prep Kit and is male gDNA from a single source. Each genetic marker included in the ForenSeq™ Prep Kit has been genotyped for this sample. The ANEG controls for both sequencing workflows were nuclease free water. Both APOS controls and ANEG controls were taken through the entire experimental process from amplification onwards.

3.2 MTDNA REFERENCE HAPLOTYPE GENERATION

3.2.1 Laboratory workflow

Each reference buccal swab was sampled following the method outlined in Section 2.4.1, with any gDNA extracted using the DNA IQ™ System as described in Section 2.5. All mitochondrial reference samples were processed in one extraction batch alongside an EPOS and ENEG control. The extracted nuclear DNA was quantified using Quantifiler™ Trio following the method in Section 2.6.

The custom amplification method described in Section 2.7.1 was followed, with each reference sample amplified separately by the two primer pools included in the Precision ID Whole Mitochondrial DNA Panel. The ENEG control was also amplified, alongside an APOS control (9947A) and ANEG control. The APOS control and all the reference samples were added at varying input volumes to ensure they reached the recommended input amount of 0.1 ng. This input volume was determined using the calculation stated in Section 2.7.1. The negative controls were added at the maximum volume input of 7.5 µL. Each sample was amplified using Primer sets A and B, resulting in two multiplexes.

After amplification, 2 µL of each sample was diluted by a factor of 100. 2 µL of each diluted sample was then run on the Fragment Analyzer following the method in Section 2.7.2.1. Resulting electropherograms were used to determine the success of amplification and quantify both the entire sample and successfully amplified mtDNA fragments. Multiplex A and B for each sample were then able to be pooled together (Section 2.7.2.2) in equal concentrations. All negative controls were pooled together in equal amounts.

Library preparation was undertaken using the KAPA HyperPrep Kit as outlined in Section 2.7.3. Libraries were then normalised and pooled together (Section 2.7.5) based on results from a library quantification and a second Fragment Analyzer (Section 2.7.4). Up to 32 samples were pooled together, in preparation for sequencing. The pooled normalised libraries were then denatured and diluted before being sequenced on the MiSeq FGx™ using a 600 cycle MiSeq Reagent Kit v3, following the methods in Section 2.7.6 and Section 2.7.7.

3.2.2 Sequencing performance

The seven reference samples were sequenced in three sequencing runs. There were two replicates of each participants' sample processed, with one of the replicates sequenced in two of the sequencing runs. This resulted in replication occurring at the amplification stage and then replication occurring at the sequencing stage. Replication at the amplification stage was required for the identification of true mitochondrial variation, and replication at the sequencing stage was required to ensure that those sequencing runs would

not fail due to the presence of low DNA quantity samples. The UAS research mode quality metrics for each of these sequencing runs is provided in the table below:

Table 3.1 Quality metrics for the mtDNA sequencing runs that contained reference samples. SeqM001 contained only reference samples, alongside controls. Both SeqM003 and SeqM004 contained the same replicate of reference samples and a mixture of Phase Three cartridge samples alongside controls.

Sequencing run ID	Input concentration (pM)	Cluster density (K/mm2)	Clusters passing filter (%)	Q Score ≥ 30 all cycles (%)
SeqM001	6	809	89.6	58.4
SeqM003	6	891	91.8	46.1
SeqM004	6	361	97.9	59.6

All three sequencing runs had a lower cluster density, compared to the recommended 1200-1400 K/mm^2 range (138), indicating that the runs were under-clustered. The Q score average for all sequencing run cycles was lower than the recommended 70% (138). In comparison to SeqM001, SeqM003 had a slightly higher cluster density and clusters passing filter. However, this sequencing run had a lower Q Score ≥ 30 value, and the reference haplotypes in this run experienced mitogenome dropouts (Section 3.2.5.1). Where previous research has successfully sequenced 32 mtDNA high quality samples in one sequencing run (97), these results indicate that high quality samples should be processed separately to lower quality samples in order to obtain full mitogenome coverage and haplotype profiles from high-quality samples. SeqM004 had the lowest cluster density, indicating that the flow cell was extremely under clustered. As the same reference samples were run in SeqM003, this lower cluster density is likely due to the cartridge samples also included in this sequencing run (further discussed in Section 5.1.3.1). Although the sequencing quality metrics did not meet the recommended values, sufficient sequencing data for subsequent bioinformatic analysis was still obtained.

3.2.3 Bioinformatic processing

The use of short amplicons assays, such as in the Precision ID panel, is known to increase the likelihood of NUMT co-amplification (114). The fragment length of the majority of known NUMTs is between 100 to 500 bp in length (94), and the length of targeted mtDNA fragments amplified by short amplicon assays also fall within this range. As short amplicon assays comprise of hundreds of different primer pairs, some of these can bind to homologous nDNA sequences (94), resulting in NUMT amplification. It is essential to take this into account for haplotype generation, especially in high nDNA quantity reference samples (see Section 3.2.4).

Bioinformatic processing for mtDNA reference samples was carried out by GeneMarker® HTS (GM-HTS). FastQ files generated from MiSeq FGx™ sequencing were directly uploaded into the GM-HTS software. Sequenced reads were aligned using a Burrows-Wheeler hash alignment and trimmed based on the default GM-HTS motif file (139), before being mapped to the rCRS. Variants were identified and recorded in a variant call format

(VCF) file. Any minor variants (PHPs) underwent further analyses (97,114,147) to ensure they were 'true variants' and mitochondrial in origin, opposed to nuclear.

3.2.3.1 GeneMarker HTS analysis pipeline

A pipeline was developed and evaluated by Forsythe (97) for the analysis of Precision ID amplified, Illumina generated sequencing data using GeneMarker HTS v2.4.1. The final pipeline avoided any NUMT interference while allowing for the detection of low-level heteroplasmy and generation of accurate haplotypes. This pipeline used three different alignments for the bioinformatic analysis of Precision ID generated mtDNA sequencing data (97). The first and second alignments are described in Section 2.7.9. The third alignment (without the motif alignment) was used to help resolve dropouts consistently seen at nucleotide positions 100 to 120 and 234 to 252 (97). This alignment used the same values for the Mito Variant Filter Setting parameters and the same Alignment Setting parameters as seen in the first alignment (Section 2.7.9), except the motif alignment (using the default GM-HTS motif file) was unselected.

3.2.3.2 Evaluation of the GM-HTS analysis pipeline

Since the above analysis pipeline's development, GeneMarker HTS v2.6.0 has been released and was the version of GeneMarker HTS (GM-HTS) software used in this research. The above analysis pipeline was trialled on this latest available version of GM-HTS software to determine overall concordance and whether the same pipeline would be suitable without any modifications.

Three APOS samples were used for this evaluation. The haplotype for the APOS control sample used in this study has been previously sequenced on both the Ion Torrent PGM™ and Illumina MiSeq FGx™ platforms (146). The MiSeq FGx™ generated haplotype for 9947A is summarised in Table 3.2.

Table 3.2 9947A haplotype and MiSeq consensus sequence variants compared to the rCRS (146)*. 9947A haplotype variants were determined based on ISFG and IUPAC nomenclature guidelines.*

Nucleotide position	rCRS reference sequence	MiSeq consensus variant	9947A haplotype
93	A	G	93G
195	T	C	195C
214	A	G	214G
263	A	G	263G
309	-	CC	309.1C, 309.2C
315	-	C	315.1C
750	A	G	750G
1393	G	G/A	1393R
1438	A	G	1438G
3242	G	G/A	3242R
4135	T	C	4135C
4769	A	G	4769G
7645	T	C	7645C
7861	T	T/C	7861Y
8448	T	C	8448C
8860	A	G	8860G
9315	T	C	9315C
13572	T	C	13572C
13759	G	A	13759A
15326	A	G	15326G
16311	T	C	16311C
16519	T	C	16519C

Using the analysis pipeline, VCF files were generated for three 9947A APOS controls. One APOS sample (APOS_reference) was from a sequencing run completed in this research project and the other two (APOS_041 and APOS_044) were APOS samples that were used in previous research (97). Each APOS control was generated using the same method, including custom amplification using the Precision ID whole mtDNA Panel, as outlined in Section 2.7. Each variant call file was then compared to the known 9947A haplotype outlined in Table 3.2, to determine overall compatibility of the analysis pipeline on GM-HTS v2.6.0.

Table 3.3 Known 9947A haplotype (146) *and the haplotypes generated for three APOS controls using the GM-HTS analysis pipeline developed by Forsythe* (97). *PHPs are bolded.*

9947A Haplotype	APOS_reference			APOS _041			APOS_044		
	Alignment One	Alignment Two	Alignment Three	Alignment One	Alignment Two	Alignment Three	Alignment One	Alignment Two	Alignment Three
93G	93G	93G	93G	93G	93G	93G	-	93G	
195C	-	195C (99.89%)	195C	-	195C	195C	-	195C	195C
214G	-	214G (99.66%)	214G	-	214G	214G	-	214G	214G
263G	-	263G (89.87%)	-	-	263G	-	-	263G	-
309.1C	-	309.1C	-	-	309.1C	-	-	309.1C	-
309.2C	-	309.2C	-	-	309.2C	-	-	309.2C	-
315.1C	-	315.1C	-	-	315.1C	-	-	315.1C	-
750G	750G	750G	750G	750G	750G	750G	750G	750G	750G
1393R	-	-	-	-	-	-	-	-	-
1438G	1438G	1438G	1438G	1438G	1438G	1438G	1438G	1438G	1438G
3242R	-	-	-	-	-	-	-	-	-
4135C	4135C	4135C	4135C	4135C	4135C	4135C	4135C	4135C	4135C
4769G	4769G	4769G	4769G	4769G	4769G	4769G	4769G	4769G	4769G
7645C	7645C	7645C	7645C	7645C	7645C	7645C	7645C	7645C	7645C
7861Y	**7861Y (17.55%)**	**7861Y (16.88%)**	**7861Y (17.55%)**	**7861Y (18.82%)**	**7861Y (17.11%)**	**7861Y (18.82%)**	**7861Y (17.85%)**	**7861Y (16.62%)**	**7861Y (17.85%)**
8448C	8448C	8448C	8848C	8448C	8448C	8448C	8448C	8448C	8448C
8860G	8860G	8860G	8860G	8860G	8860G	8860G	8860G	8860G	8860G
9315C	9315C	9315C	93165C	9315C	9315C	9315C	9315C	9315C	9315C
13572C	13572C	13572C	13572C	13572C	13572C	13572C	13572C	13572C	13572C
13759A	13759A	13759A	13759A	13759A	13759A	13759A	13759A	13759A	13759A
15326G	15326G	15326G	15326G	15326G	15326G	15326G	15326G	15326G	15326G
16311C	16311C	16311C	16311C	16311C	16311C	16311C	16311C	16311C	16311C
16519C	16519C	16519C	16519C	16519C	16519C	16519C	16519C	16519C	16519C

The 9947A haplotype contains three known PHPs that can be detected using MPS: 1393R, 3242R and 7861Y (146). As seen in Table 3.3, all three alignments for each APOS control were able to accurately report 7861Y (average variant frequency of 17.67 ± 0.26), however, 1393R and 3242R were not reported in any alignment. This was also observed by Faccinetto et al. (123) who was able to correctly report 7861Y (with a variant frequency of 17.1%), but unable to detect 1393R and 3242R with a variant threshold of 10%. They found that lowering this variant threshold to correctly call both 1393R and 3242R consequently resulted in numerous false positives. Similarly, Cihlar et al. (148) found in their research that 1393R had a lower average variant frequency of 2.8 (± 0.75) and was not able to be called with a PHP variant threshold of 10%.

Table 3.4 Regions of dropout in the mitogenome for each alignment of the APOS controls in the GM-HTS evaluation

APOS_reference			APOS_041			APOS_044		
Alignment One	**Alignment Two**	**Alignment Three**	**Alignment One**	**Alignment Two**	**Alignment Three**	**Alignment One**	**Alignment Two**	**Alignment Three**
120 - 384	-	249 - 384	120 - 384	-	249 - 384	82 - 118 120 - 384	-	82 - 118 249 - 384

The first alignment used a strict sort paired-end and motif alignment. This removed background noise and sequencing errors (97) as it requires the sequencing of amplicons in both the forward and reverse direction, with a minimum read depth of 100 reads, to have variants called. As seen in Table 3.4, Alignment One saw dropout from nucleotides 120 to 384 for all three APOS samples. Strand bias ratios were used to determine why this dropout was consistently seen in these samples. Strand bias is a ratio of amplicon reads in both the forward and reverse direction. When the strand bias ratio is one, this indicates that there is an equal amount of reads in both directions. A strand bias ratio that is less than one shows that more reads have been sequenced in one direction compared to the other. If present, sequencing error will only occur in one sequencing direction. Using a strand bias threshold ensures that only variants sequenced in both directions are included in the final haplotype, removing sequencing error.

Table 3.5 Strand bias ratios for Precision ID amplicons that span regions of dropouts in alignment one of APOS controls. Strand bias was calculated by dividing the number of forward reads by the number of reverse reads for that amplicon. Amplicons either include the HV2 homopolymer cytosine stretch (Y) or not (N).

Precision ID amplicon region	Region includes HV2 homopolymer C stretch	APOS_reference	APOS_041	APOS_044
119 - 248	N	0.8687	0.8590	0.8827
248 - 329	Y	0.1828	0.2510	0.1795
299 - 411	Y	0.0016	0.0023	0.0049

As drop out was consistently seen from nucleotides 120 to 384 during the first alignment, strand bias ratios for the amplicons that span across this region were calculated. As seen in Table 3.5, a majority of the sequencing reads for these amplicons were sequenced in one direction. This meant that these amplicons were unable to meet the minimum total read depth threshold of 100 reads in both sequencing directions, resulting in dropout. As these regions of dropout are consistent in all three APOS sample, we are confident that strand bias was due to being a challenging region to sequence rather than due to sequencing error. Both amplicons 248 – 329 and 299 – 411, encompass the homopolymer cytosine region in HV2 which is located from nucleotide positions 300 to 315. It is known that this particular HV2 region is difficult for MPS technology to sequence, typically resulting in low coverage and high strand bias, due to the alignment of a circular mitogenome to a linear reference sequence (89). In comparison, amplicon 119 – 248, spans the HV2 region but does not include the homopolymer cytosine stretch, therefore, only saw slight strand bias in all three APOS samples (Table 3.5).

The second alignment saw the removal of the strict sort paired-end alignment, and resulted in the calling of all variants, regardless of if they were only sequenced in one direction. The second alignment resolved all regions of dropout and provided full mitogenome coverage for all APOS controls (Table 3.4). This subsequently allowed for accurate detection of the 9947A haplotype variants, 195C, 214C, 263G, 309.1C, 309.2C and 315.1C, all located within the dropout region of 120 to 384.

The use of the second alignment did not see the introduction of any new false positive calls at any areas in the mitogenome. During initial development of this GM-HTS analysis pipeline, it was determined that the second alignment repeatedly introduced six false PHPs at reproducible frequencies in several APOS replicates (97). This ultimately influenced the final decision to restrict the second alignment to two regions in the mitogenome; nucleotides 248 to 329 and then 299 to 411 as these regions are amplicons in the Precision ID panel that span across the cytosine stretch in HV2. As no false positive calls were made during the second alignment in any of the APOS samples in this evaluation, these restrictions seem redundant using v2.6.0 of GM-HTS. However, the removal of the strict sort paired-end setting impacts the ability to determine if minor variants are due to sequencing error (and present in one sequenced direction) or true variation (and present in both directions). Consequently, for this research, Alignment Two was restricted to resolving regions of dropout seen in the first alignment, with any minor variants only called based on the first alignment.

As mentioned above, Alignment Two provided full mitogenome coverage for all APOS controls. This is also a point of difference between this evaluation and previous research. Forsythe found that both alignment one and two saw consistent dropouts at nucleotide positions 100 – 120 and 234 – 252 due to use of the motif alignment setting (97). A third alignment (strict sort paired end, without motif) was required to resolve these areas of dropout. Because motif alignment sees correct forensic alignment of all reads to the rCRS, and its use is invaluable in correct haplotype generation, the third alignment was restricted to nucleotides 100 – 120 and 234 - 252 only (97). The use of the third alignment in this evaluation consistently saw dropout from nucleotide position 249 – 384 and was unable to be used to call variants from nucleotide positions 234 – 252. Also, any dropout seen in first alignment was resolved in second, only to be reintroduced in the third alignment contradicting its use. Based on this, the third alignment was not used for haplotype generation.

3.2.3.3 Evaluation of the minor variant threshold

The minor variant threshold was set at 10% in Forsythe's work to help mitigate NUMT and background noise influence on PHP detection for high-quality buccal reference samples (97). Previous studies, typically favouring a large amplicon assay approach, have lowered this threshold (146,149) and successfully generated accurate haplotypes. A small evaluation on the impact of minor variant detection in data generated from a

short amplicon assay was carried out to determine whether it would be appropriate to lower the minor variant threshold in this research (Table 3.6).

Table 3.6 Known 9947A PHPs and whether they were able to be detected in GM-HTS alignment one with different minor variant thresholds. Red font indicates that although the PHP was correctly called, false positive variants were called also observed.

Known 9947A PHPs	Minor variant threshold	APOS_reference	APOS_041	APOS_044
1393R	10 %	-	-	-
	5 %	-	-	-
	2.5 %	1393R	1393R	1393R
3242R	10 %	-	-	-
	5 %	-	-	-
	2.5 %	-	-	-
7861Y	10 %	7861Y	7861Y	7861Y
	5 %	7861Y	7861Y	7861Y
	2.5 %	7861Y	7861Y	7861Y

Lowering the variant threshold to 5% did not see the introduction of false positive calls in the 9947A mitogenome, however, both 1393R and 3242R remained undetected. While a 2.5% variant threshold could call the 1393R, multiple incorrect minor variants were also called in all three controls. Based on this, lowering the variant threshold to 2.5% was deemed unsuitable for generating accurate haplotypes.

The decision was made to keep the variant threshold set to 10%. Where these results show that a 5% threshold can call minor variants correctly in a high-quality positive control (9947A) and avoid NUMT influence, this might not be the same for the reference samples processed. As high quantity samples amplified by a short amplicon approach increase the likelihood of NUMTs, the use of a higher threshold (e.g., 10%) helps minimise NUMT impact on true heteroplasmy detection (150). Furthermore, previous research has shown that a 10% threshold eliminates DNA damaged sites from impacting low-level heteroplasmy detection (4). It is expected that DNA damage will occur to touch DNA deposited on metallic surfaces, however the extent of this on the samples in this research project will not be determined until analysis. It was decided that a 10% threshold would be used for all haplotype generation to maintain consistency throughout this research.

3.2.3.4 Conclusion

As seen in Table 3.3, accurate and concordant 9947A haplotypes were able to be generated using GM-HTS (v2.6.0). These haplotypes were able to be determined using two of the three alignments in the analysis pipeline developed by Forsythe (97), removing the need for the third alignment (strict sort paired-end, no motif alignment). GM-HTS has undergone three software updates since v2.4.1. These updates have seen major, moderate, and minor changes to the software. Major changes are defined as significant changes to algorithm that could change alignment and variant call detection differences. Moderate changes are defined

as minor changes to algorithm that could affect sample grouping and final reports. These software changes have likely resulted in the third alignment (strict sort paired-end, no motif alignment) being no longer required.

The final GM-HTS analysis pipeline used for all haplotype generation consisted of only the first and second alignment of the original pipeline by Forsythe (97). Specifically, minor PHP variants were only called in Alignment One (strict sort paired-end read alignment). Only major variants and major length variants found in areas of dropout due to low coverage in Alignment One, were then called in Alignment Two (without strict sort paired-end alignment). The final analysis pipeline is described in Section 2.7.9.

3.2.4 Determining true mtDNA variants

The increased sensitivity of MPS has seen an increase in the number of detected of low-level mitochondrial variants (146). It becomes important to correctly identify that any GM-HTS detected low-level variants are true mitochondrial variation and not NUMTs, sequencing error, contamination or PCR induced stochastic effects.

In order to minimise false mtDNA variants, each reference buccal swab was processed in replicate. As explained in Section 3.2.2, one replicate was sequenced in two sequencing runs; however, only partial haplotypes were able to be recovered in both instances due to the presence of low-quality samples within these runs. This meant that while three lots of sequencing data were generated for each participant, minor variants were initially called if they appeared in at least two of the three haplotypes, given that the reason they were not seen in the last haplotype was due to drop out.

Additionally, each minor variant had to be sequenced in both directions (e.g., be called in GM-HTS Alignment One) with a quality score of greater than 30, to avoid any sequencing errors. Where PCR induced stochastic errors occur randomly throughout the mitogenome like PHPs, these errors are not reproducible (94). Therefore, the use of reference sample amplification replicates helped avoid false calls of PHP. Each cartridge and cartridge case sample was taken through the mtDNA sequencing workflow only once, due to the time and resource constraints of this project. However, as each participant's reference haplotype was also generated, any deviations were further analysed to determine if they were a result of DNA damage or one of the possible causes specified above.

The use of extraction and amplification controls helped monitor for contamination. Furthermore, if any sample contained more than three PHPs (94), this indicated that contamination had possibly occurred. These PHPs were then manually examined to determine if they were called as major variants in other samples

processed at the same time. The analyst's haplotype was also generated, to determine if self-contamination occurred during any experimental work.

To help avoid the false calling of NUMTs as minor mtDNA variation, all minor GM-HTS variants underwent a series of bioinformatic and filtering approaches (104,114). Bioinformatically, stringent mapping parameters were put in place to help avoid NUMT alignment (94). As determined by Forsythe (97), the GM-HTS identity score remained set at 90% to ensure that all sequencing reads included in alignment were at least 90% similar to the rCRS. The GM-HTS minor variant threshold was set to 10% and any minor variants called in Alignment One were searched against a database of known NUMT variants complied by Li et al. (147).

Minor variants were checked to determine if they were in phase with other variants. Firstly, the GM-HTS VCF was analysed to see if any other variants were called from the same Precision ID amplicon. Secondly, the raw GM-HTS pile-up was checked to determine if other variants were seen in the same sequencing read but were not called due to a low-level frequency. In both cases, if the minor variant was in phase with other minor variants that were also present in NUMT database, this provided evidence that these variants were of a nuclear source. If the minor variant was in phase with a homoplasmic variant of that haplotype (a major variant), this provided evidence that the minor variant was of a mitochondrial source. Furthermore, each minor variant was analysed on Haplogrep3 (151) to determine the haplotypes estimated haplogroup and whether that variant was phylogenetically expected for that haplogroup. EMPOP was used to determine if the minor variant had been observed in that estimated haplogroup or further sub-nodes of the estimated haplogroup via the haplogroup browser tool. Finally, the read sequence that contained the minor variant was searched in BLAST, to determine if it was similar to any part of the nDNA genome.

3.2.5 Sequencing results

The total number of reads generated for each reference sample in each sequencing run was analysed. In all three sequencing runs, a majority of the reads aligned to the rCRS (Table 3.7).

Table 3.7 Average sequencing performance of only the seven reference samples in three different sequencing runs. Controls were not included in this analysis.

	Total reads	Aligned reads	Aligned reads (%)	Unaligned reads
SeqM001	3141384 ± 495580	2236557 ± 393258	70 ± 3	904828 ± 123398
SeqM003	331172 ± 87654	211942 ± 54652	65 ± 3	119230 ± 33372
SeqM004	285410 ± 61710	178478 ± 35909	64 ± 4	106932 ± 26672

SeqM001 had the greatest average coverage per sample, which was likely due to this sequencing run containing the least total number of samples. In comparison, SeqM004 had the lowest average coverage per sample, which was likely due to the extremely low cluster density recorded for this sequencing run (Table 3.1). SeqM001 also had the highest percentage of the total sequencing reads align to the rCRS. Both SeqM003 and SeqM004 had lower percentage of aligned reads; this was likely due to the various low-quality mtDNA samples processed alongside these reference samples (Section 5.1.3.1).

3.2.5.1 Mitogenome coverage

The low coverage text file report generated by GM-HTS was analysed to determine whether any regions of dropout were observed in the reference samples. As seen in Table 3.8, SeqM001 generated full mitogenome coverage for all seven reference samples. Both SeqM003 and SeqM004 were unable to obtain full mitogenome coverage for a majority of the reference samples within these sequencing runs. This indicated that reference samples in SeqM003 and SeqM004 might not have variants typed that were seen in SeqM001, if they were located in an area of low coverage. This was taken into account during reference haplotype generation.

Table 3.8 Percent (%) of mitogenome coverage obtained for each buccal sample replicate. Mitogenome coverage was calculated by the number of nucleotide positions with > 100X coverage divided by the length of the rCRS (16,569). Any area of the mitogenome with less than 100X coverage was not typed.

Participant	SeqM001	SeqM003	SeqM004
One	100.0	95.1	97.5
Two	100.0	77.9	92.9
Three	100.0	99.8	99.3
Four	100.0	98.3	99.4
Five	100.0	99.6	99.6
Six	100.0	100.0	99.3
Seven	100.0	98.6	98.2

3.2.5.2 Evaluation of experimental controls

The sequencing data generated by GM-HTS for the experimental controls was analysed. As described in Section 3.1.1, two sets of controls were used to help monitor contamination levels and ensure that the experimental workflow had worked as expected.

3.2.5.2.1 Negative control evaluation

Two negative controls (ENEG and ANEG) were processed alongside a replicate of the reference samples and a positive control, in each sequencing run. For each ANEG and ENEG sample, the percentage of sequencing reads that aligned to the rCRS was recorded from the GM-HTS Alignment Statistic text file. Additionally, the

total mitogenome coverage was calculated from data given in the GM-HTS Consensus Statistic text file and used to determine the average read depth across the mitogenome. Both of these values are summarised in Table 3.9.

Table 3.9 Analysis of negative controls processed alongside the reference samples. Average read depth across the mitogenome was determined through the total coverage of a sample divided by the number of bases in the rCRS (16,569).

Sequencing run	Control	Reads aligned to rCRS (%)	GM-HTS alignment	Average read depth across mitogenome
SeqM001	ANEG	0.01	1	1.5X (± 1.3)
			2	1.6X (± 0.8)
	ENEG	0.04	1	4.7X (± 0.7)
			2	5.0X (± 0.7)
SeqM003	ANEG	0.05	1	0.6X (± 0.1)
			2	0.7X (± 0.1)
	ENEG	0.11	1	0.7X (± 0.1)
			2	1.1X (± 0.3)
SeqM004	ANEG	0.05	1	0.4X (± 0.1)
			2	0.5X (± 0.1)
	ENEG	0.10	1	1.1X (± 0.2)
			2	1.9X (± 0.4)

Table 3.9 shows the highest percentage of sequencing reads generated that aligned to the rCRS was 0.11% in the ENEG sample in SeqM003. Of these aligned reads, Alignment One (strict sort, paired-end) for this sample saw an average read depth of 0.7X (± 0.1) across the mitogenome. In Alignment Two (no strict sort, paired end), this increased to an average read depth of 1.1X (± 0.1) across the mitogenome. The increase in average read depth between Alignment One and Alignment Two was a consistent trend seen for all samples. This was expected, as by removing the strict sort, paired-end setting in Alignment Two, all reads were called regardless of if they were sequenced in only one direction.

The maximum read depth across the mitogenome was 5.0X (± 0.7). While this meant that all average read depths across the entire mitogenome were below the 40X variant allele and 100X total read depth thresholds, not all regions of the mitogenome had reads align (Table 3.10). This means that some regions of the mitogenome had read depths over both these thresholds.

Table 3.10 Reads seen in the negative controls of Phase One. Minimum and maximum number of reads shows the range of read depth in regions of the mitogenome that had reads align. Red font indicates that the maximum number of reads passes the 100X read depth threshold. The percentage of mitogenome with read alignment was calculated by the number of nucleotide positions that had reads align divided by the length of the rCRS.

Sequencing run	Control	GM-HTS alignment	Minimum number of reads	Maximum number of reads	Percent of mitogenome with read alignment
SeqM001	ENEG	One	2	232	14.2
		Two	1	238	16.2
	ANEG	One	2	110	4.8
		Two	1	119	9.5
SeqM003	ENEG	One	2	22	8.7
		Two	1	55	10.2
	ANEG	One	1	12	12.1
		Two	1	12	14.0
SeqM004	ENEG	One	2	32	10.7
		Two	1	97	12.7
	ANEG	One	2	10	9.4
		Two	1	10	12.1

As seen in Table 3.10, the maximum read depth was under 100X for all negative controls in SeqM003 and SeqM004 indicating that there were no contaminating reads at a read depth high enough to impact haplotype generation for the reference samples in these sequencing runs. However, the maximum read depth was over 100X for both negative controls in SeqM001. Where neither of these negative controls had any variants called, this does indicate that some contaminating reads in SeqM001 are present at read depths high enough to go over the set GM-HTS thresholds and could impact reference haplotype generation. Since no variants were called, all generated sequencing reads with a read depth ≥ 100X for the negative controls were the same as the rCRS. This means that the contaminating reads could not be traced back to an individual sample that was processed alongside these controls (97). No contamination was seen in the positive control for SeqM001, which generated a fully concordant profile in comparison to the known GM09947A haplotype (Table 3.11).

The relative read depth of the negative controls in SeqM001 was calculated to determine if these contaminating reads would impact haplotype generation for the reference samples also processed in this run. The relative read depth for each nucleotide position was calculated by determining how many reads are at that nucleotide position for all negative controls and positive samples in the same sequencing run (97,152). This is summarised in the below equation:

$$Relative\ read\ depth\ (\%) = \left(\frac{Neg\ control\ read\ depth\ at\ position\ \mathrm{z}}{Pos\ sample\ read\ depth\ at\ position\ \mathrm{z}}\right) \times 100$$

The relative read depth of the negative controls can be compared to the minor variant threshold, providing an estimation of if background noise or contamination is present at levels that would affect the calling of minor variants (97).

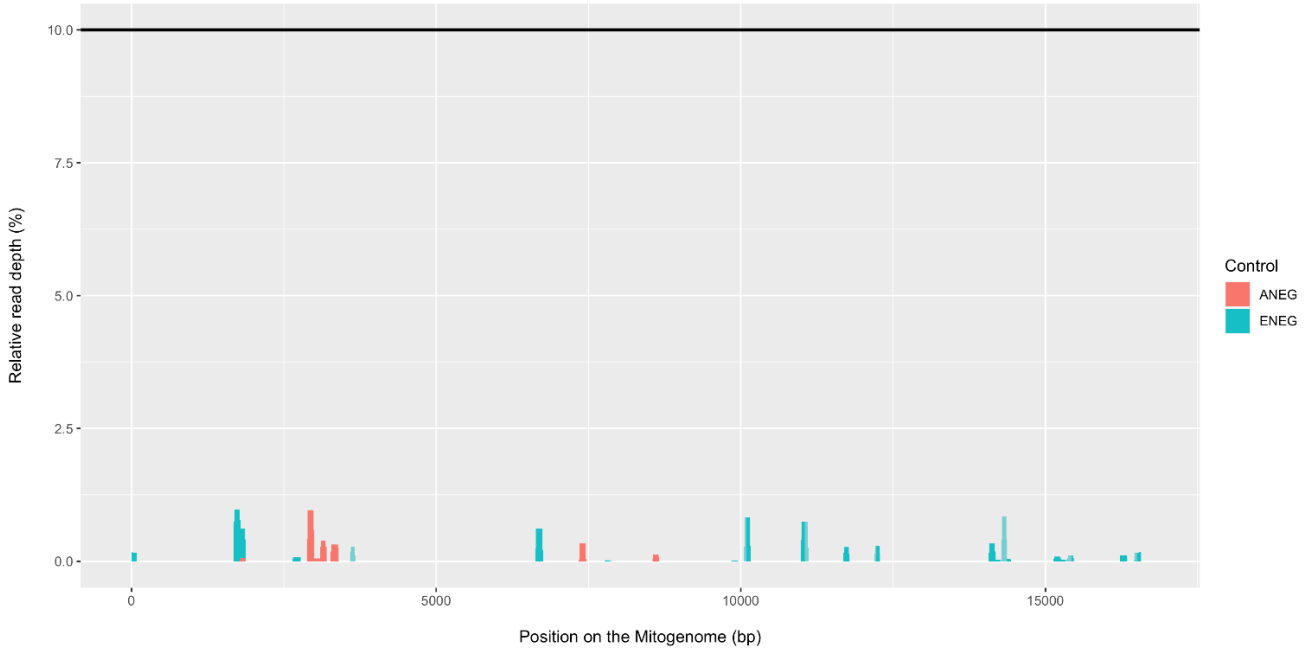

Figure 3.1 Relative read depth for negative controls processed in SeqM001. Black horizontal line at 10% represents the minor variant threshold.

As seen in Figure 3.1, no relative read depths for both negative controls in SeqM001 were high enough to pass the 10% minor variant threshold used in this research. The maximum relative read depth was 0.97% for the ENEG control, showing that all relative read depths were well below this threshold. This indicated that the low-level contamination observed in SeqM001 would not affect haplotype generation for the reference samples within this run.

3.2.5.2.2 Positive control evaluation

One positive control (APOS) was processed alongside a replicate of the reference samples and the negative controls, in each sequencing run. As seen in Table 3.11, full concordant haplotypes were generated for each APOS sample.

Table 3.11 Generated haplotypes for the APOS controls run alongside the reference samples. Concordance to the known 9947A haplotype was determined by dividing the number of concordant variants by the number of expected variants in the 9947A haplotype. Mitogenome coverage was determined by the number of bases out of the whole mitogenome (16,569 bp) with a read depth ≥ 100X.

Sequencing run	Haplotype	Missing variants	Concordance to haplotype (%)	Mitogenome coverage (%)
SeqM001	93G 195C 214G 263G 309.1C 309.2C 315.1C 750G 1438G 4135C 4769G 7645C **7861Y** 8448C 8860G 9315C 13572C 13759A 15326G 16311C 16519C	NA	100	100
SeqM003	93G 195C 214G 263G 309.1C 309.2C 315.1C 750G 1438G 4135C 4769G 7645C **7861Y** 8448C 8860G 9315C 13572C 13759A 15326G 16311C 16519C	NA	100	98.9
SeqM004	93G 195C 214G 263G 309.1C 309.2C 315.1C 750G 1438G 4135C 4769G 7645C **7861Y** 8448C 8860G 9315C 13572C 13759A 15326G 16311C 16519C	NA	100	99.6

While only SeqM001 saw full mitogenome coverage, no variants were located in the regions of the mitogenome that dropped out in SeqM003 and SeqM004. This dropout likely occurred due to the presence of low-quality samples within SeqM003 and SeqM004 (Section 5.1.3.1). As the number of nucleotide positions in the mitogenome with no coverage was extremely minimal, full haplotypes were still able to be generated in both of these APOS controls. The minor PHP variant 7861Y was correctly identified in all three APOS controls with a frequency of 18.7% ± 1.1. Additionally, no false positive major or minor variants called in any of the APOS controls further indicating that this sequencing and bioinformatic workflow was able to generate accurate haplotypes.

3.2.5.3 Analysis of GM-HTS generated haplotypes

The variant call files of the reference buccal samples were then analysed, with any minor variants called in GM-HTS Alignment One flagged for further analysis.

3.2.5.3.1 Quality control of GM-HTS generated haplotypes

The GM-HTS variant data was converted into a text file with an .hsd extension to allow for it to be uploaded directly into Haplogrep3 (v3.2.1). Haplogrep3 was used to estimate the haplogroup of each sample and to determine if the GM-HTS detected variants were expected to be seen in that haplogroup. Haplogrep is a haplogroup classification platform that allows for the efficient analysis of multiple mtDNA samples simultaneously (151). The graphical web service of the most recent version, Haplogroup3 v3.2.1, was used for this research.

Haplogrep applies the Kulczynski measure to each haplogroup within the used PhyloTree to calculate a quality score (153). PhyloTree 17 – Forensic Update 1.2 was the PhyloTree used for this research (151). How well each haplogroup matches the uploaded samples is reflected by a quality score, and the haplogroup with the highest quality score is referred to as the tophit. All the top haplogroup estimates for the reference samples in this research were coloured green indicating that the tophit estimated haplogroup had a quality score greater than 90%.

Haplogrep3 shows the tophit haplogroups expected mutations and whether they were seen in the uploaded sample haplotype. Each expected mutation is either coloured green or red; green indicating the mutation was included in the input samples haplotype, red indicating that the mutation is expected in that haplogroup but not included in the input sample. Remaining variants that are seen in the sample haplotype but not expected in that haplogroup, are also listed and coloured. Purple indicates that the variant is a hotspot mutation, where there is a high number of occurrences of that mutation in the used phylogenetic tree. Orange indicates the variant is a local private mutation, where that mutation is not associated with the tophit haplogroup but is seen in other haplogroups in the phylogenetic tree. Blue indicates that the variant is a global private mutation and not seen in the phylogenetic tree, indicating that it is possibly a genotyping error.

In the seven haplotypes, three major variants were identified as a global private mutation in Haplogrep3. These were 12127T in Participant 4's haplotype and variants 3639G and 3666C in Participant 6's haplotype. Further examination in GM-HTS data showed that all three variants were called with high quality scores of 38 and had minimal strand bias (were > 0.8). Additionally, all three variants were called in all three reference sample replicates at reproducible frequencies: 12127T at 99.18% (± 0.30), 3639G at 99.60% (± 0.09) and 3666C at 99.76% (± 0.04). This provided confidence that these major variants were called correctly by GM-HTS, and all three major variants were included within those corresponding participant haplotypes.

Due to some dropout experienced in reference haplotype generation (Table 3.8), Haplogrep3 was used to help determine if a major variant that was not seen in all three replicates was expected in the final haplotype. For example, in Participant 2's haplotype, 8273T was called as a major variant in only two of the two of the three replicates. Haplogrep3 showed that this variant was expected in the tophit haplogroup for this haplotype and further analysis showed that the region 8229 to 8329 dropped out in the third replicate, accounting for why it was not called. Based on this, 8273T was included in the final haplotype.

3.2.5.3.2 PHP detection and authentication

Out of the seven participants' haplotypes, three had minor variants detected via GM-HTS. This was 43% of haplotypes, a slightly higher proportion than expected (93) and could be due to the small sample size of

participants. Within these three haplotypes, five PHPs were detected in total. One haplotype (Participant 1) had three PHPs detected, with the other two haplotypes having only one PHP detected each. This was as expected (94), helping further indicate that no mixtures or low-level contamination was present.

Firstly, each PHP was compared to the PHPs detected in a population database representative of the New Zealand population that contained 479 full haplotypes (97). None of the five GM-HTS PHPs in this research had been previously recorded, therefore, each PHP had to be analysed further. Table 3.12 summarises the results obtained for the series of analyses (Section 3.2.4) each of these apparent PHP variants went through in order to determine if they were mitochondrial in origin:

3.12 Apparent PHP variants detected in Phase One. Each low-level variant was evaluated to determine if they were of true mitochondrial origin. Quality scores are provided for both the forward (F) and reverse (R) read.

Apparent PHP variant	Particip -ant	Frequency (%)	Seen in replicate ?	Strand bias (minor variant)	Quality score (F;R)	Known site affected by NUMTs	In phase in Precision ID amplicon	In phase in raw GM-HTS data	Expected in haplogroup ?	Seen in haplogroup on EMPOP?	Blast
1700Y *T minor*	7	11.89	No	0.99	38;37	Yes, C is detected in 2 NUMTs	No	Yes	No, private global mutation	No	Mitochondrial and nDNA origin
6962R *G minor*	1	37.15 ± 0.90	Yes	1.00	38;38	Yes, G Is detected in 0	No	No	No, private global mutation	No	Mitochondrial origin
6962R *A minor*	4	11.88 ± 0.18	Yes	1.00	38;38	Yes, A is detected in 21 NUMTs	No	No	No, private global mutation	No	Mitochondrial origin
8269R *A minor*	1	12.02 ± 1.31	Yes	1.00	38;38	Yes, A is detected in 1 NUMT	Yes, with homoplasmic variant	No	No, private global mutation	No	Mitochondrial origin
16286Y *C minor*	1	10.66 ± 0.35	Yes	1.00	38;38	No	Yes, with homoplasmic variant	No	No, private global mutation	No	Mitochondrial origin

As seen in Table 3.12, four of the apparent PHPs were seen in at least two of the participant samples at reproducible frequencies. 1700Y was only called once in Participant Seven's three haplotypes, with 1700C called as a major variant in the other two. All the GM-HTS identified minor variants had a high-quality score of > 30 in both the forward and reverse direction and showed no strand bias. Four of the PHPs were located at a nucleotide position known to be affected by NUMTs (147). Each GM-HTS VCF was checked to determine if any other variants were called in the same amplicon used to sequence the region containing the potential PHP. It was determined that none of the PHPs were in phase with any other GM-HTS detected minor variants, providing evidence that they were not of nuclear origin. Two of the PHPs (8269R and 16286Y) were in phase with a homoplasmic variant, providing evidence towards being of mitochondrial origin. In the raw data pile-up on GM-HTS, the PHPs were analysed to see if any other low-level variants (not present at frequencies high enough to be called as a minor variant by GM-HTS) were present in that same read. Only 1700Y was determined to be in-phase with other low-level variants in the raw pile-up. Approximately half of the sequencing reads containing the minor 1700T variant in Participant Seven were homologous to the nuclear genome, while the other half were homologous to the mitogenome depending on the presence and type of low-level variants also sequenced in that read. It was determined that 1700Y was not a true mitochondrial variant present in frequencies over 10% and 1700C was called as a major variant instead. BLAST searches of the sequencing reads containing all the other potential PHPs showed no homology to the nuclear genome. It was determined that the other four PHPs were all true mitochondrial heteroplasmy.

3.2.5.3.3 LHP detection

All reference haplotypes saw the insertion of a cytosine in HV2 at nucleotide position 315. Additionally, Participants Two and Six also saw an insertion of a cytosine at nucleotide position 309. No other instances of LHP were observed in any haplotype.

As the homopolymer C stretch in the HV2 is challenging for the MiSeq FGx™ to sequence (89), this region dropped out in most samples during the first GM-HTS alignment. HV2 LHP variants were called during the second GM-HTS alignment. As the second alignment does not require reads to be sequenced in both directions for variant calling, only major LHP variants were called and included in the final haplotype.

3.2.5.4 Final reference haplotypes

The resulting final haplotypes in Table 3.13 were determined for each participant:

Table 3.13 Final reference haplotypes and the corresponding haplogroup estimation from Haplogrep3 for each participant in this research. PHP variants are bolded.

Participant	Final haplotype	Haplogroup
One	73G 143A 189G 192C 194T 195C 196C 204C 207A 263G 315.1C 709A 750G 1243C 1438G 2706G 3505G 3531A 4769G 5046A 5460A **6962R** 7028T 8251A **8269R** 8860G 8994A 11674T 11719A 11947G 12414C 12705T 14766T 15326G 15884C 16223T **16286Y** 16519C	W4a1
Two	72C 152C 195C 263G 309.1C 315.1C 750G 1438G 2706G 4769G 7028T 8273T 8860G 15326G 16093C 16298C	HV0+195
Three	263G 315.1C 750G 1438G 3010A 4769G 8860G 14506G 15326G 16209C 16519C	H1
Four	73G 185A 188G 228A 263G 295T 315.1C 462T 489C 750G 1438G 2706G 3010A 4216C 4226C 4769G **6962R** 7028T 8860G 10044G 10398G 11251G 11719A 12127T 12612G 13359A 13708A 14766T 14798C 15326G 15452A 15930A 16069T 16126C	J1c2
Five	73G 150T 263G 315.1C 750G 1438G 1811G 2294G 2706G 3010A 4703C 4769G 6518T 7028T 8860G 9266A 10310A 10506G 11467G 11719A 12308G 12372A 13934T 14139G 14766T 15326G 15454C 16301T 16343G 16356C 16390A 16519C	U3a1c1
Six	73G 185A 188G 228A 263G 295T 309.1C 315.1C 462T 489C 750G 1438G 2706G 3010A 3639G 3666C 4216C 4769G 7028T 8860G 10398G 11251G 11719A 12612G 13708A 14766T 14798C 15064G 15326G 15452A 16069T 16126C 16519C	J1c2
Seven	73G 263G 315.1C 750G 1438G 1700C 2706G 3197C 4769G 5495C 6216C 7028T 8860G 9477A 11467G 11719A 12308G 12372A 13617C 14766T 14793G 15218G 15326G 15924G 16172C 16256T 16270T 16274A 16399G	U5a1a1

3.3 FORENSEQ™ REFERENCE PROFILE GENERATION

The following section describes the generation of gDNA reference profiles using the ForenSeq™ sequencing workflow.

3.3.1 Laboratory workflow

Each reference buccal swab was sampled following the method outlined in Section 2.4.1, with any gDNA extracted using the DNA IQ™ System as described in Section 2.5. All DNA reference samples were processed in one extraction batch alongside an EPOS and ENEG control. The extracted nuclear DNA was quantified using Quantifiler™ Trio following the method in Section 2.6.

Library preparation was carried out using the ForenSeq™ DNA Signature Prep Kit. PCR1, PCR2 and library purification were carried out following the methods described in Section 2.8.1. The ENEG and EPOS controls were amplified and processed alongside an APOS control (2800M) and an ANEG control.

Reference samples were prepared for normalisation following the experiments outlined in Section 2.8.2 before being stored in a -20 °C freezer, until further processing (as described in Section 4.3.1). It was decided these high DNA quantity samples would be included in the same sequencing run as the expected lower quantity touched cartridge libraries, in order to actively avoid sequencing run failure.

3.3.2 Sequencing performance

The seven reference samples were sequenced in triplicate over three separate sequencing runs, alongside low DNA quantity Phase Two samples. The UAS quality metric values obtained for each run is provided in Table 3.14.

Table 3.14 UAS quality metrics for ForenSeq™ sequencing runs containing reference samples.

Sequencing Run ID	Input concentration (pM)	Cluster density (K/mm^2)	Clusters passing filter (%)	Phasing (%)	Prephasing (%)
SeqD001	12.1	989	87.79	0.171	0.053
SeqD002	10.2	844	89.39	0.125	0.020
SeqD003	10.6	555	94.53	0.201	0.023

Each ForenSeq™ sequencing run passed the recommended UAS quality metrics (144). An average cluster density of 796 K/mm^2 (± 128) was seen over the three sequencing runs, which is within the recommend range of 400 – 1650 K/mm^2. The lower cluster density seen in SeqD003 was perhaps due to the impact of storing these samples in the fridge for an extended period of time, as this was the last sequencing run performed using these samples. Regardless, 555 K/mm^2 was still within the recommended cluster density range. As a result, all sequencing data was sufficient for further analysis, with the presence of low-quality DNA samples within these sequencing runs not impacting sequencing performance.

3.3.3 Sequencing results

As each reference sample was sequenced three times in three sequencing runs, all replicates were placed in the same project (LibD001) on the UAS for ease of analysis. All reference samples were reviewed on the UAS server and the Sample Details Report (.xlsx) was downloaded for each sample within this project. This report detailed how many aSTR, ySTR, xSTR and iiSNP loci were typed and what they were called as.

Based on an internal validation of the MiSeq FGx Forensic Genomic System by Jäger et al (57), final STR genotype calls were made using the following criteria:

- Reads were called as alleles when the intensity for that allele was greater than the analytical threshold (AT) and not identified by the UAS as stutter.

- Alleles were called homozygous if a single autosomal allele had an intensity greater than the interpretation threshold (IT).
- If a single autosomal allele was in between the analytical and interpretation thresholds, this allele was called an “ambiguous genotype”.
- If the highest signal (read count) was less than the AT, this allele was not called.

Final SNP genotype calls were then made if a read count greater than 30X was seen at an allele.

During the generation of ForenSeq™ reference profiles the following definitions were established and used to describe all subsequent ForenSeq™ results:

- Concordant loci = allele/s were typed that were expected at that locus, giving the correct result.
- Allele dropout = the dropout of one allele led to an apparent non-concordant genotype. The second allele was typed correctly, giving a homozygous genotype when a heterozygous genotype was expected.
- Non-concordant loci = an allele was typed that was not expected at that locus, giving a wrong result.
- Locus dropout/no result = none or insufficient sequencing reads aligned to that locus

In total, 1974 iiSNPs were sequenced for the reference samples, and 1966 genotypes were generated (99.6% of iiSNPs resulted in a genotype). Four iiSNPs were due to the entire SNP locus dropping out (0.20% of SNP genotypes). Of these four dropouts, rs1736442 and rs7041158 both dropped out twice. Both rs1736442 and rs7041158, have previously been recorded as some of the lowest performing markers in the kit (9,61). A manual examination of the aligned read counts for these SNP loci showed that all five samples had low allele read counts, ranging from 18X to 27X, not meeting the 30X read count threshold and dropping out. Three of these dropouts occurred in the third sequencing run, which was unsurprising given this was the last run completed using these samples, and they experienced the longest time stored at 4 ^{o}C between library normalisation and sequencing. Additionally, two dropouts were from the same sample, Participant Six (replicate three). As seen in Figure 3.2, Participant Six’s sample in SeqD003, was the only reference sample that did not reach the UAS total read guideline of 85,000 total aligned reads (144), therefore, it was not surprising low coverage and dropout of some loci was seen in this particular sample.

The other four iiSNPs experienced an allele dropout (0.20% of SNP genotypes). This resulted in the calling of a homozygous genotype when a heterozygous genotype was expected. Specifically, Participant Two saw an allele dropout in rs729172, Participant Three saw an allele dropout in rs7041158, and Participant Four saw allele dropouts in both rs1357617 and rs1736442. In all cases, the allele called was expected in that genotype and the allele that dropped out did not reach the 30X read count threshold. All of these SNP dropouts were seen in the samples run in SeqD003, again suggesting that a longer library storage time negatively impacts genotype calling.

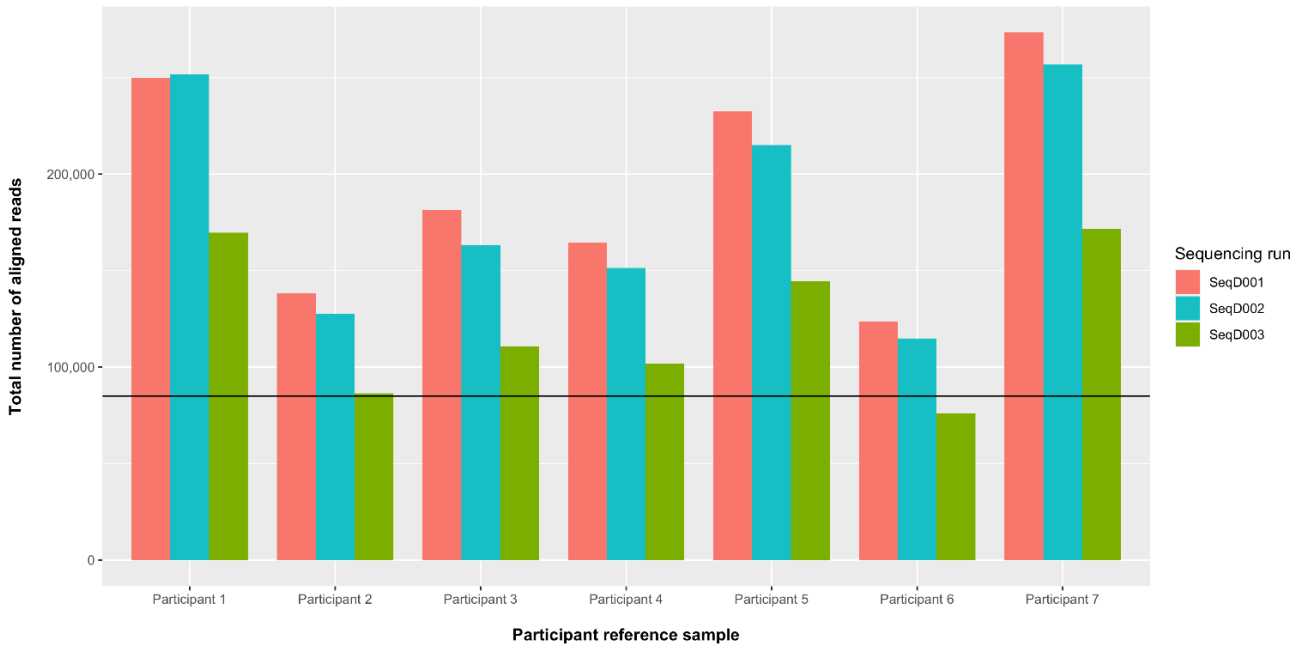

Figure 3.2 Number of aligned reads generated for each reference sample using the ForenSeq™ sequencing workflow over three sequencing runs. The black horizontal line represents the 85,000 read count guideline.

There was a total of 807 STR loci sequenced for the reference samples, of which 803 STRs were typed (99.5% of STRs resulted in a genotype). Six out of seven participants were female, therefore, ySTRs were not sequenced in these reference samples. Only four STR loci dropped out in all the reference samples. In the only male participant (Participant Six), all three replicates saw the drop out of the ySTR locus DYS389II. DYS389II has been previously found to be one of the STR loci most likely to dropout in samples with DNA inputs of ≥ 250 pg (61). It has been suggested that the dropout susceptibility of this marker is due to the long alleles within this locus (61). As this locus was unable to be typed for this participant, DYS389II was not considered for all subsequent profile generation. In SeqD003, the PentaE locus gave an inconclusive result in Participant Three due to a low read count of 27X. This level of coverage saw the PentaE locus reach both the analytical and stutter filter thresholds but not the interpretation threshold, therefore was classified as an 'ambiguous genotype'. However, as the PentaE locus had been correctly called as 13,13 for Participant Three in both SeqD001 and SeqD002, a comparison to the loci called in SeqD003 could still be made.

A majority of the SNP and STR genotypes called were concordant between the three replicates for each participant. As described above, some samples experienced the dropout of an entire SNP or STR locus due to low coverage in one replicate. In most cases, the other two replicates saw that STR or SNP locus called concordantly, allowing for a full profile to still be obtained. Only one y-STR (DYS389II) was unable to be called in all three replicates for Participant 6, and this locus was not included in any subsequent profile generation. Allele dropouts occurred in some iiSNP loci, resulting in a homozygous genotype when a heterozygous genotype was expected. Again, this only occurred in one replicate and the correct genotype was able to be

called from the other two replicates. No non-concordant alleles were typed in all the sequenced reference samples.

3.3.3.1 Control sample evaluation

The seven reference samples were processed alongside negative controls (ENEG, ANEG) and a positive control (APOS). The negative controls were sequenced only in SeqD001 and the APOS sample was sequenced in all three sequencing runs. Both types of negative control did not have any sequencing reads align to the ForenSeq™ loci targets at a high enough coverage to result in the calling of a genotype. As a result, all STR and SNP loci were not typed within these samples. Additionally, no reference profile was flagged by the UAS as being not of a single source indicating that no detectable level of contamination occurred within these samples whether being cross sample contamination, or contamination introduced during laboratory processing. This provided confidence that extremely little contamination occurred during the processing of all ForenSeq™ reference samples.

Each APOS sample was compared to the known 2800M ForenSeq™ profile (74). In all three sequencing runs all 94 iiSNP loci were typed correctly. In SeqD001 and SeqD002, all STR loci were typed, however, stutter impacted the concordance of the D9S1122 in SeqD001 seeing this locus being incorrectly called as 11,12,12. A manual examination on the UAS server showed that stutter's coverage was higher than the stutter threshold and was called. In SeqD003, 58 of the 59 STR loci were typed. DYS389II in SeqD003 was called as inconclusive. Further analysis showed that this locus had a low coverage of 40X resulting in locus dropout. Stutter was also typed in two STR loci (D6S1043 and D21S11) in SeqD003.

3.4 FINDINGS

The purpose of this first phase of experimental work was to generate reference mtDNA haplotypes and DNA profiles for each participant in this research using MPS sequencing workflows either developed (97) or evaluated and optimised (72) previously at ESR. This chapter has shown that full haplotypes and full ForenSeq™ profiles were successfully able to be generated.

Slight technical differences were noticed between the two MPS workflows. The mtDNA workflow was able to generate full coverage of the mitogenome when only high quantity samples were sequenced together. In comparison, the ForenSeq™ workflow was more robust, with high quantity samples able to generate full DNA profiles even when low quantity samples were included in the same sequencing run. Furthermore, the ForenSeq™ workflow was more streamlined and easier to perform.

4 PHASE TWO: ANALYSIS OF CARTRIDGE SAMPLES

The second phase of experiments saw the application of mtDNA and ForenSeq™ sequencing workflows to touch DNA deposited on .223 Rem cartridges. This was to determine if these sequencing workflows were sensitive enough to obtain DNA profiles from an expected trace level amount of touch DNA. Each sample was processed once. Replication occurred at the sampling level, with each participant handling eight .223 Rem cartridges (Section 2.2.2.1). Prior to sampling (Section 2.4), the eight touched .223 Rem cartridges from each participant were randomly allocated into either the mtDNA workflow or the ForenSeq™ workflow. This saw four cartridges selected to be processed through each workflow.

4.1 PRESENCE OF BACKGROUND DNA ON UNFIRED CARTRIDGES

Four cartridges were left over following the creation of the participant deposition packs (Section 2.1.3). These cartridges were swabbed (Section 2.4.2) and used as controls to determine if there was any background DNA present on the cartridges deposited before their acquisition for this research. These four cartridges were also cleaned with 70% ethanol, and then swabbed again (Section 2.4.2) to act as controls for the cleaning method that was applied to all cartridges allocated into the participant deposition packs.

Two of the background and clean control samples underwent the same methodology outlined in Section 4.2.1 to determine if any mtDNA was present and would impact haplotype generation. The other two background and clean control samples underwent the same methodology outlined in Section 4.3.1 to determine if any DNA was present and would impact ForenSeq™ DNA profile generation.

Table 4.1 Quantification of any gDNA recovered from the background and clean controls using Quantifiler™ Trio. UD is short for undetermined concentration. Background controls were cartridges swabbed before being cleaned, clean control were cartridges swabbed after being cleaned.

Cartridge	Sequencing workflow	Control	Quantifiler™ concentration (ng/µL)
1	mtDNA	Background	UD
		Clean	UD
2	mtDNA	Background	UD
		Clean	UD
3	DNA	Background	UD
		Clean	UD
4	DNA	Background	UD
		Clean	UD

The internal positive control cycle threshold (IPC C_T) values for these samples were all within the range of 25.8 to 29. This C_T range is used for casework at ESR Ltd. and shows that no PCR inhibition or reaction failure occurred within these samples. No gDNA was detected in any of the background or clean control extracts (Table 4.1).

4.1.1 Background mtDNA sequencing results

The background controls taken through the mtDNA sequencing workflow (Cartridge 1 and Cartridge 2) were sequenced over two runs (SeqM003, SeqM004). Sequencing reads that aligned to regions of the mitogenome were generated in both background and clean cartridge swabs. The read depth of each sample was analysed to determine if there would be any impact on haplotype generation in this research. The read depth of these aligned reads is shown in Figure 4.1:

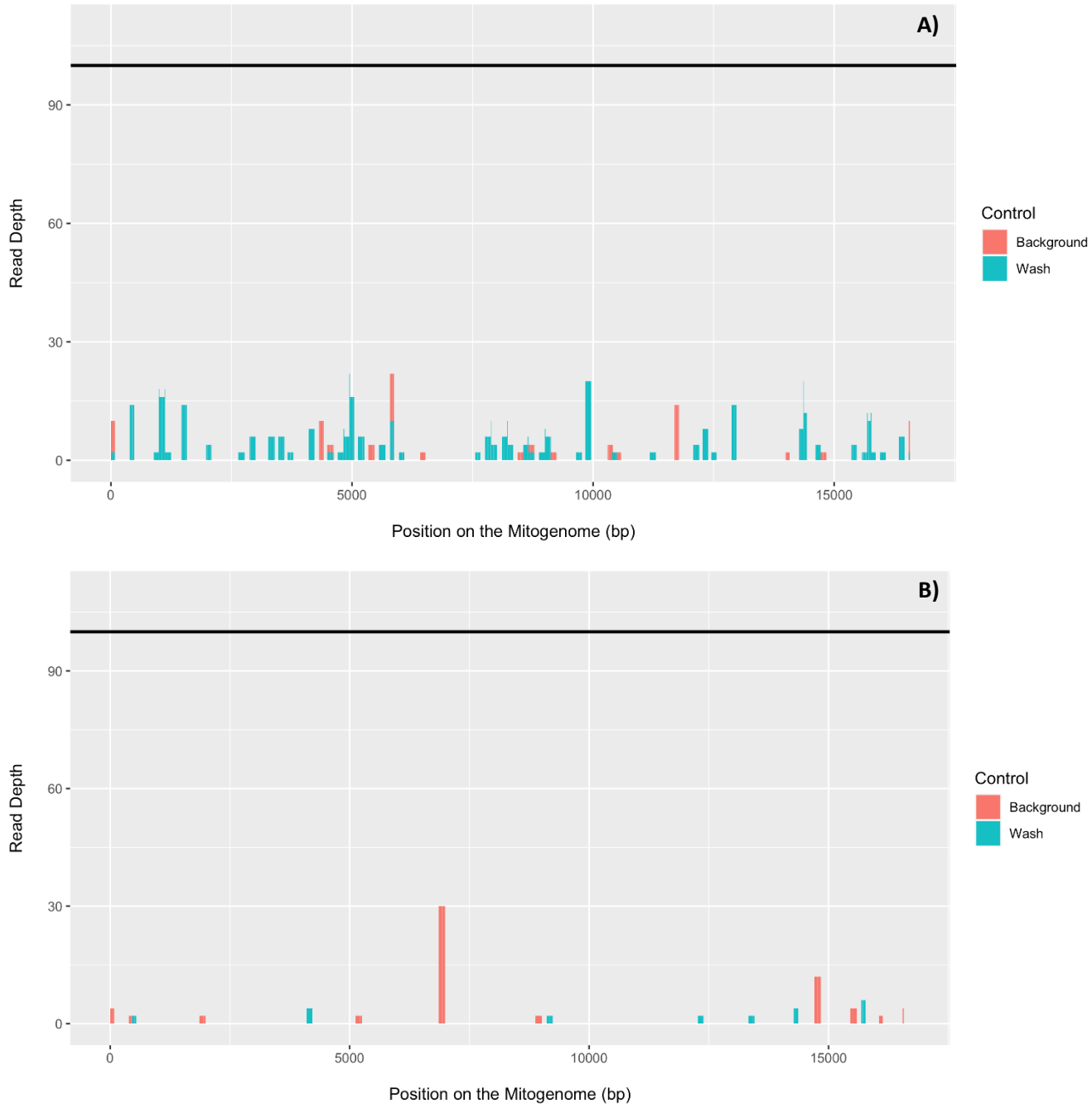

Figure 4.1 Read depth at each position in the mitogenome for background and clean controls in Phase Two. ***A)*** *Read depth for Cartridge 1.* ***B)*** *Read depth for Cartridge 2. The black horizontal line at 100X indicates the minimum total read depth threshold.*

None of the aligned reads in any control sample reached the minimum read depth threshold of 100X at any position in the mitogenome, resulting in no coverage of the mitogenome. Cartridge 1 had a maximum read depth of 22X in both the background control and clean control sample (Figure 4.1 A). Cartridge 2 had a maximum read depth of 30X in the background control, which decreased to 6X in the clean control sample (Figure 4.2 B). This shows that any reads resulted in extremely low levels of background noise, indicating that presence of background DNA on the cartridges used in this research, would not impact participant sample haplotypes. Both negative experimental control samples (ANEG, ENEG) processed alongside these samples

gave maximum read depths around the same value: 26X for the ANEG and 28X for the ENEG. This provided confidence that the aligned sequencing reads are likely low levels of background noise and/or contamination rather than background DNA.

4.1.2 Background nDNA sequencing results

The two background and clean control samples that were taken through the ForenSeq™ sequencing workflow were analysed using the UAS. No aligned sequencing reads were generated for any of these controls; therefore, no target loci were typed, indicating no background DNA on the cartridges used at levels high enough to be amplified using the ForenSeq™ Kit DMPA. This also indicates that the employed method of cleaning the cartridges did not introduce any contamination.

4.1.3 Findings

As MPS offers a more sensitive approach for typing DNA profiles, there was the possibility that any background DNA deposited on the cartridges during the manufacturing process or during the physical transfer of the ammunition to the researcher would be amplified alongside the participant's touch DNA. These results indicate that no amplifiable level of background DNA was present on the cartridges used for this research, with no DNA reads generated using the ForenSeq™ Kit and only low levels of mtDNA reads generated. The aligned mtDNA sequencing reads were indistinguishable from background noise and would not interfere with the haplotypes generated from participant touch DNA samples.

4.2 MTDNA HAPLOTYPE GENERATION FROM UNFIRED CARTRIDGES

The following experimental workflow was carried out to determine if mtDNA haplotypes could be successfully recovered from unfired cartridges using the Precision ID Whole mtDNA Genome Panel.

4.2.1 Laboratory workflow

Sampling of each touched cartridge was performed following the method in Section 2.4.2. Three separate extractions (Batches ExtM003, ExtM004, and ExtM005) were undertaken to extract all gDNA from four cartridge samples for each participant as outlined in Section 2.5. Each extraction batch included an EPOS and ENEG control. Once the gDNA was extracted, all samples underwent quantification to determine if any trace levels of touch DNA were able to be detected using Quantifiler™ Trio following the method in Section 2.6.

The 2 x Qiagen Multiplex MM and the Precision ID Whole mtDNA Panel was used for the amplification of cartridge samples and extraction controls following the custom method outlined in Section 2.7.1. An APOS

and ANEG control were added into the workflow. All cartridge samples and negative controls were added at the maximum input volume of 7.5 µL. The APOS control (9947A) had a genomic concentration of 0.1 ng/µL, 1 µL of APOS and 6.5 µL of Ultrapure water was used. Each sample was amplified once with Precision ID panel primer set A and once with Precision ID panel primer set B, creating two multiplexes.

A dilution of 1 in 100 was performed for each sample prior to quantification on the Fragment Analyzer. Based on Fragment Analyzer results, multiplex A and B were pooled together at equal concentrations for each sample, following the method stated in Section 2.7.2.2. Libraries were created using the KAPA HyperPrep Kit as outlined in Section 2.7.3. 33 samples were prepared in one library preparation batch. This included 28 cartridge samples and five controls (three ENEGs from the three separate extraction batches, one APOS and one ANEG). For library preparation, most samples had the required DNA input concentration of 125 ng; cartridge samples and the APOS control were diluted with Tris-HCl (pH 8.5), and the negative controls were left undiluted. Only one cartridge sample (mtDNA1_1) did not reach the required input concentration so was left undiluted and added at the maximum volume (25 µL).

Quantification and a second Fragment Analyzer were performed on the prepared libraries following Section 2.7.4. Libraries were normalised, pooled together, denatured, and diluted before being sequenced on the MiSeq FGx™ using a 600 cycle MiSeq Reagent Kit v3, following the methods in Section 2.7.5 to Section 2.7.7. The final sequencing run contained 33 samples: 28 cartridge samples, three ENEGs, one APOS and one ANEG sample.

4.2.2 Results

4.2.2.1 Quantification of cartridge samples

Following extraction, DNA quantification provided an estimation of the quantity of gDNA recovered from each touched .223 Rem cartridge.

Table 4.2 Quantifiler™ Trio small autosomal marker genomic concentrations of touch DNA recovered from cartridges chosen to go through the mtDNA sequencing workflow. UD stands for undetermined quantity. The degradation index is determined by dividing the small autosomal target concentration by the long autosomal target concentration. The internal positive control cycle threshold (IPC C_T) was determined by the number of cycles required for the fluorescent signal to pass a set threshold.

Participant	Cartridge ID	Quantifiler™ Trio concentration (ng/µL)	Degradation Index	IPC C_T
1	mtDNA 1_1	0.00572	2.87582	27.80341
	mtDNA 1_2	UD	-	27.90755
	mtDNA 1_3	0.00474	0.90219	27.98018
	mtDNA 1_4	UD	-	28.13536
2	mtDNA 2_1	0.00400	1.40704	28.06866
	mtDNA 2_2	0.00051	3.00472	27.93615
	mtDNA 2_3	0.00042	1.12089	27.95393
	mtDNA 2_4	0.00074	1.56544	27.81573
3	mtDNA 3_1	UD	-	27.78723
	mtDNA 3_2	UD	-	27.91922
	mtDNA 3_3	0.00033	-	27.77913
	mtDNA 3_4	UD	-	UD
4	mtDNA 4_1	UD	-	27.67338
	mtDNA 4_2	UD	-	27.73452
	mtDNA 4_3	UD	-	27.85962
	mtDNA 4_4	UD	-	27.82761
5	mtDNA 5_1	0.00046	1.09715	27.96646
	mtDNA 5_2	0.00058	1.76440	27.78381
	mtDNA 5_3	UD	-	28.12335
	mtDNA 5_4	UD	-	27.91135
6	mtDNA 6_1	UD	-	27.55440
	mtDNA 6_2	UD	-	27.41642
	mtDNA 6_3	UD	-	27.67660
	mtDNA 6_4	UD	-	27.79234
7	mtDNA 7_1	0.00059	2.27099	27.82774
	mtDNA 7_2	0.00563	3.47806	27.72426
	mtDNA 7_3	UD	-	27.95535
	mtDNA 7_4	0.00102	3.46380	27.87080

The short autosomal Quantifiler™ marker was used to determine the concentration of gDNA in each sample (ng/µL). As seen in the above table, 12 of the 28 samples (42.9%) had detectable levels of nDNA. Table 4.2 shows that the IPC cycle threshold (C_T) values for all unfired samples, excluding mtDNA 3_4, were within the range of 25.8 to 29, indicating that no PCR inhibition or reaction failure occurred within these samples. Sample mtDNA 3_4 had an undetermined IPC C_T value. This sample was quantified again in a separate reaction and gave the same result. This suggests that PCR inhibitors were present within this sample, rather than failure of the PCR reaction (127).

Some of the cartridge samples only had the large autosomal assay target detected (data not shown). This was because the large autosomal assay target has a higher copy number within human gDNA, in comparison to

the small autosomal assay target (127). Additionally, the Quantifiler™ Trio user guide warns that amplification of only the larger autosomal target can show that there is not a sufficient amount of DNA present for STR analysis (127).

Quantifiler™ Trio also allows for the quality of a DNA sample to be determined through a Degradation Index (DI). As displayed in the below equation, the DI is the ratio of the small autosomal to large autosomal assay targets and it can be used to determine if degradation has occurred within a sample (127).

$$Degradation\ Index\ (DI) = \frac{Small\ Autosomal\ (ng/\mu L)}{Large\ Autosomal\ (ng/\mu L)}$$

Degradation indices can only be calculated if both autosomal targets are detected in the Quantifiler™ assay. In this study, 11 of the 28 cartridge samples (39.3%) had a DI calculated. Of these 11 degradation indices, four are classified as being non-degraded as they fall within the range of 0 – 1.5 (154). The other seven fall within the range of 1.5 – 4, indicating that they are mildly degraded (154). Degradation was expected in this study, as touch DNA was deposited on a metallic .223 Rem surface that contained copper ions, and water was used during the double swabbing method (4). An aqueous environment on copper and brass surfaces has been found to accelerate DNA degradation (4).

As three extraction batches were used to process all the cartridge samples for mtDNA profiling, there were three sets of ENEG and EPOS controls used and quantified. As seen in Table 4.3, no DNA was detected in all the ENEG controls.

Table 4.3 Quantifiler™ Trio small autosomal marker genomic concentration results for mtDNA extraction batch controls

Extraction batch	Control	Quantifiler™ concentration (ng/µL)	IPC C_T
ExtM003	EPOS	0.78276	28.01482
	ENEG	UD	27.81617
ExtM004	EPOS	0.44054	28.02695
	ENEG	UD	27.47680
ExtM005	EPOS	0.41575	28.02655
	ENEG	UD	27.80301

4.2.2.2 Amplification of cartridge samples

How well the Precision ID panel amplified any mtDNA present on the Phase 2 cartridge samples was determined through visualising electropherograms (EPGs) produced by the Fragment Analyzer. Peaks from 100 to 250 bp, were observed and indicated that mtDNA has been successfully amplified using the Precision ID panel (Table 4.4). This showed that while nDNA was extremely limited in quantity, enough mtDNA was

deposited through touch for successful amplification to occur, most likely due to the high copy number of mtDNA per cell (85).

Table 4.4 Average concentrations of Precision ID amplified Phase Two mtDNA cartridge samples, excluding controls. All concentration values were obtained from a Smear Analysis of all fragments from 100 – 250 bp on the ProSize software v4.0.2.7.

Precision ID primer reaction	Total reaction (ng/µL)	Target amplicon (ng/µL)	Target amplicon (%)	Min target amplicon (ng/µL)	Max target amplicon (ng/µL)	Average amplicon length (bp)
Multiplex A	55.62 ± 6.65	14.94 ± 4.04	26.90	0.35	90.36	169 ± 1
Multiplex B	74.03 ± 8.71	22.69 ± 7.41	30.60	0.42	190.62	166 ± 1

Each sample was amplified with two multiplexes of primers (Multiplex A and Multiplex B). As displayed in Table 4.4, on average it appeared that Multiplex B (Precision ID Primer Pool B) performed slightly better than Multiplex A for these samples, as it gave a higher concentration of target amplicons. However, the performance of each primer multiplex on the unfired cartridge samples was determined to not be significantly different (p = 0.3635, alpha = 0.05, Welch two sample t-test). The concentration of target amplicons was determined by the number of fragments present between 100 to 250 bp in length. This range avoids the concentration of any unbound primers or primer dimers, which are all < 100 bp in length. Where this range is quite conservative considering the average amplicon length of the Precision ID panel is 163 bp, the low quantity of nDNA within these samples provided confidence that any off-target amplification e.g., NUMTs, which are most commonly 100 to 500 bp in length (94), would not have occurred or be included in the final target amplicon concentration.

The minimum target amplicon concentrations for each multiplex were 0.35 and 0.42 ng/µL. Both concentrations were from the same sample (mtDNA 3_4). This same sample had the undetermined IPC C_T value (Table 4.2), further indicating that this sample was severely inhibited or there was no mtDNA present.

A slight trend was seen between participants and target amplicon concentration, with some participants having lower target amplicon and total reaction concentrations than others. Specifically, Participants 3, 4 and 6 consistently gave low concentrations for all four of their samples. This is reflective of the natural inter-variation that is seen between people and their ability to deposit touch DNA 15 minutes after handwashing (28).

4.2.2.3 Library preparation

Following library preparation, a second Fragment Analyzer was carried out and showed how well adapter ligation and library conversion worked. With a 125 ng DNA input amount, it was expected that 25 to 50% of target amplicons would convert into libraries (131). The quality of each library was determined by assessing

the success of library conversion. Any peaks from 120 to 200 bp were target amplicons that had not been converted, with peaks from 280 bp onwards indicating that the amplicons were successfully converted into libraries. By dividing the average percentage of fragments from 280 to 700 bp, by the average percentage of fragments from 120 to 200 and 280 to 700 bp, this gave an approximation of how well library conversion occurred. Based on the data in Table 4.5, it was determined that approximately 48.7% of the target amplicons successfully converted into libraries.

Table 4.5 Average lengths of fragment peaks seen in a Fragment Analyzer EPG after the library preparation of the Phase Two samples taken through the mtDNA sequencing workflow.

Fragment type	Fragment lengths (bp)	Average representation in the sample (%)	Min (%)	Max (%)
Unbound primers/ primer dimers	1 – 100	65.6 ± 3.7	24.6	96.8
Unconverted amplicons	120 - 200	9.0 ± 1.6	1.4	29.7
Adapter-dimer	200 - 279	16.0 ± 2.0	0.9	36.9
Library	280 - 700	8.5 ± 2.2	0.5	49.1

As seen in Table 4.5, unbound primers made up a majority of each Phase Two sample's concentration. This suggests that the number of mtDNA copies present within these samples was low, with not all primers able to bind to their target regions. While the amount of unbound primers was dominant within these samples, primers will not go on to be sequenced as they do not have any adapters bound. Of more concern was the presence of adapter-dimers, as these would go on to be sequenced. On average, adapter-dimers were present in concentrations that were approximately double the library concentration.

Adapter-dimers are also quantified with the KAPA Library Quantification Kit. In this research, Phase Two libraries were individually normalised to 4 nM, before being pooled to create a PNL. The Qubit™ concentration of the PNL should have also been 4 nM, however, a concentration of 2.6 nM was obtained instead. This suggested that library normalisation was impacted by the presence of adapter-dimers, as they overestimated library concentration during quantification. One way to mitigate any downstream consequences of a low concentrated PNL (e.g., the premature stopping of a sequencing run), would be to increase the input volume of the PNL used to create the DNL so that the molarity of the final loading volume was still 6 pM. This would then require the amount of NaOH to also increase, potentially raising its concentration above 0.001 mM and increasing the likelihood of NaOH interference with flow cell binding, leading to under-clustering (72). Tris-HCL must then be added, in a volume that is double the volume of the NaOH, to avoid this (155).

The decision was made to continue onto sequencing without trialling an increased PNL volume. As this was the first time this custom method had been trialled on touch DNA recovered from cartridges, it was important to obtain baseline data. These results could then be used to determine if any modifications to the mtDNA sequencing workflow, such as increasing the PNL volume input based on the final Qubit™ concentration, would further improve sequencing results.

4.2.2.4 Sequencing performance

All 28 cartridge samples were processed in the same sequencing run, alongside experimental controls. The following run quality metrics were obtained from the UAS and provided an indication of the sequencing run performance (Table 4.6).

Table 4.6 Phase Two mtDNA sequencing run quality metrics.

Sequencing run ID	Target input concentration (pM)	Cluster density (K/mm2)	Clusters passing filter (%)	Q Score ≥ 30 all cycles (%)
SeqM002	6	762	91.6	44.8

Compared to SeqM001, which contained only reference and control samples (Table 3.7), SeqM002 gave a lower cluster density and saw a lower percentage of bases with a Q score ≥ 30. This was unsurprising, given that the quality and quantity of mtDNA within these samples was lower than the reference buccal samples. Irrespective of this, enough sequencing data was obtained for further bioinformatic analysis with GM-HTS.

Firstly, the total amount of sequencing reads, and the number of aligned reads generated for each cartridge sample was analysed. This data was obtained from the automatically generated GM-HTS Alignment Statistic text file (.txt). The total amount of reads for each mtDNA sample on average was 933777 ± 75184. The number of these reads that aligned to the mitogenome was on average was 123964 ± 31376 or 12.8% for each sample. Out of 28 samples, only two generated no aligned reads (3_4 and 6_4). This was expected as both samples had extremely low amplicon concentrations prior to library preparation, indicating that they likely contained no mtDNA.

As seen in Figure 4.2 the total number of reads varied for each sample, indicating that each sample was not equally represented in the PNL. This is possibly due to poor library normalisation as a result of pipetting inaccuracy and/or the high presence of adapter-dimers. Interestingly, samples that generated more aligned reads, did not have the highest number of total reads. For example, Sample 4_2 generated 2042306 total reads of which 124928 or 6.1% aligned to the mitogenome. In comparison Sample 2_1 generated 784206 total reads, of which 430548 or 54.9% aligned to the mitogenome.

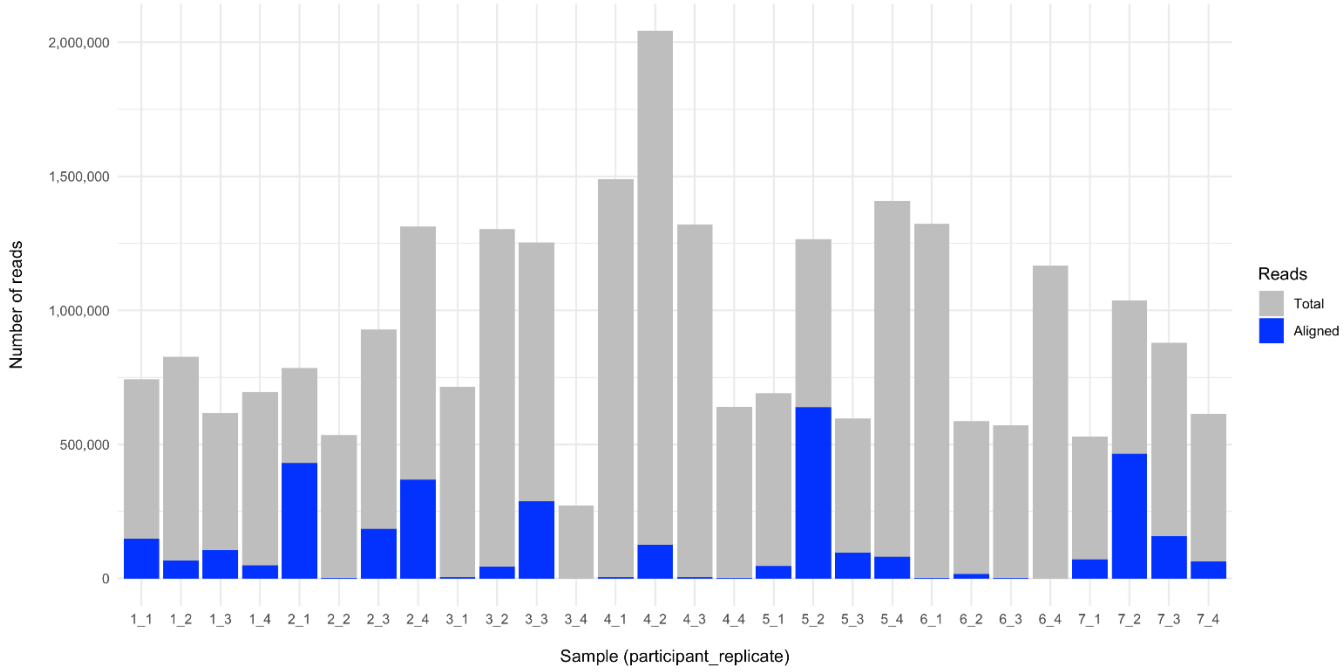

Figure 4.2 Total and aligned sequencing reads generated for unfired cartridge samples taken through the mtDNA sequencing workflow.

A majority of the total reads did not align to the rCRS, as shown in Figure 4.2. This was on average 809813 reads per sample. There are three metrics provided by GM-HTS, that can indicate why some of the sequencing reads did not align to the rCRS. Reads do not align if they are less than 90% similar to the rCRS (low identity reads), are not long enough to pass the minimum read length requirements (short, aligned reads) or have had greater than 10% of the read trimmed due to poor quality (low match proportion reads). The average number of reads for Phase Two mtDNA samples for each of these metrics is provided in Table 4.7.

Table 4.7 Average GM-HTS alignment metrics for mtDNA samples in Phase Two. Data was obtained from alignment one.

Total unaligned reads	Low identity reads	Short aligned reads	Low match proportion reads	Other
809813 ± 74074	21 ± 6	4 ± 1	15297 ± 3596	794492 ± 74867

As seen in Table 4.7, most of the unaligned reads were not filtered from the alignment using one of the metrics explained above. This indicates that unaligned reads are most likely to be adapter-dimer sequences. Out of the three metrics, the low match proportion filter saw the removal of the most reads from those that aligned, possibly due to low-quality template touch DNA present within these samples. The sequencing reads that did align to the rCRS were present in enough coverage to see the generation of haplotype profiles.

4.2.3 Bioinformatic analysis

4.2.3.1 Control evaluation

None of the negative controls processed alongside the Phase Two cartridge samples generated enough reads to reach the minimum read depth threshold of 100X at any position in the mitogenome, with the highest recorded read depth being 32X in an ENEG sample. This indicated that there were no contaminating reads at a read depth high enough to impact haplotype generation for the cartridge samples in SeqM002.

4.2.3.1.1 Positive controls

One APOS sample was run alongside the Phase Two cartridge samples. While this APOS sample had coverage greater than 100 reads for 98.0% of the mitogenome, eight regions did not reach this 100X read depth threshold and dropped out. Two variants; 1438G and 1373C were located in two of these low coverage regions and were not able to be called for this APOS sample (Table 4.8).

Table 4.8 Generated haplotype for the APOS control run in Phase Two. Concordance to the known 9947A haplotype was determined by dividing the number of concordant variants by the number of expected variants in the 9947A haplotype. Mitogenome coverage was determined by the number of bases out of the whole mitogenome (16,569 bp) with a read depth ≥ 100X.

Sequencing run	Haplotype	Missing variants	Concordance to haplotype (%)	Mitogenome coverage (%)
SeqM002	93G 195C 214G 263G 309.1C 309.2C 315.1C 750G 4135C 4769G 7645C **7861Y** 8448C 8860G 9315C 13759A 15326G 16311C 16519C	1438G 13572C	90.5	98.0

It was expected that a full haplotype would be able to be generated for the positive control. As previous research was able to repeatedly generate fully concordant haplotypes for an APOS control within a sequencing batch of 32 high mtDNA quantity samples (97), this indicated that the presence of lower quality and quantity libraries within SeqM002 impacted sequencing performance and the ability to obtain full haplotypes. This same APOS sample was sequenced again during optimisation of the workflow and generated a full haplotype (discussed further in Section 5.1.4.2).

4.2.3.2 Unfired cartridge haplotypes

Each Phase Two sample result was compared to the known reference profile (Table 4.9). The number of observed variants concordant to the reference profile was divided by the expected number of variants in the reference profile. This is summarised in the below equation:

$$Profile\ obtained\ (\%) = \left(\frac{Number\ of\ observed\ concordant\ variants}{Number\ of\ expected\ reference\ sample\ variants}\right) \times 100$$

The number of extra variants typed that were not expected to be seen when compared to the reference profile was also recorded. These are referred to as non-concordant variants and were not included in the above calculation (are further discussed in Section 4.2.3.4).

Table 4.9 Summary of mtDNA profiling results in Phase Two. Amount of the expected profile obtained (%) was determined through comparing the observed variants to the variants expected in that known reference haplotype. Haplogroup estimates and quality scores were assigned to the concordant variants using Haplogrep3.

Sample participant_replicate	Profile obtained (%)	Mitogenome coverage (%)	Haplogroup estimate	Haplogroup quality score (%)	Number of non-concordant variants
1_1	78.9	78.3	W4a	90	1
1_2	86.8	92.4	W4a1	98	9
1_3	89.5	93.2	W4a1	95	1
1_4	81.6	83.3	W4a	93	1
2_1	100.0	99.4	HV0+195	94	-
2_2	6.3	3.0	H*2	52	-
2_3	100.0	94.6	HV0+195	94	-
2_4	100.0	99.3	HV0+195	94	7
3_1	-	1.6	-	-	-
3_2	63.6	91.8	H2	77	9
3_3	100.0	98.5	H1	91	7
3_4	-	-	-	-	-
4_1	2.9	4.3	H32	54	1
4_2	87.5	97.8	J1c2	89	15
4_3	8.8	2.7	R2'JT	55	-
4_4	2.9	0.5	H32	54	-
5_1	93.8	94.0	U3a1c	96	-
5_2	100.0	99.1	U3a1c1	99	6
5_3	100.0	97.6	U3a1c1	99	1
5_4	56.3	98.1	U	97	23
6_1	-	-	-	-	-
6_2	39.4	34.2	J1	65	1
6_3	-	-	-	-	-
6_4	-	-	-	-	-
7_1	86.2	93.7	U5a1a1	91	3
7_2	100.0	99.8	U5a1a1	97	-
7_3	93.1	97.7	U5a1a1	93	7
7_4	96.6	94.6	U5a1a1	94	-

Of the 28 cartridge samples collected, 23 samples generated haplotypes (82.1% of the samples). The extent of haplotype generation ranged from extremely partial, with only one variant correctly called (Samples 4_1 and 4_4), to complete haplotypes successfully generated in 25% of the total samples. It was found that there

was a strong positive correlation between the amount of mitogenome coverage achieved and the percentage of variants recovered for a sample that were concordant to the reference profile ($r = 0.97$, $p = < .001$, Pearson's correlation coefficient).

Five of the 28 samples (17.9% of total samples) had no variants called. As mentioned prior, two samples (3_4 and 6_4) did not generate any sequencing reads that aligned to the rCRS. Two samples (6_1 and 6_3) saw the generation of some aligned sequencing reads, however, these were unable to reach the minimum read depth threshold of 100X at any place in the mitogenome. This resulted in the entire mitogenome dropping out. One sample (3_1), passed this minimum read depth threshold in some areas of the mitogenome, however, no variants were located in these typed regions.

4.2.3.3 Haplogroup analysis

As seen in Table 4.9, haplogroups were assigned to the 23 samples that generated haplotypes using Haplogrep3 (151). The haplogroup estimates for five samples were highlighted red indicating that the tophit haplogroup was of low quality (156). This was due to the extremely partial haplotypes generated, with less than 40% of the reference haplotype recovered for these samples. Two samples, Sample 3_2 and 4_2, were highlighted yellow indicating that the tophit haplogroup was of moderate quality (156). Sample 3_2 had 63% of Participant Three's haplotype recovered and was estimated into the haplogroup H2. This subclade classification was incorrect, as Participant Three's reference haplotype was estimated as haplogroup H1. In comparison, Sample 4_2 had 87% of Participant Four's reference haplotype recovered and was correctly estimated into the haplogroup J1c2.

16 samples (57.1% of the total samples) gave a tophit haplogroup that were highlighted green, indicating that the quality score was greater than 90%. All these samples were either estimated to belong to the same haplogroup as the reference sample or estimated to belong to a haplogroup one node higher than the reference haplotype in the phylogenetic tree. For example, Sample 1_4 had the tophit haplogroup W4a. W4a then further differentiates into the subclade W4a1 which was the haplogroup estimated to Participant 1's reference sample.

These results suggest that:

- If Haplogrep gives a red warning, there is not enough mitogenome coverage for an accurate haplogroup estimation. It is recommended that the probability of that haplogroup's frequency in the global population is not able to be considered and any comparisons are instead made based on the typed major variants alone.

- If Haplogrep gives a yellow warning, the estimated haplogroup has to be interpreted with caution. The amount of mitogenome coverage obtained should be considered as well as the quality score assigned to the tophit haplogroup. It appeared that any sample with a haplogroup quality score over 80% and over 80% mitogenome coverage was likely either assigned the correct haplogroup estimation or the correct node (with the true haplogroup located in a sub-node of that node).
- If Haplogrep gives a no warning, this provides a high confidence in the assigned haplogroup estimation.

4.2.3.4 Non-concordant variants

Fifteen of the Phase Two samples had variants called that were not seen in that samples' corresponding reference haplotype (Table 4.9). All of these non-concordant variants were called as minor variants during GM-HTS alignment one. Each of these variants were manually analysed to investigate if they were present due to contamination, stochastic errors, or were possibly due to mtDNA template damage.

Eight of these samples had over three minor variants detected (28.6% of samples). As it is expected for a maximum of three PHPs to be seen in one individual (94), these eight samples were firstly analysed to determine if contamination had occurred during the experimental workflow. Four samples showed the exact same pattern of minor variants at nucleotide positions 2775, 2776, 2777, 2778, 2782, and 2784 (Table 4.10). These nucleotide positions were all amplified by the same amplicon in the Precision ID panel, which spanned from 2773 to 2888. The GM-HTS raw data pile-up was investigated, and it was found that all of these variants occurred in the same sequencing read. This amplicon for these four samples did not show any signs of strand bias with an average ratio of 0.94 ± 0.02. Additionally, the coverage for this amplicon within these samples ranged from 314 reads (Sample 1_2) to 2346 reads (Sample 5_2), indicating that this pattern of variants was not restricted to a specific level of coverage. As this pattern has not been observed before, it is hypothesised that this is a damage pattern that can appear with mtDNA recovered from metal surfaces that has been amplified with the Precision ID panel. These variants were subsequently removed from that sample's haplotype. After removal, all four samples had less than three PHPs detected providing confidence that no contamination had occurred.

Table 4.10 Pattern of minor variants spanning nucleotide positions 2275 to 2784 in the mitogenome. These were observed in four Phase Two samples and each minor variant was seen at a reproducible frequency.

Minor Variant	rCRS	IUPAC	Minor variant frequency (%)				
			1_2	2_4	3_3	5_2	Average
2775T	2775A	2775W	15.28	14.34	12.29	12.53	13.61 ± 0.72
2776C	2776G	2776S	15.28	14.34	12.37	12.36	13.59 ± 0.73
2777C	2777G	2777S	15.28	14.34	12.37	12.44	13.61 ± 0.72
2778A	2778T	2778W	15.28	13.72	12.45	12.36	13.45 ± 0.68
2782T	2782A	2782W	15.28	14.13	12.29	12.10	13.45 ± 0.76
2784del	2782A	a2784	15.28	14.13	12.29	12.10	13.45 ± 0.76

There were four samples with more than three PHPs that did not contain this pattern of variants. All four samples were analysed for possible contamination. None of the minor variants matched the analyst's haplotype, however, one sample from Participant Three, showed signs of contamination with another sample. Six of the nine minor variants detected by GM-HTS in Sample 3_2 were major variants that were called in Participant One's haplotype. During experimental work, the plate layout of all samples saw Participant One's four replicates directly to the right of Participant Three's four replicates. It is likely that cross-contamination between Sample 3_2 and one of Participant One's replicates occurred before the addition of UDI adapters, leading to the calling of some of Participant One's major variants as minor variants in Sample 3_2'S haplotype. These variants were able to be removed from the final haplotype. All other samples with non-concordant minor variants showed no signs of cross-contamination with another sample. Furthermore, all negative controls processed in Phase Two did not show any high levels of contamination (Section 4.2.3.1), providing confidence that contamination was restricted to Sample 3_2.

In most cases, a non-concordant minor variant was called at a nucleotide position where a major variant was called in the known reference haplotype. The major variant was called concordantly, however, the calling of another nucleotide with frequencies above the 10% minor variant threshold also occurred and saw the variant called as a PHP. One option could be to increase minor variant threshold in GM-HTS to avoid the incorrect calling of a minor variant. In these cases, the minor variant was called at frequencies ranging from 10.35 to 33.33% suggesting that a threshold as high as 35% would be required. Consequently, the ability to identify true PHPs and/or mixtures/contamination within these samples would be impacted, with previous research arguing that increasing the minor variant threshold is not an effective solution and recommended increasing the template for PCR amplification or a incorporating a DNA repair procedure instead (130).

If any minor variants were not able to be removed from the final haplotype following these two filtering approaches, there was the possibility they were sites of damage or stochastic PCR error (118). Currently, there are only databases of likely damage sites in the mtDNA CR available for comparison, therefore each site was investigated further. It was found that each site was called with a Q score > 30 and showed no amplicon bias, providing confidence they were not due to sequencing error.

There are two main types of hydrolytic DNA damage: deamination and depurination. Deamination sees an amine group removed from either a cytosine or adenine base leading to the formation of uracil or hypoxanthine, which are analogues of thymine and guanine (157). Consequently, this sees the transition of a C to T or the transition of a G to A (157). Depurination then sees the loss of a purine base resulting in an abasic site, which during PCR enrichment typically results in the incorporation of an A base but can also lead to a transversion into a pyrimidine base. Oxidative damage can then be seen as a G transverses into a T (or a C to an A) (4). One previous study, restricted to analysis of only the CR, have found that deamination (C to T transitions) made up a majority of the damaged lesions in mtDNA recovered from unfired cartridges (4). As seen in Table 4.11, 28 instances of potential DNA damage and stochastic error were identified in the Phase Two samples. Pyrimidine transitions (C to T or T to C) were the most commonly observed. 14 instances of damage consistent with deamination patterns were observed, of which 9 were C to T transitions aligning with previous results (4).

Table 4.11 Identified minor variants in all Phase Two samples that were possibly due to mtDNA damage or stochastic PCR errors. Variants were determined to be either potential sites of damage or PCR error based on (118,130).

Nucleotide position	Variant	IUPAC	Seen more than once?	Frequency (%)	Potential reason	Haplogrep3 ID colour
1903	C1903T	1903Y	-	10.86	Deamination	Blue
1958	G1958A	1958R	-	10.51	Deamination	Blue
3197	T3197C	3197Y	-	19.63	PCR error	Blue
4216	T4216C	4216Y	-	17.93	PCR error	Blue
5442	T5442C	5442Y	-	15.08	PCR error	Blue
6272	A6272G	6272G	-	15.54	PCR error	Blue
6335	C6335T	6335Y	-	13.01	Deamination	Blue
7768	A7768G	7768R	-	17.86	PCR error	Blue
8613	A8613G	8613R	-	14.02	PCR error	Blue
9625	C9625T	9625Y	-	23.15	Deamination	Blue
9797	T9797C	9797Y	-	15.47	PCR error	Blue
9824	T9824C	9824Y	-	15.66	PCR error	Blue
12091	T12091C	12091Y	-	21.36	PCR error	Blue
12775	G12775A	12775R	-	11.86	Deamination	Blue
13954	C13954T	13954Y	-	16.20	Deamination	Blue
14178	T14178C	14178Y	-	15.37	PCR error	Blue
14182	T14182C	14182Y	-	15.22	PCR error	Blue
14368	C14368T	14368Y	-	10.27	Deamination	Blue
14371	T14371C	14371Y	-	10.70	PCR error	Blue
14398	A14398T	14398W	-	10.38	PCR error	Blue
14404	C14404T	14404Y	-	10.27	Deamination	Blue
14418	C14418T	14418Y	-	10.27	Deamination	Blue
14766	C14766T	14766Y	-	11.12	Deamination	Blue
15043	G15043A	15043R	-	24.26	Deamination	Blue
15261	G15261A	15261R	-	23.89	Deamination	Blue
16270	C16270T	16270Y	-	33.33	Deamination	Blue
16311	T16311C	16311Y	-	33.33	PCR error	Yellow
16336	G16336A	16336R	-	36.00	Deamination	Blue

All haplotypes with the non-concordant variants listed in Table 4.11 were uploaded to Haplogrep3 to determine whether the presence of these non-concordant variants was expected in the phylogenetic tree. All except one were identified as being a global private mutation (gave a Haplogrep ID colour of blue) indicating that that variant had not been observed before and that possibly a genotyping error had occurred. Excluding the identified damage pattern, all instances of potential damage and stochastic error were only seen once within all 28 samples. This indicated that damage and stochastic error was not reproducible between the four

replicates for each participant and that the use of replicates can help identify sites of damage and error. For application of this workflow into casework, it is recommended that replicates of each sample occur at the amplification level as it has been proven that no site of DNA damage is replicated in duplicate amplification (4,157).

4.2.4 Findings

The results from Phase Two show that the mtDNA sequencing workflow can successfully amplify and sequence mtDNA from touch DNA deposits on .223 Remington cartridges. The targeting of mtDNA is a more sensitive approach for this evidence type, as samples with undetermined quantities of nDNA contained sufficient quantities mtDNA to be amplified, successfully converted into a library, and sequenced. The sequencing performance of this workflow on unfired cartridge samples was lower in comparison to high-quantity buccal samples with a low cluster density and low Q Score ≥ 30 suggesting that optimisation should focus on improving these parameters, especially cluster density to ensure that the highest possible coverage of the mitogenome is able to be obtained.

There were several non-concordant minor variants observed in the unfired cartridge samples that had to be identified and removed from the final haplotype. This same mtDNA sequencing workflow has previously been tested on telogen hair shaft fragments and rootless hair fragments samples that contain limited nDNA (97). Extra variants that were non-concordant to the reference haplotype were also recorded, however, these were consistently seen between replicates from the same participant and were identified as authentic PHP variants (97). In comparison, the non-concordant variants observed in the unfired cartridge samples randomly occurred and were not replicated - excluding the pattern of variants from nucleotide position 2775 to 2784. This suggests that mtDNA recovered from unfired cartridges can be more damaged or prone to stochastic errors compared to other low nDNA content samples (e.g. telogen hair shafts).

Holland et al. (4) identified 44 sites of damage when using the PowerSeq™ CRM Nested System kit (Promega) to amplify mtDNA from unfired cartridges. Most of these sites were present in 1 to 2% of total reads generated for that nucleotide position. In the present study, potential sites of damage and stochastic errors were seen at higher frequencies of 17.2% ± 1.4. Additionally, non-concordant minor variants observed at positions where a major variant was expected to be seen had frequencies ranging from 10.35 to 33.33%. Positions with mixed base calls that are not true heteroplasmy are also commonly referred to as miscoding lesions (149). It has been found that the frequency of a minor variants identified as miscoding lesions increases as storage and damage conditions worsen (149). The high frequencies of minor variants arising from potential damage and stochastic errors suggest that the mtDNA in this study experienced damaging conditions, possibly due to the

cartridges not being sampled within 96 hours after touch DNA deposition (12), and the use of water during sampling (4).

To identify potential damage and anomalies for each sample, two filtering steps were used. If non-concordant variants remained following the below steps, these were identified as being a potential site of damage.

1. **Each minor variant was compared to other samples processed at the same time.** If over three minor variants were identified as major variants in another sample, this indicated that cross-contamination had occurred, and these minor variants were removed from the final haplotype.
2. **Each minor variant was compared to the major variants detected for that sample**. If a nucleotide position had both a major and minor variant detected this indicated the presence of PHP. As a known reference haplotype was available for comparison, it was determined that in all cases the major variant was called correctly, and the minor variant was not expected. In this research, these variants (both the minor and major) were removed from the final haplotype for ease of analysis. However, as the major variant was called concordantly to the reference profile, future work should focus on determining how to confidently call the major variant in this instance. One option could be to increase the minor variant threshold and then call haplotypes based on the major variants only, however this would restrict any ability to identify low-level contamination or mixtures. Separately, the use of replicates at the amplification level should help distinguish between true PHP and other artefacts appearing as PHP (149) as only true PHPs would be present in reproducible frequencies. Therefore, it would be expected that only the major variant would be called in a second replicate.

4.3 FORENSEQ™ PROFILE GENERATION FROM UNFIRED CARTRIDGES

The following experimental workflow was carried out to determine if nuclear DNA profiles could be successfully recovered from unfired cartridges using the ForenSeq™ Kit.

4.3.1 Laboratory workflow

Sampling of each touched cartridge was performed following the method in Section 2.4.2. Three separate extraction batches (ExtD002, ExtD003, and ExtD004) were undertaken to extract all gDNA from four cartridge samples for each participant using the method outlined in Section 2.5. All extracted nDNA was quantified using Quantifiler™ Trio (Section 2.6). ForenSeq™ libraries were prepared and sequenced following Section 2.8. Each sequencing run included a mixture of controls, reference samples and touched cartridge samples.

4.3.2 Results

4.3.2.1 Quantification of cartridge samples

DNA quantification provided an estimation of the quantity of gDNA recovered from each touched .223 Rem cartridge chosen to undergo the ForenSeq™ sequencing workflow (Table 4.12).

Table 4.12 Quantifiler™ Trio small autosomal marker genomic concentrations of touch DNA recovered from cartridges chosen to go through the ForenSeq™ sequencing workflow. UD stands for undetermined quantity. The degradation index is a ratio of the small to large autosomal target.

Participant	Cartridge ID	Quantifiler™ Trio concentration (ng/µL)	Degradation Index	IPC C_T
1	DNA 1_1	UD	-	27.95469
	DNA 1_2	0.00292	1.94565	27.89340
	DNA 1_3	UD	-	28.09097
	DNA 1_4	0.00205	1.63463	28.16219
2	DNA 2_1	0.00079	3.48297	28.00282
	DNA 2_2	0.00057	-	28.06262
	DNA 2_3	0.00245	5.22265	27.98347
	DNA 2_4	UD	-	27.89332
3	DNA 3_1	UD	-	28.18231
	DNA 3_2	0.00015	-	28.01461
	DNA 3_3	UD	-	28.46222
	DNA 3_4	UD	-	28.23310
4	DNA 4_1	UD	-	27.76669
	DNA 4_2	UD	-	27.95490
	DNA 4_3	UD	-	27.67636
	DNA 4_4	UD	-	27.75567
5	DNA 5_1	UD	-	27.90248
	DNA 5_2	0.00051	2.68327	27.84307
	DNA 5_3	UD	-	28.14161
	DNA 5_4	0.00027	1.45386	27.83392
6	DNA 6_1	UD	-	27.71493
	DNA 6_2	UD	-	27.89829
	DNA 6_3	UD	-	27.80629
	DNA 6_4	0.00022	-	27.82134
7	DNA 7_1	0.00068	2.42106	27.50372
	DNA 7_2	0.00100	-	31.17970
	DNA 7_3	0.00015	2.09925	28.05130
	DNA 7_4	UD	-	27.64617

Of the 28 .223 Rem cartridge samples, 12 samples (42.9% of sample) had the short autosomal Quantifiler™ assay target detected, with the largest quantity being 0.00292 ng/µL. The rest of the samples had an undetermined amount of nDNA. Of these 12 samples, eight also had the large autosomal target detected, allowing a degradation index (DI) to be calculated. One sample was classified as non-degraded as it was within the DI range of 0 – 1.5, six were classified as moderately degraded as they fell between 1.5 – 4, and one sample was classified as degraded as was between 4 – 10 (154). Again, this was unsurprising given the use of water on a brass metallic surface (4) during the double-swab collection method. All samples had an IPC C_T between 25.8 and 29.0 except Sample DNA 7_2, which had the IPC C_T flag triggered. This indicated that the DNA within this sample was likely affected by PCR inhibition (127).

As seen with the mtDNA unfired cartridge samples, three extraction batches were also used to process all touched cartridge samples for ForenSeq™ DNA profiling. A positive (EPOS) and negative (ENEG) control was processed alongside each batch. Each ENEG control had an undetermined amount of DNA indicating that no contamination occurred during sampling, extraction, and quantification of the touched cartridge samples within these three batches and each EPOS control successfully returned a DNA quantity between 0.2 to 0.4 ng/µL.

4.3.2.2 Library Preparation

The ForenSeq™ kit was used for library preparation of these samples. After library preparation each sample was run on the Fragment Analyzer to determine whether ForenSeq™ targets were able to be successfully amplified and converted into libraries.

As seen in Table 4.13, most of the peaks for these samples were seen from 160 to 180 bp. Peaks within this range represent the presence of adapter-dimers, which like the mtDNA sequencing workflow, will go on to be sequenced. Very little peaks were seen from 200 bp onwards, indicating that extremely low quantities of ForenSeq™ libraries were generated. While this kit has increased sensitivity through the use of shorter amplicons compared to amplification kits that target conventional STR markers for CE-STR analysis, these results in conjunction with the above quantification results, show that the samples recovered from cartridges in this research contain extremely little nDNA.

Table 4.13 Average lengths of fragment peaks seen in a Fragment Analyzer EPG after the library preparation of the Phase Two samples taken through the ForenSeq™ sequencing workflow.

Fragment Type	Fragment lengths	Average representation in the sample (%)	Minimum value (%)	Maximum value (%)
Unbound primers/primer dimers	1 – 99	19.3 ± 2.6	6.9	57.9
Adapter-dimer	160 – 180	72.5 ± 2.5	34.9	86.1
Library	200 – 700	7.0 ± 0.5	2.9	12.9

Based on these results, and previous recommendations in the literature (72,73), it was decided that the ForenSeq™ reference samples, would be sequenced alongside the ForenSeq™ unfired cartridge samples. Running lower quality samples alongside higher quality samples on a MiSeq FGx™ instrument increases the chances of a sequencing run successfully completing, rather than prematurely stopping due to a high presence of adapter-dimers and low-quality libraries (72).

4.3.2.3 Sequencing performance

Three sequencing runs were used to sequence the ForenSeq™ unfired cartridge samples. All of these runs passed the UAS run quality metrics as previously discussed in Section 3.3.2.

Figure 4.3 shows the total number of sequencing reads and the number of reads that aligned to ForenSeq™ reference sequences for each Phase Two ForenSeq™ sample. The total number of reads for each sample was obtained from a FastQC report, and the number of aligned reads was obtained directly from the UAS server. None of the cartridge samples in Phase Two, had enough aligned reads to reach the UAS minimum total read count guideline of 85,000 reads, indicating that any typed loci should be interpreted with caution (144).

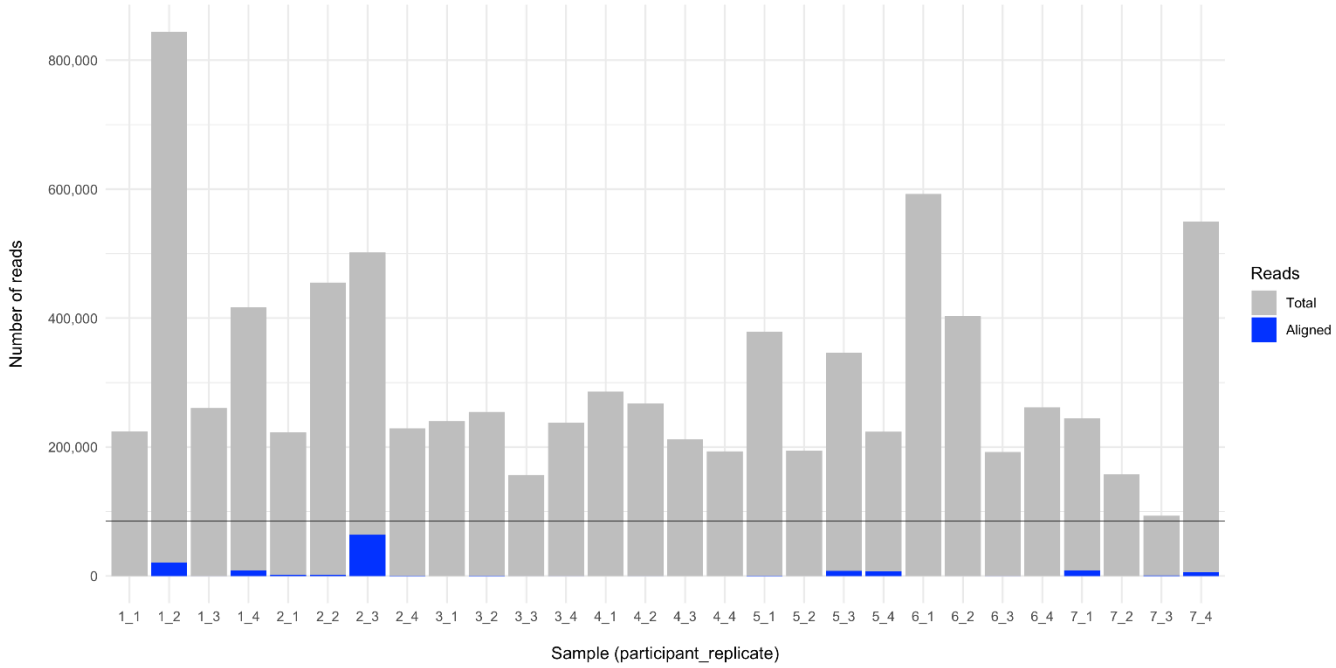

Figure 4.3 Generated sequencing reads for the unfired cartridge samples taken through the ForenSeq™ sequencing workflow. The black horizontal line at 85,000 reads represents the UAS read count guideline. Samples were sequenced over three independent sequencing runs.

The number of aligned reads ranged from zero to 64,098 reads in these 28 cartridge samples. Specifically, 20 of the 28 samples generated reads (71.4%), with eight samples generating no reads at all (28.6%). Of these 20 samples, on average only 1.5% (± 0.6) of the total sequenced reads were aligned reads. In comparison, to the cartridge samples that went through the mtDNA sequencing workflow, 12.8% (± 2.9) of the total sequenced reads were reads that aligned to the mitogenome. This percentage then actually increases to 13.8% (± 3.1) when removing the data for the two mtDNA samples that generated no aligned reads. Based on this, it appears that targeting the whole mitogenome opposed to various nDNA loci is more sensitive for

obtaining informative MPS data from biological samples recovered from unfired cartridges. This is due to mtDNA being more robust and having a higher copy number per cell compared to nDNA (78).

The total number of reads varied between samples sequenced within the same sequencing batch (Figure 4.3), indicating that there was an unequal representation of samples during cluster generation, a result of inaccurate library normalisation. This possibly occurred due to two reasons. Firstly, pipetting errors could have occurred during manual library quantification. As each sample was serially diluted to 1:10,000, any slight deviation in the volume of either the sample or TE buffer would impact the final dilution factor and affect quantification. Similarly, if an inaccuracy in pipetting occurred when loading 4 µL of the diluted sample on the qPCR plate, the final quantification value obtained will not be reflective of the true concentration of the sample. Secondly, the results from library preparation indicated that these Phase Two samples had large amounts of adapter-dimers present and minimal ForenSeq™ libraries. As adapter-dimers contain complete adapter sequences (158), they contribute to library quantification concentrations. Library normalisation, therefore, would have occurred largely based on the concentration of these adapter-dimers.

4.3.3 Bioinformatic analysis

4.3.3.1 Control evaluation

One ANEG and three ENEG samples were processed alongside the 28 Phase Two ForenSeq™ samples. The ANEG control and two of the ENEG samples saw the typing of no loci indicating that no contamination had occurred. One of the ENEG controls saw the typing of one iiSNP locus - rs722290 with a read depth of 50X. In a developmental validation study, Jäger et al. (57) considered ANEG controls with zero to four loci typed (called above the AT threshold) to be blank.

One APOS control was processed alongside the 28 Phase Two ForenSeq™ samples. All STR loci were called concordantly compared to the 2800M reference profile. 91 out of 94 iiSNP loci were typed. The three loci that dropped out (rs1357617, rs7041158, rs1736442) all had low read counts and did not pass the read count threshold. Of the 91 typed loci, three had allele dropouts resulting in the incorrect call of a homozygous genotype when the expected reference genotype was heterozygous. In each instance, the interpretation threshold was flagged and one of the expected alleles did not have enough coverage to be called in the final UAS generated genotype. iiSNP dropout due to low coverage in the APOS 2800M control has recorded before (55).

4.3.3.2 STR profiles recovered from unfired cartridges

As each participant provided a reference sample, each Phase Two ForenSeq™ profile could be compared to the known corresponding reference genotypes for a-STR, x-STR and y-STR loci. This comparison saw the number of STR loci concordant to the reference profile divided by the total number of ForenSeq™ target a-STR and x-STR loci for female participants. For the one male participant, the number of STR loci concordant to the reference profile was divided by the total number of ForenSeq™ STR loci. These are both summarised in the following equations:

$$Female\ STR\ profile\ (\%) = \frac{Number\ of\ concordant\ typed\ STR\ loci}{(28\ aSTR + 7\ xSTR)}$$

$$Male\ STR\ profile\ (\%) = \frac{Number\ of\ concordant\ typed\ STR\ loci}{(28\ aSTR + 7\ xSTR + 24\ ySTR)}$$

The total number of STR loci typed for each sample was recorded. Table 4.14 shows that there were discordant loci typed, they will be discussed in the following section.

Table 4.14 Summary of ForenSeq™ STR profiling results in Phase Two. Concordant STR loci were determined through a comparison to a known reference profile. Loci with allele dropouts gave a homozygous genotype when a heterozygous genotype was expected.

Sample	Aligned reads	Number of typed loci	Concordant loci	Loci with allele dropout	% concordant STR profile
DNA 1_1	-	-	-	-	-
DNA 1_2	20913	32	15	14	42.9
DNA 1_3	62	1	1	-	2.9
DNA 1_4	8731	17	6	11	17.1
DNA 2_1	2258	9	2	3	5.7
DNA 2_2	2136	6	1	4	2.9
DNA 2_3	64098	16	-	2	-
DNA 2_4	616	1	-	1	-
DNA 3_1	-	-	-	-	-
DNA 3_2	192	-	-	-	-
DNA 3_3	28	-	-	-	-
DNA 3_4	79	-	-	-	-
DNA 4_1	-	-	-	-	-
DNA 4_2	56	-	-	-	-
DNA 4_3	-	-	-	-	-
DNA 4_4	-	-	-	-	-
DNA 5_1	216	1	-	1	-
DNA 5_2	697	4	1	3	2.9
DNA 5_3	7701	6	2	4	5.7
DNA 5_4	7544	8	4	3	11.4
DNA 6_1	-	-	-	-	-
DNA 6_2	25	-	-	-	-
DNA 6_3	107	1	-	1	-
DNA 6_4	-	-	-	-	-
DNA 7_1	8835	16	5	9	14.3
DNA 7_2	-	-	-	-	-
DNA 7_3	910	6	1	5	2.9
DNA 7_4	6082	4	4	-	11.4

In 15 of the 28 Phase Two samples (53.6% of samples) STR loci were typed. A total of 128 STR loci were typed of which, 107 were a-STRs, two were y-STRs and 19 were x-STRs. A lack of male participants accounted for the low number of typed y-STR loci.

As seen in Table 4.14, Sample 2_3 generated the highest number sequencing reads that aligned to ForenSeq™ targets, however, none of the typed STR loci were concordant with Participant Two's reference profile. This was an abnormal result and suggested that contamination had occurred. No y-STR loci were typed indicating that this sample was not contaminated by any male sample (APOS or any replicate from Participant Six). All other participant's, the analyst and EPOS STR profiles were compared to Sample 2_3. All the observed non-concordant loci contained alleles that were seen in Participant One's STR profile. A comparison between Sample 2_3 and all four of Participant One's replicates showed they were also typed in either Sample 1_2 or

1_4. Two loci saw allele dropouts. Based on this it appeared that contamination occurred between Sample 2_3 and either Samples 1_2 or 1_4 and this sample was excluded from any further analysis.

Of the 112 total STR loci typed (this total excludes Sample 2_3), 42 loci in 11 samples were fully concordant to that samples corresponding reference profile. This was 37.5% of all STR loci typed. Of the 42 concordant loci, 22 were heterozygous, and 20 were homozygous. In this study, some STR loci were more likely to be concordantly typed than others. D3S1358 and D4S2408 were typed concordantly in three samples, Amelogenin was typed concordantly in four samples and D9S1122 was typed concordantly in five samples. The amplicon lengths for these STR are less than 200 bp (74).

The allele coverage ratio (ACR) was determined for all a-STR and x-STR loci typed that were expected to be heterozygous. ACR was determined through dividing the lowest number of reads by the highest number of reads at a specific locus. The range of ACR seen in this research was from 0 to 0.98. Some loci experienced allele dropouts (Section 4.3.3.2.1) accounting for why ACRs of 0 were obtained. An ACR between 0.6 – 1.0 shows that a heterozygous STR genotype is balanced (159). Heterozygous STR loci with both alleles typed concordantly ranged from 0.22 to 0.98 with an average of 0.55 ± 0.04. This showed that most heterozygote genotypes typed from unfired cartridges were imbalanced. This was excepted given the low nDNA recovered and lower DNA inputs are known to decrease ACR (159).

A correlation analysis was performed to test if there was a relationship between the number of aligned sequencing reads and the percentage of a concordant STR profile recovered. As shown in Figure 4.4, a Pearson correlation coefficient (R) of 0.97 indicated that there was a strong positive linear correlation between these variables.

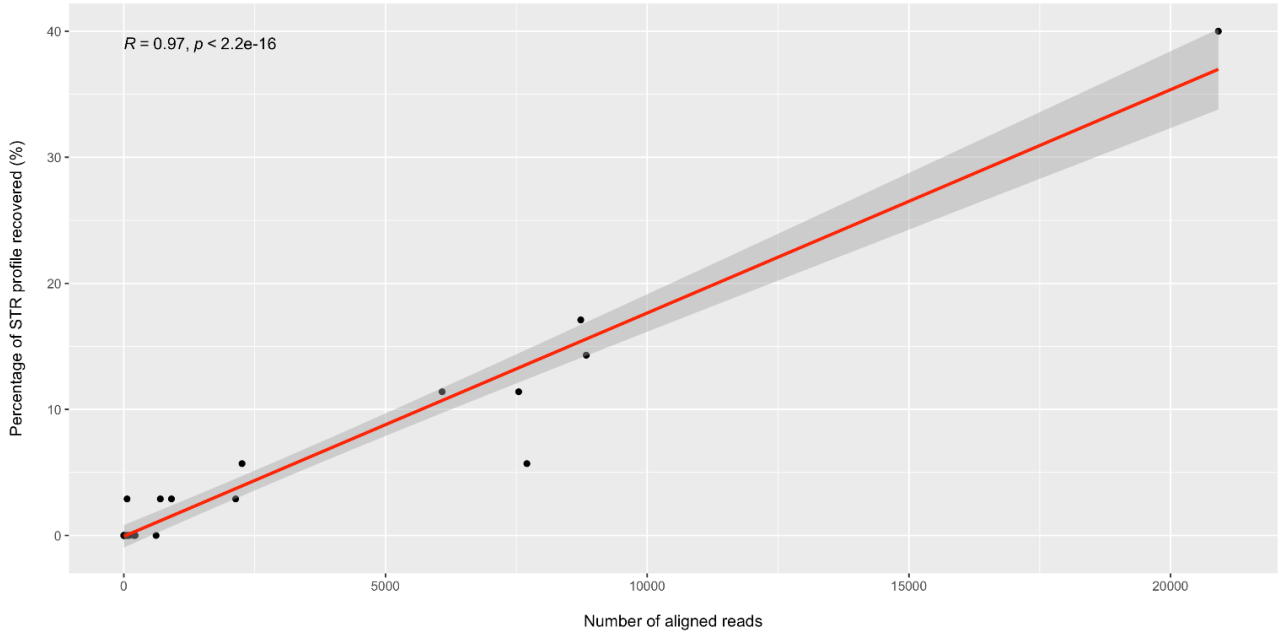

Figure 4.4 Scatter plot showing the relationship between the number of aligned sequencing reads and the percentage of the ForenSeq™ STR profile that was recovered for Phase Two cartridge samples. The strong positive correlation was significant with a p-value of < 0.05.

It was hypothesised that as the number of aligned reads increased, the percentage of the typed loci that were concordant would increase. A second correlation analysis was performed to determine the relationship between these variables, still excluding Sample 2_3. Due to the presence of outliers, Spearman's rank correlation coefficient was used and showed a weak positive relationship of 0.32 (p-value = 0.18, alpha = 0.05) indicating that there was no significant correlation between the number of aligned reads and percentage of typed loci that were fully concordant. This shows that while more aligned reads were correlated with a higher percentage of a concordant profile, there was no correlation between number of aligned reads and the percentage STR loci typed concordantly. This was evident when looking at individual samples. For example, Sample 7_4 had 6082 aligned reads and typed four STR loci all concordantly. Sample 5_4 then had 7544 aligned reads and typed eight STR loci, however, only four were fully concordant. This indicated that the presence of fully concordant STR loci was extremely random across the partial ForenSeq™ DNA profiles recovered from unfired cartridges.

46.4% of the samples had no STR loci typed. This was expected as these samples had low amounts of aligned reads ranging from zero to 616X (Table 4.14).

4.3.3.2.1 STR loci with allele dropouts

There were 59 STR loci typed that experienced allele dropouts. This was 52.7% of the total STR loci typed in the Phase Two samples. In all instances, one allele was typed concordantly but dropout of the second allele

saw the calling of a homozygous genotype when a heterozygous genotype was expected. The read depths for the typed allele ranged from 33X to 1595X. In comparison, the read depths for the untyped allele ranged from 0X to 104X. Of the 59 loci with allele dropouts, 12 loci saw the typing of reads that aligned to the expected allele, however, these were either low coverage or the UAS identified the typed allele as stutter, resulting in dropout and exclusion from the final genotype. The other 47 instances saw no reads typed for the allele that dropped out. Consequently, the range of ACRs for all typed STR loci with allele dropouts ranged from 0 to 0.22, indicating heterozygous imbalance (159).

The main limitation appears to be the low coverage of any successfully typed loci and how this impacts the analyst's ability to confidently call STR loci accurately. Several instances of allele dropouts were seen, resulting in the incorrect calling of homozygous genotype. As 32 samples were sequenced in each run, running less samples could possibly increase the sequencing coverage for each sample and subsequently increase the read depth of each typed loci, providing more confidence in whether a locus was called correctly as homozygous or not. It is recommended that further research investigates this.

4.3.3.2.2 Non-concordant STR loci

There were 11 non-concordant genotypes typed in five of the unfired cartridge samples, these are described in Table 4.15.

Table 4.15 Non-concordant STR loci observed in Phase Two samples. Red font indicates that allele was not expected at that locus.

Sample	Loci	STR	Expected genotype	Genotype called	Parent allele/s read counts	Read count of allele/s that dropped in
1_2	D13S317	aSTR	12,13	12,12.3,13	82,69	111
1_2	D21S11	aSTR	31.2,31.2	30.2,30.2	-	32
1_2	DXS7423	xSTR	15,16	15,15,16	234,360	307
2_1	FGA	aSTR	23.2,24	18,18	-	32
2_1	D9S1122	aSTR	12,13	11,11	-	77
2_1	TH01	aSTR	7,9	6,9	78	37
2_1	DYS481	ySTR	-	14	-	52
2_2	D20S482	aSTR	13,14	15,15	-	688
5_4	DYS391	ySTR	-	11	-	736
7_1	FGA	aSTR	19,22	21,22	295	74
7_1	DXS7132	xSTR	11,13	12,13	205	48

Firstly, each sample was analysed to try to determine whether contamination had occurred, resulting in the non-concordant genotype. There was only one male participant in this research (Participant Six), therefore, two genotypes were the incorrect calling of a y-STR locus in a female sample. In Sample 5_4, DYS391 was called, and the genotype was not concordant to any possible sources of male contamination in this research:

the APOS (2800M) sample or Participant Sixes reference profile. This indicates that an alternative source of contamination may have occurred. These could have been another researcher that was using the laboratory, non-self-DNA present on the participants hands during deposition, or a possible low-level of background DNA on the cartridge that remain after cleaning. Where Section 4.1 showed that no reads were generated from any background swabs taken from the cartridges, this was only a proportion of the cartridges provided for this research. In Sample 2_1, DYS19 was called as 14 with a read depth of 52X. This matched both Participant Six and the APOS sample genotype at this locus. However, Participant Six did not type any STRs in any of its four Phase Two replicates, providing confidence that cross contamination between these samples during experimental procedures did not occur. As Sample 2_1 saw the typing of three other non-concordant loci, these were compared to all known reference profiles. None of the non-concordant genotypes matched the APOS reference profile, any other participant, or the analysts reference profile. Given the low read counts for four non-concordant loci in Sample 2_1, this also suggests that alternative source of low-level contamination occurred. Similarly, in Sample 2_2, a non-concordant allele with a high read count was typed, this could have also been due to one of the suggested alternative sources of contamination.

In two instances, stutter was seen at the n-1 position. These both occurred in Sample 7_1 and saw the typing of an allele at a read depth smaller than the parent allele, but high enough to pass the stutter threshold and be called. Three non-concordant loci were generated for Sample 1_2. A final genotype of 15,15,16 at DXS7423 was called. Further analysis showed that the 15 allele at this locus experienced a base change at one nucleotide position from a C to an A. The 15 allele with the A base was seen in more reads (307X) compared to the 15 allele with the C base (234X). This C to A nucleotide transversion is consistent with oxidative damage (4). Additionally, sequencing error whereby two T nucleotides were sequenced instead of three was seen in D13S317, resulted in the calling of the 12.3 allele. Finally, the 30.2 allele that dropped in at locus D21S11 barely passed the read count threshold (32X). In comparison, the reference profile did see sequencing reads align to 30.2 however, this was n-1 stutter position and did not have a high enough read count to pass the stutter threshold.

4.3.3.3 iiSNP profiles recovered from unfired cartridges

Each ForenSeq™ iiSNP profile could also be compared to the known corresponding reference iiSNP genotype. This comparison saw the number of iiSNP concordant to the reference profile divided by the total number of ForenSeq™ target iiSNP loci and is summarised in the below equation:

$$iiSNP\ profile\ (\%) = \frac{Number\ of\ concordant\ typed\ iiSNP\ loci}{94}$$

Table 4.16 Summary of ForenSeq™ iiSNP profiling results in Phase Two. Concordant loci were determined through a comparison to a known reference profile. Loci with allele dropout gave a homozygous genotype when a heterozygous genotype was expected.

Sample	Number of typed loci	Concordant loci	Loci with allele dropout	% concordant iiSNP profile
DNA 1_1	-	-	-	-
DNA 1_2	67	45	22	47.9
DNA 1_3	-	-	-	-
DNA 1_4	47	36	11	38.3
DNA 2_1	14	12	2	12.8
DNA 2_2	9	6	3	6.4
DNA 2_3	70	-	-	-
DNA 2_4	3	1	2	1.1
DNA 3_1	-	-	-	-
DNA 3_2	4	4	-	4.3
DNA 3_3	-	-	-	-
DNA 3_4	1	-	1	-
DNA 4_1	-	-	-	-
DNA 4_2	1	-	1	-
DNA 4_3	-	-	-	-
DNA 4_4	-	-	-	-
DNA 5_1	1	-	1	-
DNA 5_2	3	2	1	2.1
DNA 5_3	11	7	4	7.4
DNA 5_4	27	15	12	16.0
DNA 6_1	-	-	-	-
DNA 6_2	-	-	-	-
DNA 6_3	-	-	-	-
DNA 6_4	-	-	-	-
DNA 7_1	19	9	10	9.6
DNA 7_2	-	-	-	-
DNA 7_3	6	1	5	1.1
DNA 7_4	12	5	7	5.3

As seen in Table 4.16, 16 of the 28 unfired cartridge samples (57.1% of samples) saw the typing of 295 iiSNP loci. It was presumed the 295 typed iiSNPs would be those with the shortest sized amplicons and therefore the loci more likely to be recovered from low quantity and quality DNA samples. This assumption was investigated through determining if there was a relationship between amplicon size and overall number of times each iiSNP loci was typed. Spearman's rank correlation coefficient (R) gave a value of -0.29, indicating a weak negative relationship between these two variables: as the iiSNP amplicon size increased, the number of times that locus was typed slightly decreased (Figure 4.5). This result was significant with a *p* value of 0.011.

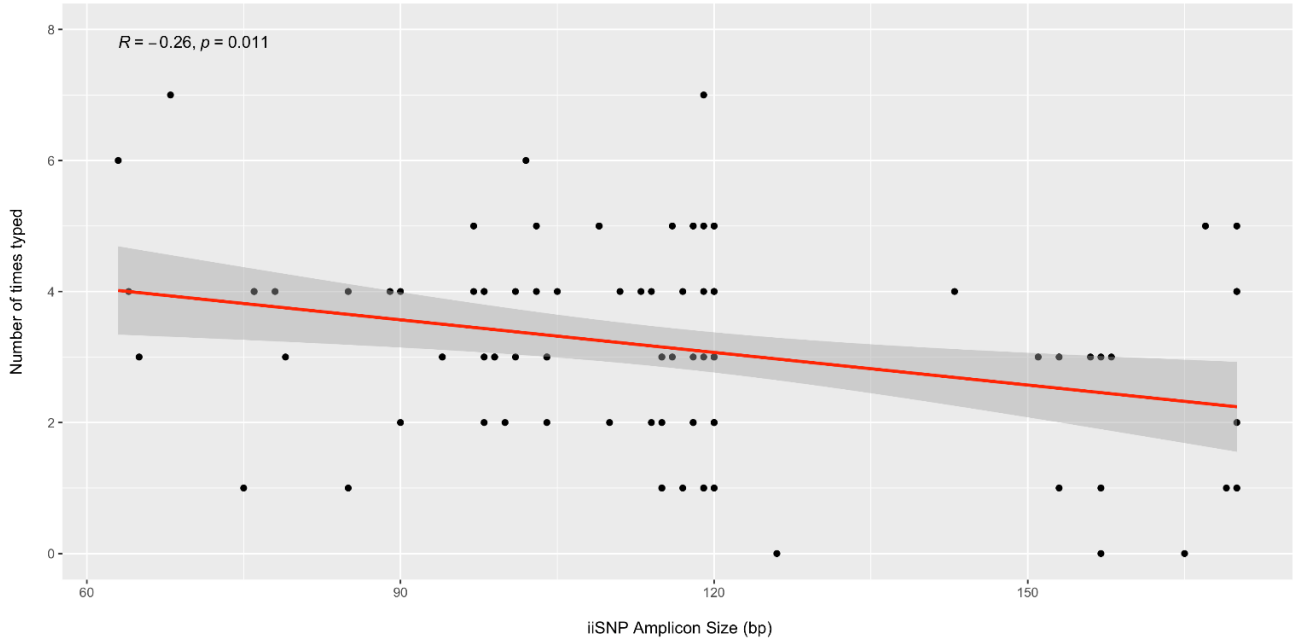

Figure 4.5 Scatter plot showing the relationship between ForenSeq™ iiSNP amplicon size and the number of times that iiSNP was typed in Phase Two cartridge samples.

Of the 295 total iiSNP loci typed, 70 were from Sample 2_3. As it was determined this sample was contaminated (Section 4.3.3.2), it was excluded from further analysis. This meant that 143 iiSNP loci of the 225 typed were concordant to the corresponding reference profile (Table 4.16).

Partial iiSNP profiles were able to be obtained from 12 of these samples. These widely ranged from 1.1% to 47.9% of the 94 iiSNPs concordantly typed. Previous research found that SNP loci are more likely to be typed compared to STR loci in degraded samples (57). In the Phase Two, 28 cartridge samples saw 35 concordant aSTR loci and 143 concordant iiSNPs typed. Where numerically, more iiSNPs were typed, when considering the number of each type of marker present in the ForenSeq™ Kit, both marker types were typed at similar rates. There was a possibility of 28 aSTR loci and 94 iiSNPs able to be typed for each sample. This meant that 35 out of 784 aSTRs were typed (4.5%) and 143 out of 2632 iiSNPs (5.4%) were typed. However, more iiSNP loci were concordant to the corresponding reference profile in comparison to STR loci. Specifically, 63.6% of the total iiSNP loci typed were concordant, where only 32.8% of the total STR loci typed were concordant. This shows that iiSNPs are more likely to be accurately typed in low quantity touch DNA samples.

As Sample 1_2 saw the typing of 45 concordant iiSNP loci, this sample was selected for further analysis. Of the 45 concordant loci, 35 were homozygous and 10 were heterozygous. In all samples, the accurate calling of a homozygous genotype was more frequent than a heterozygous genotype. The read counts of the concordant homozygous alleles in Sample 1_2 ranged from 32X to 745X as displayed in Figure 4.6.

4.3.3.3.1 iiSNP loci with allele dropout

82 loci experienced allele dropouts. As seen in Table 4.17, it appeared that the T allele was the most prone to dropping out. When accounting for the number of times each allele was expected to be typed, all alleles had very similar rate of dropping out.

Table 4.17 Frequency of how often each nucleotide base dropped out at an iiSNP locus in Phase Two.

Allele	Number of times dropped out	Proportion in total number of allele dropouts (%)	How many times allele was expected to be typed	Rate of dropping out (%)
A	20	24.4	40	50.0
C	18	22.0	39	46.2
G	19	23.2	37	51.4
T	25	30.5	48	52.1

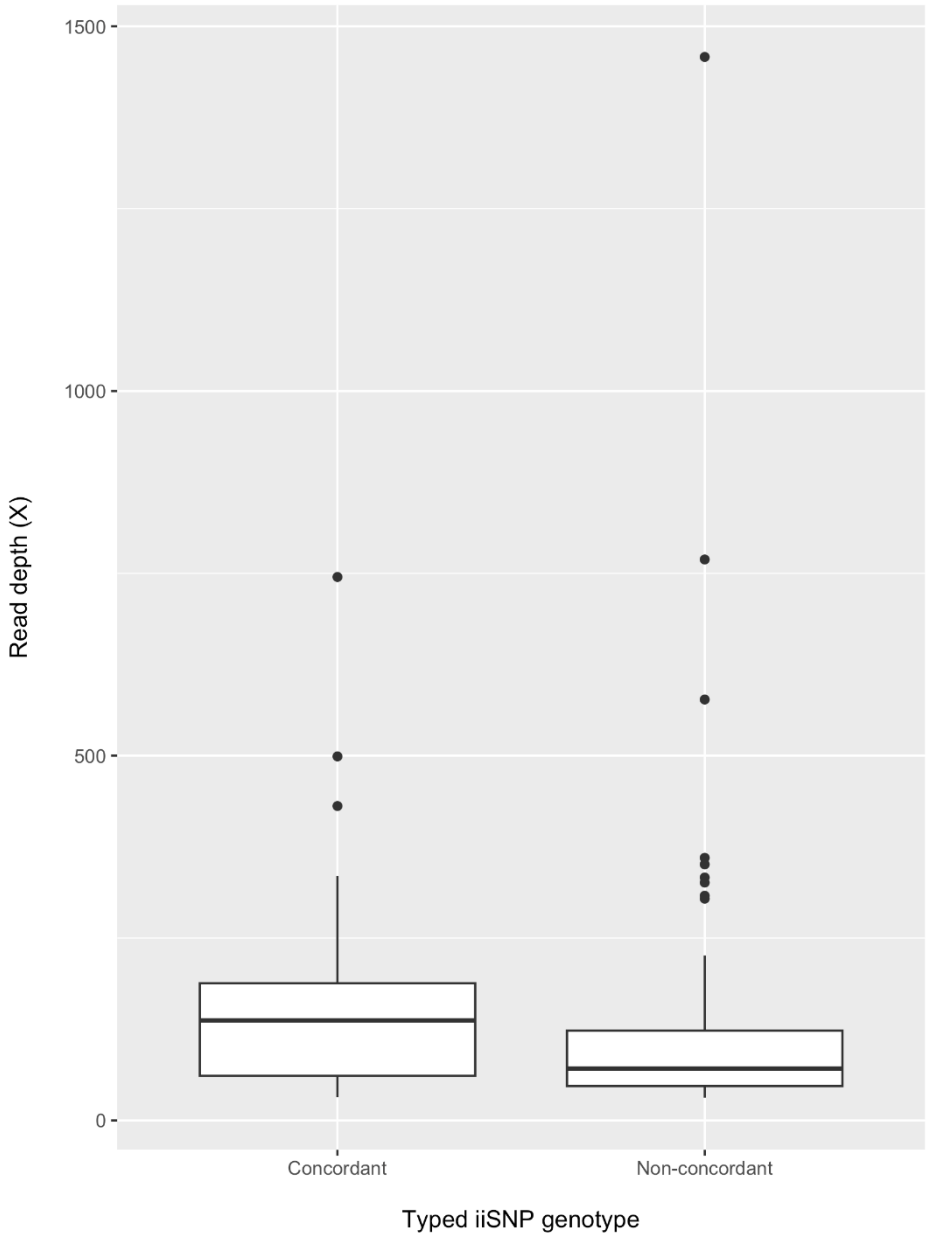

Figure 4.6 Range of read depths (X) for iiSNP alleles typed as homozygous in Phase Two. Concordant indicates that the homozygous genotype for Sample 1_2 was expected in comparison to Participant Ones reference profile. Non-concordant indicates that a homozygous genotype was called due to allele dropout when a heterozygous genotype was expected.

As seen in Figure 4.6, there was a large range of read depths for recorded for iiSNP loci called as homozygous when a heterozygous genotype was expected. The minimum and maximum read depths observed were 31X and 1458X respectively. When compared to the read depths of typed iiSNP loci concordantly typed as being homozygous, these results show that it is near impossible to distinguish between a true homozygous and a homozygous genotype called due to allele dropout as the range of read depth overlap.

Where a majority of the untyped alleles had a read depth of 0X, there were six instances where sequencing reads did align in the untyped alleles, with the maximum read depth being 29X. This was below the allele read

depth threshold of 30X used for this research. It should be noted that a 30X threshold was set on the conservative side, given that it was expected that the nDNA would be of low quality and quantity, with the rate of allele drop-ins initially unknown. As no iiSNP allele drop-ins were seen for any of the touch DNA profiles recovered from unfired cartridges, there is the opportunity to lower this minimum read threshold for this evidence type. One study investigating mock forensic case-type samples set their minimum read threshold to 10X and did not see the typing of any iiSNP allele drop-ins (55). Where this would see the typing of more concordant alleles, the inability to distinguish between true homozygous calls and allele dropouts would still remain a limitation.

There were some iiSNP loci that were more prone to allele dropouts. In this research rs2040411, rs914165, rs763869 and rs876724 had alleles dropout in three samples. rs987640, rs10776839 had alleles dropout in four samples.

4.3.4 Findings

The application of the ForenSeq™ sequencing workflow to touch DNA recovered from an unfired .223 Rem cartridge was able to type 225 iiSNP and 112 STR loci. This allowed for the comparison of these typed loci to a known reference profile. 63.6% and 37.5% of the typed iiSNP and STR loci were fully concordant to the expected reference genotype. 36.4% and 52.7% of the typed iiSNP and STR loci called one allele correctly and had the second allele dropout due to low coverage. These results show that partial STR and iiSNP profiles were able to be successfully obtained from this evidence type.

The main limitations with the application of the ForenSeq™ sequencing workflow to low-quantity samples appears to be the limited amounts of template nDNA. This led to a large presence of adapter-dimers, and the low coverage of typed target loci. A high proportion of adapter-dimers relative to the target library will see a majority of the sequencing reads align to adapter sequence (158). Decreasing the amount of adapter-dimers could improve the sequencing coverage of the target loci. This could be achieved through implementing a second purification after PCR2 (158). There is the risk that an additional purification step could decrease the already low amount of target DNA libraries, therefore, further research would have to trial these suggestions.

Sharma et al. (55) tested the ForenSeq™ kit on mock casework samples and found that the sample with the lowest DNA input (600 pg) had the most allele dropouts. Separately, Guo et al. (71) found that a 100 pg DNA input could produce full STR profiles but iiSNP profiles experienced allele dropouts. Substantial dropouts of both markers were then seen with a DNA input of 20 to 10 pg input DNA, with at a maximum of 64% and 44% of a STR and iiSNP profile able to be detected (71). In the present research, none of the samples reached the target DNA input of 0.2 ng/µL (or 1 ng). The largest quantity of DNA recovered was in Sample 1_2 (Table 4.14)

with 0.00292 ng/µL (or 14.6 pg). This sample had the most concordant STR and iiSNP loci typed, but also experienced the highest number of allele dropouts. While more iiSNPs concordant to the corresponding reference profile were typed compared to STRs, both marker types saw allele dropouts resulting in the calling of a homozygous genotype when a heterozygous genotype was expected. Without knowing the true source it would be impossible to discern between a true homozygous and a homozygous called due to allele dropout as the read depths of the typed allele in both instances are similar. While allele dropouts are not as ideal as a fully concordant genotype, they were expected given the low-quantity of input DNA. Importantly, the allele called was concordant to the reference profile. One allele recovered in the context of firearm evidence would not result in an exclusion to a reference profile and can instead still offer useful information to forensic practitioners. This indicates that for low input DNA samples homozygous genotypes should be considered as a possible partial genotype.

Of more of a concern was the presence of non-concordant loci. While no non-concordant iiSNP loci were observed, 11 instances of non-concordant STR loci were typed. These were a combination of low-level contamination, incorrectly typed stutter, potential DNA template damage and sequencing error. These were able to be identified due to comparison to a known reference profile, however, this will not be possible in forensic casework where a sample comes from an unknown source. Further research should trial optimising Stutter Thresholds for low-quantity DNA samples or using replicates at the amplification level to try to resolve the number of non-concordant loci.

5 FIRED CARTRIDGE CASE SAMPLES

5.1 PHASE THREE: OPTIMISATION OF MPS WORKFLOWS

As sequencing reads that aligned to either the rCRS or ForenSeq™ targets could be generated from touched cartridges using previously developed and evaluated sequencing workflows (72,97), optimisations were carried out to determine the final sequencing workflows to be applied to fired cartridge cases. Two optimisations were trialled with the aim of improving the quality of sequencing data generated through the mtDNA sequencing workflow, as it showed greater profiling potential for low quality DNA samples. Three factors were used to monitor sequencing data quality; the number of sequencing reads generated that aligned to the rCRS, the amount of mitogenome that was called, and the concordance of any generated haplotype to the corresponding reference haplotype.

5.1.1 Optimisation One: application of the 'Full' mtDNA method

The Precision ID user guide (128) provides a 2-in-1 and conservative methodology for the PCR amplification and library preparation of mtDNA samples. The 2-in-1 method sees the amplification of each primer pool separately, before being pooled together in equal volumes prior to library preparation in a new well. This method is recommended for low copy number samples (128). This is similar to the custom amplification method used in this research (Section 2.7.1.1), which instead pooled the two primer pools together in equal concentrations. The conservative method sees amplification of each pool in a half-volume reaction and then the transfer of one pool into the other, before library preparation. This method is recommended for non-degraded samples (128), and was not practical for the unfired cartridge and fired cartridge case samples used in this research.

In an evaluation of the Precision ID whole mtDNA genome panel by Strobl et al. (121), a third PCR method was trialled for the amplification and library preparation of mtDNA samples. This method, the "Full" method, sees the separate amplification of each primer pool, with both amplified pools then processed independently until sequencing. They found that the full method gave the highest number of total sequencing reads and coverage, compared to the 2-in-1 and conservative methods, however used twice the amount of reagents making it less cost-effective and consequently was restricted to samples with extremely low quantities of mtDNA. Comparably, the ForenSeq™ mtDNA Whole Genome Kit utilises a similar method for amplification and library preparation (100).

As low coverage was seen in various regions of the mitogenome, for the cartridge samples from Phase Two, a variation of the "Full" method (Section 5.1.1.1) was trialled to determine if the read depth of these areas could be increased to provide full mitogenome coverage for more samples.

5.1.1.1 Methodology

Two Phase Two extracts from each participant, were taken through the following "Full" method workflow (Figure 5.1). These extracts were selected so that they were the same as the samples selected to trial an additional purification on (Section 5.1.2.1).

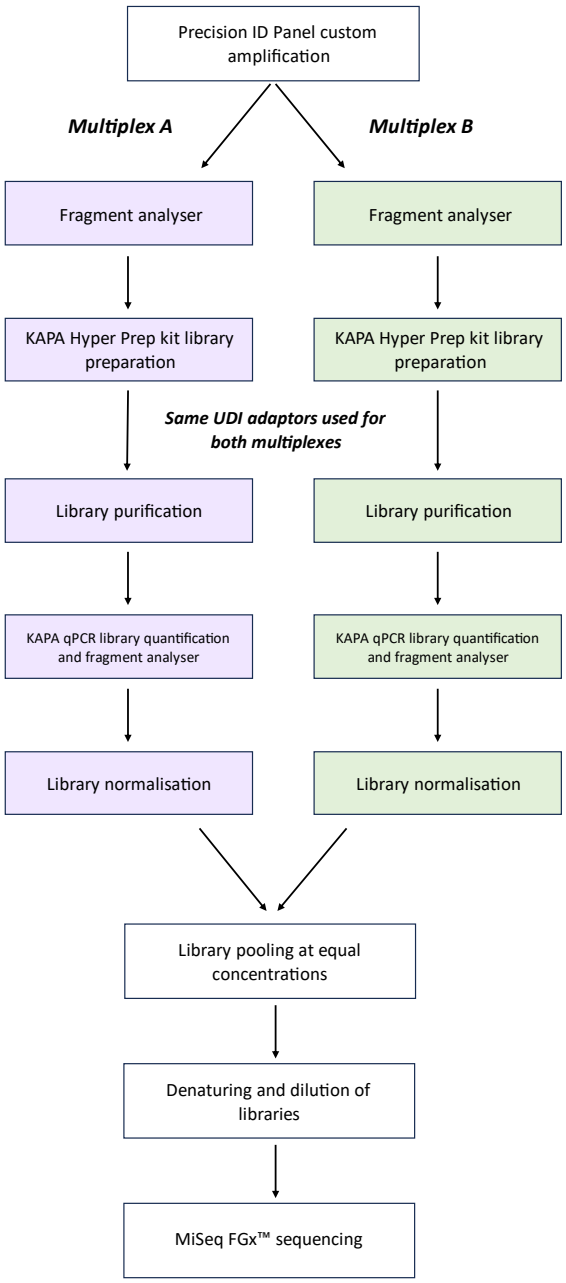

Figure 5.1 The 'Full' method mtDNA sequencing workflow.

All samples were amplified by Precision ID primer pools in two multiplex reactions (Section 2.7.1.1), giving two amplification products for each sample. These amplicons were not pooled together. Instead, the total concentration of each amplicon pool was obtained using the Fragment Analyzer, as described in Section 2.7.2.1 and this concentration was used to calculate the volume of each pool required for input into library preparation.

Library Preparation was carried out following the method in Section 2.7.3. As each sample had been amplified twice, the same KAPA UDI adapter was added to both amplified versions of the same sample. Libraries underwent preparation for normalisation as described in Section 2.7.4 and then 2.5 µL of each library was pooled together to create a Pooled Normalised Library (PNL). Since the same adapters were used for both amplification products, each sample had a final volume of 5 µL in the PNL. The PNL was denatured and diluted to give a final denatured normalised library (DNL), that was sequenced on the MiSeq FGx™ platform following the methods described in Section 2.7.6 and 2.7.7.

5.1.2 Optimisation Two: additional sample purification

A common limitation seen in Phase Two was the high presence of adapter-dimers in both mtDNA and DNA libraries. Following targeted amplification, any unused primers present in a sample can have adapter indices bind during adapter ligation/PCR2, forming an adapter-dimer. Adapter-dimers are non-ideal as they are known to impact qPCR normalisation, especially in low input samples, through overestimating the total library amount and underestimating DNA fragments containing target amplicons (160). Adapter-dimers can also impact sequencing run performance; they have a shorter fragment size compared to target DNA library fragments, leading to a higher bridge amplification efficiency, and interference with cluster generation (73,158).

Low quality samples can also increase the formation of adapter-dimers (158). As there is less target DNA template, less primers are used during targeted amplification. This increases the presence of unused primers and facilitates the formation of adapter-dimers. More adapter-dimers within a sample increases the chance that some will remain after a single purification step. Previous research has shown that a second repetition of the purification step at the end of library preparation decreases the number of adapter-dimers present in challenging forensic casework samples (73). Similarly, a second purification is strongly recommended in the ForenSeq™ mtDNA Whole Genome kit for samples with low levels of genomic DNA (100). A second purification step was trialled to determine if the amount of adapter-dimers from cartridge samples could be decreased and whether this would improve sequencing performance.

5.1.2.1 Methodology

The 'low coverage report' file generated by GM-HTS was manually checked for each mtDNA sample from Phase Two. Two out of four samples from each participant were selected to trial a second purification on. These samples were chosen based on having the lowest total number of dropouts, with dropouts spanning the shortest nucleotide length in the mitogenome. All samples from Phase Two were in the same library preparation batch and stored in a -20 °C freezer following library normalisation (Section 2.7.5). The samples selected for a second purification, were transferred to a new 96 well plate and the library purification process outlined in Section 2.7.3.3 was repeated. After this second purification, these samples were taken through the rest of the mtDNA sequencing workflow (Section 2.7.4 to Section 2.7.9).

5.1.2.2 mtDNA library quality

In Figure 5.2A, large peaks are seen at 44, and 59 bp, indicating the presence of unbound primers following targeted mtDNA amplification. Similarly, peaks can be seen around 200 to 260 bp indicating the presence of adapter-dimers. These have all remained following a single purification step at the end of library preparation. The addition of an extra library purification step resulted in a massive decrease in the number of unbound primers and primer-dimers, as seen by the minimal peaks from 1 to 200 bp in Figure 5.2B. The presence of adapter-dimers also decreased in this sample, as the average length of fragments from 200 to 700 bp increased from 296 bp to 330 bp. However, the use of an additional purification saw a decrease in the concentration of successfully converted mtDNA libraries from 0.2620 ng/µL to 0.0783 ng/µL. This was a similar trend seen across all repurified samples. To determine whether this decrease in the amount of adapter-dimers or mtDNA library concentration would influence mtDNA profiling, these samples were sequenced.

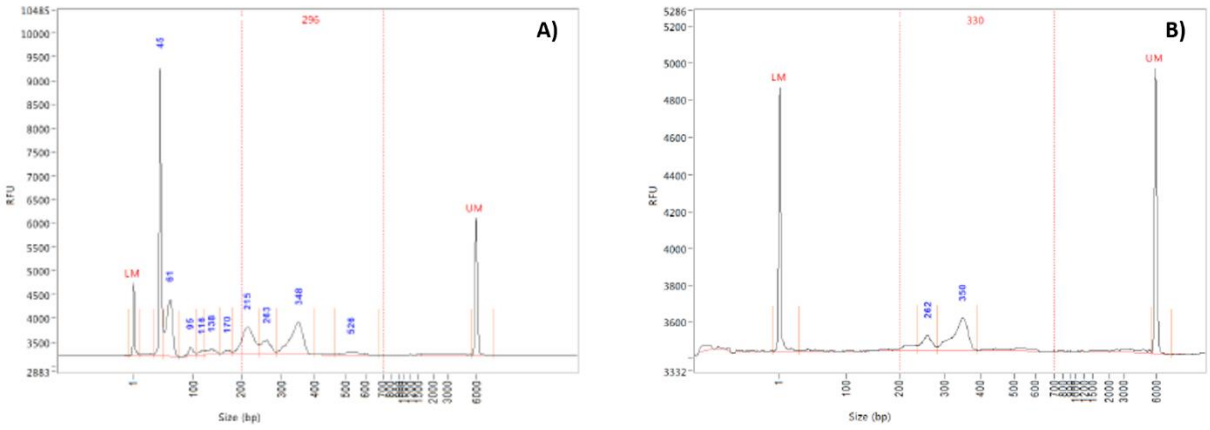

Figure 5.2 Fragment Analyzer EPGs for mtDNA Sample 3_3 following library purification. A) Single purification step. B) Additional purification step. Red lines from 200 bp to 700 bp indicate DNA fragments that will be sequenced by the MiSeq FGx™ platform.

5.1.3 Overall findings

The Phase Two samples that underwent the above optimisations were sequenced over two sequencing runs: SeqM003 and SeqM004. As the same samples for each participant underwent both optimisations, a direct comparison could be made with the results originally obtained for these samples in Phase Two (SeqM002).

5.1.3.1 Sequencing performance

As mentioned in Section 3.2.2, a second replicate of participant reference samples was included in both SeqM003 and SeqM004. After the completion of SeqM003, it became apparent that running 31 mtDNA samples in the same sequencing run was not optimal as only partial haplotypes were obtained for reference buccal samples that had previously generated full haplotype profiles in a sequencing run with only 10 samples. The presence of lower quality and quantity mtDNA samples, e.g., touch DNA on firearm evidence, reduces the number of samples that should be sequenced in one run, regardless of whether higher quality samples are also present. The optimal number of lower quality samples to be included in one sequencing run was not able to be determined due to both time and resource constraints; future research should prioritise investigating this. As SeqM003 was run before SeqM004, it was decided to continue to use a batch of 32 samples in the SeqM004 sequencing run. This would allow for the success of the trialled optimisations to be determined, rather than being confounded by any improvements resulting from running less samples in one sequencing run. Consequently, for a direct comparison between Phase Two samples, repurified samples, and 'full' method samples, over three sequencing runs could be made. A summary of these sequencing runs is provided in Table 5.1.

Table 5.1 Summary of sequencing runs containing Phase Two samples that underwent different optimisations including quality metrics obtained from the UAS.

Sequencing run ID	Total number of samples	Sample Types	Cluster density (K/mm2)	Clusters passing filter (%)	Q Score ≥ 30 all cycles (%)
SeqM002	33	Controls, Phase Two samples	762	91.6	44.8
SeqM003	31	Controls, Reference samples, Phase Three samples ('Full method')	891	91.8	46.1
SeqM004	32	Controls, Reference samples, Phase Three samples ('Repurified')	361	97.9	59.6

The extremely low cluster density seen in SeqM004 was likely due to the repurified samples, as a second purification decreased the amount of adapter-dimers present within these samples. This was noticeable during library quantification, as each sample gave a decreased diluted concentration. On average the average diluted concentration for the 14 Phase Two samples decreased from 2.635 ± 0.449 pM in SeqM002 to 1.208 ± 0.416 pM in SeqM004. In contrast, the average fragment length for these samples increased from 274 ± 9 bp in SeqM002 to 304 ± 6 bp in SeqM004. The high-quality reference samples in Phase One gave an average fragment length of 345 ± 13 bp and went on to generate accurate haplotypes with full mitogenome coverage, indicating that a library with an average fragment length in this range will provide optimal sequencing results. An increase in the average fragment length of SeqM004 samples illustrated that by removing adapter-dimers, a more purified and higher quality mtDNA library could be obtained.

As a result of the decreased adapter-dimers concentration, most of the repurified samples did not have a concentration high enough to require a dilution to 4 nM during library normalisation and were instead added at maximum volume into the PNL. Consequently, the molarity of the final PNL was 1.6 nM, quite far from the target PNL molarity of 4 nM. The rest of the sequencing workflow was carried out without any changes, regardless of this low molarity, with the sequencing run still able to be completed successfully. The extremely low cluster density was a direct consequence of the low quantity of mtDNA within these samples.

SeqM004 had the highest percentage of bases with a Q Score ≥ 30. As adapter-dimers can impact sequencing data quality (158), it was unsurprising that a second purification improved the Q Score value. In comparison to SeqM002, SeqM003 gave a higher run quality metric values (Table 5.1). At this stage, it could not be determined if this was due to the 'Full' amplification samples, or the presence of the reference samples included within this sequencing run, so both sequencing and mitogenome coverage were further analysed.

5.1.3.2 Sequencing coverage

The number of total and aligned reads generated for each sample was investigated to determine if either optimisation could efficiently increase the amount of sequencing reads that aligned to the mitogenome (Table 5.2). Compared to SeqM002, both SeqM003 and SeqM004 had a lower average of total sequencing reads. SeqM004 had the lowest average total reads, which is likely due to the extremely low cluster density recorded for this sequencing run.

Table 5.2 Average amount of sequencing reads generated for 14 Phase Two samples. The same samples did not go through any optimisations (SeqM002), were amplified using the 'Full' method (SeqM003) and underwent a second library purification (SeqM004).

Sequencing Run	Total Reads	Aligned Reads	Aligned Reads (%)
SeqM002	1152046 ± 105044	199670 ± 54313	19 ± 5
SeqM003	839579 ± 114596	197140 ± 70146	21 ± 6
SeqM004	391815 ± 65485	190145 ± 46285	48 ± 9

Where both optimisations increased the average percentage of sequencing reads that aligned to the rCRS, SeqM004 showed the greatest improvement (Table 5.2). A one-way ANOVA was used to determine if there was a statistical difference between the three different sequencing runs and the mean percentage of sequencing reads that aligned to the mitogenome. It was found that there was a significant difference between the mean percentage of aligned sequencing reads (p = 0.00724, alpha = 0.05). A post hoc Tukey HSD test (alpha 0.05) showed that there was no significant difference between SeqM003 and SeqM002 (p = 0.9787) indicating that the 'Full' method did not improve the amount of aligned sequencing reads. It was found that there was a significant difference between SeqM004 and SeqM002 (p = 0.01293) and also SeqM004 and SeqM003 (p = 0.02130). Based on this, it can be concluded that a second purification (SeqM004) increased the percentage of aligned reads, through the removal of adapter-dimers. The full distribution of data for each sequencing run can be visualised in Figure 5.3.

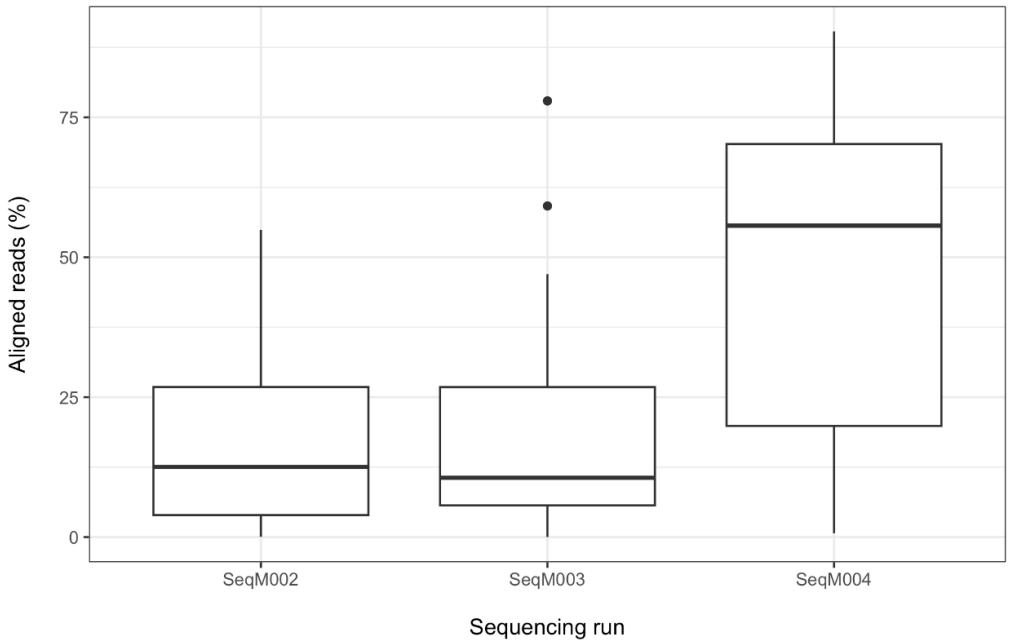

Figure 5.3 The percentage of aligned sequencing reads for the same Phase Two samples over three different sequencing runs. These Phase Two samples did not go through any optimisations (SeqM002), were amplified using the 'Full' method (SeqM003) and underwent a second library purification (SeqM004).

5.1.3.3 Mitogenome coverage

As it was previously determined that no Phase Two sample was able to provide full mitogenome coverage (Table 4.9), SeqM003 and SeqM004 were further investigated to determine if either optimisation could increase the percentage of the mitogenome that was typed (Figure 5.4). In this research, 100X coverage was used at each nucleotide position, for typing to occur.

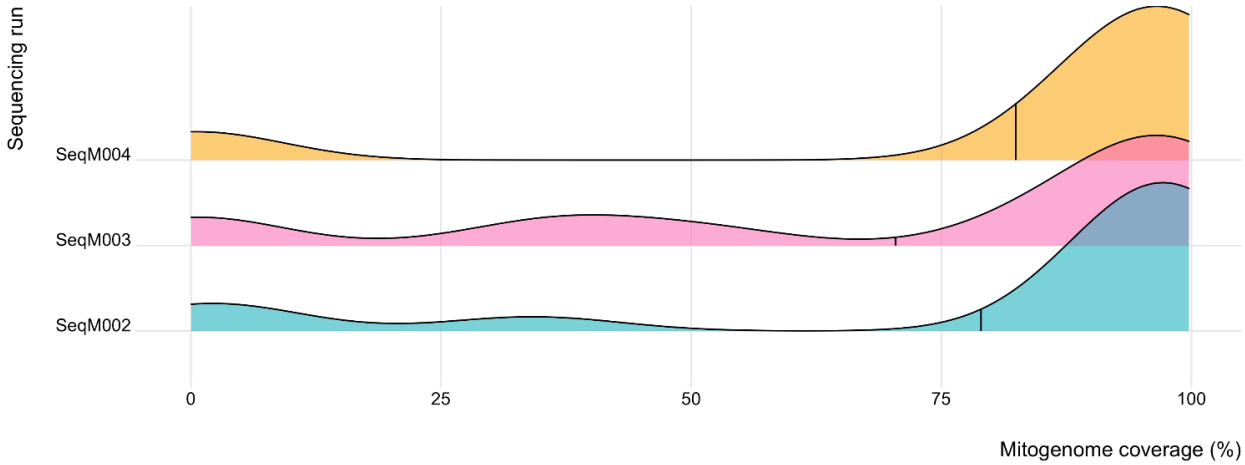

Figure 5.4 Ridgeline graph showing how much of the whole mitogenome was recovered for the same fourteen Phase Two samples under different conditions. These conditions were no optimisations (SeqM002), amplification using the 'Full' method (SeqM003) and a second library purification (SeqM004). Any nucleotide position that had < 100X read depth was not typed and resulted in dropout. Black vertical line represents the mean.

On average, SeqM002 saw 79.0% (± 9.8) of the whole mitogenome typed for the fourteen samples used to trial these optimisations. In comparison, SeqM003 saw 70.4% (± 10.0) of the mitogenome typed and SeqM002 saw 82.4 % (± 9.3). This indicates that the 'Full' method saw more of the mitogenome dropout and that a second library purification slightly increased the amount of mitogenome typed in comparison to SeqM002. However, a one-way ANOVA showed that the difference between the mean percentage of the mitogenome typed and all three of these optimisations was not statistically significant (p = 0.669, alpha = 0.05, one-way ANOVA).

5.1.4 Control Evaluation

While Phase Three samples were sequenced alongside different types of samples (e.g., reference samples), they were processed independently until sequencing. This means that the 'Full' method and repurified samples had their own sets of controls.

5.1.4.1 Negative controls

Both ENEG controls had less than 100X coverage recorded at each nucleotide position in the mitogenome, resulting in entire mitogenome dropout (Table 5.3). In comparison, both ANEG controls had more than 100X coverage at particular regions in the mitogenome. This indicated that there were contaminating reads at a read depth high enough to impact haplotype generation for the cartridge samples in both SeqM003 and SeqM004. Where the ANEG sample in SeqM004 saw the calling of no variants, the ANEG sample in SeqM003 saw the calling of the major variant 417A with 126X coverage. An investigation into possible sources of contamination showed that 417A was not recorded in any of the participants, APOS control or the analysts' reference haplotypes. Furthermore, this variant was not called in any of the samples processed alongside this control. As low-level contamination in this ANEG sample resulted in the calling of a major variant, each SeqM003 sample was investigated to determine if any incorrect major variants were called. No major variants were called in any of the cartridge samples that were not expected in the corresponding reference haplotype. Any minor variants typed that were not expected in the reference haplotype were then able to be identified and removed following the process outlined in Section 4.2.4. The opportunity for low-level contamination to occur was exacerbated in the "Full" method, as each sample was processed in two individual multiplexes until library pooling. This meant that each sample experienced double the amount of physical handling, increasing the chance of contamination. The 417A variant observed in this ANEG sample was low-level contamination likely introduced by low-levels of background DNA present during laboratory processing, emphasising the sensitivity of this workflow and the importance of strict anti-contamination measures for mtDNA analysis.

Table 5.3 Sequencing reads seen in the negative controls of Phase Three samples. Red font indicates that the maximum number of reads passed the 100X read depth threshold. The percentage of mitogenome with read alignment was calculated by the number of nucleotide positions that had reads align divided by the length of the rCRS.

Sequencing run	Control	Reads aligned to rCRS (%)	Average read depth across mitogenome	Maximum number of reads	Percentage of mitogenome with read alignment (%)
SeqM003	ENEG	0.001	0.3X (± 0.1)	8	4.6
	ANEG	0.030	18.0X (± 2.1)	464	14.2
SeqM004	ENEG	0.031	1.7X (± 0.6)	50	6.2
	ANEG	0.030	5.6X (± 0.5)	126	16.0

5.1.4.2 Positive controls

As seen in Table 5.4, both APOS samples processed alongside the optimised samples were able to generate full concordant haplotypes. This was an improvement from the APOS control sequenced in Phase Two, which experienced several regions of unresolvable dropouts and was unable to provide a full haplotype (Section 4.2.3.1.1). Where SeqM004 was unable to provide full mitogenome coverage, this was only due to the dropout of a single nucleotide position; nucleotide position 120.

Table 5.4 APOS control samples processed alongside Phase Three samples, in two sequencing runs; SeqM003 and SeqM004. SeqM003 samples underwent the "Full" amplification method and SeqM004 samples underwent a second purification.

Sequencing run	Haplotype	Missing variants	Concordance to haplotype (%)	Mitogenome coverage (%)
SeqM003	93G 195C 214G 263G 309.1C 309.2C 315.1C 750G 1438G 4135C 4769G 7645C **7861Y** 8448C 8860G 9315C 13572C 13759A 15326G 16311C 16519C	NA	100.00	100.00
SeqM004	93G 195C 214G 263G 309.1C 309.2C 315.1C 750G 1438G 4135C 4769G 7645C **7861Y** 8448C 8860G 9315C 13572C 13759A 15326G 16311C 16519C	NA	100.00	99.99

As the same APOS sample was sequenced in both SeqM002 and SeqM004, with the only differences being the number of library purifications that this APOS sample underwent and the types of samples it was sequenced alongside, this shows that the quality of samples within a sequencing run impacts sequencing performance.

5.1.5 mtDNA haplotype generation

The GM-HTS generated haplotype for each sample was analysed to determine how much of the reference haplotype was able to be obtained. 13 of the 14 samples were able to produce either a partial or full haplotype in all three sequencing runs (Table 5.5). While the difference between the means of each sequencing run was not significant (p = 0.678, alpha = 0.05, one-way ANOVA), on average the addition of a second purification recovered a higher proportion of variants concordant to the corresponding reference profile.

Table 5.5 Percentage (%) of corresponding reference haplotype obtained for fourteen Phase Two samples. These same samples did not go through any optimisations (SeqM002), were amplified using the 'Full' method (SeqM003) and underwent a second library purification (SeqM004).

Sample	SeqM002	SeqM003	SeqM004
1_2	86.8	23.7	86.8
1_3	89.5	78.9	89.5
2_1	100.0	100.0	100.0
2_4	100.0	93.8	100.0
3_2	63.6	27.7	63.6
3_3	100.0	100.0	100.0
4_1	2.9	5.8	2.9
4_2	87.5	50.0	85.3
5_2	100.0	96.9	100.0
5_4	56.3	68.8	65.6
6_1	-	-	-
6_2	39.4	72.7	81.8
7_2	100.0	100.0	100.0
7_3	93.1	86.2	89.7
Average	**72.8 ± 9.5**	**64.6 ± 9.7**	**76.1 ± 9.0**

Only Sample 6_1 was unable to generate a haplotype in any of the sequencing runs. In both SeqM002 and SeqM003 the maximum coverage recorded at any nucleotide position in the mitogenome was 88X and 22X respectively, consequently resulting in the entire mitogenome dropping out. In SeqM004, Sample 6_1 saw a slightly higher coverage, with one region of the mitogenome able to be called (142X coverage). As no variants were seen for this participant in that particular region, a haplotype was still unable to be generated for this sample.

5.1.6 Conclusions

It was determined that a second purification would be implemented into the final mtDNA sequencing workflow for the fired cartridge cases. The addition of a second purification successfully removed unbound primers, primer-dimers, adapter-dimers, overall improving the percentage of sequencing reads that aligned to the rCRS. Where the concentration of successfully converted libraries did slightly decrease, this was not large enough to negatively impact mitogenome coverage, with no significant difference recorded between all three sequencing runs.

It was determined that the 'Full' method would not be implemented into the final sequencing workflow. Where this method showed no difference in the percentage of sequencing reads that aligned to the rCRS in comparison to Phase Two samples, it required double the amount of reagents and time for experimental processing. This made the 'Full' method not only more expensive to perform, but also increased opportunities for the analyst to make experimental errors and introduce contamination.

5.2 PHASE FOUR: MTDNA HAPLOTYPE GENERATION FROM FIRED CARTRIDGE CASES

The fourth phase of experiments firstly saw the application of the optimised mtDNA sequencing workflow to any touch DNA recovered from .223 Rem cartridges cases after being fired from a bolt-action rifle. The purpose of this experiment was to determine if this sequencing workflow was able to obtain mtDNA haplotypes from the most common type of evidence found at firearm-related crime scenes, a fired cartridge case. It is well-known that cartridges are subjected to a stressful environment, with high temperatures and pressures during the firing process, exposing any touch DNA to the highest amount of possible degradation.

5.2.1 Laboratory workflow

Seven participants handled four .223 Rem cartridges each. These touched cartridges were fired from a bolt-action rifle and then collected as described in Section 2.2.2.2. Sampling for each fired cartridge case was performed following the method in Section 2.4.2. Three separate extractions (Batches ExtF001, ExtF002, and ExtF003) were undertaken to extract all gDNA from the fired cartridge cases using the method outlined in Section 2.5. All samples then underwent quantification using Quantifiler™ Trio following the method in Section 2.6.

Samples were amplified using the custom Precision ID Panel method outlined in Section 2.7.1.1. All fired samples and negative controls were added at the maximum input volume of 7.5 µL.

A dilution of 1 in 10 was performed for all fired samples and negative controls, and a 1 in 100 dilution was performed on both positive controls prior to quantification on the Fragment Analyzer. Multiplex A and B were pooled together following the method stated in Section 2.7.2.2. Library preparation was carried out as described in Section 2.7.3. A second library purification was performed by repeating the process outlined in Section 2.7.3.3.

Library normalisation was carried out as described in Section 2.7.4 and Section 2.7.5. After library pooling, the 4 nM PNL was denatured and diluted to a final loading concentration of 10 pM. This saw 300 µL of the 20 pM DNL, 30 µL of 20 pM PhiX control v3 Library (Illumina, Inc), and 270 µL of HT1 combined into a new 1.5 mL Lo-bind tube (Eppendorf) to create 600 µL of a 10 pM DNL. The rest of the denature and diluting stage was performed as stated in Section 2.7.6. The DNL was then sequenced on the MiSeq FGx™ using a MiSeq Reagent Kit v3, as described in Section 2.7.7. The final sequencing run contained 33 samples: 28 fired cartridge case samples, three ENEGs, one APOS and one ANEG sample.

5.2.2 Results

5.2.2.1 Quantification of fired cartridge samples

Following extraction, DNA quantification provided an estimation of the quantity of gDNA recovered from each fired cartridge case (Table 5.6).

Table 5.6 Quantifiler™ Trio small autosomal marker genomic concentrations of touch DNA recovered from fired cartridge cases. UD stands for undetermined quantity. The degradation index is a ratio of the small autosomal target concentration to the long autosomal target concentration.

Participant	Cartridge ID	Quantifiler™ Trio concentration (ng/µL)	Degradation Index	IPC C_T
1	Fired 1_1	UD	-	28.12991
	Fired 1_2	UD	-	28.14551
	Fired 1_3	UD	-	28.25577
	Fired 1_4	0.00072	9.49237	28.22701
2	Fired 2_1	UD	-	28.12765
	Fired 2_2	0.00032	-	28.03568
	Fired 2_3	0.00184	1.30234	27.83387
	Fired 2_4	UD	-	27.87398
3	Fired 3_1	0.00032	-	27.89229
	Fired 3_2	0.00102	3.29154	27.87309
	Fired 3_3	0.00018	-	27.99703
	Fired 3_4	UD	-	28.07313
4	Fired 4_1	UD	-	28.15439
	Fired 4_2	UD	-	27.73405
	Fired 4_3	UD	-	27.74101
	Fired 4_4	UD	-	28.30382
5	Fired 5_1	UD	-	27.73153
	Fired 5_2	UD	-	27.83425
	Fired 5_3	UD	-	27.67900
	Fired 5_4	UD	-	27.82962
6	Fired 6_1	UD	-	27.97945
	Fired 6_2	UD	-	27.97939
	Fired 6_3	UD	-	27.96624
	Fired 6_4	UD	-	UD
7	Fired 7_1	UD	-	27.56612
	Fired 7_2	UD	-	27.80432
	Fired 7_3	UD	-	27.91248
	Fired 7_4	0.00071	1.87683	27.94555

The small autosomal Quantifiler™ assay target was detected in only 7 of the 28 fired cartridge extracts. This was fewer samples compared to the unfired cartridge samples in Phase Two, whereby concentrations were detected in 12 samples for both the mtDNA and ForenSeq™ sequencing workflows. As seen in Table 5.6, Sample Fired 2_3 gave the largest concentration with 0.00184 ng/µL. In comparison, the largest concentration obtained for all the unfired cartridge samples (Table 4.2 and Table 4.12) was 0.00563 ng/µL (Sample mtDNA

7_2). These observations suggest that, as expected, there is less recoverable gDNA remaining on cartridge cases post-firing.

In this experimental phase, only four of the 28 fired cartridge samples (14.3%) had both the large autosomal and short autosomal marker detected and were able to provide a degradation index (DI). Of these four degradation indices, one falls between the range of 0 – 1.5, and can be classified as non-degraded, two fall between 1.5 and 4 and can be classified as moderately degraded and one falls between 4 – 10 and can be classified as degraded (154). This sample (Fired 1_4), gave the largest DI in this whole research, further evidence that the firing process degrades any DNA present on a cartridge. IPC C_T values for all samples, excluding Fired 6_4, were between the range of 25.8 to 29.0, indicating these reactions worked as expected. Sample Fired 6_4 had an undetermined IPC C_T value. As this sample was not quantified again, this could have been either due to reaction failure or the presence of PCR inhibitors (127).

As three extraction batches were used, there were three sets of EPOS and ENEG controls also analysed. All of the ENEGs had an undetermined quantity of gDNA, indicating that no contamination occurred during extraction or quantification. All of the EPOS samples had concentrations detected, and no IPC C_T flag was triggered, indicating that all extraction and quantification experiments worked as expected.

5.2.2.2 Amplification of fired cartridge samples

It was determined that a 1 in 10 dilution was required for accurate quantification of the amplified fired cartridge samples using Fragment Analyzer. Analysis of the resulting EPGs showed that there were DNA fragments between 100 and 250 bp for some samples, indicating that amplification of mtDNA had occurred. Importantly, this showed that mtDNA was able to be recovered from fired cartridge samples.

In comparison to the unfired samples (Table 4.4), the average concentration of the target amplicons and the percentage of target amplicon fragments in the total reaction were both lower. Specifically, a Welch two sample t-test showed that there was a significant difference between the unfired samples and fired samples target amplicon concentrations obtained using primer pool A (p-value = 0.005873, alpha = 0.05) and primer pool B (p-value = 0.007748, alpha = 0.05). This further indicates that there is less amplifiable mtDNA present on a cartridge case after it has been fired.

Table 5.7 Average concentrations of Phase Four mtDNA fired cartridge case samples, excluding controls. All concentration values were obtained using the Smear Analysis of all fragments from 100 – 250 bp on ProSize v4.0.2.7.

Precision ID primer reaction	Total Reaction (ng/µL)	Target Amplicon (ng/µL)	Target Amplicon (%)	Min Target Amplicon (ng/µL)	Max Target Amplicon (ng/µL)	Amplicon length (bp)
Multiplex A	24.90 ± 2.17	2.55 ± 1.07	10.20	0.02	28.76	177 ± 3
Multiplex B	20.85 ± 3.00	1.36 ± 0.40	6.50	0.05	8.35	168 ± 1

Compared to the unfired mtDNA amplified samples, Precision ID primer reaction A (multiplex A) gave a slightly higher average concentration of target amplicon fragments. However, as seen with the unfired samples, the actual performance of each primer multiplex was determined to not be significantly different for the fired samples (p-value = 0.3005, alpha = 0.05, Welch two sample t-test). This shows that for low-quality mtDNA samples, both primer pools within the Precision ID panel perform equally as well as one another.

5.2.2.3 Library preparation

A second Fragment Analyzer run was carried out to determine how well adapter ligation and library conversion worked for the Precision ID amplicons. Furthermore, the success of a second purification was also analysed, based on the presence of remaining primer and adapter-dimers (Table 5.8).

Table 5.8 Average lengths of fragment peaks seen in a Fragment Analyzer EPG after the library preparation of the Phase Four samples taken through the mtDNA sequencing workflow.

Fragment type	Fragment lengths (bp)	Average representation in the sample (%)	Min (%)	Max (%)
Unbound primers/ primer dimers	1 – 100	15.9 ± 2.5	0.4	42.9
Unconverted amplicons	120 - 200	11.1 ± 2.4	0.3	56.3
Adapter-dimer	200 - 279	23.8 ± 2.4	5.6	46.8
Library	280 - 700	42.9 ± 5.2	9.4	92.0

Unbound primers and primer-dimers were present in lower frequencies compared to the Phase Two unfired samples (Table 4.5) as a result of a second purification. Where adapter-dimers were present in a higher frequency in comparison to Phase Two, they were only half the concentration of the mtDNA library. This resulted in a more purified sample, where on average a majority of the fragments detected by Fragment Analyzer fell between 280 to 700 bp as shown in Table 5.8. This also indicated that successful library conversion of mtDNA amplicons in fired cartridge case samples had occurred.

5.2.2.4 Sequencing preparation

All fired samples were pooled together after being normalised to a target concentration of 4 nM. During normalisation it became apparent that the concentration of a majority of these samples was lower than 4 nM. Consequently, these samples were added into the PNL directly, forgoing normalisation. As a direct result of this, the PNL had a final concentration of 0.85 nM.

Previously, the lowest PNL concentration obtained in this research was 1.6 nM for the samples run in SeqM004 (Section 5.1.3.1). While SeqM004 was sequenced successfully with enough data obtained for subsequent analysis, an extremely low cluster density of 361 K/mm^2 was recorded indicating that the flow cell was extremely under clustered. To help avoid under clustering, given that an even lower PNL concentration was obtained for the fired samples, the input volume of the 20 pM DNL was increased so that the final DNL concentration was 10 pM instead of 6 pM (as described in Section 5.2.1).

5.2.2.5 Sequencing performance

The following run quality metrics in Table 5.9 were obtained from the UAS at the completion of the sequencing run containing the fired cartridge case samples (SeqM005).

Table 5.9 Phase Four mtDNA UAS sequencing run quality metrics

Sequencing Run ID	Target input concentration (pM)	Cluster density (K/mm2)	Clusters passing filter (%)	Q Score ≥ 30 all cycles (%)
SeqM005	10	538	96.8	64.5

This sequencing run obtained the highest percentage of bases with a Q Score ≥ 30 in comparison to all other mtDNA sequencing runs undertaken in this research. This was likely due to all samples within this run undergoing a second purification, suggesting that a second purification should also be implemented for high mtDNA quantity samples to obtain a more optimal sequencing performance.

The cluster density value of 538 K/mm^2 was below the recommended range of 1200 – 1400 K/mm^2 indicating that the flow cell was under clustered. However, this value was higher than the cluster density obtained for SeqM004 (361 K/mm^2). This showed that while the PNL was far below the target concentration of 4 nM, increasing the final DNL concentration improved the number of clusters generated during sequencing. As a result, enough sequencing data was obtained from the fired cartridge samples for further bioinformatic analysis with GM-HTS.

The total amount of sequencing reads for each mtDNA sample was on average 691650 ± 153078. The number of these reads that then aligned to the mitogenome was on average was 424468 ± 147591 for each sample. Out of the 28 fired samples, only one sample (6_4) generated no aligned reads. Similar to Sample 3_4 in Phase Two (Table 4.2), this sample also gave an undetermined IPC C_T value. This shows that samples failing the IPC C_T are very unlikely to give any information when processed through the mtDNA workflow. Five of the other samples generated less than 2000 aligned reads, ranging from one to 1936 reads.

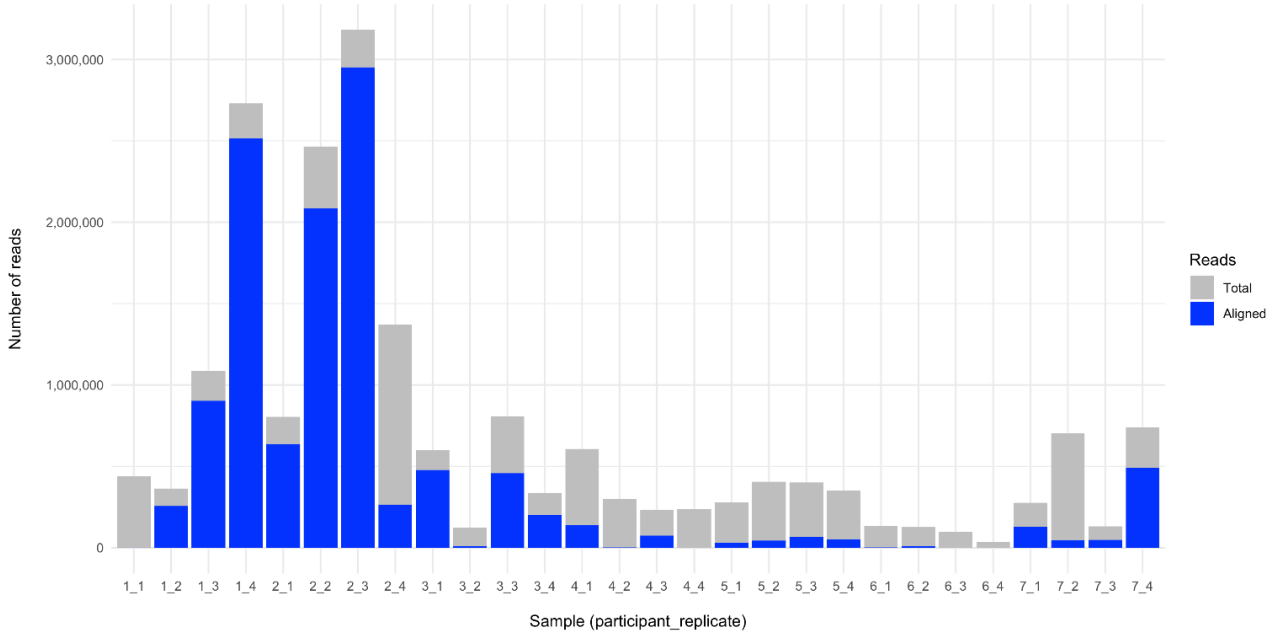

Figure 5.5 Total and aligned sequencing reads generated the fired cartridge case samples taken through the mtDNA sequencing workflow.

As seen in Figure 5.5, the total number of reads varied for each sample indicating that each sample was not equally represented in the PNL. This was expected, given that a majority of the samples were present in concentrations lower than 4 nM and did not require normalisation. Future work could look at lowering the target concentration of the PNL from 4 nM to a concentration as low as 0.85 nM. A lower target concentration will see a portion of the low-quantity samples require normalisation in order to obtain this concentration, creating a more uniform representation of samples within a sequencing run.

The removal of the adapter dimers saw a larger amount of the total sequencing reads align to the rCRS. This was evident comparing the average amount of aligned sequencing reads obtained in all Phase Four samples to all Phase Two samples (SeqM002). On average, the amount of aligned sequencing reads in Phase Four was 35.6% ± 6.4, where in Phase Two this was only 12.8% ± 2.9.

As discussed in Section 4.2.2.4, there are three metrics that can help indicate why some of the generated sequencing reads do not align to the rCRS. Interestingly, there were slightly higher amounts of all three metrics in this sequencing run in comparison to SeqM002 (the sequencing run that processed the Phase Two samples) even though there were less average total unaligned reads (Table 5.10). This suggests that mtDNA recovered from a fired cartridge cases was more degraded and/or damaged, as more reads can be filtered out due to being low in quality.

Table 5.10 Average GM-HTS alignment metrics for mtDNA samples in Phase Four. Data was obtained from GM-HTS alignment one.

Total unaligned reads	Low identity reads	Short aligned reads	Low match proportion reads	Other
267182 ± 40745	260 ± 91	11 ± 3	29408 ± 9612	237774 ± 41907

5.2.3 Bioinformatic analysis

5.2.3.1 Control evaluation

A full haplotype with entire mitogenome coverage was generated for the APOS control processed alongside the fired mtDNA samples.

5.2.3.1.1 Negative controls

As shown in Table 5.11, ENEG 1 and ENEG 2 had less than 100X coverage recorded at each nucleotide position in the mitogenome, resulting in entire mitogenome dropout, indicating that any contaminating reads would not impact haplotype generation. ENEG 3 and ANEG had more than 100X coverage at various regions in the mitogenome, indicating that contaminating reads were at high enough levels to be called.

Table 5.11 Reads seen in the negative controls of Phase Four. Red font indicates that the maximum number of reads seen at one nucleotide position passes the 100X read depth threshold. Percentage of mitogenome with read alignment was calculated by the number of nucleotide positions that had reads align divided by the length of the rCRS (16,569).

Sequencing Run	Control	Reads aligned to rCRS (%)	Average read depth across mitogenome	Maximum number of reads	% of mitogenome with read alignment
SeqM005	ENEG 1	0.291	5.51X (± 0.52)	84	13.9
	ENEG 2	0.004	0.014X	2	0.7
	ENEG 3	0.432	8.36X (± 0.65)	126	19.0
	ANEG	0.353	4.47X (± 0.81)	188	15.0

ENEG3 saw the calling of two major variants: 16362C and 16390A, both with a coverage of 118X. When compared to all the known haplotypes of samples processed alongside this control and the analyst's

haplotype, both of these variants were not seen in any haplotype. The frequencies for these variants were 98.30% and 100% respectively, indicating that possibly low-level background DNA from another source (e.g. another analyst in the laboratory) contaminated this sample. The ANEG sample saw the calling of a variant at 6971 with a coverage of 188X. A mixture of a C (55.55% of reads) and T (44.44% of reads) was seen at this position resulting in the calling of a PHP. When compared to all the known haplotypes of samples processed alongside this control and the analyst's haplotype, no variant was observed at this nucleotide position, again suggesting that low-level background DNA from an alternative source contaminated this sample. One approach to deal with level of contamination would be to increase the minimum total read depth GM-HTS threshold from 100X to 200X to avoid the influence of any low-level contamination in the analysis of the fired samples. A minimum read count of 200X has been used in other studies that have investigated the influence of damage in the mtDNA CR and were able to generate full and partial haplotypes (4,157). If contaminating reads were present above the minimum threshold level in a negative control in casework, all samples in that batch would have to undergo laboratory processing again. This was not enforced in this instance due to time and resource constraints.

5.2.3.2 Fired cartridge haplotypes

The same calculation in Section 4.2.3.2 was used to determine how much of each known haplotype was able to be obtained from a fired cartridge case sample. 22 of the 28 Phase Four samples (78.6% of total samples) generated partial to full haplotypes. Specifically, five of these samples were able to provide complete haplotypes (17.9% of total samples).

Table 5.12 Summary of mtDNA profiling results in Phase Four. The concordance of variants within the recovered haplotype was determined through a comparison to a reference profile. Haplogroups estimates were assigned with a quality score to the concordant variants using Haplogrep3. Bolded non-concordant variants indicate that the damage pattern from nucleotide position 2775 to 2784 was observed in that sample.

Sample participant_replicate	Concordance to reference haplotype (%)	Mitogenome coverage (%)	Haplogroup estimate	Haplogrep Quality	Number of non-concordant variants
1_1	-	-	-	-	-
1_2	94.7	97.9	W4a1	99	60
1_3	97.4	99.1	W4a1	101	**12**
1_4	97.4	100.0	W4a1	101	1
2_1	93.8	99.1	HV0+195	89	**13**
2_2	75.0	99.8	HV	98	**29**
2_3	100.0	100.0	HV0+195	94	1
2_4	93.8	98.0	HV0+195	89	**78**
3_1	100.0	99.1	H1	91	7
3_2	18.2	16.1	K1c2a	54	-
3_3	100.0	99.3	H1	91	**6**
3_4	100.0	98.7	H1	91	**9**
4_1	82.4	92.3	J1c2	90	**36**
4_2	-	1.0	-	-	3
4_3	50.0	92.5	J	78	33
4_4	-	-	-	-	-
5_1	28.1	55.2	R	70	**41**
5_2	21.9	78.2	R	78	**57**
5_3	40.6	76.4	U3a'c	68	**61**
5_4	78.1	80.7	U3a1c	91	37
6_1	-	3.1	-	-	4
6_2	21.1	23.2	J1	62	2
6_3	-	-	-	-	-
6_4	-	-	-	-	-
7_1	96.6	98.1	U5a1a1	95	**11**
7_2	69.0	87.3	U5a1a1	84	**30**
7_3	79.3	89.9	U5a1a1	87	23
7_4	100.0	99.1	U5a1a1	97	**6**

Six of the 28 fired cartridge samples had no variants called. These were the same six samples that generated the least amount of aligned reads (as displayed in Figure 5.5). In four of these samples, the entire mitogenome dropped out. In the other two samples, some regions had greater than 100X coverage but did not contain any variants.

5.2.3.3 Haplogroup analysis

The 22 fired samples with variants typed that were expected in comparison to the reference profile were uploaded to Haplogrep3 for haplogroup estimation. Three of the estimated tophit haplogroups were highlighted red, indicating a low-quality score (156). Sample 3_2, which had the lowest percentages of mitogenome coverage and reference haplotype recovered, unsurprisingly differentiated into the wrong node

of haplogroups. Phylogenetically, haplogroup R can diverge into haplogroup U, which can then diverge into haplogroup K (where the Haplogrep3 assigned haplogroup is located). Haplogroup R can separately also diverge into haplogroup H, which is where Participant Three's expected haplogroup, H1, is located. Both Samples 5_3 and 6_2 were then assigned haplogroups that were less specific. These samples had less haplogroup resolution due to drop out and were estimated to belong to an ancestor haplogroup several nodes before the reference samples haplogroup estimation. As concluded in Section 4.2.3.3, any sample assigned a haplogroup with a low-quality score in Haplogrep indicates that there is not enough mitogenome coverage for an accurate haplogroup estimation and any comparisons to reference profiles should be made on the typed major variants alone.

Eight of the assigned haplogroups were highlighted yellow indicating that the tophit haplogroup had a moderate quality score (156). These eight samples had from 21.9% to 93.8% of the reference haplotype recovered. Five of these samples (Sample 2_1, 2_4, 4_1, 7_2 and 7_3) were correctly estimated into their expected haplogroup and had both mitogenome coverage greater than 80% and a haplogroup quality score greater than 80%. The other three samples were estimated into a haplogroup located at least four nodes in the phylogenetic tree before their expected haplogroup estimate. These samples had at least either their quality score or mitogenome coverage under 80% indicating that their estimated haplogroup should be interpreted with caution (Section 4.2.3.3).

11 of the fired samples were highlighted green, indicating that the tophit haplogroup estimate had a high-quality score. Nine of these samples gave same haplogroup that was expected for that sample. Sample 5_4 had only 80.7% mitogenome coverage and was assigned into U3a1c, one node before U3a1c1 (Participant Fives expected haplogroup estimate). Sample 2_2, which had over 80% mitogenome coverage only had 75% of the reference profile correctly typed. As a result, Sample 2_2 was assigned into haplogroup HV, a haplogroup two nodes before Participant Two's expected haplogroup estimate (HV0+195). As suggested in Section 4.2.3.3, no warning from Haplogrep indicated that either the correct haplogroup or correct node of haplogroups had been estimated.

5.2.3.4 Non-concordant variants

There were 560 non-concordant variants (as defined in Section 4.2.3.4) recorded for the 28 fired cartridge case samples. In comparison, the unfired samples had 92 non-concordant variants. Additionally, 19 of the fired samples had more than three minor variants called, where only eight unfired samples had more than three minor variants called. Each of these 19 samples were analysed to determine if cross-contamination had occurred. No evidence of self-contamination by the analyst was identified. One sample (Sample 2_2) had 44.8% of its non-concordant variants match major variants in Participant Sevens haplotype, indicating that

cross-contamination had occurred. These minor variants were able to be removed from Sample 2_2's final haplotype. All other samples showed no signs of any cross-contamination with another sample.

79 of the non-concordant minor variants (14.1%) were at a nucleotide position where a major variant was also seen. These variants were removed from the final haplotype, leaving a total of 469 non-concordant variants that underwent further analysis.

5.2.3.4.1 Non-concordant major variants

Unlike the unfired samples, where all the non-concordant variants were identified as minor variants, two of the fired samples saw the calling of non-concordant major variants. In Sample 5_2, 14244T was called as a major variant with a frequency of 88.60% and a read depth of 272X. In Sample 5_4 both 4810T and 12549A were called as major variants with frequencies of 94.00% and 70.58% respectively. The read depths at these two nucleotide positions were 100X and 102X. As low-level background contamination was seen in two of the negative controls this indicated that a 200X minimum read threshold should have been used during GM-HTS alignment (Section 5.2.3.1.1). Had this been implemented, both 4810T and 12549A would have not been called.

As the calling on non-concordant major haplotypes can impact haplogroup assignment, both samples' haplotypes with the non-concordant major variants were uploaded to Haplogrep3. Both assigned tophit haplogroups were the same as the tophit haplogroups assigned without including the incorrect major variants (Table 5.12). All three non-concordant major variants were classified as global private mutations in Haplogrep3, warning that possibly a genotyping error had occurred. Additionally, these haplotypes were also uploaded to *EMPcheck* for a secondary quality check. *EMPcheck*, is a tool provided by EMPOP that can identify both new transversions or ambiguous variants, for an additional quality check. For upload, haplotypes were compiled into a *.emp text file based on the requirements specified in (161). New transversion messages appeared for both 4810T and 12549A, providing further evidence that these variants were possibly genotyping errors. As none of these non-concordant major variants were reproducible in the other replicates from Participant Five, it is likely that they occurred due to damage or stochastic errors. It was unlikely that these errors appeared during sequencing, as strand bias ratios for the amplicons containing the affected nucleotide positions were all greater than 0.9. The processing of replicates of the same sample at the amplification level, would provide greater confidence in casework practices, as it is likely that these variants would not be replicated at this level (4,157).

5.2.3.4.2 Non-concordant minor variants

Of the 560 total non-concordant variants called, 465 non-concordant minor variants remained after the above filtering processes. In 14 of the fired samples, the same pattern of variants identified at nucleotide positions 2275, 2776, 2777, 2778, 2782 and 2784 in the unfired samples was observed. These variants accounted for 15% of the total non-concordant variants. After this, 372 non-concordant minor variants remained. These were all possibly sites of either stochastic errors, mtDNA damage or low-level background contamination. In comparison, the unfired mtDNA samples had 27 non-concordant variants that were possible sites of error or damage. As there was the possibility that low-level background contamination was present, these results could only suggest that the firing process potentially resulted in more damage to the mtDNA template.

A random selection of fired samples and all their remaining non-concordant variants were further analysed. In total 43 non-concordant variants were investigated. These sites were identified in samples that were processed in the same extraction batch as ENEG controls that had no contamination (e.g., not in the same extraction batch as the contaminated ENEG sample), and also processed at least four rows of wells away from the contaminated ANEG sample in a 96-well plate. The frequencies of these variants ranged from 10.0 to 48.64% with an average frequency of 17.9% ± 1.3. The amplification of damaged DNA has previously led to damaged bases being called at a frequency up to 25% (130). It was suggested that using 25 – 30 PCR cycles might help lessen stochastic amplification of damage, therefore, reducing the number of PCR cycles in the custom Precision ID amplification method to 25 may help.

Three of the non-concordant variants were all located at position 16519. In all cases the T base (seen in the rCRS) transitioned to a C. This occurred in Sample 2_1, 2_2 and 2_3, three of the four replicates from Participant Two. The frequencies of this transition were not reproducible; with values of 10.0, 14.9 and 35.8%. The other 40 non-concordant variants were only observed once at each nucleotide position.

Unlike in Phase Two where only damage consistent with deamination was observed, there were nine instances consistent with oxidative DNA damage whereby a G base transversed into a T or a C base transversed into an A (4). There was also an instance whereby a C base transversed into a G, and a T base transversed into a G. These are also mutations that tend to occur under oxidative conditions (162). All potential signs of oxidative damage were recorded in the same sample (Sample 5_1), suggesting that this sample was particularly degraded, a hypothesis that could be confirmed with a mtDNA quantification (163). The remaining 30 non-concordant variants were then all transition mutations; 11 were consistent with PCR errors and 19 were consistent with deamination. Similar to Phase Two, pyrimidine transitions were the main change observed, with a C to T base change (deamination damage) observed in 12 samples.

5.3 PHASE FOUR: FORENSEQ™ PROFILE GENERATION FROM FIRED CARTRIDGE CASES

The 28 fired cartridge extracts taken through the mtDNA sequencing workflow (as described in Section 5.2.1) were also processed through the ForenSeq™ sequencing workflow so a direct comparison between the performance of both MPS workflows on the same sample could be undertaken.

5.3.1 Laboratory workflow

The 28 fired cartridge cases, three ENEG and three EPOS extracts, after being quantified using Quantifiler™ Trio (in Section 5.2.1.1), were taken through the ForenSeq™ sequencing workflow by a second analyst. ForenSeq™ libraries were prepared and sequenced as per Section 2.8. Up to 32 libraries were pooled together for sequencing in the same run. Like in Phase Two, each sequencing run included a mixture of controls, reference samples and fired cartridge case samples to ensure that the sequencing run would not fail. The pooled libraries were denatured, diluted, and sequenced following methods described in Section 2.8.4 and Section 2.8.5.

5.3.2 Results

Quantification of the gDNA recovered from 28 fired cartridge samples gave concentrations ranging from undetermined to 0.00018 ng/µL, as described in Section 5.2.2.1.

5.3.2.1 Library preparation

After library preparation, each fired cartridge case sample was run on the Fragment Analyzer to determine whether ForenSeq™ STR and SNP targets were able to be successfully recovered, amplified, and converted into libraries.

Table 5.13 Average lengths of fragment peaks seen in a Fragment Analyzer EPG after the library preparation of the Phase Four samples taken through the ForenSeq™ sequencing workflow.

Fragment Type	Fragment lengths	Average representation in the sample (%)	Minimum value (%)	Maximum value (%)
Unbound primers/primer dimers	1 – 99	16.0 ± 2.1	8.9	46.7
Adapter-dimer	160 – 180	69.7 ± 2.3	38.6	84.1
Library	200 – 700	12.7 ± 0.8	5.7	22.2

As with the Phase Two unfired cartridge samples, most of the peaks seen in the Fragment Analyzer EPGs for the Phase Four fired samples were located from 160 to 180 bp, indicating a high presence of adapter-dimers. This was unsurprising, given that the addition of a second purification was not trialled and implemented into the ForenSeq™ sequencing workflow due to time constraints. When comparing the mean library concentrations obtained from the Fragment Analyzer, there was a significant difference between the fired libraries and the unfired libraries (p = 0.00839, alpha = 0.05, Welch two sample t-test). As the unfired samples had a higher mean library concentration (0.29 ng/µL) compared to the unfired samples (0.16 ng/µL), this shows there were lower concentrations of library in the fired samples. This was unsurprising, given that there was on average less nDNA recovered from the fired samples (Section 5.2.2.1).

5.3.2.2 Sequencing performance

The ForenSeq™ fired cartridge case libraries were sequenced alongside several high-quality libraries over two separate sequencing runs. Both sequencing runs passed the recommended UAS quality metrics (144) as seen in Table 5.14.

Table 5.14 UAS quality metrics for ForenSeq™ sequencing runs containing Phase Four samples.

Sequencing Run	Cluster density (k/mm^2)	Cluster passing filter (%)	Phasing (%)	Pre-phasing (%)
SeqD004	1194	89.23	0.112	0.020
SeqD005	1209	89.07	0.095	0.010

As seen in Figure 5.6, none of the fired cartridge case samples in Phase Four had enough aligned reads to reach the UAS minimum total read count guideline. This indicated that all typed loci were to be interpreted with caution. The number of aligned reads ranged from zero to 34,693 reads. Sequencing reads that aligned to ForenSeq™ target regions were generated for 20 of the samples. In these 20 samples, on average only 1.7% ± 1.0 of the total sequencing reads were aligned sequencing reads which was very similar to the average seen for the unfired samples (Section 4.3.2.3)

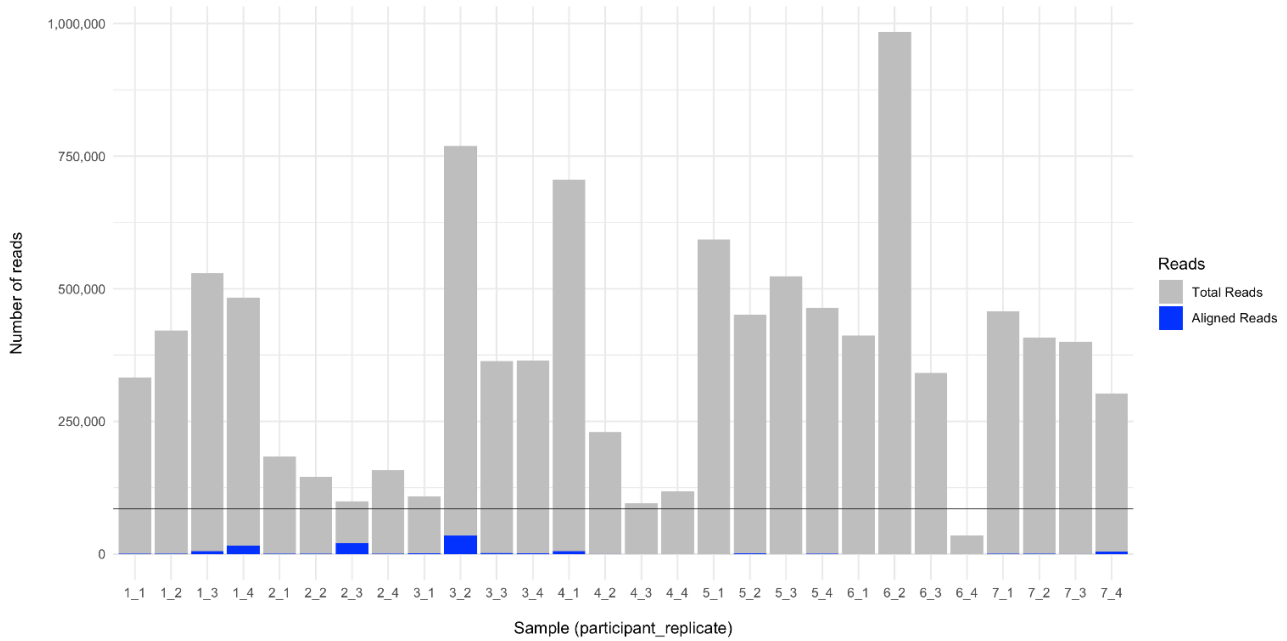

Figure 5.6 Generated sequencing reads for fired cartridge case samples taken through the ForenSeq™ sequencing workflow. The black horizontal line at 85,000 reads represents the UAS read count guideline. Phase Four samples were sequenced over two independent sequencing runs.

5.3.3 Bioinformatic analysis

5.3.3.1 Control evaluation

One ANEG, three ENEGS and one APOS sample were processed alongside the ForenSeq™ Phase Four samples. All negative controls generated no sequencing reads that aligned to ForenSeq™ targets, indicating no contamination occurred. The APOS control (2800M) gave a full and concordant STR and iiSNP profile.

5.3.3.2 STR profiles recovered from fired cartridge cases

The same methods of comparison, as described in Section 4.3.3.2, were used to determine how much of a reference STR profile could be recovered from a fired cartridge case.

Table 5.15 Summary of ForenSeq™ STR profiling results in Phase Four. Concordant loci were determined through a comparison to a known reference profile. Loci with allele dropout gave a homozygous genotype when a heterozygous geneotype was expected.

Sample (participant_replicate)	Aligned reads	Number of typed loci	Concordant loci	Loci with allele dropout	% concordant STR profile
1_1	215	-	-	-	-
1_2	718	1	-	1	-
1_3	5059	3	1	2	2.9
1_4	15429	10	1	7	2.9
2_1	270	1	-	1	-
2_2	513	3	-	3	-
2_3	20065	28	11	12	31.4
2_4	399	-	-	-	-
3_1	1209	6	1	5	2.9
3_2	34693	14	7	7	20.0
3_3	1571	2	1	-	2.9
3_4	1394	1	1	-	2.9
4_1	5293	2	-	2	-
4_2	78	1	-	-	-
4_3	-	-	-	-	-
4_4	-	-	-	-	-
5_1	-	-	-	-	-
5_2	1148	1	-	-	-
5_3	-	-	-	-	-
5_4	275	-	-	-	-
6_1	-	-	-	-	-
6_2	-	-	-	-	-
6_3	-	-	-	-	-
6_4	-	-	-	-	-
7_1	180	-	-	-	-
7_2	166	-	-	-	-
7_3	74	-	-	-	-
7_4	4234	7	1	6	2.9

14 of the 28 fired samples (50% of samples) saw the typing of 80 STR loci. This was a lower amount of total STR loci compared to the 28 unfired samples and was unsurprising given that unfired samples had on average a higher library concentration (Section 5.3.2.1). Eight of the 28 samples saw the calling of 24 STR genotypes that were concordant with the corresponding reference profile. This generated partial STR profiles ranging from 2.9% to 31.4%. Similarly, to the unfired cartridge samples where 32.8% of typed STR loci were concordant, in the fired samples 30% of typed STR loci were concordant. These results suggest that for ForenSeq™ STR profiles recovered from firearm evidence, the amount of typed STR loci expected to be concordant is around 30%.

Of the 24 concordant loci, 12 were homozygous and 12 were heterozygous. Interestingly, the STR loci that were more likely to be concordantly typed from fired cartridge case samples differed from unfired cartridges.

In this phase of experimental work, TPOX was correctly typed four times and D2S441 was correctly typed three times. Both of these loci had amplicon lengths of less than 180 bp (74) and TPOX has previously been identified as a robust locus, able to be typed in samples with low amounts of nDNA (164). The 12 concordant heterozygous loci were seen in two samples, Samples 2_3 and 3_2. Both of these samples had the highest amount of nDNA recovered as indicated by the Quantifiler™ Trio results (Table 5.6). The ACRs of all expected a-STR and x-STR heterozygous genotypes in this research that had at least one allele typed were between 0 to 0.89. The loci that experienced allele dropout, then had ACRs ranging from 0 to 0.17, which slightly overlapped with the ACR range seen for loci that had both alleles typed concordantly. These specifically ranged from 0.08 to 0.89, with an average of 0.52 ± 0.08. These results indicate that a majority of the heterozygous STR loci recovered from fired cartridge cases were below the 0.60 threshold and imbalanced (159).

5.3.3.2.1 STR loci with allele dropouts

46 STR loci had one allele typed concordantly with the second allele then dropping out. The read depth of the typed allele ranged from 32X to 1,788X. The read depths for the untyped allele then ranged from 0X to 146X, with 42 untyped alleles having no reads generated. This resulted in these loci being called as homozygous when a heterozygous genotype was expected. Four loci then saw typing of reads to the expected allele, however these were either low coverage or identified by the UAS as stutter. As a result, these alleles were excluded from the final genotype.

5.3.3.2.2 Non-concordant STR loci

There were 10 non-concordant STR loci typed in five of the Phase Four samples as described in Table 5.16.

Table 5.16 Non-concordant STR loci observed in Phase Four samples. Red font indicates that allele was not expected in that locus.

Sample	Loci	STR	Expected genotype	Genotype called	Parent allele/s read counts	Read count of allele/s that dropped in
1_4	aSTR	D9S1122	11,13	10,11,13	1199,150	322
1_4	aSTR	D20S482	13,14	12,14	218	2470
2_3	aSTR	FGA	23.2,24	23,23.2,24	49,172	90
2_3	aSTR	D16S539	12,13	11,12,13	33,84	180
2_3	aSTR	D17S1301	12,13	11,12	454	100
2_3	aSTR	D21S11	30,30.2	30,30.2,30.2	149,123	35
2_3	xSTR	DXS7423	15,15	14,15,15,15	289	373,192,407
3_3	aSTR	D2S1338	16,19	18,19	384	276
4_2	aSTR	TH01	7,8	9.3,9.3	-	78
5_2	ySTR	DYS391	-	10	-	1017

One sample (Sample 5_2) saw the calling of allele 10 at the y-STR locus DYS391 with a read depth of 1017X. This was unexpected, given that Participant Five was known to be female (Section 3.3.3). The called genotype matched the 2800M APOS sample's expected genotype, however, given that the APOS sample was of high DNA quantity it would be expected that if cross-contamination between these samples occurred more y-STR would be typed. As discussed in Section 4.3.3.2.2, an alternative source of contamination may have occurred, this was also the case for the calling of TH01 in Sample 4_2. D20S482 (in Sample 1_4) did have the expected 13 allele typed, however this was identified as stutter, and the 12 allele was typed. As this allele had a high read depth (relative to the other read depths obtained for this evidence type) it is possible that this was contamination.

In Sample 2_3, two of the non-concordant loci (D21S11 and DXS7423) saw the calling of an A nucleotide base when a C was expected. In DXS7423 this base change occurred twice at two different nucleotide locations in the 15 allele, resulting in three different counts of the 15 allele. C to A nucleotide transversions are consistent with oxidative damage patterns (4). The 14 allele was then in a n-1 stutter position but was able to be called at a higher read depth than the true 15 allele. Three of the other non-concordant loci (D9S1122, D17S1301 and D2S1338) also saw stutter coverage pass the stutter threshold resulting in an incorrect genotype.

Locus FGA in Sample 2_3 incorrectly called allele 23. This was possibly a result of sequencing error as in the STR sequence for this locus, one T base was called in allele 23 when in comparison to the same location in the 23.3 parent allele, three T bases were expected. A replicate at the sequencing level would be able to confirm and resolve this type of error. Finally, D16S539 in Sample 2_3 saw the typing of expected n-1 stutter (allele 11) with a read count higher than the parent allele.

5.3.3.3 iiSNP profiles recovered from fired cartridge cases

Using the same calculation described in Section 4.3.3.3, each iiSNP profile was compared to the known corresponding reference iiSNP genotype.

Table 5.17 Summary of ForenSeq™ iiSNP in Phase Four. Concordant iiSNP loci were determined through a comparison to a known reference profile. Loci with allele dropout gave a homozygous genotype when a heterozygous genotype was expected.

Sample	Number of typed loci	Concordant loci	Loci with allele dropout	% concordant iiSNP profile
1_1	1	1	-	1.1
1_2	2	1	1	1.1
1_3	8	6	2	6.3
1_4	24	11	13	11.6
2_1	2	1	1	1.1
2_2	3	1	2	1.1
2_3	67	49	18	51.6
2_4	1	1	-	1.1
3_1	9	5	3	5.3
3_2	35	23	11	24.2
3_3	2	-	2	-
3_4	7	5	2	5.3
4_1	4	2	2	2.1
4_2	-	-	-	-
4_3	-	-	-	-
4_4	-	-	-	-
5_1	-	-	-	-
5_2	-	-	-	-
5_3	-	-	-	-
5_4	3	2	1	2.1
6_1	-	-	-	-
6_2	-	-	-	-
6_3	-	-	-	-
6_4	-	-	-	-
7_1	1	-	1	-
7_2	1	1	-	1.1
7_3	1	-	1	-
7_4	12	6	6	6.4

Of the 183 iiSNPs typed, 115 were typed concordantly (62.8%). This was very similar in comparison to Phase Two which saw 63.6% of the total iiSNP loci typed concordantly. These results further show that iiSNPs are more likely to be typed concordantly compared to STRs. Partial iiSNP profiles were obtained for 15 samples and ranged from 1.1% to 51.6%. Of the 115 concordant iiSNPs, 98 were homozygous and 17 were heterozygous. The read counts for the iiSNPs called correctly as homozygous ranged from 34X to 2233X (Figure 5.7). The ACR for the concordant heterozygous iiSNP loci widely ranged from 0.11 to 0.83, with an average of 0.51 ± 0.07. For SNPs, a threshold of 0.5 indicates acceptable allele balance (159). There were some iiSNPs that were more likely to be typed concordantly than others; rs2107612, rs1335873 and rs1028528 were typed in four samples and rs9905977 was typed in five samples.

5.3.3.3.1 iiSNPs with allele dropouts

There were 66 iiSNPs that were typed as homozygous due to allele dropout, when a heterozygous genotype was expected. All four nucleotide bases had similar probabilities of dropping out (Table 5.18).

Table 5.18 Frequency of how often each nucleotide base dropped out at iiSNP loci in Phase Four samples.

Allele	Number of times dropped out	Proportion in non-concordant loci	How many times allele was expected to be typed	Chance of dropping out (%)
A	16	24.2	32	50.0
C	17	25.8	35	48.6
G	16	24.2	32	50.0
T	17	25.2	33	51.5

These read depths for the expected allele that was typed ranged from 31X to 1423X. As seen in Figure 5.7, these read depths overlap with the read depths seen for a concordant iiSNP homozygous genotype. A majority of the alleles that dropped out had a read depth of 0X. In three instances, sequencing reads did align to the untyped allele, however, these were all under 30X leading to allele dropout.

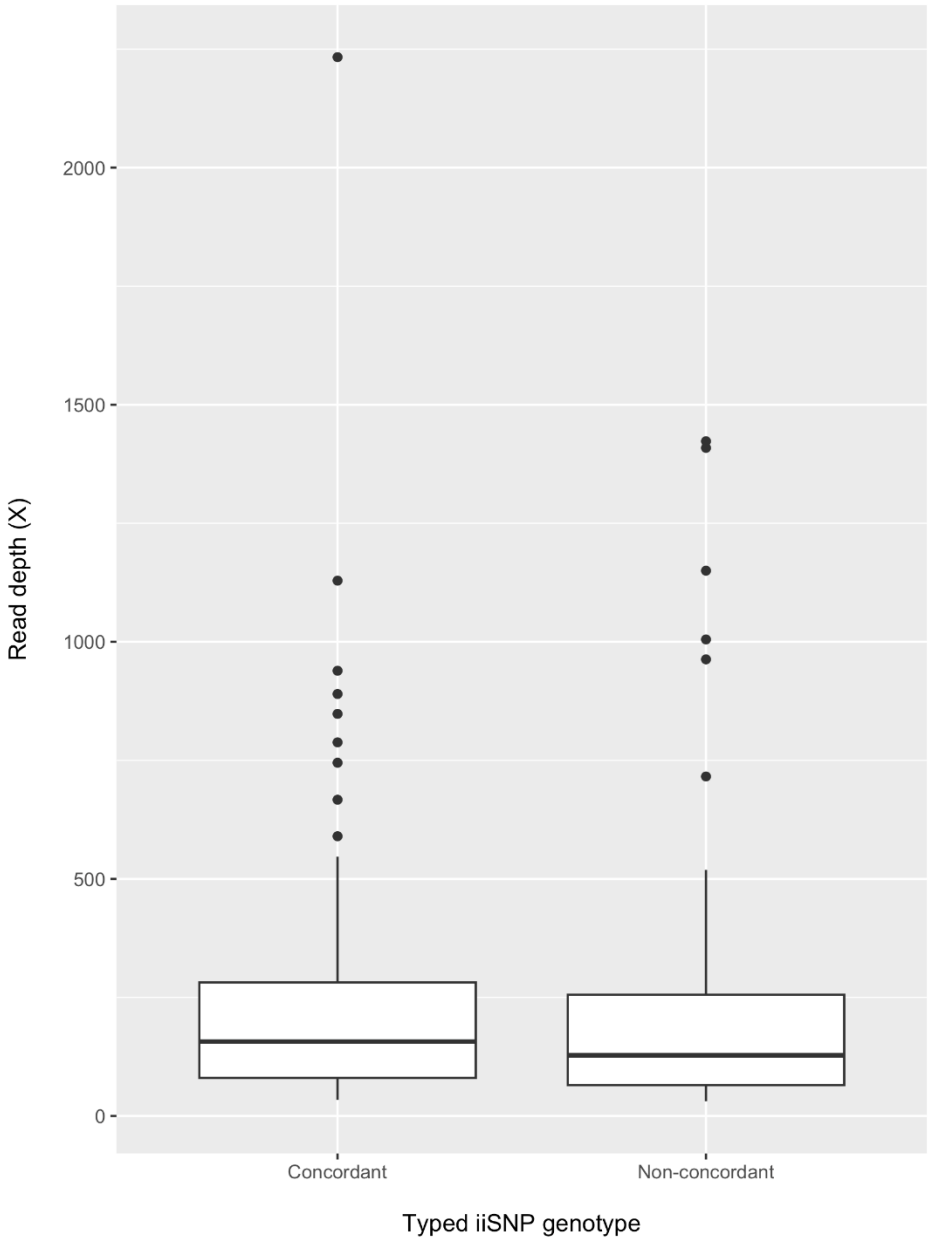

Figure 5.7 Range of read counts (X) for each iiSNP allele typed as a homozygous iiSNP in Phase Four. Concordant indicates that the homozygous genotype was expected in comparison to the known reference profile. Non-concordant indicates that a homozygous genotype was called due to allele dropout when a heterozygous genotype was expected for that locus.

5.3.3.3.2 Non-concordant iiSNP loci

Two non-concordant iiSNPs were seen in the Phase Four samples due to an allele dropping in. In accordance with a previous study (164), the rate of allele drop-in was lower for iiSNPs compared to STRs. Allele drop-ins resulted in an expected homozygous loci being typed as a heterozygous. As seen in Table 5.19, both instances saw the non-concordant allele typed with a read count higher than the expected allele. In both samples, all other iiSNP and STRs typed were either fully concordant or experienced allele dropouts, therefore, it was not obvious if cross-contamination occurred. As iiSNP allele drop-in was not observed in the unfired samples, iiSNP drop-ins could be more likely to be seen in fired cartridge case samples, however, a larger study with

more samples would be required to confidently confirm this. The ACR for both loci were within the ACR range seen for concordant heterozygous iiSNPs (0.11 to 0.83), indicating that being able to distinguish between an expected heterozygous genotype and a false heterozygous genotype due to allele drop-in is impossible.

Table 5.19 Non-concordant alleles that dropped in at an iiSNP locus in two of the Phase Four samples.

Sample	Locus	Concordant		Non-concordant		ACR
		Allele	Read Count	Allele	Read count	
3_1	rs221956	T	59	C	71	0.83
3_2	rs729172	C	72	A	189	0.38

5.3.4 Comparison between both MPS workflows

As all Phase Four extracts underwent both MPS workflows, a direct comparison between their ability to generate DNA profiles was carried out. For this comparison, the amount of mitogenome coverage and the number of concordant iiSNP and STR alleles were determined for each sample. As seen in Table 5.20:

- Twelve samples had genetic information obtained for all three genetic markers.
- Seven samples had only coverage of the mitogenome obtained.
- Five samples had both iiSNPs typed and coverage of the mitogenome obtained.
- Three samples had no genetic information obtained.
- One sample saw only one iiSNP loci typed.

Table 5.20 A comparison of genetic information obtained from Phase Four samples using both MPS workflows. Mitogenome coverage represents how much of the mitogenome (16,569 bp) had > 100X read depth and was typed. STR alleles represents how many of the 70 (for females), or 89 (for males) STR alleles were concordant to the known reference profile. iiSNP alleles represents how many of the 188 iiSNP alleles were concordant to the known reference profile. Each concordant homozygous and heterozygous genotype counted as two alleles. Incorrectly called homozygous genotypes due to allele dropout/drop-in were counted as one allele if that allele was concordant to the reference profile.

Sample	Quantifiler™ nDNA concentration	Mitogenome coverage (%)	STR alleles (%)	iiSNP alleles (%)
1_1	UD	-	-	1.1
1_2	UD	97.9	1.4	1.6
1_3	UD	99.1	5.7	7.4
1_4	0.00072	100.0	17.1	18.6
2_1	UD	99.1	1.4	1.6
2_2	0.00032	99.8	4.3	2.1
2_3	0.00184	100.0	61.4	61.7
2_4	UD	98.0	-	1.1
3_1	0.00032	99.1	10.0	7.4
3_2	0.00102	16.1	30.0	30.9
3_3	0.00018	99.3	4.3	1.1
3_4	UD	98.7	2.9	6.4
4_1	UD	92.3	2.9	3.2
4_2	UD	1.0	-	-
4_3	UD	92.5	-	-
4_4	UD	-	-	-
5_1	UD	55.2	-	-
5_2	UD	78.2	-	-
5_3	UD	76.4	-	-
5_4	UD	80.7	-	2.7
6_1	UD	3.1	-	-
6_2	UD	23.2	-	-
6_3	UD	-	-	-
6_4	UD	-	-	-
7_1	UD	98.1	-	0.5
7_2	UD	87.3	-	1.1
7_3	UD	89.9	-	0.5
7_4	0.00071	99.1	11.4	9.6

mtDNA haplotypes were more likely to be recovered and successfully sequenced from fired cartridge case samples (Table 5.20). There were only two instances where using the ForenSeq™ Kit to target nDNA provided more genetic information: in Sample 1_1, where only one iiSNP locus had both alleles concordantly typed and in Sample 3_2 where 16.1% of the mitogenome was sequenced. Given that Sample 3_2 had a nDNA concentration of 0.00102 ng/µL and had approximately 30% of STR and iiSNP alleles concordantly typed, this appeared to be an abnormal result, especially considering that a higher mitogenome coverage was obtained for samples with a lower nDNA quantity (e.g. Sample 7_1). The Fragment Analyzer EPG after mtDNA amplification for Sample 3_2 showed high peaks from 100 to 250 bp indicating Precision ID amplification of mtDNA successfully occurred in both multiplexes. The Fragment Analyzer EPG after mtDNA library

preparation showed very small peaks from 280 to 700 bp, indicating that library preparation did not work well, accounting for why low mitogenome coverage obtained. If this were to occur with a casework sample, it would be recommended to repeat from library preparation onwards.

Table 5.20 also shows that iiSNP alleles are more likely to be recovered compared to STR alleles, as STR alleles were only observed in samples that had iiSNP alleles recovered. Additionally, these results show that an undetermined Quantifiler™ Trio quantity can still see the typing of the mitogenome, STR alleles and iiSNP alleles.

To compare the performance of both workflows in terms of the amount of a DNA profile that was obtained, the amount of the total ForenSeq™ profile recovered was determined for each sample. This saw the total number of concordant STR and iiSNP alleles for each sample divided by the total number of expected ForenSeq™ alleles (258 for females, 277 for males). Initial tests found that the percentage of DNA profiles obtained violated assumptions for a parametric test such as ANOVA. Specifically, this data significantly deviated from a normal distribution (p = <0.001, alpha = 0.05, Shapiro-wilk test) and there was a significant difference between the variances (p = <0.001, alpha = 0.05, Levene's test). A descriptive statistical analysis was performed on this instead. A density plot was used to visualise the distribution of how much of a DNA profile was able to be recovered from both workflows (Figure 5.8)

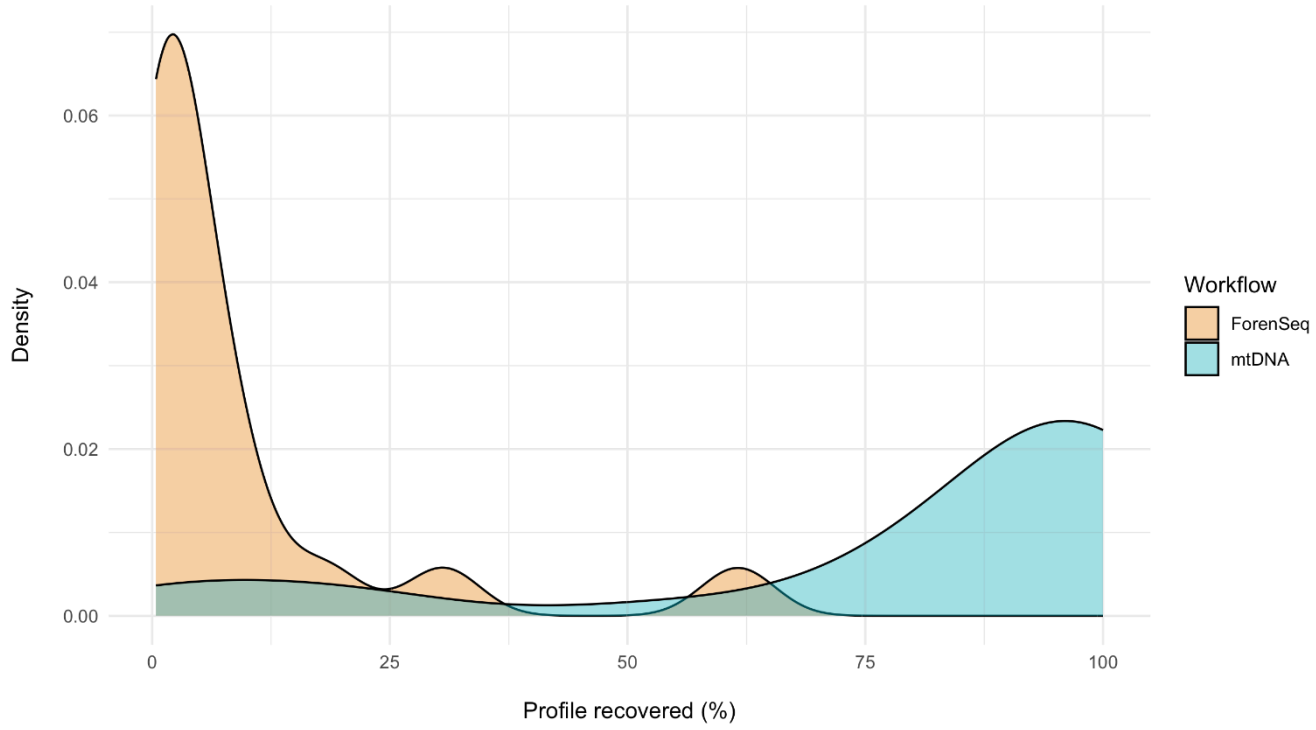

Figure 5.8 Density plot showing the distribution of how much of a DNA profile was able to be obtained from fired cartridge cases using both MPS workflows.

Most of the ForenSeq™ profiles obtained were 0 to 10% of the total expected profile, as seen by the large sharp orange peak seen in Figure 5.8. In comparison, a majority of the mtDNA haplotypes obtained were from 80% to 100% of the total mitogenome as seen by the smaller, broader blue peak. For both workflows the data was skewed due to outliers, however, in opposite directions. As a result, the mean percentage recovered for each workflow was impacted and did not accurately reflect sequencing performance (Table 5.21). Instead, the median provided a better indicator of how well each workflow worked. The ForenSeq™ workflow had a median of 0.8% and the mtDNA workflow had a median of 91.9%. These results further prove that more of a mtDNA haplotype compared to an nDNA profile is able to be successfully recovered from a fired cartridge case.

Table 5.21 Measures of central tendency and dispersion for the amount of a DNA profile obtained from a fired cartridge case using both MPS workflows.

Workflow	Mean	Standard deviation	Minimum	Lower quartile	Median	Upper Quartile	Maximum
ForenSeq™	5.6 ± 7.8	12.9	0.0	0.0	0.8	3.7	61.6
mtDNA	67.3 ± 2.4	41.3	0.0	21.4	91.1	99.1	100.0

6 DISCUSSION AND CONCLUSIONS

Fired cartridge cases and unfired cartridges are common evidence types recovered from firearm crime scenes. Conventional methods used for forensic DNA typing of such evidence have had limited success in obtaining reliable DNA profiles, both in controlled research experiments (6,8) and samples recovered from crime scenes (41,165). The main limitation restricting successful DNA profiling is the low amounts of recoverable nDNA. Low template DNA results from a combination of several factors; the initial low transfer of touch DNA, inhibitory interactions between the metallic cartridge and DNA, and the highly intensive environment created during the firing process (4,8,26,34). As the touch DNA samples used in this study had limited amounts of detectable nDNA, it is unlikely that conventional CE-STR DNA typing would have produced many interpretable results. Previous research at ESR Ltd., has investigated methods to improve conventional DNA profiling results and a similar amount of DNA was recovered from touch DNA deposited on cartridge cases post firing (166). However, when using Identifiler™ Plus, no alleles were typed under standard PCR conditions (28 cycles) and only one allele was typed using LCN conditions (34 cycles). Massively parallel sequencing (MPS) technologies can efficiently target multiple genetic markers and generate information of whole genomes or selected regions of DNA. Commercial MPS amplification kits have been developed for forensic use; these utilise short amplicons to allow for processing of highly degraded, low DNA quantity samples. This research aimed to determine if using such technologies was a suitable alternative for obtaining DNA profiles from cartridge cases. Two kits, the ForenSeq™ DNA Signature Prep Kit and the Precision ID Whole mtDNA Genome Panel, were utilised in custom sequencing workflows (72,97) to generate nDNA and mtDNA profiles.

6.1 OVERALL FINDINGS

The first objective was determining whether these MPS workflows could generate DNA profiles from unfired cartridges. Of the 28 samples taken through the mtDNA sequencing workflow, 24 generated sequencing reads that aligned to the rCRS. This saw seven samples produce full haplotypes, with partial haplotypes generated for the 17 others. Importantly, enough genetic information was obtained in 13 of the 24 samples to see assignment into the same haplogroup as the corresponding reference sample when uploaded to Haplogrep3. Where non-concordant minor variants were also typed, a series of filtering steps could be undertaken to determine whether these were due to contamination, sequencing errors or potentially damage. It is recommended that for application into casework, replicates are used from amplification onwards, as true mitochondrial variants are reproducible across PCR events.

A further 28 samples were then taken through the ForenSeq™ sequencing workflow. 20 of these samples had sequencing reads generated that aligned to ForenSeq™ targets and saw the typing of both STR alleles and iiSNP alleles. From this, 11 samples typed 42 STR loci, and 13 samples typed 143 iiSNP loci that were fully concordant to the corresponding reference profile. This was 37.5% of the total STRs typed and 63.6% of the total iiSNPs typed. One sample, Sample 1_2, saw the typing of 15 STR loci and 45 iiSNPs, creating a partial profile that was compatible with loci used for conventional DNA profiling at ESR Ltd. It was found that the main limitation of the ForenSeq™ workflow was the low-quantity of template DNA recovered from the unfired cartridges. This saw a majority of the typed loci experienced allele dropouts (52.7% of the typed STRs and 36.4% of the typed iiSNPs). Consequently, expected heterozygous loci were typed incorrectly as homozygous. Where more iiSNPs were typed concordantly compared to STR loci, in instances where the reference profile is unknown, it would be hard to distinguish between true concordant genotypes and a homozygous genotype called due to an allele dropping out, for both types of genetic marker. As the allele that was typed was concordant to the reference profile, this suggests that any homozygous genotype should be treated as a partial genotype. Non-concordant STR alleles were typed that without a known reference profile, would be unable to be identified and removed, suggesting that iiSNPs have greater potential for obtaining profiles from unfired cartridges.

The second objective was to optimise the custom mtDNA sequencing workflow to improve mtDNA haplotype quality. Two different optimisations were trialled. It was found that adding a second purification after library preparation created a more purified, higher-quality mtDNA library. This increased the percentage of generated sequencing reads that aligned to the rCRS and improved the amount of the known reference profile that could be recovered. The “Full method” mtDNA sequencing workflow, where each primer multiplex was processed separately until normalisation, was also trialled. This method has previously shown to increase mitogenome coverage, giving high total read and mean coverage values (121). Conversely, in this research, the ‘Full’ method saw a lower average total reads per sample compared to the samples taken through the workflow with no modifications. Specifically, Strobl et al. (121) achieved an average of 932,779 total reads using the “Full” method and 436,755 total reads using the 2-in-1 method. In the present research, average total reads of 839,579 were obtained using the “Full” method and 1,152,046 were obtained using the 2-in-1 method, with no significant difference in the amount of mitogenome coverage between the two methods (Section 5.1.3.3). There were variations in methodology that could have contributed to these observed differences: in the current research a MiSeq FGx™ platform was used and up to 33 samples were included in a run. In comparison Strobl et al (121) used the Ion Torrent PGM™ machine and only six to ten samples were including in each sequencing run. As Multiplex A and B for each sample were processed individually, the “Full” method required twice the number of reagents, making it more expensive and time-consuming to carry out.

Based on these results, it was concluded that only a second purification would be implemented into the final mtDNA sequencing workflow.

The third objective of this research was to determine if these alternative MPS technologies could generate DNA profiles from fired cartridge cases. At the time of this research, this was the first known attempt to sequence whole mitochondrial haplotypes from this evidence type. The final mtDNA MPS workflow was applied to 28 samples recovered from fired .223 Rem cartridge cases. As observed in previous studies (8,50), it was found that firing impacted the quantity of DNA recovered as less gDNA was recovered from fired cartridge cases compared to unfired cartridges. As no mtDNA quantification was performed, it was unknown if the amount of mtDNA was affected by the firing process, however, Fragment Analyzer results post-amplification suggested that less mtDNA was present on a cartridge case after firing. 22 of the 28 fired cartridge case samples had enough of the mitogenome to reach ≥ 100X read depth, resulting in the generation of a haplotype. 14 samples were assigned a haplogroup estimate the same as their corresponding reference sample. Haplogrep3 was utilised as a QC tool, and when combined with the percent of the mitogenome with coverage, was able to provide confidence about how accurate the estimated haplogroup was compared to the expected reference haplogroup estimate.

In comparison to the unfired samples where 28 potential sites of damage or PCR error were identified, 372 sites of either potential damage, stochastic PCR error or low-level contamination were identified in the fired samples. As contamination was observed in two of the negative controls it could not be ruled out as a contributing factor to the large number of non-concordant minor variants. 43 of the non-concordant sites were selected for further analysis. These sites were identified in samples that were not processed in the same extraction batch as the contaminated ENEG sample, and also processed at least four rows of wells away from the contaminated ANEG sample in a 96-well plate, providing some confidence that contamination would not have occurred in these samples. The 28 potential sites of damage and PCR error in the unfired samples had an average minor variant frequencies of 17.2 ± 1.4. As the frequency of minor variants increases as storage and damage conditions worsen (149), it was expected that the fired samples would have a higher average minor variant frequency. In the 43 potential damage/PCR error sites identified in fired samples, the average minor variant frequency was 17.9% ± 1.3. This small difference is probably due to a combination of two factors. Firstly, the fired samples were sampled within one to five days after deposition, minimising DNA damage inflicted by surfaces containing copper ions (12) whereas the unfired samples were sampled 30 days after deposition. Secondly, as it appeared that less mtDNA was recovered from a fired cartridge case, this would mean there was a lower mtDNA template input into amplification, with low template samples more susceptible to damage (157).

This research was also the first known attempt to apply the ForenSeq™ Kit to nDNA recovered from fired cartridge cases. The same 28 samples were taken through the ForenSeq™ workflow and saw the typing of iiSNP and STR alleles. Compared to the unfired samples, fewer STR and iiSNP alleles were typed in the fired samples due to the lower amounts of nDNA recovered post-firing. From the alleles that were typed, eight samples saw the calling of 24 STR genotypes, and 15 samples saw the calling of 115 iiSNP genotypes fully concordant to the corresponding reference profile. This was only 30% of the total number of STRs typed and 62.8% of the total iiSNPs typed, again showing that iiSNPs were more likely to be typed concordantly. Compared to the unfired samples a similar rate of allele dropout was observed with 57.5% of the total typed STRs and 36.1% of the total typed iiSNPs experiencing allele dropouts. This meant that the same limitations that appeared when analysing the STR and iiSNP profiles obtained from unfired cartridges were also encountered with the fired samples, further restricting an analyst's ability to confidently call a correct genotype when a reference profile is unknown. This strengthens the conclusion that a homozygous genotype should be treated as a partial profile at that locus. Unlike with the mtDNA workflow, where a clear increase in the amount of non-concordant minor variants typed was observed, the rate of non-concordantly typed STRs and iiSNPs was similar between unfired and fired samples.

6.1.1 Targeting of mtDNA compared to nDNA

This research has shown that if quantification results for samples recovered from cartridges and cartridge cases show minimal amounts of nDNA or give an undetermined quantity, useable genetic information can still be obtained using MPS. Where the ForenSeq™ workflow was more straightforward and streamlined to perform compared to the custom mtDNA workflow, interpretation of the generated sequencing data was difficult. This was largely due to the dropout of alleles in heterozygous genotypes as result of low coverage. As mtDNA has a high copy number per cell and a double membrane (78), this offers mtDNA more protection in degrading environments compared to nDNA. Consequently, more informative profiles from fired cartridge cases could be obtained through the targeting of mtDNA. Additionally, sequencing data obtained from the mtDNA workflow was easier to interpret. This was largely due to mtDNA haplotypes being reported as major and minor variants. Where non-concordant mtDNA variants were often seen, 99.6% of these were minor variants. This suggests that calling haplotypes based only off major variants would result in an accurate mtDNA profile.

Where STRs are the universal gold standard genetic marker targeted for forensic DNA profiling and are well established into casework protocols, the targeting of mtDNA through MPS is a more successful method for obtaining DNA profiles from firearm evidence. Previous work has identified the Precision ID custom workflow as being the most suitable for implementation into casework in New Zealand and additionally, this workflow was used to create a database of haplotypes from New Zealand populations (97). Based on current and

previous work, mtDNA profiling should be considered for further validation and implementation into forensic casework, especially for challenging samples unable to provide useable nDNA profiles.

6.2 RECOMMENDATIONS FOR FUTURE WORK AND IMPLEMENTATION INTO CASEWORK

This research has demonstrated that MPS technologies can successfully be used for DNA profiling of touch DNA recovered from fired cartridge cases. However, before application into casework it is necessary to fully optimise both MPS workflows and address any of the apparent limitations that arose in this study.

6.2.1 Recommendations applicable to both workflows

The methods of sampling and extraction used in this research project were selected as they are the current methods implemented into casework at ESR Ltd. Past research attempts, restricted by the lack of profiling success through conventional CE-STR methods, have largely focused on improving the recovery of DNA from firearm evidence (41,44,45,50,52). For example, the BTmix solution and the rinse and swab method as developed by Bille et al (52) was found to improve quality and quantity of DNA recovered from fired cartridge cases in a direct comparison to using a double-swab method and water. Future work could investigate different methods of sampling and extraction in conjunction with the MPS workflows used in this research. For instance, as an aqueous environment on copper or brass surfaces is known to accelerate DNA degradation, it has been recommended that 0.5M EDTA should be used to collect DNA from metallic surfaces instead of water (4). This mitigates degradation and maximises the amount of DNA recovered. 0.5M EDTA should be trialled during the double swab sample collection method to determine if a higher percentage of a DNA profile can be obtained and whether the number of non-concordant variants can be decreased, especially in fired cartridge case samples.

One limitation of the current study was that mtDNA quantification was not performed due to time and resource constraints. Where the quantity of mtDNA present in an extracted sample was not required for the final mtDNA MPS workflow, it would allow for the determination of what the lower limit of this workflow is. A tetraplex real-time qPCR assay was developed that combines the quantification of both mtDNA and nDNA targets into the same reaction (163). Importantly, this assay includes two sized mtDNA targets (69 bp and 143 bp), allowing for mtDNA degradation to be determined (163). This would offer additional information on whether a higher DI corresponds with more non-concordant variants.

6.2.2 Recommendations for the mtDNA workflow

Full mtDNA haplotypes could be recovered from fired cartridge cases using the mtDNA MPS workflow. By increasing the loading DNL concentration the sequencing performance of the final mtDNA sequencing run improved. Cluster density can be used to monitor sequencing performance, as the target for a well-balanced libraries using the 600 cycle MiSeq Reagent Kit v3 is between 1200 to 1400 K/mm^2 (138). All of the cluster densities obtained in this research were under clustered with no sequencing run reaching this target range, minimising the total data output able to be obtained (167). Further optimisation of the loading concentration is therefore recommended.

Optimisation of the number of PCR cycles should be undertaken and may help minimise stochastic amplification of mtDNA damage (130). As 30 cycles was used in this research, decreasing the number of samples could minimise the high number of non-concordant minor variants sequenced in the fired cartridge case samples.

In this research, non-concordant variants were more likely to result from a nucleotide transition. In both fired and unfired samples, this was most commonly a pyrimidine transition (either a C to T or T to C change). C to T transitions are characteristic of cytosine deamination and can be repaired by the NEBNext® FFPE DNA repair mix (New England BioLabs®, Ipswich, Massachusetts, United States) (130,168). Implementing a DNA repair step after DNA extraction could reduce the number of non-concordant minor variants. A different approach would be to ignore minor variants and call haplotypes based solely on major variants. Only three non-concordant major variants were detected throughout this entire research (4810T, 12549A, 14244T) that when uploaded to Haplogrep3 were classified as global private mutations and did not impact haplogroup estimation. Furthermore, when using replicates at the amplification level these variants would not be reproducible; therefore, it is recommended that at least two replicates for each sample are processed, with any major variant seen in only one replicate (and was not in a region of dropout in the second), then removed from the final haplotype. It would be up to each independent forensic laboratory to determine whether to include or exclude minor variants in casework practices.

Previous research has recorded potential damage hotspots (nucleotide positions that had damage recorded more than three times) in the mitogenome CR (157). As the targeting of the whole mitogenome as a genetic marker has become more accessible with the advent of MPS (90), future research could identify nucleotide positions that are potential damage hotspots throughout the entire mitogenome. This would be a powerful tool for helping forensic practitioners with the differentiation of heteroplasmy and damage, especially since damage hotspots were the only instance of damage replicated in duplicate amplification (157).

6.2.3 Recommendations for the ForenSeq™ workflow

Future research could test Primer Mix B (DPMB) in the ForenSeq™ Signature Prep Kit to determine whether the use of additional sets of SNPs (56 aiSNPs and 22 piSNPs) would see the typing of more concordant SNP loci in comparison to a reference profile. The amplicon length (bps) of each target SNP in all three kinds of SNPs are of relatively similar sizes (74), and in lower input DNA concentrations DPMB has been found to produce higher read counts compared to DPMA (73). Jäger et al. (57) found that in gDNA inputs of 7.82 pg and 15.625 pg, genotype concordance of typed aiSNPs and piSNPs were higher compared to iiSNPs. As gDNA inputs into ForenSeq™ workflow ranged from UD to 14.6 pg, this suggests that more concordant results could be obtained using the aiSNP and piSNPs in DPMB.

As the majority of the DNA fragments in these library samples were adapter-dimers (Section 4.3.2.2 and 5.3.2.1), a second purification should be tested in the ForenSeq™ workflow. Additionally, as a-STR loci were typed in both the unfired and fired samples, future research should trial uploading these profiles to a probabilistic genotyping software such as STRmix™ NGS (169). This would be especially useful to help determine if non-concordant STR loci (resulting from allele drop-ins and sequencing errors) can be resolved.

6.3 CONCLUDING REMARKS

Where further optimisation and validation of both sequencing workflows is still required for this evidence type, these results suggest that the whole mitogenome and STR/iiSNP alleles could be targeted in conjunction. Where targeting mtDNA resulted in a higher percentage of the reference haplotype, mtDNA is only discriminatory based on the maternal line (87). Discrimination between maternal relatives is possible through low-levels of heteroplasmy (109), however, this is not present in a majority of the population (93). Additionally, 99.6% of non-concordant variants obtained from fired cartridge cases were minor, resulting in calls of heteroplasmy that were not expected when compared to the reference profile. For this evidence type it is recommended that minor variants are ignored for haplotype generation. As partial iiSNPs and STR genotypes were able to be obtained, these can be used to help differentiate between individuals of the same maternal line.

In summary, MPS technologies are a successful alternative for obtaining DNA profiles from unfired cartridges and fired cartridge cases. Targeting both STR and iiSNPs, and the whole mitogenome through a multiplex of short amplicons is highly sensitive and able to obtain genetic information from extremely low quantities of DNA. Based on the results of this research, forensic practitioners should strongly consider implementing MPS technologies for the biological analysis of unfired cartridges and cartridge cases.

7 APPENDICES

7.1 ETHICS DOCUMENTS

7.1.1 Participant Information Sheet

SCHOOL OF CHEMICAL SCIENCES
Faculty of Science Centre,
Building 301
23 Symonds St,
Auckland
Ph. (09) 373 7599

PARTICIPATION INFORMATION SHEET (PIS)

Alternative DNA technologies for obtaining DNA profiles from cartridge cases

Research Team

Student Researcher: Teigan Tarapi - MSc Candidate in Forensic Science, School of Chemical Sciences at the University of Auckland.

Principal Investigator: Dr SallyAnn Harbison - Director of the Forensic Science Programme at the University of Auckland and Senior Science Leader/DNA Technical Leader, Forensic Biology Group at ESR Ltd.

Co-supervisor: Bethany Forsythe - Senior Technician, Forensic Biology Team at ESR Ltd.

Invitation

You are invited to participate in a University of Auckland, Master of Science research project in collaboration with the Institute of Environmental Science and Research Ltd. (ESR), Mt Albert, Auckland. You have received this invitation after reading the summary that was emailed to you where you have decided that you have an interest in participating. Please read the following information carefully, and if you have any questions or you wish to have further information, please contact myself or one of my supervisors.

Background

Firearm-associated crime in New Zealand has increased in recent years. One common type of forensic evidence that can be found at a firearm crime scene is a fired cartridge case. Being able to obtain a DNA profile from a fired cartridge case is desirable in forensic investigations to link an individual to an offence. However, the use of standard nuclear DNA profiling using short tandem repeats (STRs) and capillary electrophoresis (CE) technologies has proven to be unsuccessful. This is due to low amounts of DNA transferred to the

SCHOOL OF CHEMICAL SCIENCES
Faculty of Science Centre,
Building 301
23 Symonds St,
Auckland
Ph. (09) 373 7599

cartridge case upon handling, degradation through exposure to high temperatures during the firing process and inhibitory chemical interactions between nucleic acids and metallic cartridge surfaces.

Massively parallel sequencing (MPS) technologies allow the analysis of many short amplicons simultaneously with greater sensitivity, outperforming CE workflows for heavily degraded DNA samples, such as DNA recovered from fired cartridge cases. The Precision ID mtDNA Whole Genome Panel is a recently developed amplification panel from Thermo Fisher Scientific (Waltham, MA, USA). The panel is specifically designed for the amplification and massively parallel sequencing (MPS) of whole mitochondrial genomes from degraded samples. Similarly, The ForenSeq DNA Signature Prep Kit is part of an MPS workflow by Verogen Inc. (San Diego, CA, USA) that sees the amplification of both STRs and single nucleotide polymorphisms (SNPs), using shorter amplicons than CE workflows. The ForenSeq DNA Signature Prep Kit has been successfully validated for forensic application, and it has been shown that ForenSeq outperforms standard CE kits for degraded samples.

The use of the Precision ID Whole mtDNA Genome Panel or the ForenSeq DNA Signature Prep Kit on fired cartridge cases has not yet been investigated at ESR or in literature. This research project will help determine whether these alternative DNA technologies can successfully generate DNA profiles from both unfired and fired cartridge cases.

Participant Procedures

This research requires both a buccal sample and a DNA sample. Buccal cells will be collected by the participant swabbing the inside of their mouth. DNA samples will be collected from .223 Remington cartridge cases after having been handled by the participant. You will be asked to repeat the sample if necessary.

By participating in this study and signing the consent form you are agreeing to handle the provided .223 Remington cartridge cases and provide a buccal sample. To maintain your anonymity, samples will be collected and coded alphanumerically by the third-party representative Leah Tottey before being stored securely. This is to ensure that the link between you and your DNA profile remains completely anonymous.

SCHOOL OF CHEMICAL SCIENCES
Faculty of Science Centre,
Building 301
23 Symonds St,
Auckland
Ph. (09) 373 7599

Participation will involve attendance to an initial sample collection session where reference buccal swabs will be collected, and the handling of cartridge cases will occur. You will be provided with a deposition instruction sheet, buccal swab, and .223 Remington cartridges. Kindly follow the instructions on the instruction sheet. After the procedure, please return the samples as instructed.

Depending on the success of the research, participants may be invited back for a maximum of four sample collection sessions. The participants will handle .223 Remington cartridges and place them into the chamber of a bolt action rifle under the supervision of a trained ESR firearm examiner. A deposition instruction sheet and the .223 Remington cartridges will be provided. Again, please follow the instructions on the instruction sheet and after the procedure please return the samples as instructed.

The collection of all samples should take less than 2 hours during any of the collection sessions.

Participant rights

Each participant in the study will be randomly assigned an alphanumeric identification code to de-identify their donated samples. This information will only be accessible by Leah Tottey. Leah Tottey is an ESR employee and is suitably independent to you (the participant in this study) as she is not a part of the research team and therefore will not be involved in any of the experimental analyses associated with your samples/data. Consent forms will be kept in a locked cabinet at the Institute of Environmental and Science Research Ltd. for six years. At the end of the six years, they will be destroyed by the Principal Investigator (Dr SallyAnn Harbison).

Information and samples given by you will not be used for any additional research purposes beyond the scope of this project. Any samples that are supplied by you will be destroyed during the course of the research, no later than December 2024. If you would like your samples returned to you at the end of the research project, please advise Leah Tottey. Data generated in this research will be used for the completion of a MSc thesis by Teigan Tarapi and may subsequently be published in a scientific journal and/or be used in the development of a commercial service but will be untraceable to you. Your DNA profile, if successfully generated during this study, will not be uploaded to any public DNA databases.

SCHOOL OF CHEMICAL SCIENCES
Faculty of Science Centre,
Building 301
23 Symonds St,
Auckland
Ph. (09) 373 7599

Participation in this research project does involve the handling of live ammunition. This will be done under the close supervision of a trained ESR firearm examiner. Participation does not involve you firing the firearm nor is there any way the live ammunition can be fired while being handled; therefore, this research does not carry any significant risk.

If you have an adverse reaction to any of the surfaces, halt the experiment immediately and wash your hands with soap and water. Contact details of Leah Tottey (the third-party representative) are provided below if you need further assistance.

Both the nuclear genome and mitochondrial genome contain potential personal health information sites that can be associated with disease. This information is not needed in forensic investigations or sought after in this study. The nuclear STRs and SNPs targeted during the ForenSeq MPS workflow were selected and designed by the manufacturer to avoid any personal health information. The regions of the mitochondrial genome that are known and confirmed to be linked to disease will be bioinformatically masked during alignment of the raw sequencing data regardless of whether a disease linked variant is present or not. This will mean that any personal health information in an individual's mitochondrial haplotype will not be visible to the researchers at any point. If any other unforeseen incidental findings are observed, the students' supervisors will seek guidance appropriate for the finding and take appropriate action. It is important to note that if you do not want to be made aware of any incidental findings then you should not participate in this study.

As a voluntary participant, you have the right to withdraw from this project at any stage without explanation **up to three months** after sample collection. If you do this, donated sample(s) and any subsamples such as DNA extracts will be destroyed. All information/data acquired from the sample(s) will also be destroyed. If you would like your samples/any data from the samples returned to you upon withdrawing, please contact Leah Tottey.

Participation is voluntary, so non-participation will in no way affect your relationship with the Institute of Environmental and Science Research Ltd or the University of Auckland (if you are a student). This has been assured in writing by both the General Manager Forensic, John Bone and the Head of School of Chemical Sciences, Professor Duncan McGillivray.

SCHOOL OF CHEMICAL SCIENCES
Faculty of Science Centre,
Building 301
23 Symonds St,
Auckland
Ph. (09) 373 7599

Upon completion of this research study, the summary of the results will be available to all participants, and you may choose to request a copy by ticking the appropriate box and providing your email address on the consent form. In addition, results will be found in the final thesis of Teigan Tarapi, which will be available for viewing at both the University of Auckland and the library located at the Institute of Environmental and Science Research Ltd., Mt Albert.

Spiritual/Cultural Beliefs:

You may hold beliefs about a sacred and shared value of all or any tissue samples you agree to provide, and the data generated from them. The cultural issues associated with this should be discussed with your family/whanau as appropriate. There are a range of views held by Māori around these issues; some iwi disagree with the collection and storage of samples citing whakapapa and advise their people to consult before participating in research where this occurs. However, it is acknowledged that individuals have their right to choose. If you are Māori and would like to request a specific tikanga (Māori custom) process, please feel free to talk to the third-party representative (Leah Tottey)

Thank you for taking the time to read over this participation information sheet and for your consideration in voluntarily taking part in this research. In the event you have any further questions or concerns regarding your participation in this study, please feel free to contact one of the researchers below.

Researcher: Teigan Tarapi

MSc student (Forensic Science) University of Auckland

Email: ttar053@aucklanduni.ac.nz
Teigan.Tarapi@esr.cri.nz

SCHOOL OF CHEMICAL SCIENCES
Faculty of Science Centre,
Building 301
23 Symonds St,
Auckland
Ph. (09) 373 7599

Co-supervisor: Bethany Forsythe

The Institute of Environmental Science and Research, Mt Albert Science Centre, Auckland
Email: Bethany.Forsythe@esr.cri.nz

Principal investigator: Dr SallyAnn Harbison

The Institute of Environmental Science and Research, Mt Albert Science Centre, Auckland
Email: SallyAnn.Harbison@esr.cri.nz

Head of School of Chemical Sciences: Professor Duncan McGillivray

University of Auckland
Ph: +64 9 923 8338
Email: d.mcgillivray@auckland.ac.nz

Third party representative: Leah Tottey

The Institute of Environmental Science and Research, Mt Albert Science Centre, Auckland
Email: Leah.Tottey@esr.cri.nz

For concerns of an ethical nature, you can contact the Chair of the Auckland Health Research Ethics Committee at **ahrec@auckland.ac.nz** or at **09 373 7599 extn 83711**, or at **Auckland Health Research Ethics Committee, The University of Auckland, Private Bag 92019, Auckland 1142.**

Approved by the Auckland Health Research Ethics Committee on 27/04/2023 for three years. Reference number AH25804.

7.1.2 Consent form

SCHOOL OF CHEMICAL SCIENCES
Faculty of Science Centre,
Building 301
23 Symonds St,
Auckland
Ph. (09) 373 7599

PARTICIPANT CONSENT FORM

[This will be held for a period of 6 years]

Alternative DNA technologies for obtaining DNA profiles from cartridge cases

Research Team

Student Researcher: Teigan Tarapi - MSc Candidate in Forensic Science, School of Chemical Sciences at the University of Auckland.

Principal Investigator: Dr SallyAnn Harbison - Director of the Forensic Science Programme at the University of Auckland and Senior Science Leader/DNA Technical Leader, Forensic Biology Group at ESR Ltd.

Co-supervisor: Bethany Forsythe - Senior Technician, Forensic Biology Team at ESR Ltd.

Consent is sought for the following:

1) Consent for a reference buccal sample.
2) Consent for a DNA and a mtDNA touch sample to be collected (after the handling of a .223 Remington cartridge case), sampled and used in research.

Please read the following and fill in the information if that applies to you

- I have read the Participant Information Sheet describing this research project, and I understand its contents.
- I have had the opportunity to ask questions and have had my questions answered to my satisfaction.
- I understand that I can ask further questions and have my questions answered by any of the following researchers listed above.

SCHOOL OF CHEMICAL SCIENCES
Faculty of Science Centre,
Building 301
23 Symonds St,
Auckland
Ph. (09) 373 7599

- I understand that my participation is completely voluntary and if appropriate, I have consulted with my family/whānau.
- I understand that participation and non-participation as a staff member or student will in no way affect my relationship with the Institute of Environmental Science and Research Ltd and the University of Auckland. The General Manager Forensic, John Bone (ESR Ltd.) and Head of School of Chemical Sciences, Professor Duncan McGillivray, both assure this in writing.
- I have been informed of my right to withdraw my participation without explanation **up to three months** after sample collection, and I have been given information on how to make an application to withdraw my sample and any data associated with my samples.
- I understand that by withdrawing my participation, my donated sample(s) and any subsamples such as DNA extracts will be destroyed. All information/data acquired from the sample(s) will also be destroyed. I understand that I can contact the third-party representative (Leah Tottey) if I would like any samples/data obtained from the samples to be returned to me.
- I understand that this research will require the collection of the specified biological materials noted below and that DNA profiles produced from this will be used in this research.
- I understand that this research requires the donation of buccal samples and the handling of a .223 Remington cartridge case. I understand that the DNA and mtDNA profiles produced from this will be used in this research.
- I understand that my identity as a participant will remain confidential and that all of my samples will be destroyed at the completion of the thesis unless I request to the third-party representative (Leah Tottey) for my samples to be returned to me.
- I understand that the samples obtained in this research will be used for only this named project.
- I understand that the data generated from these samples may be used for the publication of scientific papers in peer-reviewed journals, at conferences and the production of MSc thesis directly related to this project. However, I understand that no data or information will be published that is identifiable back to me.
- I understand that any samples submitted for this research will not be used for any future research.
- I understand that the data obtained from my samples for this research project will be stored electronically at ESR indefinitely.

SCHOOL OF CHEMICAL SCIENCES
Faculty of Science Centre,
Building 301
23 Symonds St,
Auckland
Ph. (09) 373 7599

Please tick the box if you consent to the following:

☐ I want to receive a summary of the findings, which can be emailed to me at the address provided below.

☐ I consent to being contacted via the email address provided below, if there are further samples needed for this research project only.

For participants

Full Name (printed clearly):

Date:

Signature:

Email address (if you have ticked the box):

Approved by the Auckland Health Research Ethics Committee on 27/04/2023 for three years. Reference number AH25804.

7.2 RECOVERY OF FIRED CARTRIDGE CASES

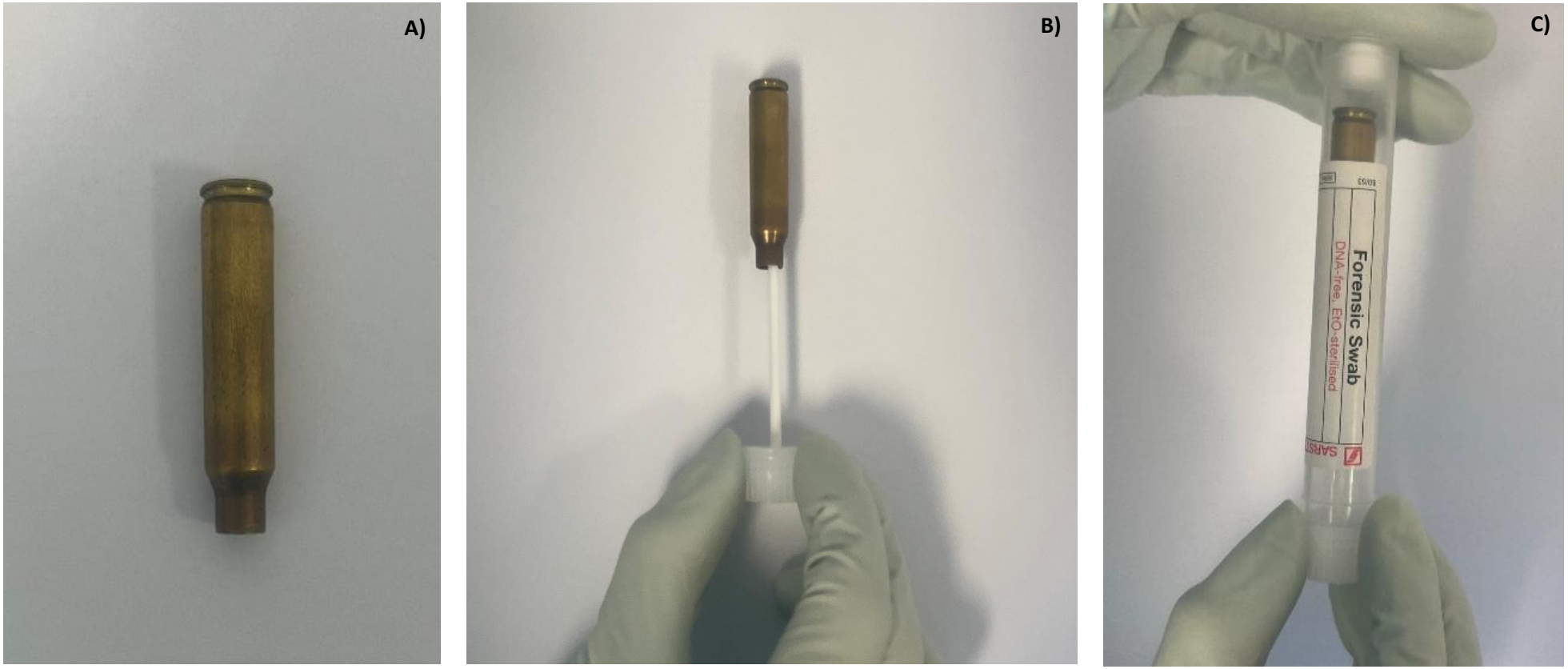

Figure 7.1 Recovery of a fired .223 Rem cartridge case using a Forensic Swab. ***A)*** *A fired cartridge case was deposited onto a clean white piece of paper.* ***B)*** *The swab head was inserted into the fired cartridge case and lifted the cartridge case off the paper by a gloved firearms expert.* ***C)*** *The swab casing was placed back over the swab head, sealed and the entire swab containing recovered cartridge case was placed into an envelope.*

7.3 CONFIRMED PERSONAL HEALTH INFORMATION SITES

Personal Health Information (PHI) sites			
583	3902	8313	12258
616	3904	8340	12276
1494	3905	8344	12294
1555	3906	8356	12315
1606	3908	8363	12316
1630	4171	8528	12706
1644	4298	8851	13042
3243	4300	8969	13051
3256	4308	8993	13094
3258	4332	9035	13513
3260	4450	9155	13514
3271	5521	9176	14459
3280	5537	9185	14482
3291	5650	9205	14484
3302	5690	9206	14487
3303	5703	10010	14495
3376	5728	10158	14568
3460	7445	10191	14674
3635	7471	10197	14709
3697	7497	10663	14710
3700	7510	11777	14849
3733	7511	11778	14864
3890	8306	12147	15579

REFERENCES

1. New Zealand Police. Firearms Information Summary - 1 June 2023 [Internet]. 2023 Jun [cited 2023 Jun 12]. Available from: https://www.police.govt.nz/about-us/publication/firearms-data
2. Redstall S. Shop workers threatened with gun, hatchet, during man's weeks-long crime spree. Stuff [Internet]. 2023 May 17 [cited 2023 Jun 13]; Available from: https://www.stuff.co.nz/national/crime/132069384/shop-workers-threatened-with-gun-hatchet-during-mans-weekslong-crime-spree
3. RNZ. Man shot, search underway after serious incident in central Christchurch. RNZ [Internet]. 2022 Aug 30 [cited 2023 Jun 13]; Available from: https://www.rnz.co.nz/news/national/473754/man-shot-search-underway-after-serious-incident-in-central-christchurch
4. Holland MM, Bonds RM, Holland CA, McElhoe JA. Recovery of mtDNA from unfired metallic ammunition components with an assessment of sequence profile quality and DNA damage through MPS analysis. Forensic Sci Int Genet. 2019 Mar 1;39:86–96.
5. Butler JM. Advanced topics in forensic DNA typing: methodology. Academic press; 2011.
6. Horsman-Hall KM, Orihuela Y, Karczynski SL, Davis AL, Ban JD, Greenspoon SA. Development of STR profiles from firearms and fired cartridge cases. Forensic Sci Int Genet. 2009 Sep;3(4):242–50.
7. Jansson L, Forsberg C, Akel Y, Dufva C, Ansell C, Ansell R, et al. Factors affecting DNA recovery from cartridge cases. Forensic Sci Int Genet. 2020 Sep 1;48.
8. Polley D, Mickiewicz P, Vaughn M, Miller T, Warburton R, Komonski D, et al. An investigation of DNA recovery from firearms and cartridge cases. Journal of the Canadian Society of Forensic Science. 2006;39(4):217–28.
9. Churchill JD, Schmedes SE, King JL, Budowle B. Evaluation of the Illumina® Beta Version ForenSeq™ DNA Signature Prep Kit for use in genetic profiling. Forensic Sci Int Genet. 2016 Jan 1;20:20–9.
10. Arms Act 1983 (N.Z.).
11. Saferstein R. Criminalistics: An Introduction to Forensic Science, 11th Edition. Pearson; 2015.
12. McElhoe J, Mandracchia T, Bille T, Holland M. Impact of storage conditions and time on DNA yield from ammunition cartridges. Int J Legal Med. 2023;
13. Barnes FC, Skinner S. Cartridges of the World: 10th Edition, Revised and Expanded. 2003.
14. Lee HC, Ladd C, Scherczinger CA, Bourke MT. Forensic Applications of DNA Typing: Part 2: Collection and Preservation of DNA Evidence. Am J Forensic Med Pathol [Internet]. 1998;19(1). Available from: https://journals.lww.com/amjforensicmedicine/Fulltext/1998/03000/Forensic_Applications_of_DNA_Typing__Part_2_.2.aspx

15. Fan H, Chu JY. A Brief Review of Short Tandem Repeat Mutation. Genomics Proteomics Bioinformatics. 2007 Jan 1;5(1):7–14.
16. Butler JM. Fundamentals of Forensic DNA Typing [Internet]. Butler JM, editor. Fundamentals of Forensic DNA Typing. San Diego: Academic Press; 2010 [cited 2023 Jun 18]. Available from: https://www.sciencedirect.com/science/article/pii/B9780123749994000023
17. Gymrek M. A genomic view of short tandem repeats. Curr Opin Genet Dev. 2017 Jun 1;44:9–16.
18. Wickenheiser RA. Trace DNA: a review, discus-sion of theory, and application of the transfer of trace quantities of DNA through skin contact. 2002.
19. van Oorschot RAH, Jones MK. DNA fingerprints from fingerprints. 1997;
20. Locard E. L'enquête criminelle et les méthodes scientifiques. E. Flammarion; 1920.
21. Burrill J, Daniel B, Frascione N. A review of trace "Touch DNA" deposits: Variability factors and an exploration of cellular composition. Forensic Sci Int Genet. 2019 Mar 1;39:8–18.
22. van Oorschot RAH, Szkuta B, Meakin GE, Kokshoorn B, Goray M. DNA transfer in forensic science: A review. Vol. 38, Forensic Science International: Genetics. Elsevier Ireland Ltd; 2019. p. 140–66.
23. Sessa F, Salerno M, Bertozzi G, Messina G, Ricci P, Ledda C, et al. Touch DNA: Impact of handling time on touch deposit and evaluation of different recovery techniques: An experimental study. Sci Rep. 2019;9(1).
24. Quinones I, Daniel B. Cell free DNA as a component of forensic evidence recovered from touched surfaces. Forensic Sci Int Genet. 2012 Jan 1;6(1):26–30.
25. Burrill J, Rammenou E, Alawar F, Daniel B, Frascione N. Corneocyte lysis and fragmented DNA considerations for the cellular component of forensic touch DNA. Forensic Sci Int Genet. 2021 Mar 1;51.
26. Bonsu DOM, Higgins D, Austin JJ. Forensic touch DNA recovery from metal surfaces – A review. Vol. 60, Science and Justice. Forensic Science Society; 2020. p. 206–15.
27. Kita T, Yamaguchi H, Yokoyama M, Tanaka T, Tanaka N. Morphological study of fragmented DNA on touched objects. Forensic Sci Int Genet. 2008 Dec 1;3(1):32–6.
28. Lowe A, Murray C, Whitaker J, Tully G, Gill P. The propensity of individuals to deposit DNA and secondary transfer of low level DNA from individuals to inert surfaces. 2002.
29. Martin B, Blackie R, Taylor D, Linacre A. DNA profiles generated from a range of touched sample types. Forensic Sci Int Genet. 2018 Sep 1;36:13–9.
30. Lee LYC, Tan J, Lee YS, Syn CKC. Shedder status—An analysis over time and assessment of various contributing factors. J Forensic Sci. 2023;
31. Phipps M, Petricevic S. The tendency of individuals to transfer DNA to handled items. Forensic Sci Int. 2007 May 24;168(2–3):162–8.

32. Goray M, Fowler S, Szkuta B, Van Oorschot RAH. Shedder status - An analysis of self and non-self DNA in multiple handprints deposited by the same individuals over time. Forensic Sci Int Genet. 2016 Jul 1;23:190–6.
33. Szkuta B, Ballantyne KN, van Oorschot RAH. Transfer and persistence of DNA on the hands and the influence of activities performed. Forensic Sci Int Genet. 2017 May 1;28:10–20.
34. Montpetit S. Obtaining DNA from ammunition: A review. WIREs Forensic Science [Internet]. 2020 Mar 1;2(2):e1352. Available from: https://doi.org/10.1002/wfs2.1352
35. Buckingham AK, Harvey ML, van Oorschot RAH. The origin of unknown source DNA from touched objects. Forensic Sci Int Genet. 2016 Nov 1;25:26–33.
36. Daly DJ, Murphy C, McDermott SD. The transfer of touch DNA from hands to glass, fabric and wood. Forensic Sci Int Genet. 2012 Jan;6(1):41–6.
37. Wood I, Park S, Tooke J, Smith O, Morgan RM, Meakin GE. Efficiencies of recovery and extraction of trace DNA from non-porous surfaces. Forensic Sci Int Genet Suppl Ser. 2017;6:e153–5.
38. Fonneløp AE, Egeland T, Gill P. Secondary and subsequent DNA transfer during criminal investigation. Forensic Sci Int Genet. 2015;17:155–62.
39. Tobias SHA, Jacques GS, Morgan RM, Meakin GE. The effect of pressure on DNA deposition by touch. Forensic Sci Int Genet Suppl Ser. 2017 Dec 1;6:e12–4.
40. Moore D, Beaumont D, Brown M, Clayton T, Coleman K, Subhani Z, et al. An investigation of two methods of DNA recovery from fired and unfired 9 mm ammunition. Science and Justice. 2021 Mar 1;61(2):160–9.
41. Dieltjes P, Mieremet R, Zuniga S, Kraaijenbrink T, Pijpe J, De Knijff P. A sensitive method to extract DNA from biological traces present on ammunition for the purpose of genetic profiling. Int J Legal Med. 2011 Jul;125(4):597–602.
42. Kuffel A, Gray A, Daeid NN. Impact of metal ions on PCR inhibition and RT-PCR efficiency. Int J Legal Med. 2021 Jan 1;135(1):63–72.
43. Prasad E, Barash M, Hitchcock C, van Oorschot RAH, Raymond J, McNevin D, et al. Evaluation of soaking to recover trace DNA from fired cartridge cases. Australian Journal of Forensic Sciences [Internet]. 2021 Sep 3;53(5):512–22. Available from: https://doi.org/10.1080/00450618.2020.1757758
44. Prasad E, Hitchcock C, Raymond J, Cole A, Barash M, Gunn P, et al. DNA recovery from unfired and fired cartridge cases: A comparison of swabbing, tape lifting, vacuum filtration, and direct PCR. Forensic Sci Int [Internet]. 2020 Dec 1 [cited 2023 Jun 12];317:110507. Available from: https://www.sciencedirect.com/science/article/pii/S0379073820303698
45. Elwick K, Gauthier Q, Rink S, Cropper E, Kavlick MF. Recovery of DNA from fired and unfired cartridge casings: comparison of two DNA collection methods. Forensic Sci Int Genet. 2022 Jul 1;59:102726.

46. Czado N, Houston R, Hughes S. Comparison of metal ions recovered during DNA analysis of brass ammunition and effects of copper and zinc ions on DNA profiling. Forensic Sci Int Genet Suppl Ser. 2022 Dec 1;8:162–4.
47. Burkitt MJ. Copper-DNA adducts. In: Methods in enzymology. Elsevier; 1994. p. 66–79.
48. Govindaraju M, Shekar HS, Sateesha SB, Vasudeva Raju P, Sambasiva Rao KR, Rao KSJ, et al. Copper interactions with DNA of chromatin and its role in neurodegenerative disorders. J Pharm Anal. 2013;3(5):354–9.
49. Gashi B, Edwards MR, Sermon PA, Courtney L, Harrison D, Xu Y. Measurement of 9mm cartridge case external temperatures and its forensic application. Forensic Sci Int. 2010 Jul;200(1–3):21–7.
50. Montpetit S, O'Donnell P. An optimized procedure for obtaining DNA from fired and unfired ammunition. Forensic Sci Int Genet. 2015;17:70–4.
51. Hartless S, Walton-Williams L, Williams G. Critical evaluation of touch DNA recovery methods for forensic purposes. Forensic Sci Int Genet Suppl Ser. 2019 Dec 1;7(1):379–80.
52. Bille TW, Fahrig G, Weitz SM, Peiffer GA. An improved process for the collection and DNA analysis of fired cartridge cases. Forensic Sci Int Genet. 2020 May 1;46:102238.
53. Prasad E, Van der Walt L, Cole A, van Oorschot RAH, Barash M, Gunn P, et al. The effects of soaking for DNA recovery on the striation patterns of fired cartridge cases. Australian Journal of Forensic Sciences. 2019 Jul 29;51(sup1):S35–8.
54. Barrot C, Moreno C, Sánchez C, Rodríguez M, Ortega M, Huguet E, et al. Comparison of Identifiler®, Identifiler Plus® and Minifiler® performance in an initial paternity testing study on old skeletal remains at the forensic and legal medicine area of the Government of Andorra (Pyrenees). Forensic Sci Int Genet Suppl Ser. 2011 Dec;3(1).
55. Sharma V, van der Plaat DA, Liu Y, Wurmbach E. Analyzing degraded DNA and challenging samples using the ForenSeq™ DNA Signature Prep kit. Science and Justice. 2020 May 1;60(3):243–52.
56. Xavier C, Parson W. Evaluation of the Illumina ForenSeq™ DNA Signature Prep Kit – MPS forensic application for the MiSeq FGx™ benchtop sequencer. Forensic Sci Int Genet. 2017 May 1;28:188–94.
57. Jäger AC, Alvarez ML, Davis CP, Guzmán E, Han Y, Way L, et al. Developmental validation of the MiSeq FGx Forensic Genomics System for Targeted Next Generation Sequencing in Forensic DNA Casework and Database Laboratories. Forensic Sci Int Genet. 2017 May 1;28:52–70.
58. Köcher S, Müller P, Berger B, Bodner M, Parson W, Roewer L, et al. Inter-laboratory validation study of the ForenSeq™ DNA Signature Prep Kit. Forensic Sci Int Genet. 2018 Sep 1;36:77–85.
59. Børsting C, Morling N. Next generation sequencing and its applications in forensic genetics. Forensic Sci Int Genet. 2015 Sep 1;18:78–89.
60. Hollard C, Ausset L, Chantrel Y, Jullien S, Clot M, Faivre M, et al. Automation and developmental validation of the ForenSeq ™ DNA Signature Preparation kit for high-throughput analysis in forensic laboratories. Forensic Sci Int Genet. 2019 May 1;40:37–45.

61. Hussing C, Huber C, Bytyci R, Mogensen HS, Morling N, Børsting C. Sequencing of 231 forensic genetic markers using the MiSeq FGx™ forensic genomics system – an evaluation of the assay and software. Forensic Sci Res. 2018 Apr 3;3(2):111–23.
62. Sobrino B, Carracedo A. SNP typing in forensic genetics: a review. In: Forensic DNA typing protocols [Internet]. Springer; 2005 [cited 2023 Jun 18]. p. 107–26. Available from: https://link.springer.com/protocol/10.1385/1-59259-867-6:107
63. Budowle B, Van Daal A. Forensically relevant SNP classes. Vol. 44, BioTechniques. 2008. p. 603–10.
64. Mehta B, Daniel R, Phillips C, McNevin D. Forensically relevant SNaPshot® assays for human DNA SNP analysis: a review. Vol. 131, International Journal of Legal Medicine. Springer Verlag; 2017. p. 21–37.
65. Butler JM, Coble MD, Vallone PM. STRs vs. SNPs: Thoughts on the future of forensic DNA testing. Forensic Sci Med Pathol. 2007 Sep;3(3):200–5.
66. Ballard D, Winkler-Galicki J, Wesoły J. Massive parallel sequencing in forensics: advantages, issues, technicalities, and prospects. Vol. 134, International Journal of Legal Medicine. Springer; 2020. p. 1291–303.
67. Rothberg JM, Hinz W, Rearick TM, Schultz J, Mileski W, Davey M, et al. An integrated semiconductor device enabling non-optical genome sequencing. Nature. 2011 Jul 21;475(7356):348–52.
68. Illumina. An introduction to Next-Generation Sequencing Technology [Internet]. 2017. Available from: www.illumina.com/technology/next-generation-sequencing.html
69. England R, Harbison S. A review of the method and validation of the MiSeq FGx™ Forensic Genomics Solution. Wiley Interdisciplinary Reviews: Forensic Science. 2020;2(1):e1351.
70. Verogen. Verogen MiSeq FGx Forensic Genomics Solution [Internet]. 2018. Available from: www.verogen.com/products
71. Guo F, Yu J, Zhang L, Li J. Massively parallel sequencing of forensic STRs and SNPs using the Illumina® ForenSeq™ DNA Signature Prep Kit on the MiSeq FGx™ Forensic Genomics System. Forensic Sci Int Genet. 2017 Nov 1;31:135–48.
72. England R. The development and validation of massively parallel sequencing marker panels for use within a New Zealand population in forensic science [Doctoral dissertation, ResearchSpace@ Auckland]. 2022.
73. Senst A, Caliebe A, Scheurer E, Schulz I. Validation and beyond: Next generation sequencing of forensic casework samples including challenging tissue samples from altered human corpses using the MiSeq FGx system. J Forensic Sci. 2022 Jul 1;67(4):1382–98.
74. Verogen. ForenSeq DNA Signature Prep Kit Reference Guide. Document # VD2018005 Rev. D. 2022.
75. Thanakiatkrai P, Phetpeng S, Sotthibandhu S, Asawutmangkul W, Piwpankaew Y, Foong JE, et al. Performance comparison of MiSeq forensic genomics system and STR-CE using control and mock IED samples. Forensic Sci Int Genet Suppl Ser. 2017 Dec 1;6:e320–1.

76. Lutz S, Weisser -H.-J, Heizmann • J, Pollak • S. mtDNA as a tool for identification of human remains Identification using mtDNA. Vol. 109, Int J Legal Med. 1996.
77. Parsons TJ, Coble MD. Increasing the forensic discrimination of mitochondrial DNA testing through analysis of the entire mitochondrial DNA genome. Croat Med J. 2001;42(3):304–9.
78. Holland MM, Parsons TJ. Mitochondrial DNA Sequence Analysis-Validation and Use for Forensic Casework. Forensic Sci Rev. 1999;11(1):21–50.
79. Nass MM, Nass S. INTRAMITOCHONDRIAL FIBERS WITH DNA CHARACTERISTICS. I. FIXATION AND ELECTRON STAINING REACTIONS. J Cell Biol [Internet]. 1963;19:593–611. Available from: http://europepmc.org/abstract/MED/14086138
80. Budowle B, Allard MW, Wilson MR, Chakraborty R. Forensics and Mitochondrial DNA: Applications, Debates, and Foundations. Annu Rev Genomics Hum Genet [Internet]. 2003 Sep 1;4(1):119–41. Available from: https://doi.org/10.1146/annurev.genom.4.070802.110352
81. Butler JM. Chapter 14 - Mitochondrial DNA Analysis. In: Butler JM, editor. Advanced Topics in Forensic DNA Typing: Methodology [Internet]. San Diego: Academic Press; 2012. p. 405–56. Available from: https://www.sciencedirect.com/science/article/pii/B9780123745132000142
82. Forsythe B, Melia L, Harbison S. Methods for the analysis of mitochondrial DNA. WIREs Forensic Science [Internet]. 2021 Jan 1;3(1):e1388. Available from: https://doi.org/10.1002/wfs2.1388
83. Stoneking M. Hypervariable Sites in the mtDNA Control Region Are Mutational Hotspots. The American Journal of Human Genetics. 2000 Oct 1;67(4):1029–32.
84. Greenberg BD, Newbold JE, Sugino A. Intraspecific nucleotide sequence variability surrounding the origin of replication in human mitochondrial DNA. Gene. 1983 Jan 1;21(1–2):33–49.
85. Satoh M, Kuroiwa T. Organization of multiple nucleoids and DNA molecules in mitochondria of a human cell. Exp Cell Res. 1991;196(1):137–40.
86. Melton T, Holland C, Holland M. Forensic mitochondria DNA analysis: current practice and future potential. Forensic Sci Rev. 2012;24(2):101.
87. Giles RE, Blanc H, Cann HM, Wallace DC. Maternal inheritance of human mitochondrial DNA. Proceedings of the National academy of Sciences. 1980;77(11):6715–9.
88. Canale LC, Parson W, Holland MM. The time is now for ubiquitous forensic mtMPS analysis. Wiley Interdisciplinary Reviews: Forensic Science. 2022;4(1):e1431.
89. King JL, Larue BL, Novroski NM, Stoljarova M, Seo SB, Zeng X, et al. High-quality and high-throughput massively parallel sequencing of the human mitochondrial genome using the Illumina MiSeq. Forensic Sci Int Genet. 2014 Sep 1;12:128–35.
90. Chaitanya L, Ralf A, van Oven M, Kupiec T, Chang J, Lagacé R, et al. Simultaneous Whole Mitochondrial Genome Sequencing with Short Overlapping Amplicons Suitable for Degraded DNA Using the Ion Torrent Personal Genome Machine. Hum Mutat. 2015 Dec 1;36(12):1236–47.

91. McElhoe JA, Holland MM, Makova KD, Su MSW, Paul IM, Baker CH, et al. Development and assessment of an optimized next-generation DNA sequencing approach for the mtgenome using the Illumina MiSeq. Forensic Sci Int Genet. 2014;13:20–9.
92. Parson W, Strobl C, Huber G, Zimmermann B, Gomes SM, Souto L, et al. Evaluation of next generation mtGenome sequencing using the Ion Torrent Personal Genome Machine (PGM). Forensic Sci Int Genet. 2013;7(5):543–9.
93. Just RS, Irwin JA, Parson W. Mitochondrial DNA heteroplasmy in the emerging field of massively parallel sequencing. Forensic Sci Int Genet. 2015 May 5;18:131–9.
94. Marshall C, Parson W. Interpreting NUMTs in forensic genetics: Seeing the forest for the trees. Vol. 53, Forensic Science International: Genetics. Elsevier Ireland Ltd; 2021.
95. Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. Nat Genet. 1999;23(2):147.
96. SWGDAM. SWGDAM Interpretation Guidelines for MtDNA Analysis. Scientific Working Group on DNA Analysis Methods; 2019.
97. Forsythe BK. The forensic use of whole mitochondrial genome sequencing in New Zealand [Doctoral dissertation, ResearchSpace@ Auckland]. 2023.
98. Eichmann C, Parson W. "Mitominis": Multiplex PCR analysis of reduced size amplicons for compound sequence analysis of the entire mtDNA control region in highly degraded samples. Int J Legal Med. 2008 Sep;122(5):385–8.
99. Thermo Fisher Scientific. Precision ID mtDNA Whole Genome Panel [Internet]. [cited 2023 Aug 3]. Available from: https://www.thermofisher.com/order/catalog/product/A30938
100. Verogen. ForenSeq mtDNA Whole Genome Kit Reference Guide. Document # VD2020006 Rev. B. 2020.
101. Promega Corporation. PowerSeq ® Whole Mito System Technical Manual [Internet]. 2023. Available from: www.promega.com
102. Gorden EM, Sturk-Andreaggi K, Marshall C. Capture enrichment and massively parallel sequencing for human identification. Vol. 53, Forensic Science International: Genetics. Elsevier Ireland Ltd; 2021.
103. Templeton JEL, Brotherton PM, Llamas B, Soubrier J, Haak W, Cooper A, et al. DNA capture and next-generation sequencing can recover whole mitochondrial genomes from highly degraded samples for human identification [Internet]. 2013. Available from: http://www.investigativegenetics.com/content/4/1/26
104. Ring JD, Sturk-Andreaggi K, Alyse Peck M, Marshall C. Bioinformatic removal of NUMT-associated variants in mitotiling next-generation sequencing data from whole blood samples. Electrophoresis. 2018 Nov 1;39(21):2785–97.
105. Payne BAI, Wilson IJ, Yu-Wai-Man P, Coxhead J, Deehan D, Horvath R, et al. Universal heteroplasmy of human mitochondrial DNA. Hum Mol Genet. 2013;22(2):384–90.
106. Melton T. Mitochondrial DNA heteroplasmy. Forensic Sci Rev. 2004;16(1):1–20.

107. Marshall C, Sturk-Andreaggi K, Ring JD, Dür A, Parson W. Pathogenic variant filtering for mitochondrial genome haplotype reporting. Genes (Basel). 2020 Oct 1;11(10):1–10.
108. Gallimore JM, McElhoe JA, Holland MM. Assessing heteroplasmic variant drift in the mtDNA control region of human hairs using an MPS approach. Forensic Sci Int Genet. 2018 Jan 1;32:7–17.
109. Holland MM, Makova KD, McElhoe JA. Deep-coverage MPS analysis of heteroplasmic variants within the mtgenome allows for frequent differentiation of maternal relatives. Genes (Basel). 2018 Mar 1;9(3).
110. King JL, Sajantila A, Budowle B. MitoSAVE: Mitochondrial sequence analysis of variants in Excel. Forensic Sci Int Genet. 2014;12:122–5.
111. Sturk-Andreaggi K, Parson W, Allen M, Marshall C. Impact of the sequencing method on the detection and interpretation of mitochondrial DNA length heteroplasmy. Forensic Sci Int Genet. 2020 Jan 1;44.
112. Parson W, Gusmão L, Hares DR, Irwin JA, Mayr WR, Morling N, et al. DNA Commission of the International Society for Forensic Genetics: Revised and extended guidelines for mitochondrial DNA typing. Forensic Sci Int Genet. 2014;13:134–42.
113. Dayama G, Emery SB, Kidd JM, Mills RE. The genomic landscape of polymorphic human nuclear mitochondrial insertions. Nucleic Acids Res. 2014 Nov 10;42(20):12640–9.
114. Cihlar JC, Strobl C, Lagacé R, Muenzler M, Parson W, Budowle B. Distinguishing mitochondrial DNA and NUMT sequences amplified with the precision ID mtDNA whole genome panel. Mitochondrion. 2020 Nov 1;55:122–33.
115. Santibanez-Koref M, Griffin H, Turnbull DM, Chinnery PF, Herbert M, Hudson G. Assessing mitochondrial heteroplasmy using next generation sequencing: A note of caution. Mitochondrion. 2019 May 1;46:302–6.
116. Woerner AE, Cihlar JC, Smart U, Budowle B. Numt identification and removal with RtN! Bioinformatics. 2020 Oct 15;36(20):5115–6.
117. McElhoe JA, Holland MM. Characterization of background noise in MiSeq MPS data when sequencing human mitochondrial DNA from various sample sources and library preparation methods. Mitochondrion. 2020 May 1;52:40–55.
118. Gilbert MTP, Binladen J, Miller W, Wiuf C, Willerslev E, Poinar H, et al. Recharacterization of ancient DNA miscoding lesions: Insights in the era of sequencing-by-synthesis. Nucleic Acids Res. 2007 Jan;35(1):1–10.
119. Pereira V, Longobardi A, Børsting C. Sequencing of mitochondrial genomes using the Precision ID mtDNA Whole Genome Panel. Electrophoresis. 2018 Nov 1;39(21):2766–75.
120. Woerner AE, Ambers A, Wendt FR, King JL, Moura-Neto RS, Silva R, et al. Evaluation of the precision ID mtDNA whole genome panel on two massively parallel sequencing systems. Forensic Sci Int Genet. 2018 Sep 1;36:213–24.

121. Strobl C, Eduardoff M, Bus MM, Allen M, Parson W. Evaluation of the precision ID whole MtDNA genome panel for forensic analyses. Forensic Sci Int Genet. 2018 Jul 1;35:21–5.
122. Strobl C, Churchill Cihlar J, Lagacé R, Wootton S, Roth C, Huber N, et al. Evaluation of mitogenome sequence concordance, heteroplasmy detection, and haplogrouping in a worldwide lineage study using the Precision ID mtDNA Whole Genome Panel. Forensic Sci Int Genet. 2019 Sep 1;42:244–51.
123. Faccinetto C, Sabbatini D, Serventi P, Rigato M, Salvoro C, Casamassima G, et al. Internal validation and improvement of mitochondrial genome sequencing using the Precision ID mtDNA Whole Genome Panel. Int J Legal Med. 2021;135:2295–306.
124. Buś MM, Lembring M, Kjellström A, Strobl C, Zimmermann B, Parson W, et al. Mitochondrial DNA analysis of a Viking age mass grave in Sweden. Forensic Sci Int Genet. 2019 Sep 1;42:268–74.
125. Templeton JEL, Linacre A. DNA profiles from fingermarks. Biotechniques. 2014 Nov 1;57(5):259–66.
126. Pang BCM, Cheung BKK. Double swab technique for collecting touched evidence. Leg Med. 2007 Jul;9(4):181–4.
127. Thermo Fisher Scientific. Quantifiler HP and Trio DNA Quantification Kits User Guide. 4485354 Rev. H. 2018.
128. Thermo Fisher Scientific. Precision ID mtDNA Panels with the HID Ion S5/HID Ion GeneStudio S5 System Application Guide. MAN0017770 C.0. 2021.
129. QIAGEN®. QIAGEN® Multiplex PCR Handbook [Internet]. 2010. Available from: www.qiagen.com
130. Gorden EM, Sturk-Andreaggi K, Marshall C. Repair of DNA damage caused by cytosine deamination in mitochondrial DNA of forensic case samples. Forensic Sci Int Genet. 2018 May 1;34:257–64.
131. KAPA BioSystems Inc. KAPA HyperPrep Kit Technical Data Sheet. KR0961 - v7.19. 2019.
132. KAPA Biosystems Inc. KAPA HyperPlus Kit Technical Data Sheet. KR1145 - v3.16. 2016.
133. KAPA BioSystems Inc. KAPA Dual-Indexed Adapter Kit Technical Data Sheet. KR1318 – v3.19. 2019.
134. Beckman Coulter Inc. Agencourt AMPure XP PCR purification. Instructions For Use [Internet]. 2016. Available from: www.beckmancoulter.com/customersupport/support.
135. KAPA Biosystems Inc. KAPA Library Quantification Kit Illumina ® Platforms Technical Data Sheet. KR0405 - v11.20. 2020.
136. Thermo Fisher Scientific. User Guide: Qubit dsDNA HS Assay Kits. 2015.
137. Illumina. What is the PhiX Control v3 Library and what is its function in Illumina Next Generation Sequencing [Internet]. [cited 2024 Feb 27]. Available from: https://knowledge.illumina.com/library-preparation/general/library-preparation-general-reference_material-list/000001545
138. illumina. Specifications for the MiSeq System [Internet]. [cited 2023 Dec 12]. Available from: https://www.illumina.com/systems/sequencing-platforms/miseq/specifications.html
139. Holland MM, Pack ED, McElhoe JA. Evaluation of GeneMarker® HTS for improved alignment of mtDNA MPS data, haplotype determination, and heteroplasmy assessment. Forensic Sci Int Genet. 2017 May 1;28:90–8.

140. SoftGenetics, LLC. GeneMarker ® HTS Mitochondrial NGS Analysis Software User Manual [Internet]. 2017. Available from: www.softgenetics.com
141. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics. 2010 Jan 15;26(5):589–95.
142. Bandelt HJ, Parson W. Consistent treatment of length variants in the human mtDNA control region: A reappraisal. Int J Legal Med. 2008 Jan;122(1):11–21.
143. Mehta B, Venables S, Roffey P. Comparison between magnetic bead and qPCR library normalisation methods for forensic MPS genotyping. Int J Legal Med. 2018 Jan 1;132(1):125–32.
144. Verogen. ForenSeq™ Universal Analysis Software Guide. Document # VD2018007 Rev. A. 2018.
145. Levin BC, Cheng H, Reeder DJ. A human mitochondrial DNA standard reference material for quality control in forensic identification, medical diagnosis, and mutation detection. Genomics. 1999;55(2):135–46.
146. Riman S, Kiesler KM, Borsuk LA, Vallone PM. Characterization of NIST human mitochondrial DNA SRM-2392 and SRM-2392-I standard reference materials by next generation sequencing. Forensic Sci Int Genet. 2017 Jul 1;29:181–92.
147. Li M, Schroeder R, Ko A, Stoneking M. Fidelity of capture-enrichment for mtDNA genome sequencing: Influence of NUMTs. Nucleic Acids Res. 2012 Oct;40(18).
148. Cihlar JC, Peters D, Strobl C, Parson W, Budowle B. The lot-to-lot variability in the mitochondrial genome of controls. Forensic Sci Int Genet. 2020 Jul 1;47.
149. Rathbun MM, McElhoe JA, Parson W, Holland MM. Considering DNA damage when interpreting mtDNA heteroplasmy in deep sequencing data. Forensic Sci Int Genet. 2017 Jan 1;26:1–11.
150. Sturk-Andreaggi K, Ring JD, Ameur A, Gyllensten U, Bodner M, Parson W, et al. The value of whole-genome sequencing for mitochondrial DNA population studies: Strategies and criteria for extracting high-quality mitogenome haplotypes. Int J Mol Sci. 2022;23(4):2244.
151. Schönherr S, Weissensteiner H, Kronenberg F, Forer L. Haplogrep 3-an interactive haplogroup classification and analysis platform. Nucleic Acids Res. 2023 Jul 5;51(W1):W263–8.
152. Cihlar JC, Amory C, Lagacé R, Roth C, Parson W, Budowle B. Developmental validation of a MPS workflow with a PCR-based short amplicon whole mitochondrial genome panel. Genes (Basel). 2020 Nov 1;11(11):1–31.
153. Haplogrep 3. Kulczynski Metric [Internet]. [cited 2024 Jan 29]. Available from: https://haplogrep.readthedocs.io/en/latest/kulczynski/
154. Vernarecci S, Ottaviani E, Agostino A, Mei E, Calandro L, Montagna P. Quantifiler® Trio Kit and forensic samples management: A matter of degradation. Forensic Sci Int Genet. 2015;16:77–85.
155. Wu WW, Phue JN, Lee CT, Lin C, Xu L, Wang R, et al. Robust Sub-nanomolar Library Preparation for High Throughput Next Generation Sequencing. BMC Genomics. 2018 May 4;19(1).

156. Haplogrep 3. Errors and Warnings [Internet]. [cited 2024 Feb 25]. Available from: https://haplogrep.readthedocs.io/en/latest/errors-warnings/

157. Holland CA, McElhoe JA, Gaston-Sanchez S, Holland MM. Damage patterns observed in mtDNA control region MPS data for a range of template concentrations and when using different amplification approaches. Int J Legal Med. 2021 Jan 1;135(1):91–106.

158. Illumina Inc. Adapter dimers causes, effects, and how to remove them [Internet]. [cited 2023 Dec 29]. Available from: https://knowledge.illumina.com/library-preparation/general/library-preparation-general-troubleshooting-list/000001911

159. Silvia AL, Shugarts N, Smith J. A preliminary assessment of the ForenSeq™ FGx System: next generation sequencing of an STR and SNP multiplex. Int J Legal Med. 2017 Jan 1;131(1):73–86.

160. England R, Nancollis G, Stacey J, Sarman A, Min J, Harbison SA. Compatibility of the ForenSeq™ DNA Signature Prep Kit with laser microdissected cells: An exploration of issues that arise with samples containing low cell numbers. Forensic Sci Int Genet. 2020 Jul 1;47.

161. EMPOP. EMPOP mtDNA Database Directions for Use [Internet]. 2019 [cited 2024 Feb 1]. Available from: https://empop.online/use

162. Kino K, Sugiyama H. Possible cause of G·C → C·G transversion mutation by guanine oxidation product, imidazolone. Chem Biol. 2001;369–78.

163. Xavier C, Eduardoff M, Strobl C, Parson W. SD quants—Sensitive detection tetraplex-system for nuclear and mitochondrial DNA quantification and degradation inference. Forensic Sci Int Genet. 2019 Sep 1;42:39–44.

164. Zhang Q, Zhou Z, Liu Q, Liu L, Shao L, Zhang M, et al. Evaluation of the performance of Illumina's ForenSeq™ system on serially degraded samples. Electrophoresis. 2018 Nov 1;39(21):2674–84.

165. Radojicic V, Keckarevic Markovic M, Puac F, Kecmanovic M, Keckarevic D. Comparison of different methods of DNA recovery and PCR amplification in STR profiling of casings—a retrospective study. Int J Legal Med. 2018 Nov 1;132(6):1575–80.

166. Suresh M. Improving methods to obtain DNA profiling results from cartridge cases [Masters thesis, ResearchSpace@ Auckland]. 2022.

167. Illumina. Optimizing Cluster Density on Illumina Sequencing Systems. 2016.

168. Bioke. NEBNext FFPE DNA Repair Mix [Internet]. [cited 2024 Feb 8]. Available from: https://www.bioke.com/webshop/neb/m6630.html

169. Cheng K, Bright JA, Kelly H, Liu YY, Lin MH, Kruijver M, et al. Developmental validation of STRmix™ NGS, a probabilistic genotyping tool for the interpretation of autosomal STRs from forensic profiles generated using NGS. Forensic Sci Int Genet. 2023 Jan 1;62.