The development of enhanced experimental strategies for the DNA analysis of low-template or compromised forensic sample types

Victoria Ann Barlow

A thesis submitted in partial fulfilment of the requirements of the University of Northumbria at Newcastle for the degree of Doctor of Philosophy

Research undertaken in the Faculty of Health & Life Sciences

November 2015

Abstract

Single-cell DNA analysis is not routinely carried out in a forensic setting as it is considered unreliable due to challenges associated with DNA amplification, contamination and profile interpretation. In light of the development of increasingly sensitive techniques, the question of the reliability of single-cell DNA analysis in terms of both processing and interpretation is addressed in the first part of this thesis. Optimising all stages of the DNA analysis process has provided a sensitive method which facilitates the successful outcome of a useable profile from single-cells. Although no consensus profile can be generated for this sample type, interpretation guidelines have been set to enable the robust analysis of single cells. It has been concluded that single-cells can be reliably amplified and profiled for forensic purposes.

Both DNA and textile fibres have a proven track record in forensic casework yet their analysis is rarely combined. As an application of the aforementioned single-cell DNA analysis, this project explores the possibility that when fibres are transferred from one surface to another, they could also be acting as a vector for the wearer's own DNA, through cells that have adhered to the fibre surfaces. Fluorescent staining and microscopy is used to detect the cells *in situ* on the fibre surface, which are then recovered and processed for DNA using the previously optimised single-cell analysis methods, along with a newly developed DNA assay designed for the amplification of low DNA template samples. The results of this study have demonstrated that cells can be visualised *in situ* on the fibre surface and that there is potential for cell transfer to occur. It has been concluded however, that from a casework point of view, targeting transferred fibres for cells may not be the best approach as it is time consuming and has not been shown to be effective in this study.

The final part of this thesis is focused on the efficacy of massively parallel sequencing (MPS) technology for samples that are expected to be severely degraded due to age or exposure to a hostile environment. The ability of both the recently launched Illumina ForenSeq™ DNA Signature Prep Kit for nuclear DNA markers and an in-house method for the sequencing of degraded mitochondrial DNA, have been tested to determine if MPS offers a more comprehensive evaluation of degraded material than the traditional PCR-CE methods. The results of the ForenSeq kit have demonstrated the effectiveness of its low molecular weight STR and SNP markers for amplifying low template, degraded DNA samples, with alleles amplified using less than 20 pg total DNA input. This kit has also therefore shown application in the field of bioarchaeology, as it can provide the biological sex of the sample, biogeographic ancestry information and also aids detection of sample/control contamination. The in-house mitochondrial DNA assay resulted in the successful amplification and sequencing of samples for which no nuclear DNA was amplified. The high depth of read coverage in these samples, average of 18,000, allowed for the identification of even low level variants.

Acknowledgements

This project could never have been started or completed without my supervisor Dr. Eleanor Graham, who has guided me through my PhD, both academically and personally. Your support and encouragement has made this experience one that has allowed me to develop my confidence and skills both as a researcher and also as a lecturer. I have learnt such a lot from you and want to thank you for that, and also for your seemingly unending patience with my somewhat 'flexible' deadlines. I have truly enjoyed the past three years.

I would also like to thank my other supervisors Professor Martin Evison and Dr. Kelly Sheridan for all your advice over the past three years. Martin, I very much appreciate all the work you put in to help give me the opportunity to start this project, and Kelly I could not have navigated the world of fibres without you. I would like to express my thanks to the members of the Northumbria University Forensic Research Centre, particularly Ray Palmer and Sophie Carr, who have all at one point or another offered advice or encouragement during my time here. I would like to thank Suzanne and Paul for all their help during my project, particularly for their patience whilst I was working in their lab. I would also like to thank Paul Agnew for his advice and support throughout my PhD.

To my family; Mom, Kate, Daniel, Andrew, Michael, Sarah and, of course, our little Miss Evie. I want to thank you all for your love and support over the past few years, for encouraging or distracting me as required, and for at least pretending to listen to the many hours of explanations on my research. I could not have done this without you all. For my Chris, I want to thank you for your calming influence and gentle encouragement; you have helped to keep me sane the last few months.

Life as a PhD student would not have been nearly so much fun, without the support of my friends, particularly my pentagon ladies Hayley, Francesca, Hazel and Laura, and the guys Adnan, Giles, Alex, Rory and Greg. Thank you all for keeping me entertained the last three years!

Finally I would like to express my appreciation to all the individuals who have volunteered to take part in this project, particularly Francesca and Giles for their frequently donated DNA samples. I would also very much like to thank our collaborators, Dr. Andrew Birley and Dr. Trudi Bruck at Vindolanda, Marco Guimaraes and Raffela Francisco in Brazil, and Nicola Oldroyd and Laurence Devesse at Illumina Inc. for providing the samples and the means for making much of this project possible.

Declaration

I declare that the work contained in this thesis has not been submitted for any other award and that it is all my own work. I also confirm that this work fully acknowledges opinions, ideas and contributions from the work of others.

Any ethical clearance for the research presented in this thesis has been approved. Approval has been sought and granted by the Faculty Ethics Committee on 28th November 2012.

I declare that the Word Count of this Thesis is 45850 words (original submission)

Name: Victoria Ann Barlow

Signature:

Date:

Contents

Index of figures

Index of tables

Abbreviations

ATP	Adenosine triphosphate
$CaCl_2$	Calcium chloride
CRS	Cambridge Reference Sequence
cm	Centimetre
DI	Degradation index
^{o}C	Degrees Celsius
DAPI	4’,6-diamidino-2-phenylindole
DNA	Deoxyribonucleic acid
dsDNA	Double-stranded DNA
DPBS	Dulbecco’s phosphate buffered saline
EDTA	Ethylenediaminetetraacetic acid
EtOH	Ethanol
g	G-force
g	Gram
GuSCN	Guanidine Thiocyanate
HCl	Hydrochloric acid
J	Joule
ltDNA	Low template DNA
MPS	Massively Parallel Sequencing
µg	Microgram
µm	Micrometre
µM	Micromolar
MSP	Microspectrophotometer
mg	Milligram
mm	Millimetre
mM	Millimolar

mtDNA	Mitochondrial DNA
mtGenome	Mitochondrial genome
M	Molar
MBG water	Molecular Biology Grade water
ng	Nanogram
nm	Nanometre
NGS	Next Generation Sequencing
NTC	No template control
nDNA	Nuclear DNA
PFA	Paraformaldehyde
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
POP	Perfomance optimised polymer
pg	Picogram
Pro K	Proteinase K
qPCR	quantitative PCR
rCRS	Revised Cambridge Reference Sequence
rpm	Revolutions per minute
ssDNA	Single-stranded DNA
NaCl	Sodium chloride
NaOH	Sodium hydroxide
SLS	Scientific Laboratory Supplies
TBE Buffer	Tris-Borate-EDTA buffer
TE Buffer	Tris - Ethylenediaminetetraacetic acid
UV	Ultra Violet

1. Introduction

1.1 Forensic science

Forensic science is the use of a scientific approach for the investigation, evaluation and communication of evidence pertaining to the criminal justice system. One of the first recorded cases of forensic investigation can be traced anecdotally back to early China, during which an early form of entomology was used to identify from a number of sickles the one that had been used as the murder weapon; it was the only one to attract flies, thus indicating the presence of blood (Sung, 1981).

There are many reported uses of forensic science over the following centuries, for example the use of fingerprints to prevent relatives claiming the pension of deceased military persons, and the use of toxicology to detect arsenic poisoning in a murder trial in 1883 (Tilstone, 2006). It was not until the early 20th century however, that the first police lab was opened in Lyon, France, by Edmond Locard. It is Locard's principle of exchange, whereby he describes how evidence of the perpetrator will be left at the scene or that traces from the scene will be taken away on their being (Locard, 1920), which remains the underpinning principle of forensic criminal casework.

Forensic science is continually evolving in order to meet the requirements of modern day society; the 21st Century has already seen the rapid expansion of areas such as digital based forensics and those related to terrorism. The nature of not only criminal or civil investigations in their traditional sense, but also areas such as disaster victim identification and bio-archaeological investigations ensures that a wide range of different disciplines can be applied and encompassed under the forensic science umbrella, each with its own set of analytical methods and techniques used to answer the questions posed by the investigation.

Whilst the UK courts do not have a specific set of standards for the admissibility of scientific evidence in court (Ireland, 2015), the impact and challenge to forensic science following the Daubert v. Merrell Dow Pharmaceuticals ruling (1993) and the U.S. initiative to strengthen Forensic Science (National Research Council (U.S.). Committee on Identifying the Needs of the Forensic Science Community. et al., 2009, Butler, 2015b) have been felt internationally (Ireland, 2015). The emphasis is now placed firmly on the fundamental science that underlies forensic analysis, with a comprehensive assessment of the capabilities and limitations of the varying analytical techniques required before confidence in any result, both positive and negative, can be communicated. Although forensic Deoxyribonucleic Acid (DNA) analysis has, since its inception, been subjected to rigorous testing and validation (as reviewed in (Saks and Koehler, 2005) and (Butler, 2015b)), and so is held up as the 'gold standard' of forensic disciplines in this context, the fast pace of development in this field

along with the increasingly sensitive techniques employed, demand continuous improvement and evaluation for use in forensic casework.

1.2 Forensic DNA analysis

In 1984 Sir Alec Jeffreys first demonstrated the use of multiple minisatellite regions as a method by which two individuals could be distinguished, a technique he termed 'DNA fingerprinting' (Jeffreys et al., 1985). Minisatellites are loci that can consist of up to approximately one thousand tandem repeat units, each one being made up of 10 -100 bp (Jeffreys et al., 1985). Multi-locus probes hybridised with conserved core sequences across many minisatellite loci which, when combined with southern blotting, allowed for the visualisation of the multiple DNA bands that made up the individual's DNA fingerprint (Jeffreys et al., 1985). The potential for the application of this technique in a forensic context was quickly recognised (Gill et al., 1985) and along with the developed single-locus probes, which allowed for the targeting and visualisation of a single locus at a time to facilitate interpretation, formed the earliest type of DNA analysis utilised in criminal casework; the infamous Colin Pitchfork case, which saw not only the identification and conviction of the offender in 1987, but also the first exoneration of an individual based upon DNA evidence (reviewed in (Jobling and Gill, 2004)). Efforts were made to collect data on the observed frequency of alleles within differing populations so that the discriminatory power of a DNA profile could be assessed and the advent of the polymerase chain reaction (PCR) by Kary Mullis (Saiki et al., 1988), together with the introduction of short tandem repeats (STRs) (Weber and May, 1989) and development of the Sanger sequencing (Sanger et al., 1977, Butler et al., 2004) technology, prompted rapid advancement in the area of forensic genetics.

1.2.1 Short tandem repeats

STRs differ from minisatellites in that they have fewer repeat units of lengths of between 2-6 nucleotides (Kimpton et al., 1993), producing an overall shorter length of target fragment (Figure 1.1). The shorter target size, and the ability to design primers close to the target region (mini-STRs), facilitates amplification of smaller quantities of DNA template resulting in an increase in assay sensitivity relative to the use of minisatellites; an important advantage for forensic DNA analysis given the type of material often available for analysis.

The ability to visualise amplified DNA fragments through the use of fluorescent tags added during PCR, facilitated multiplex reactions enabling additional markers to be analysed simultaneously, thus improving the discriminatory value of the assay. The use of capillary electrophoresis (CE) allows for the separation and fluorescent detection of the PCR products, with the lower molecular weight DNA fragments passing through the detection window earliest (Butler et al., 2004). The inclusion of a DNA fragment size standard with every

sample along with the use of an allelic ladder (Butler et al., 1995, Butler et al., 2004), enabled the unambiguous identification of alleles at all loci, thus enabling the application of a standardised nomenclature system (Olaisen et al., 1998) allowing the comparison of data between labs and importantly, the introduction of the National DNA Database (NDNAD) in the United Kingdom (UK) in 1995 (Werrett, 1997). The ability to search against DNA profiles both from previous offenders and crimestains from unsolved crimes, enabled perpetrators to be identified and crimes to be linked together in a way that had not previously been possible, highlighting the contribution of DNA analysis to the criminal justice system.

Approximately 3% of the human genome consists of STRs (Lander et al., 2001), often located in the promotor regions of genes (Sawaya et al., 2013) as well as at the terminal regions of the chromosomes (Misawa, 2016). The complexity of the repeat units varies between markers with loci categorised as 'simple', in that the sequence is made up of identical repeating units, 'compound', where at least two repeat units are present, and 'complex', in which multiple types of repeat units are interspersed with additional bases (Urquhart et al., 1994) as shown in Figure 1.1. The complex loci are generally associated with an increase in the distinguishing capability at that locus (Gill et al., 1997a).

An investigation into the use of fourteen STR loci, arranged into three possible multiplexes was carried out by Kimpton et al. in order to assess suitability for their inclusion in the first STR Quadplex used in forensic DNA analysis (Kimpton et al., 1993, Kimpton et al., 1994). This was based on their predicted discriminatory power, well-matched amplification conditions for multiplexing, compatible allele fragment sizes and the ability to generate distinct alleles for accurate sizing. The use of tetra- or penta- nucleotide repeat units rather than di- or tri- nucleotide repeats was also considered in order to reduce the frequency of associated stutter peaks. Stutter peaks are a common and unavoidable artifact that results in the appearance of additional peaks exactly one or more repeat units less than the true allele peak or, less frequently in the plus one repeat unit position (Walsh et al., 1996), which is thought to arise during DNA synthesis as a consequence of the polymerase 'slipping' off the template strand (see Section 3.1); reducing the occurrence of these stutter peaks aids interpretation of the resulting DNA profile.

Since the early markers were identified, the completion of studies such as the Human Genome Project, has increased the number of STR loci that may be potentially useful in a forensic context. The maximum size of the DNA target is a defining feature of newly selected loci in order to ensure that maximum information can be obtained from low quantity/ poor quality samples. This must however, be balanced with a high number of

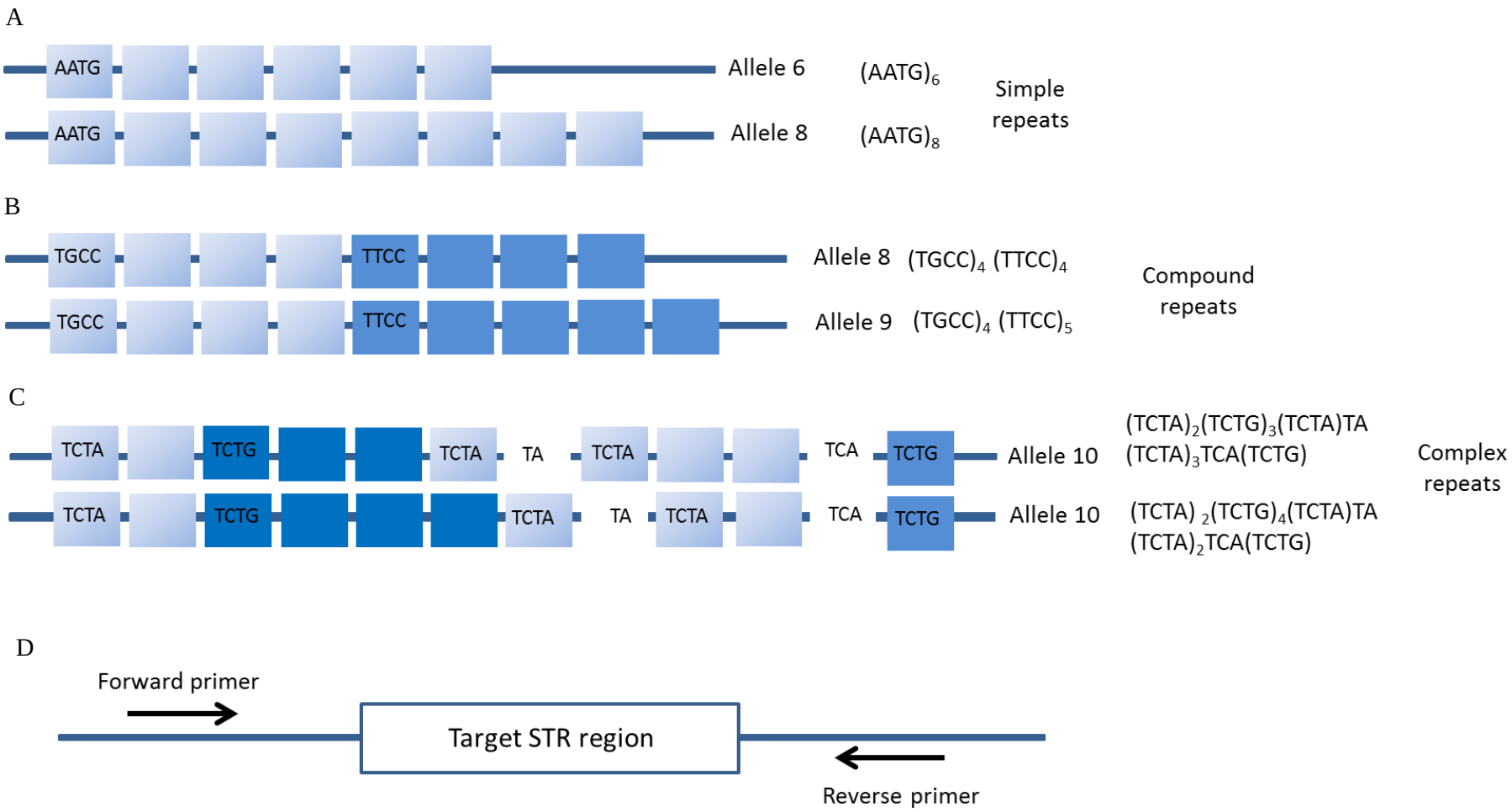


Figure 1.1 Overview of the structure of short tandem repeats. A – Simple STRs are made up of identical repeating units with alleles six and eight shown in the example above. B – Compound STRs are made up of at least two different repeat units shown above in two different shades of blue. The total number of repeat units are counted to give the allele number; alleles eight and nine above. C- Complex STRs are made up of multiple types of repeat units are interspersed with additional bases with only the various repeat units counted to give the allele number (Urquhart et al., 1994). D - Primers are used to target the region of interest with the STR between them, with the number of repeats at each locus determining the total length of the DNA fragment.

associated alleles and a high degree of heterozygosity in order to maintain the discriminatory power of the assay. A summary of this information for loci currently used in the UK is provided in Table 1-1. The mutation rates for commonly used forensic STR loci has been shown to be approximately 9.78 x 10^{-4} mutations per locus per generation, with those loci having a higher value of heterozygosity also have a higher mutation rate than those with a lower heterozygosity value. FGA for example has a heterozygosity value of 0.8671 and a mutation rate of approximately 1.98 x 10^{-3} (Caucasian), whilst TH01 has a heterozygosity value and mutation rate of 0.7802 and 0.02 x 10^{-3} respectively (Caucasian) (Sun et al., 2014). Following a recent genome wide study of STR loci, the overall germline mutation rate has been calculated as approximately 7.6 x10^{-5} mutations per locus per generation (Fungtammasan et al., 2015); the lower mutation rates in STRs selected for forensic analysis facilitate their efficacy for use in kinship analysis Other factors affecting STR mutation rates include repeat unit length and sequence, ethnicity, gender and age (Fungtammasan et al., 2015, Sun et al., 2014).

Finally, all selected loci must be amenable to multiplexing in terms of both primer compatibility and positioning on the chromosome relative to other loci in order to avoid genetic linkage. By ensuring that any loci on the same chromosome are separated by millions of base pairs the risk of both being inherited together through recombination during meiosis is eliminated. This means that loci are independent of one another and can therefore be used to calculate the profile frequency by way of the product rule (Carracedo, 1998, Coble and Butler, 2005).

1.2.2 Single nucleotide polymorphisms

In addition to the length mutations observed with STRs, single nucleotide polymorphisms (SNPs) are also utilised for forensic investigation. SNPs arise from a single point mutation in the genome resulting in a change to the expected nucleotide present. Following an in-depth study of data generated as part of the 1000 genomes project (The Genomes Project, 2015), a total of approximately 38 million SNPs have been identified throughout the genome, with an average of 3.6 million SNPs detected per individual (Sudmant et al., 2015)

Traditionally the focus of identity markers for forensic purposes has been on the use of STRs, as the biallelic nature of the majority of SNPs results in the requirement for a high number of loci to be analysed in order to obtain a comparable power of discrimination to relatively few STR loci, i.e. approximately fifty SNPs is equivalent to approximately twelve STRs (Gill, 2001). More recently, however, identity SNPs (iSNPs) are being included alongside STRs in assays developed for PCR-massively parallel sequencing (PCR-MPS)

techniques (Section 1.3.3), in order to take advantage of their short fragment sizes, which facilitates amplification in low quality forensic material.

SNPs have been demonstrated as highly informative markers for indicating both biogeographic ancestry (Frudakis et al., 2003, Shriver et al., 2003) and phenotypic characteristics, such as hair and eye colour (Liu et al., 2009, Walsh et al., 2013) of an individual under investigation. SNPs are also the marker type utilised for the analysis of mitochondrial DNA (mtDNA) as described in the following section.

Table 1-1 Summary of STR locus information. The chromosome location of each STR locus included in the NGM Amplification Kit is given, along with the minimum and maximum fragment sizes at that locus across the range of alleles observed. Information adapted from (Biosystems, 2015). All loci are shown to have high heterozygosity values of >0.7 (AllST*R – Autosomal Database (allstr.de), Qualitype GmbH, Germany) allowing for greater variability within the population.

Locus	Chromosome location	Minimum fragment size	Maximum fragment size	Allele range	Heterozygosity value
D10S1248	10q26.3	72.0	127.0	8-18	0.7470
vWA	12p13.31	149.0	214.3	11-24	0.8077
D16S539	16q24.1	223.6	277.6	5-15	0.7783
D2S1338	2q35	281.6	356.0	15-28	0.8756
D8S1179	8q24.13	117.9	174.9	8-19	0.8094
D21S11	21q11.2-q21	178.8	249.8	24-38	0.8299
D18S51	18q21.33	259.5	347.5	7-27	0.8764
D22S1045	22q12.3	76.0	120.0	8-19	0.7198
D19S433	19q12	122.3	166.3	9-17.2	0.7534
TH01	11p15.5	176.4	221.1	4-13.3	0.7802
FGA	4q28	221.6	372.0	17-51.2	0.8671
D2S441	2p14	74.5	113.4	9-16	0.7403
D3S1358	3p21.31	114.4	168.4	12-19	0.7979
D1S1656	1q42.2	170.0	224.0	9-20.3	0.8951
D12S391	12p13.2	225.0	287.0	14-27	0.8933

1.2.3 Mitochondrial DNA

It was first demonstrated in 1966 that in contrast with the linear nature of the nuclear genome (Watson and Crick, 1953), the mitochondrial genome (mtGenome), shown in Figure 1.2, has a circular structure (van Bruggen et al., 1966, Nass, 1966). The Sanger method was used by Anderson et al. in 1981 (Anderson et al., 1981) to sequence the first human mtGenome, which consisted of 16,569 bp. This published sequence, widely known as the Cambridge Reference Sequence (CRS), was then revised (rCRS) and updated by Andrews et al. in 1999 (Andrews et al., 1999) to correct for eleven original sequencing errors.

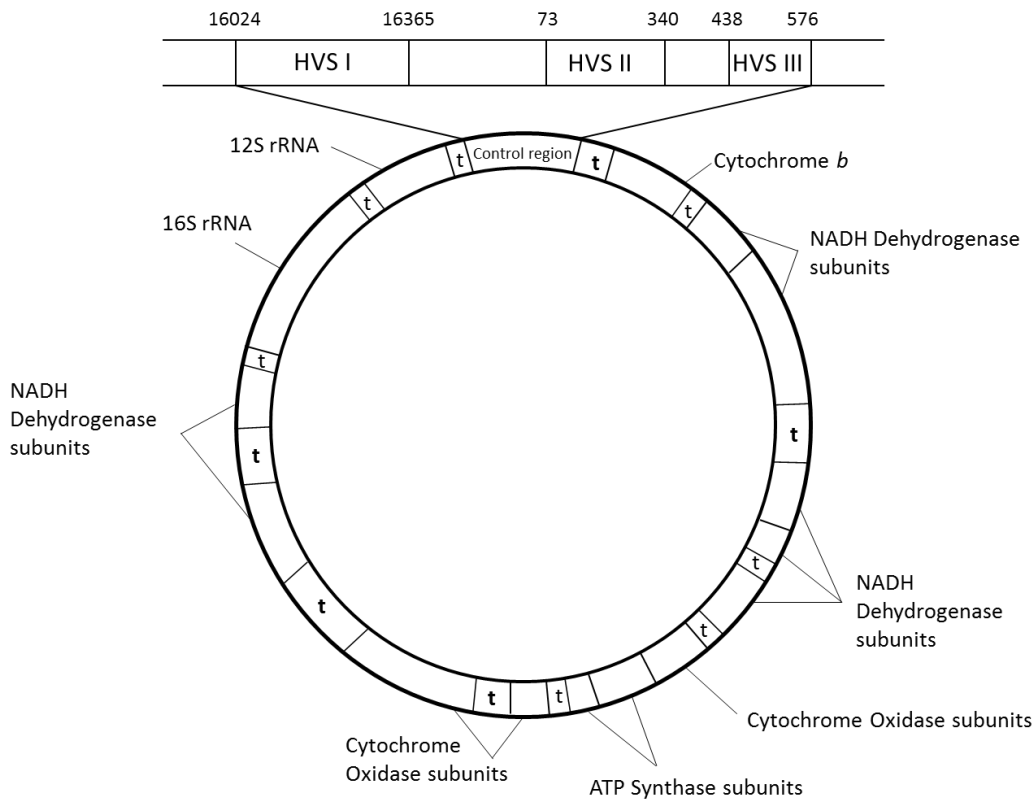


Figure 1.2 Diagram showing overview of the mtGenome. The double-stranded mtGenome contains thirty-seven genes coding for fourteen tRNAs (t) (areas encoding multiple tRNAs indicated in bold), two rRNAs and twelve polypeptides on the heavy strand, and eight tRNAs and one polypeptide on the light strand (Anderson et al., 1981). The control region contains three hypervariable sections between nucleotide positions 16024 and 576.

The mtGenome is intertwined with core mtDNA binding proteins, primarily the structural Mitochondrial Transcription Factor (TFAM) (Kaufman et al., 2007), to form spherical nucleoids (Nass, 1969). This mtDNA/protein complex is thought to form a dense core, so compactly formed that access is limited from other matrix proteins (Brown et al., 2011). Surrounding this core are multi-functional factors that interact with the core through protein-protein interactions (Brown et al., 2011). Kukat et al. utilised immunofluorescence together

with stimulated emission depletion (STED) microscopy to determine the average diameter across 38,777 mammalian nucleoids as 102 ± 0.3 nm, approximately 70 nm without the antibody attached (Kukat et al., 2011). The quantity of nucleoids per mitochondrion (1 – 2000 nucleoids) and the mtGenome copy number per nucleoid (1 – 15 copies) have been debated over the years (Brown et al., 2011, Iborra et al., 2004, Legros et al., 2004, Satoh and Kuroiwa, 1991, Kukat et al., 2011, Cavelier et al., 2000), but a consistent average of 1.4 copies of the mtGenome in each nucleoid was observed by both the 1991 paper of Satoh & Kuroiwa (Satoh and Kuroiwa, 1991) and 2011 work of Kukat et al. (Kukat et al., 2011). There was however, a five-fold difference in total estimated mtDNA copy number per cell, with averages of approximately 500 copies and 2720 respectively, presumably due to the development of superior microscopy techniques in the last two decades allowing better resolution for the 70 nm diameter nucleoid structures (Kukat et al., 2011). The high copy number and nucleoid structure contributes to its ability to better withstand the effects of DNA degradation than nuclear DNA, as discussed further in Section 1.2.4.3.

It was established in 1980 that, aside from rare incidences of paternal inheritance (Schwartz and Vissing, 2002), human mtDNA is inherited maternally (Giles et al., 1980) although the actual mechanisms behind how the mtDNA is inherited are still under investigation it has been demonstrated that mitochondria in the spermatozoa do enter the oocyte at fertilisation (Ankel-Simons and Cummins, 1996). As discussed by Carelli, there are two prevailing theories that account for the elimination of the paternal mitochondria following conception. The ‘dilution model’ works on the assumption that the ratio of copies of paternal to maternal mtDNA is approximately 1:15,860, thus resulting in the masking of the paternal mtDNA. The second theory supports the active elimination of the paternal mitochondria possibly through ubiquitination of mitochondria in the sperm, selective mitophagy or direct degradation (Carelli, 2015). As summarised by Brogenhagen, it is also not yet clear whether it is the entire nucleoid, potential containing multiple copies of the mtGenome, or an individual mtGenome within the nucleoid that actually constitutes the genetic unit of inheritance (Bogenhagen, 2012). The elucidation of these mechanisms will improve understanding on how both germline and somatic variants are inherited.

1.2.3.1 mtDNA variants

The human mtGenome is evolving at approximately six to seventeen times the rate of nDNA (Brown et al., 1979, Wallace et al., 1987, Miyata et al., 1982) resulting in the presence of both point and length mutations. Due to this relatively high rate of mutation, along with the cellular abundance of mtDNA, instances of heteroplasmy can arise whereby divergent versions of the mtGenome can be present within the same cell (Holt et al., 1988). The rate at

which heteroplasmy occurs has been shown to vary dependent upon the area of the mtGenome in question (Pesole et al., 1999), with the synonymous sites evolving at a faster rate than the non-synonymous positions (Pesole et al., 1999), and the tissue type of the host cell (Naue et al., 2015, Calloway et al., 2000). Recent studies using next generation sequencing (NGS) have recorded heteroplasmy in approximately a quarter of individuals sampled when using a minor allele detection setting of ≥ 10% of the total number of base calls at a given position (Li et al., 2010, King et al., 2014a, Ramos et al., 2013, Just et al., 2015). Setting the variant detection threshold at ≥ 0.2% however, suggests that heteroplasmy may be more prevalent than previously thought as the ultra-deep sequencing of two mtDNA amplicons revealed that all thirty-four individuals sampled, had at least one heteroplasmic site present (Payne et al., 2013); seven of these (~21 %) would have been detected had the threshold been set at 10%, which is consistent with other papers.

The improvement in the ability to detect low level variants using NGS technologies has also been illustrated recently by Bintz et al. (Bintz et al., 2014) when in addition to their target mtDNA sequence, a nuclear mitochondrial sequences (NumtS) (Lopez et al., 1994) was also detected. NumtS are fragments of the mtGenome that become incorporated into the nuclear genome during the repair of damaged DNA (Blanchard and Schmidt, 1996, Ricchetti et al., 2004). There are currently over seven hundred of these fragments documented (Calabrese et al., 2012, Dayama et al., 2014) that, once integrated into the nuclear genome, evolve at a slower rate than the mtGenome and so have use in phylogenetic studies (Zischler et al., 1995a). The potential to misinterpret a NumtS for a genuine mtDNA sequence has been demonstrated (Zischler et al., 1995b) and sample haplotypes, particularly low level heteroplasmic sites, should be checked against known NumtS variants (Calabrese et al., 2012, Dayama et al., 2014) before final designation.

Historically the control region (CR), also known as the displacement loop (D-loop), has been the target of interest for forensic geneticists due to the high mutation rate, estimated at 2.5/site/Myr (Parsons et al., 1997). The CR is the largest non-coding segment of the mtGenome (Anderson et al., 1981) and includes three hypervariable sections (HVS). Amplification and sequencing of HVSI or HVSI/II for haplotype comparisons is common practice as these regions are considered mutational hotspots (Jazin et al., 1998) and therefore highly informative, as well as offering a target that is practical for Sanger sequencing. The maternally inherited nature of the mtGenome results in an expected inability to differentiate between relatives in the maternal line. In addition to this and unlike the nuclear genome, the mtGenome is a single-linked molecule and so the frequencies observed for SNPs in the genome cannot be assessed individually, but as a whole haplotype (Parsons and Coble, 2001).

Mitochondrial haplotypes are generated by sequencing either a targeted region of the mtGenome or its entirety for a direct comparison with the rCRS. Deviations from the rCRS are recorded according to the guidelines published by the DNA Commission of the International Society of Forensic Genetics (ISFG) (Parson et al., 2014, Carracedo et al., 2000) as detailed in appendix 8.1. The mtDNA haplotype profiles of two samples can be compared with one another resulting in an inclusion, exclusion or inconclusive result (Carracedo et al., 2000). The absence of a heteroplasmic site should not be used to exclude an individual due to the potential for differences within and between cell types (Calloway et al., 2000, Naue et al., 2015), but conversely the presence of a heteroplasmic site may increase the strength of the evidence (Parson et al., 2014, Carracedo et al., 2000). Match probability and likelihood ratios can then be derived from datasets compiled in databases such as EMPOP (Parson, 2015), or those containing more relevant localised population data (Parson et al., 2014) to assess the significance of a match between a reference sample and the sample in question.

It has now been recommended by the ISFG DNA Commission that the entire control region, i.e. nucleotide positions 16024 – 576, should be sequenced as a minimum for forensic analysis (Parson et al., 2014) in order to maximise the potential to discriminate between two individuals or assign haplogroups based on markers outside of the HVSI/II fragments (Parson et al., 2014). Further to this however, it has been demonstrated that if the whole mtGenome is sequenced then distinct haplotypes can be observed in over 90% of randomly sampled individuals (Just et al., 2015, King et al., 2014a).

The uniparental inheritance of mtDNA, along with its high rate of mutation has been exploited by both population geneticists for use in tracking the worldwide migration of the human population, and by forensic geneticists for use in kinship analysis, through which familial relationships can be established. A landmark study in 1992 by Excoffier et al. (Excoffier et al., 1992) outlined a framework for mapping and establishing genetic distance between mitochondrial haplotypes. A year later, seventeen Native American populations were consigned to the first four haplogroups (A-D) based on their mitochondrial haplotype. In the last two decades the haplotypes and haplogroups of individuals from around the world have been published in hundreds of papers (most recent review (Kivisild, 2015)), which have contributed towards the assembly of a comprehensive phylogenetic tree (van Oven and Kayser, 2009, Van Oven, 2015). Based on the presence or absence of the population specific polymorphisms used to construct this tree, an individual's haplotype can be used to assign their haplogroup and thus determine their ancestral biogeographic origins. This information can be particularly useful within a forensic context in inferring the ethnicity and thus some phenotypic characteristics of an unknown suspect.

The multi-copy nature of the mtGenome within every cell offers two further avenues for use in forensic DNA analysis in the form of an alternative source of DNA in samples that either inherently contain little or no nDNA such as keratinous tissues (Parson et al., 2015, Higuchi et al., 1988, Anderson et al., 1999) or, of particular interest for this study, samples in which the quantity or quality of nuclear DNA (nDNA) is in question, i.e. nDNA is expected to be severely degraded due to spatiotemporal effects or quantification results indicate a low concentration/absence of nDNA; this is explored further in chapter five.

1.2.4 Optimal biological forensic samples

The biological samples types collected during the course of a forensic investigation can vary widely, as can the quantity and quality of DNA contained within them. An ideal sample for forensic DNA analysis will contain a high quantity of good quality DNA, such as may be found in a fresh blood or saliva sample. Once extracted, such a sample can be diluted to acquire the optimal quantity for the forensic DNA kits (approximately 0.5 ng – 2 ng DNA), and the longer DNA fragments associated with good quality DNA should result in the amplification of even high molecular weight loci, thus generating a full DNA profile. Reference samples taken directly from an individual will invariably yield optimal DNA, the exceptions being hair samples taken by force and samples taken from a decomposing corpse. A significant advantage of optimal DNA is that having a higher starting quantity allows for multiple testing to be carried out on the same extract, thereby increasing the amount of information that can be acquired.

1.2.5 Sub-optimal DNA

As reviewed recently by Butler (Butler, 2015a) forensic DNA analysis is now moving into an era of growth and into one of what he terms ‘sophistication’, whereby the use of techniques including those with higher sensitivity and greater depth of information will be introduced into casework. In order to achieve this, these technologies must be evaluated to ascertain their capabilities and limitations for use with forensic material, particularly for that which is sub-optimal.

1.2.5.1 Low template DNA

‘Touch’ DNA samples (van Oorschot and Jones, 1997) are the most prevalent form of ltDNA sample collected in forensic casework as they inherently contain very few nucleated cells (Kita et al., 2008), however sub-optimal samples that would usually be considered as a good source of DNA, but are aged or have been subjected to harsh environmental conditions, may also yield relatively low quantities of DNA. With a development in technology, along with the improvement in sensitivity of the related chemistries, there is now the capability to recover, extract and amplify DNA from even the most challenging of

forensic material. There has been much discussion amongst the forensic genetics community regarding what defines a DNA sample as being low template, with DNA quantity frequently favoured using upper threshold figures of either 100 pg (Whitaker et al., 2001, Forster et al., 2008) or 200 pg (Budowle et al., 2009a, Gill et al., 2008), equivalent of approximately thirty-four and sixty-four times the mass of a haploid genome (approximately 3 pg).

Gill and Buckleton (Gill and Buckleton, 2010) argue that it is impractical to define a threshold for low template DNA based on quantity because it does not necessarily reflect the quality of the DNA within the sample, i.e. stochastic effects of PCR may be evident for the high molecular weight (HMW) loci whilst the low molecular weight (LMW)are unaffected within the same profile, or in the case of a mixed sample, a number of individuals may all be contributing to the overall DNA quantification value.

1.2.5.2 Degradation of DNA

The degradation of DNA in apoptosis, the predominant form of programmed cell death, is characterised by the cleavage of chromosomal DNA initially into HMW fragments of 50 - 300 kb (Oberhammer et al., 1993). This is followed by subsequent cleavage into LMW fragments, producing a DNA ladder effect with intervals of approximately 200 bp when visualised on an agarose gel (Wyllie et al., 1980, Widlak et al., 2000). As reviewed by Samejima & Earnshaw (Samejima and Earnshaw, 2005) more than twenty nucleases have been identified as having varying degrees of involvement in DNA cleavage during apoptosis. One such enzyme primarily active in mammals is Caspase-activated DNase (CAD) (human homologue known as Caspase-activated nuclease), which exists as an inactive complex with it's chaperone, Inhibitor of CAD (ICAD) (Enari et al., 1998), in the nucleus of the cell (Lechardeur et al., 2000). A cell autonomous nuclease, CAD is predominantly involved in the degradation of LMW DNA (Widlak et al., 2000) but a role in HMW DNA fragmentation has also been demonstrated in some cell types (Zhang et al., 2000). Once ICAD has been removed, the scissor-like structure of the CAD homodimer (Woo et al., 2004) targets the phosphodiester bonds and cleaves the double stranded DNA (dsDNA) located in the inter-nucleosomal space (Widlak et al., 2000) creating nucleosomal units, which are approximately 200 bp in length. The mechanisms for the fragmentation of DNA through apoptosis have been extensively investigated; there is still very little understanding of the process of DNA digestion in both regulated (RCD) and accidental cell death (ACD) (Mizuta et al., 2013).

Hydrolysis of the N-glycosyl bond (Drohat and Maiti, 2014) that connects the nucleobase with the sugar/phosphate backbone, is catalysed by a monofunctional glycosylase that results in cleavage of the base to leave an abasic site (Lindahl and Karlstro.O, 1973, Lindahl

and Nyberg, 1972, Loeb and Preston, 1986). Due to the relative instability of this bond, hydrolysis at this site occurs readily with approximately 10,000 abasic sites created each day per cell (Lindahl and Nyberg, 1972). This process is often referred to as 'depurination' as it has been demonstrated that the rate of loss of cytosine and thymine ($1.5 \times 10^{-12}\ s^{-1}$) (Lindahl and Karlstro.O, 1973) is only 5% that of the loss of the purines, guanine and adenine ($3.0 \times 10^{-11}\ s^{-1}$) (Lindahl and Nyberg, 1972). Bifunctional glycosylases also act to catalyse the hydrolysis of N-glycosidic bond using an amine nucleophile from the enzyme itself creating an intermediate Schiff base, which then allows for the β-elimination of the phosphodiester backbone at the 3' end of the abasic site causing single strand nicks (Drohat and Maiti, 2014) and consequently further fragmentation of the DNA (Briggs et al., 2007, Ginolhac et al., 2012).

The formation of interstrand DNA cross-links between abasic sites and opposing nucleotides was first described in 1964 (Freese and Cashel, 1964). The reaction, which has only recently become fully understood, involves the aldehyde group of the abasic site interacting with the exocyclic N^2-amino group of an opposing guanine base (Dutta and Gates, 2006, Catalano et al., 2015) or the N^6-amino group of adenine (Price et al., 2014) to form an interstrand DNA cross-link (Dutta and Gates, 2006, Catalano et al., 2015). If left unrepaired, this type of cross-link can cause cell death to occur in a eukaryotic cell (Grossmann et al., 2001).

Hydrolytic (Orlando et al., 2013) or nitric oxide-induced (Wink et al., 1991) deamination causes a direct change in the sequence of cellular DNA through the removal of the amine group of the base. Deamination by hydrolysis is most frequently observed as a transition from cytosine to uracil ($5.8 \times 10^{-13}\ s^{-1}$) (Duncan and Miller, 1980) or 5-methylcyctosine to thymine ($2.6 \times 10^{-13}\ s^{-1}$) (Shen et al., 1994, Duncan and Miller, 1980), whilst the transition from adenine to hypoxanthine (Shapiro and Klein, 1966, Lindahl and Nyberg, 1974, Duncan and Miller, 1980) only occurs at approximately 2-3% of the purines (Karran and Lindahl, 1980). Conversely, nitric oxide-induced deamination more readily occurs in purine bases such as the transition from guanine to xanthine (Caulfield et al., 1998). The transition of one base to another in the DNA sequence can result in the incorrect nucleotide being incorporated into the daughter strand during replication. For example, the deaminated cytosine base uracil will go on to pair with adenine following which, during subsequent replication, adenine will in turn pair with thymine generating a C→T point mutation (Duncan and Miller, 1980).

Damage to DNA can also be caused by oxidation reactions involving reactive oxygen species (ROS) (Feig et al., 1994), such as the superoxide anion (O_2^{2-}), the highly reactive hydroxyl radical ($HO^{\cdot}$), Hydrogen peroxide (H_2O_2) and the singlet oxygen ($^1O_2$). These can

be formed as a by-product of normal cellular metabolism (Freeman and Crapo, 1982), a result of ionising or ultra violet (UV) radiation (Peak and Peak, 1989, Zhang et al., 1997, Ravanat et al., 2001) or through the metabolism of exogenous chemicals (Cooke et al., 2003). Oxidation can also occur by the loss of one electron from the nucleobase ring causing a charge that will make its way along the double helical structure of the DNA strand until it becomes trapped and results in damage to the DNA base (Kanvah et al., 2010). The site at which this reaction occurs is determined by the localised DNA sequence (Kanvah et al., 2010), but it has been widely shown that guanine is preferentially oxidised due to both its low ionisation potential (Steenken and Jovanovic, 1997, Fleming et al., 2015) and the steric effects of the chemical structure (Carter et al., 1996). The oxidation of guanine to form 8-oxo-7,8-dihydroguanine (8-OxoG) is the most frequently observed oxidative lesion with 1 in 40,000 8-OxoG sites present in the human genome (Shigenaga et al., 1990), 10% of which result in a G→T transversion site (Hanes et al., 2006).

In addition to this, the absorption of UV light by the double bonds in pyrimidine bases can lead to cyclobutane pyrimidine dimers (Setlow, 1966, Mouret et al., 2006) whereby adjacent pyrimidine bases are covalently linked to form TT, CT, TC or CC base pairings, or 6-4 photoproducts (Mitchell and Nairn, 1989, Ravanat et al., 2001), which involve a single covalent bond between two carbon atoms in the ring. In each type of pyrimidine dimer lesion, the intrastrand bonding of molecules results in disruption to the DNA helical structure, which will prevent DNA replication unless repaired (Rastogi et al., 2010).

In addition to the aforementioned mechanisms of DNA degradation, consideration must also be given to the potential for further degradation or loss of DNA as a direct result of laboratory processing and sample storage. During the preparation of some samples, particularly bone and teeth, it may be necessary to physically cut and grind the specimen prior to DNA extraction which may cause DNA fragments to shear or degrade due to friction or the heat produced, which is thought to cause double-strand breaks in the DNA leading to an increase in fragmentation (Purschke et al., 2010).

Although extremely difficult to quantitate, it is thought that the physical manipulation of the samples, for example through pipetting or vortexing, during DNA isolation can cause shearing of the DNA due to hydrodynamic breakage (Bowman and Davidson, 1972, Davison, 1959, Yoo, 2011). The fragmentation of DNA through use of a vortex however, has recently been refuted by Rossmanith et al. (2011) who did not observe any significant difference in the quantity of DNA recovered between two samples, one of which had been routinely vortexed over the course of a month (Rossmanith et al., 2011). The qPCR target

was however, only 274 bp and so may not reflect the level of degradation to the larger DNA fragments.

Storage of DNA extracts at -20 ˚C or -80 ˚C, along with freeze-thaw cycles over this period, results in the formation of ice crystals, which have been shown to physically shear and therefore further fragment the DNA (Ross et al., 1990, Shao et al., 2012). The effects of freeze/thaw cycles were observed in longer genomic DNA fragments (Shao et al., 2012, Rossmanith et al., 2011), i.e. greater than 100,000 bases (Shao et al., 2012), but not in 274 bp PCR amplicons, even after twenty freeze-thaw cycles (Rossmanith et al., 2011), suggesting that extracts containing degraded DNA would not be at further risk of fragmentation if stored frozen.

1.2.5.3 Mitocondrial DNA versus nuclear DNA

The multi-copy nature of the mtGenome within every cell offers two further avenues for use in forensic DNA analysis in the form of an alternative source of DNA in samples that either inherently contain little or no nDNA such as keratinous tissues (Parson et al., 2015, Higuchi et al., 1988, Anderson et al., 1999) or, of particular interest for this study, samples in which the quantity or quality of nDNA is in question, i.e. nDNA is expected to be severely degraded due to the effects of age and/or environment, or quantification results indicate a low concentration/absence of nDNA.

It has been estimated that nDNA may degrade approximately twice as fast as mtDNA (Allentoft et al., 2012), and so mtDNA can often be amplified from samples in which nDNA cannot (Schwarz et al., 2009, Krings et al., 1997). The ability of mtDNA to persist for a longer period of time than nDNA in forensic samples is thought to be due to a number of contributing factors. As previously mentioned the number of copies of the mtGenome present in human cells can be in the thousands thus resulting in an increased chance of mtDNA recovery and amplification relative to the nuclear genome, of which there is just one copy (Robin and Wong, 1988). Not only does the circular nature (van Bruggen et al., 1966, Nass, 1966) of the mtGenome provide resistance to exonuclease activity (Shokolenko et al., 2013) that linear nDNA (Watson and Crick, 1953) is susceptible to, but the compact protein-packaged core of the nucleoid (Kaufman et al., 2007, Nass, 1969, Brown et al., 2011, Lo et al., 2011) may offer a degree of protection to the mtGenome(s) within it. Support for this structural defence is provided by Foran (2006) who demonstrated that in whole tissue mtDNA showed less signs of degradation than nDNA, but when a sample of that same tissue was homogenised the mtDNA was shown to degrade at a quicker rate than the nDNA (Foran, 2006).

The early demonstrations that mtDNA persisted for a longer duration than nDNA (Krings et al., 1997), along with the phylogenetic and biogeographical information that could be obtained from this genome, made it the 'go to' marker for any sample that was suspected to be degraded; investigators were not willing to risk a negative result in the pursuit of nDNA. The entire DNA analysis workflow has however, been developed over time from DNA recovery right through to data analysis, and so the potential to extract and successfully sequence nDNA has increased. In addition to this, an improved understanding of the mechanisms behind DNA degradation has led to the design of a wide range of highly informative markers which maximise the information obtained from every sample.

1.3 Forensic DNA amplification kits

1.3.1 PCR-CE STR kits

The first reported assay was made up of just four STR markers (Kimpton et al., 1993, Kimpton et al., 1994). However, this was quickly replaced by the Second Generation Multiplex (SGM), containing six STRs and Amelogenin (Sullivan et al., 1993) for identification of sex. SGM had a match probability of $<5^{-6}$, significantly reduced from its predecessor, which had a match probability of approximately $<10^{-4}$ (Jobling and Gill, 2004). The match probability is calculated using allele frequency tables, which are based on the frequency with which a particular allele is observed at a specific locus in different populations around the world (based largely upon reference data sets). The allele frequencies for each present allele are used to calculate the genotype frequency at each locus using Hardy Weinberg. The resulting genotypes are then multiplied together (product rule) to obtain the match probability; the probability that two DNA profiles from unrelated individuals will match by random chance (Balding and Nichols, 1994). As more loci are included in an assay, the ability to discriminate between two individuals increased, resulting in a decrease in the match probability.

The SGM assay was further advanced through the addition of four new STRs (SGM Plus), which resulted in a reduction in the match probability of the assay to just $<10^{-13}$ (Jobling and Gill, 2004). The ten STR loci included in the SGM Plus kit are D3S1358, vWA, D16S539, D2S1338, DS81179, D21S11, D18S51, D19S433, TH01 and FGA. Introduced into forensic casework in 2000, the SGM Plus kit, a product of Applied Biosystems, was utilised by all forensic providers in the UK for loading to the NDNADB until 2014. Although the calculated match probability values for SGM Plus were so low that it was commonly reported as a standard 'less than 1 in billion', the Prüm Treaty and its subsequent adoption by the EU Council (Union, 2008) requires that an exchange of DNA profile information within Europe is possible and so the NDNAD was updated to include the European Standard

Set (ESS) of loci (Gill et al., 2006b, Gill et al., 2006a), and three 'DNA 17' kits were developed and approved for forensic use in England, Wales and Northern Ireland.

1.3.1.1 DNA 17 assays

In an opening of the forensic market, there are three DNA 17 kits that have been approved for loading to the NDNAD; NGM Select (Applied Biosystems) (Green et al., 2013), ESI 17 (Promega) (Tucker et al., 2011) and Investigator ESS Plex SE (Qiagen) (Qiagen, 2015a). All three of these kits incorporate Amelogenin, the same ten loci from the previous SGM Plus kit and an additional six loci, D10S1248, D22S1045, D2S441, D1S1656, D12S391 and SE33, which loci include the European standard set of twelve STR markers (Gill et al., 2006b). Although all kits utilise the same STR loci, the differing primer binding sites and chemistries result in slight variations in amplification sensitivity and also potential discordance between results for the same sample (Tomas et al., 2014). An example of a DNA 17 assay is shown in Figure 1.3.

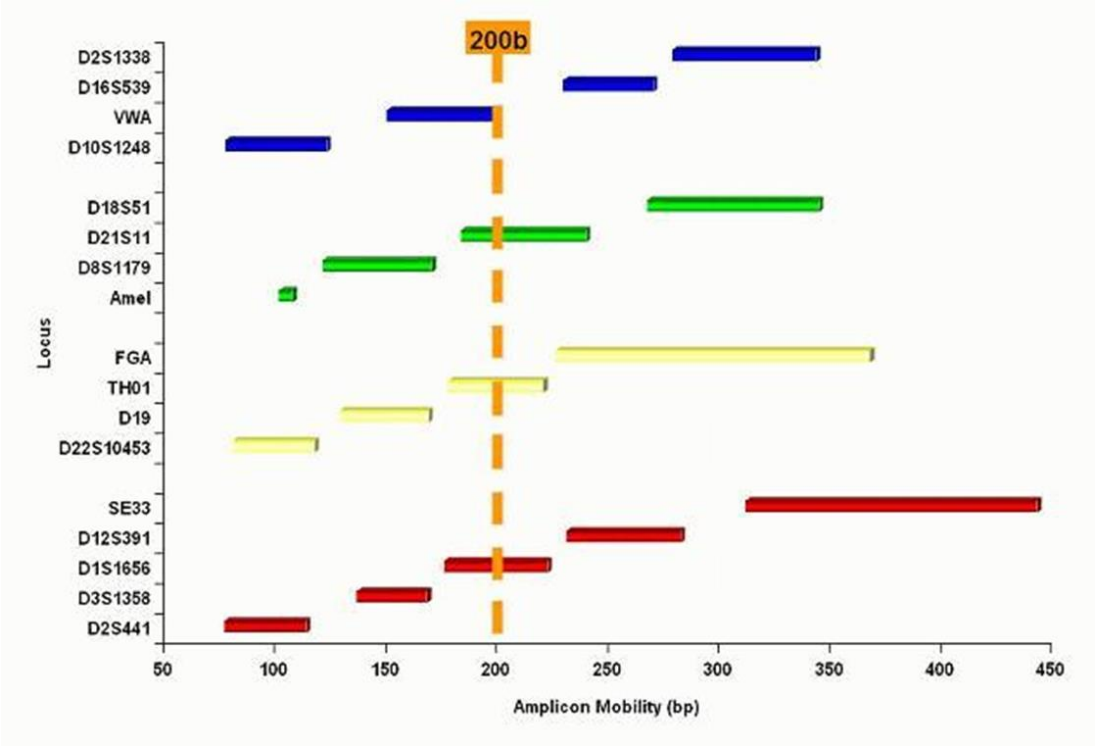


Figure 1.3 Example of a DNA 17 kit. The chart is a representation of the loci included in the Applied Biosystems NGM Select kit (Biosystems, 2015), and the allele size range for each locus. The different colours represent the colour of the fluorescent dye attached to the respective primer sets. Those loci with a discriminatory value have been designed where possible in the low molecular weight range (< 200 bp).

1.3.1.2 Alternative assays

Aside from those kits developed for loading to the NDNAD, there are numerous additional assays that have been developed over the years. Whilst many have been developed for loading to the national databases of countries outside of the UK or for non-database identification, there are kits which have been developed for use in specific circumstances. AmpFℓSTR® MiniFiler for example, includes only eight low molecular weight loci, and so is often used for amplification of samples that are expected to be degraded/inhibited (Mulero et al., 2008). PCR assays such as AmpFℓSTR® Identifiler Direct (Wang et al., 2011) and Promega's ESX/ESI systems (Tucker et al., 2012a, Tucker et al., 2011) have been developed for use in direct PCR, whereby the sample is added directly to the PCR tube without pre-isolation of the DNA (Section 3.1), and so have a higher tolerance for the potentially increased presence of PCR inhibitors. AmpFℓSTR® YFiler has been designed to target seventeen STRs on the Y-chromosome only (Mulero et al., 2006), which is particularly useful in cases with a high background level of female DNA such as in sexual assault casework. In more recent years SNP assays have been developed for compatibility with capillary electrophoresis (CE) instruments, including those for determining hair and eye colour; most notably HIrisPlex which contains a total of twenty-four SNPs (Walsh et al., 2013).

1.3.1.3 Limitations of PCR-CE

Despite the success of the aforementioned PCR-CE kits for DNA profiling in forensic science, there are some significant limitations to this method. The number of multiplexed DNA fragments of the same or overlapping size that can be analysed at any one time using CE, is restricted due to constraints on the number of different dyes (currently a maximum of six) that can be detected by the instrument itself.

Current forensic DNA amplification assays rely on the use of fragment length analysis rather than the sequence of the STR itself. It is known that SNPs can occur within the STR repeat units or flanking regions, resulting in isoalleles (Section 5.1) that could enable further distinction between two individuals or help resolve DNA mixtures.

Due to the limitations on the number of loci able to be detected using CE, along with the necessity of specific allelic ladders, it is extremely complex to multiply differing marker types for simultaneous analysis. Consequently multiple amplification reactions are required; each one consuming part of what may be a very limited DNA extract.

The introduction of massively parallel sequencing (MPS) has alleviated or resolved the aforementioned limitations associated with CE fragment analysis, thereby facilitating the

development of a new generation of multiplex amplification assays for use in forensic casework.

1.3.2 Massively Parallel Sequencing

The development of MPS was originally driven by the need to reduce both the time and financial costs associated with Sanger sequencing for large scale genomic study (Collins et al., 2003), however this multifaceted technology has found numerous additional and diverse applications across the spectrum of genetics, including in clinical diagnostics (Jones, 2015), molecular ecology (Ekblom and Galindo, 2011) and functional genomics (Morozova and Marra, 2008). The fields of forensic science and bioarchaeology are no exception to this as the potential is explored to develop both novel approaches to their work or for opportunities to make improvements on traditional techniques in order to generate a better quality and quantity of results.

1.3.2.1 Development of platforms

The first MPS platform, released by 454 Life Sciences in 2005 (Margulies et al., 2005) was based upon the use of pyrosequencing. This method uses a DNA library which is first prepared by randomly fragmenting double-stranded DNA (dsDNA) and ligating specialised adaptors to each strand; the DNA is then denatured and captured by beads so that each bead has one, single-stranded DNA fragment attached to it. The DNA fragments are then amplified in-situ using emulsion PCR, resulting in beads that have 10 million copies of their DNA fragment attached. The beads are then transferred onto a slide that contain wells 55 µm deep (Margulies et al., 2005), enough to hold a single DNA bead and smaller beads containing the immobilised enzymes need for the pyrosequencing reaction to occur. During sequencing, the reagents flow past the open side of the vertically held slide, allowing them to diffuse into the wells to surround the beads. As a dNTP is incorporated into the sequence by a DNA polymerase, a nucleotide dependant quantity of pyrophosphate is released, which is then transformed to ATP, in the presence of adenosine 5’ phophosulfate by ATP sulfurylase. ATP then converts luciferin to oxyluciferin, resulting in the generation of photons that are detected by a charge coupled device camera on the closed side of the wells. Apyrase is utilised to digest remaining dNTPs and then the reaction is repeated.

This first sequencing by synthesis technology demonstrated the capability to simultaneously sequence over 300,000 DNA fragments (approximately 100-110 bases) of a *Mycoplasma genitalium* genome, with 96% accuracy in a single run of four hours (Margulies et al., 2005), representing a 100-fold improvement on the Sanger sequencing technology. Just three years later, the extent to which MPS was having already having an impact on the field of genomics was demonstrated when the first human genome sequenced using this technology was

published, taking just two months to complete (Wheeler et al., 2008), a feat that had previously required a worldwide collaboration and over a decade in time using the Sanger technique (Collins et al., 2004).

In the years following the release of the 454 system (purchased by Roche but recently discontinued) several MPS platforms have been developed, improving performance in terms of run speed, read length, sample capacity, accuracy and quality.

Developed by the same designer, the technology underlying the Ion Torrent (now owned by Thermo Scientific) PGM™ and Proton™ series platforms, is principally the same as the previously described 454 system. The difference between these two technologies is that where the 454 system relies on photon detection resulting from the incorporation of a nucleotide, the Ion Torrent platforms uses semiconductor sequencing which measures the release of protons. The released proton causes a change in the well pH which is detected as a difference in potential across a metal oxide layer that is placed between the bottom of the well and the sensor plate, thus negating the requirement for scanning with a camera and so decreasing the time taken per run (Rothberg et al., 2011). The Ion Torrent PGM™, released in 2010, is currently the quickest bench top sequencer commercially available with the of the longest read length capacity of 400 nucleotides (van Dijk et al., 2014a). The Proton™, Ion Torrents latest benchtop platform, has been designed for high-throughput use with the capacity to produce up to 10 Gb of data on 200 bp reads in less than four hours (Technologies, 2015). Both the 454 and Ion Torrent systems have been shown to be prone to errors in homopolymeric regions of the DNA sequence (Loman et al., 2012, Margulies et al., 2005).

The Sequencing by Oligo Ligation Detection (SOLiD®) sequencing chemistry (Thermo Scientific) introduced in 2007, is again very similar to the 454 system, using beads and emulsion based PCR for target amplification. Rather than dispensing the beads into wells however, the template DNA undergoes a 3' modification which enables a covalent bond to be formed between the bead and a glass slide. Primers are hybridised to the adaptors that were added during library DNA preparation and then four fluorescently labelled di-base probes are used in a ligation reaction. When a probe is complementary to the template DNA, it hybridises and is then ligated, prior to fluorescent detection and imaging. The dye is then cleaved off with silver ions to liberate a 5' phosphate group and the process is repeated for the required number of cycles. The synthesised DNA strand is removed and a new primer is then hybridised, offset by one nucleotide base and the steps repeated so that each base across the sequence is read twice (Technologies, 2013, Mardis, 2008). The key advantage of this system is its high base calling accuracy of 99.85% after filtering (Liu et al., 2012) however

due to its short read length of 35 nucleotides and long run time this platform is not widely used for research. The introduction of the SOLiD® Wildfire system in 2013, has sought to rectify these short comings as the DNA library amplification is now carried out directly on a flow chip using isothermal amplification rather than the previous bead/emulsion PCR system (Ma et al., 2013), thus increasing the read length to 75 nucleotides whilst decreasing both the preparation/run times by a quarter (Technologies, 2012).

Illumina's sequencing by synthesis technology, acquired from Solexa in 2007, is the basis for their MiSeq, NextSeq, HiSeq and HiSeq X series sequencing platforms (Illumina, 2014b). DNA libraries are pooled, denatured and added to the flow cell where single-stranded DNA will hybridise with complimentary adaptor sequences. A DNA polymerase is used to synthesise a complimentary strand of DNA which is attached to the flow cell and the original template is then washed away. The oligo at the free end of the remaining DNA strand then hybridises with a complimentary adaptor on the flowcell to form a bridge and a new strand of DNA is synthesised. The dsDNA is denatured, resulting in two independent strands that are anchored to the flowcell, which then each undergo bridge amplification until a cluster comprising of up to one thousand copies of the original template is generated. All reverse strands are cleaved and washed away, all 3' ends on the oligo lawn and DNA fragments are blocked and a sequencing primer is added to the flowcell, which hybridises to the template DNA. All four fluorescently labelled, reversible terminators are added to the flowcell and the one which is complimentary to the template will be incorporated by the DNA polymerase, whilst the remaining terminators are washed away. The fluorescent label acts as a polymerase terminator and, using total internal reflection fluorescent imaging with two lasers, the incorporated nucleotide can be identified. The fluorescent tag is then cleaved to allow for the addition of the next nucleotide (Illumina, 2010). This cycle of incorporation, washing and cleaving repeats of up to 300 bases (Illumina, 2014b). 10 million clusters per square centimetre of the flow cell means that billions of reads can be taken per run.

Using the HiSeq X platform, Illumina is the first company to meet the $1000 human genome (van Dijk et al., 2014a), which achieved using two flowcells, each with a capacity for 3 billion reads of 2 x 150 bp lengths, generating up to 1.8 Tb in less than 3 days (Illumina, 2014b). Illumina are widely acknowledged as the current leaders in the MPS market (Metzker, 2010, Quail et al., 2012, van Dijk et al., 2014a), with >90% of all the worlds current sequencing data generated using their platforms (Illumina, 2014b). This is most likely due to the fact that although other platforms can be more accurate (Technologies, 2013), quicker (Technologies, 2015) or have longer maximum read lengths (Bayley, 2015), the Illumina platforms represent a balance between all these variables. The limitation with technology primarily lies in the use of two lasers, red and green; each of which is required to

differentiate between two dyes, A/C and G/T respectively that have similar emission profiles. Schirmer et al. (Schirmer et al., 2015) used 16S bacterial data to demonstrate that the nucleotide substitution errors occur with increasing frequency towards the end of each read and that read 2 has a significantly higher error rate. It was proposed that the increase in errors at the end of the read sequences is due to an accumulation of phasing and dephasing events, which results in a loss of cluster signal (Schirmer et al., 2015). This also means that variable clustering density on the flowcell or a DNA library lacking in complexity can result in a loss of base calling accuracy or the inability to demultiplex the sequences (van Dijk et al., 2014a, Schirmer et al., 2015).

A significant development in MPS was the introduction of technologies that no longer required a DNA amplification step as part of the protocol for whole genome or *de novo* sequencing. The removal of PCR associated artifacts and amplification bias (Hart et al., 2010) results in sequences that are truly reflective of their endogenous state, in terms of both sequence fidelity and the numbers observed.

The HeliScope™ Single Molecule Sequencer (SMS) (SeqLL, Woburn, MA, USA) was the first technology to be commercially released that does not include a PCR enrichment step in the protocol. Briefly, extracted DNA is cleaved into fragments of 100-200 bp in length and a poly-A universal priming sequence is added to the end of each strand. A fluorescently tagged adenosine nucleotide is then added to the end of the poly-A sequence and the DNA fragments are loaded onto the flow cell for hybridisation with oligo-T universal capture sites. The surface of the flow cell is illuminated and the incorporated fluorescent adenosine enables the position of every strand recorded by CCD camera. The template label is then removed and a DNA polymerase plus one type of fluorescently tagged nucleotide is added to the flow cell and then subsequently washed away to remove any excess. The flow cell is again illuminated and images capture the positions of the strand that have incorporated the additional nucleotide. The fluorescent tags are removed and the step is repeated for each type of nucleotide for the required number of cycles with every DNA strand to being sequenced independently (Hart et al., 2010). The HeliScope™ SMS was the first single molecule sequencing technology to be used to sequence a human genome in 2009 (Pushkarev et al., 2009), achieving 90% alignment to the reference genome. One of the biggest advances introduced by the HeliScope™ SMS is the ability to directly sequence RNA molecules through the use of a reverse transcriptase enzyme rather than a DNA polymerase, thus negating the requirement for prior conversion to cDNA (Ozsolak et al., 2009).

The development of the PacBio RS II Real-Time (SMRT®) DNA Sequencing System (Menlo Park, CA, USA) in 2010 represented another step forward in MPS as it was the first SMS platform with the ability to record the DNA sequence in real time (Eid et al., 2009). This is accomplished using an array of up to seventy-five thousand 29Φ polymerase molecules (Schadt et al., 2010), each one of which is immobilised on a glass surface within a nano-photonic zero-mode waveguide (ZMW) (a 43-50 nm hole within a 100 nm metal film) (Levene et al., 2003). Nucleotides, which are fluorescently labelled on the phosphate chain of the nucleotide rather than the base (Korlach et al., 2003), are passed over the ZMWs and down to the polymerase molecule by diffusion. The size of the ZMW restricts its illumination to the bottom 30 nm; therefore any nucleotides above this level will not be detected (Levene et al., 2003). As a nucleotide is incorporated by the polymerase it fluoresces and the signal is recorded before the phosphate chain and the fluorescent tag are cleaved. The use of this nanotechnology based approach enables the polymerase to work without the interruptive steps of nucleotide flushing/washing, signal recording and fluorescent tag cleaving, which results in an increase in the length of reads, approximately twenty thousand nucleotides, and also a significant reduction in the time take to complete a sequence (Levene et al., 2003, Eid et al., 2009, Schadt et al., 2010).

Nanotechnology has also been employed by Oxford Nanopore Technologies (Oxford, Oxfordshire, UK) in their three instruments the GridION™, PromethION™ and MinION™. As alluded to in the name, the recently launched MinION™ is a hand held device and is the only portable device currently commercially available. This technology utilises the negative charge of DNA to draw each strand through a 1 nm protein nanopore, which is set in a synthetic polymer membrane containing several hundred nanopores (Bayley, 2015). An enzyme sits above the nanopore, separating the DNA, which allows a single strand to then be drawn through the centre of the nanopore. The DNA strand blocks part of the ion current as it moves through the pore, causing a disruption in the current that is characteristic of the nucleotides present, thus allowing the sequence of the passing DNA strand to be identified. The DNA library is prepared so that each dsDNA molecule has a hairpin at one end so that as one strand is sequenced, the opposite strand is then drawn through the nanopore for sequencing also (Clarke et al., 2009).

The serious limitation shared by all current SMS technologies is in the high error rate of raw reads ~ 5% (HeliScope) and 15% (PacBio) in, particularly for the detection of insertions and deletions (indels), which can make sequence alignment extremely complex. In an attempt to sequence the genome of the lambda phage, just 10% of the raw reads mapped to the reference genome (Mikheyev and Tin, 2014), even when using the Basic Local Alignment with Successive Refinement (BLASR) tool, which has been developed to accommodate the

longer read lengths and the high instances of indels found in SMS technologies (Chaisson and Tesler, 2012). It has been suggested that the high number of indels observed in the MinION™ data may be due to signal errors caused by an uneven movement of the DNA strand as it passes through the pore (Mikheyev and Tin, 2014). Whilst these technologies have found application for rapid clinical tests (Quick et al., 2015, Jones, 2015), they are unlikely to replace the PCR based methods for de novo sequencing until the sequencing becomes more robust (Laver, 2015, Mikheyev and Tin, 2014).

1.3.2.2 Library preparation

Library preparation is the first step in the process irrespective of the sequencing platform to be used and is crucial in determining the quality of the sequencing data generated (van Dijk et al., 2014b, Schirmer et al., 2015). Many protocols use either mechanical shearing or enzymatic cleavage to randomly fragment the DNA strands (Knierim et al., 2011) followed by a PCR step which allows for the platform specific adaptors to be attached. Specific indexes can also be incorporated at this point in cases where multiple samples are processed together. In the case of Illumina platforms, paired-end sequencing allows for the use of a dual index system whereby DNA fragments have an index at each end of the DNA strands to be sequenced (Syed et al., 2009) increasing the number of samples that can be multiplexed together. A size selection or PCR clean up step can then be utilised to remove any unused primers/adaptors from the reaction prior to sequencing. The use of a PCR step during DNA library preparation results in amplification bias (Dabney and Meyer, 2012) that is reflected in the resulting sequence data, particularly evident in samples that are not GC neutral (van Dijk et al., 2014b). It has been shown that the choice of DNA polymerase can affect the extent of bias observed (Dabney and Meyer, 2012), as can the inclusion of additives such as betaine, which reduces the bias seen when amplifying GC rich genomes (Aird et al., 2011, van Dijk et al., 2014b).

A pair-ended, dual index system developed by Kozich et al. (Kozich et al., 2013) for sequencing of amplicons from the bacterial 16S region has circumvented the requirement for the fragmentation of DNA, instead employing a one-step PCR during which the DNA extract is amplified using primers which include one of the two platform specific flowcell adaptors (in this case the Illumina MiSeq), a unique sample index and the target region. The forward and reverse primers each carry different index sequences, which together create a unique sample identifier. Removing the requirement for a second PCR step, which would usually be required to attach the flowcell adaptors, reduces the potential for PCR artifacts and can be customised for any target sequence.

Illumina's TruSeq DNA PCR-Free Library Preparation Kit uses ligation to attach adaptors to sample DNA, thus eliminating the PCR bias at this stage of sample preparation. The suggested DNA input for this technique is relatively high at a minimum input of 25 ng required, however it has been demonstrated that this kit can be effective with as little as 6 ng of amplified STR product (Zeng et al., 2015).

1.3.2.3 Target enrichment techniques

Target enrichment techniques allow for the DNA sequencing of areas of the genome of specific interest to be analysed, thus optimising the amount of information obtained for the cost/time required for processing and data analysis. Historically, this role has been fulfilled by PCR, but difficulties in designing efficient large-scale multiplex reactions particularly non-specific and preferential amplification, limit the capability of most PCR techniques to take full advantage of MPS. Micro-droplet PCR however, alleviates some of these issues through the separation of DNA template and PCR reagents into single-plex reaction droplets (Tewhey et al., 2009).

An alternative to PCR is the use of molecular inversion probes (MIP) which are made up of a synthetic single (Nilsson et al., 1994) or double (Yoon et al., 2015) oligonucleotide strand that has a highly specific sequence region at each end. These regions hybridise both upstream and downstream of the target, which is then copied and incorporated into a circular DNA molecule by polymerase and ligase enzymatic activity (Nilsson et al., 1994). An exonuclease is then used to digest any linear background DNA, thus increasing the proportion of the target sequence, which can then be amplified prior to sequencing. This technique has been particularly successful in targeting exons in the genome for DNA analysis (Porreca et al., 2007) with up to 50,000 exons enriched simultaneously in a 55,000 multiplex reaction (Turner et al., 2009), and also in DNA methylation studies (Deng et al., 2009, Diep et al., 2012).

Hybridisation capture is a target enrichment method that uses target specific oligonucleotides which are immobilised on a microarray platform (Albert et al., 2007) or in solution using a biotinylated probes and beads system (Gnirke et al., 2009). Any non-specific DNA is then washed away, thus enriching the regions of interest and once isolated, the target sequences are then released from the probes and eluted ready for MPS. The key advantage of both MIP and hybridisation methods is the capability to enrich large target fragments of DNA in a more efficient way than PCR based techniques, in that it can be completed in a single reaction using as little as 3 µg of DNA (Mamanova et al., 2010).

Microfluidic droplet enrichment for targeted sequencing (MESA) is a novel technique recently been published by Eastburn et al. (Eastburn et al., 2015), which eliminates the requirement for either PCR or hybridisation techniques. Utilising droplets to encapsulate DNA molecules, the 50-200 kb target sequences are identified by amplification of approximately 100 bps with a highly specific Taqman assay. The Taqman identified sequences are then separated from those droplets without the target sequence by dielectrophoretic sorting when the fluorescent tag is detected. The collected DNA target sequences are then released from the oil based droplets and the amplified section is removed from the DNA fragment enzymatically, prior to sequencing (Eastburn et al., 2015). This technique allows for the enrichment of longer target sequences than PCR or hybridisation methods and resulted in a 13-fold enrichment of a 200 kb DNA target. The relatively high DNA input requirement of 100 ng however, will again need to be reduced for utilisation with forensic samples.

1.3.2.4 Use of MPS for Human Identification

The investigation governed nature of forensic DNA analysis requires a focused approach, exploiting highly informative markers in order to answer questions posed during the inquiry. Sequencing of large regions, or indeed the whole human genome is therefore a redundant endeavour in terms of the added value of the additional information, as well as the time and monetary considerations. There are two notable exceptions to this type of approach; firstly, the recently demonstrated distinguishing capability in haplotype when the whole mtGenome is sequenced (Just et al., 2015, King et al., 2014a) as discussed in Section 1.2.3 and secondly is the use of whole nuclear genome mapping by MPS to differentiate between monozygotic twins based on rare polymorphisms, who would otherwise be indistinguishable from one another (Weber-Lehmann et al., 2014).

The markers traditionally utilised in forensic casework are currently undergoing development and evaluation for use on the MPS platforms, as well as the rigorous validation processes that accompany any new technology for forensic application. In addition to a limited read capacity, the number of multiplexed DNA fragments of the same or overlapping size that can be analysed using capillary electrophoresis is restricted due to constraints on the number of different dyes that can be detected by the instrument.

For the majority of sequencing work, depth of coverage refers to the average coverage per base across a run, whereas targeted sequencing for forensic applications relies on the actual amplicon or base count for comparison between alleles/loci and for the accurate identification of the present variants. The minimum depth of coverage that can be used to accurately identify the allele/variant calls in a profile will determine the number of loci that

can be multiplexed in one reaction, which in turn impacts on the number of samples that can be processed in one run (Eduardoff et al., 2015). Investigations into the depth of coverage required for forensic analysis found that for single source profiles, 13x – 40x coverage (Quail et al., 2012, Eduardoff et al., 2015, King et al., 2014a) is sufficient to accurately identify variant mutations.

In an early study on the use of MPS for STR sequencing, Bornman et al. (Bornman et al., 2012) demonstrated a >99% confidence in allele designation using a depth of coverage of 18,500 reads. This figure however was generated using the limited 150 base single read sequences available on the Illumina MiSeq platform at the time of this study and is therefore likely to be grossly over-estimated using updated chemistry. A more recent study using 454 pyrosequencing found that a depth of coverage of 160 reads was sufficient to accurately designate alleles (Fordyce et al., 2011) whilst another set filters to 1% of the total read count (Van Neste et al., 2012). Work to determine the minimum depth of coverage required for accurate STR allele designation is still on-going (Zeng et al., 2015) and will be platform, chemistry, marker and locus dependant.

Aside from the ability to multiplex markers and samples, one of the significant advantages to using MPS for forensic analysis is that rather than a profile based on fragment size analysis, the actual sequence of the STR is obtained. Numerous isoalleles (Heinrich et al., 2005, Rockenbauer et al., 2014, Gelardi et al., 2014, Dalsgaard et al., 2014) as well as SNP sites in the flanking regions of the repeat structures (Divne et al., 2010) have already been documented, which provide the capability to distinguish between two individuals with the same allele call for a locus, thus increasing the discriminatory value of the allele and also aiding the deconvolution of mixtures. The allele sequence may also prove particularly useful for distinguishing stutter from the true allele in situations where one allele is a single repeat less than the other (Zeng et al., 2015).

Gelardi et al. (Gelardi et al., 2014) proposed a nomenclature system for isoalleles to facilitate consistency across the field. Firstly, the standard locus name is listed followed by the length of the STR divided by the length of the individual repeating unit. The sequence and number of each repeat unit(s) is then added and lastly any flanking region variants are listed, using rs numbers where possible (Gelardi et al., 2014). For example two differing versions of D3S1358 locus allele 14 would be written as D3S1358[14]TCTA[1]TCTG[3]TCTA[10] and D3S1358[14]TCTA[1]TCTG[1]TCTA[12] (Gelardi et al., 2014). In order to better accommodate variants in the flanking regions of the STR target, it has been suggested that additional information for each variant be included in the format; locus, allele, chromosome: position and reference genome (Gettings, 2015),

which carries the advantage of limiting discrepiences as reference genomes are updated. Alternatives to this bracketed approach have also been suggested by Gettings et al. (Gettings, 2015) including the use of the complete MPS generated string sequence which could also allow for recording of platform, chemistry etc. utilised for sequencing. Whilst the transition to MPS technologies requires a thorough investigation into the most apposite method for recording the generated sequence data, the development of new nomenclature systems will inevitably result in inconsistency amongst data sets as individuals use their own preference. Progress in this area is quickly needed so that publication of guidelines by a standardising body such as ISFG can be issued to ensure uniformity.

Data analysis is currently a substantial constraint on the introduction of MPS for use in routine casework. A lack of end to end analysis software and the sheer quantity of data produced by any MPS run has resulted in the utilisation of developed or adapted ad hoc workflows, which can make it very difficult to assess the data quality and also complicates data comparison between groups, due to differing parameters or nomenclature. The development of multiple Y-SNP phylogenies is one example of this, whereby the once ubiquitous Y-Chromosome Consortium tree has been superseded by several differing trees as new data sets are released (Larmuseau et al., 2015).

The ability to multiplex a variety of markers of the same size is a significant advantage of MPS over the traditional PCR-CE methodologies for the processing of degraded forensic samples. Smaller amplicons have a better success rate in samples containing sub-optimal DNA and so a larger quantity of markers will maximise the information available from these sample types.

1.3.3 MPS forensic multiplex kits

There are currently a limited number of commercially produced forensic assays available for use with MPS technology. Both Illumina and Thermo Scientific have developed kits exclusively for use with their own platforms, whilst Promega have developed a series of assays for use with the Illumina MiSeq platform. A summary of the number and type of markers utilised by each is provided in Table 1-1. The Ilumina Forenseq™ DNA Signature Prep Kit clearly is ahead of the rest of the market in terms of the variety of markers analysed and the number of markers included for analysis.

1.3.3.1 MiSeq FGx Forensic Genomics System

The release of the ForenSeq™ DNA Signature Prep Kit (Illumina, 2015a), which has been developed by Illumina for use exclusively with their MiSeq FGx Forensic Genomics System (Illumina, 2015b) is the first assay which exploits the marker multiplexing capability of

MPS. This kit contains two primer sets; set A contains autosomal, X and Y STRs and identity SNPs, whilst phenotypic and ancestry informative SNPs are additionally included in set B, as summarised in Table 1-2.

87 (56%) of the markers used in primer set A and 149 (64%) of those included in set B are below 150 bp in size (Table 1-3). If the DNA fragment length has degraded to less than 200 bp (Wyllie et al., 1980), then 119 and 190 loci in primer sets A and B respectively should still be amenable to DNA amplification, depending on the degree of degradation, representing a substantial improvement in terms of the potential for results when compared with the amplification kits currently used in forensic casework. For example, only eight out of seventeen loci in the AmpFℓSTR® NGM™ Kit and seven out of nine loci of the AmpFℓSTR® MiniFiler™ Kit loci (designed specifically for degraded samples) fall below 200 bp.

In line with the traditional amplification kits, the suggested DNA input for this kit is 1 ng (Illumina, 2015a) and although, due its recent launch date, there is very little information available on DNA sensitivity tests, Børsting & Morling (Borsting and Morling, 2015) did make reference to having obtained results from a DNA input of 50 pg. This figure was given with the caveat that data were still being processed but if this is indeed the case then, given the 5 µl volume (Illumina, 2015a) allowed in the first PCR set-up then the minimum DNA concentration that would be required is 10 pg / µl, a figure below that considered for low template DNA forensic samples. It should be noted however, that no information was provided as to quality of the resulting DNA profile at this low DNA input.

Table 1-2 Summary of commercial forensic kits currently or soon to be available for use with MPS platforms. The two kits containing mtDNA include primers for ten overlapping sections of the control region and PowerSeq™ Auto/Mito/Y is the first commercial kit to contain both STRs and mtDNA targets.

Company	Kit name	Platform	Marker type (number of targets)							Reference
			STRs	Y-STRs	X-STRs	Identity SNPs	Ancestry SNPs	Phenotypic SNPs	mtDNA	
Thermo Scientific	HID-Ion AmpliSeq™ Identity Panel	Ion Torrent PGM platform				✓(169)				(Borsting et al., 2014, Eduardoff et al., 2015, Seo et al., 2013)
Thermo Scientific	HID-Ion AmpliSeq™ Ancestry Panel	Ion Torrent PGM platform					✓(165)			(Churchill et al., 2015)
Thermo Scientific	Ion Torrent™ HID STR 10-plex assay	Ion Torrent PGM platform	✓(9)							(Fordyce et al., 2015)
Promega	PowerSeq™ Auto System	MiSeq MPS platform	✓(23)							(Zeng, 2015)
Promega	PowerSeq™ Mito	MiSeq MPS platform							✓(10)	(Downey, 2015)
Promega	PowerSeq™ Auto/Mito/Y	MiSeq MPS platform	✓(22)	✓(23)					✓(10)	(Downey, 2015)
Illumina Inc.	ForenSeq™ DNA Signature Prep Kit	MiSeq FGx Genomics System	✓(27)	✓(24)	✓(7)	✓(95)	✓(56)	✓(22)		(Illumina, 2015a)

Table 1-3 Maximum amplicon size of markers included in ForenSeq™ DNA Signature Prep Kit adapted from (Illumina, 2015a).

	Maximum amplicon size (bp)				
Marker type	<100	100-150	151-200	>200	Total
Identity SNPs	24	53	17	1	95
Autosomal STRs	0	8	10	9	27
Y STRs	0	2	4	18	24
X STRs	0	0	1	6	7
Phenotypic SNPs	3	8	4	7	22
Ancestry SNPs	18	33	5	0	56

1.3.3.2 Mitochondrial DNA analysis

Illumina have also made available a protocol for mtDNA D-loop amplification (Illumina, 2013) or use with the MiSeq benchtop platform, including a list of suggested primers that have been adapted for duplex tagging, whereby two versions of each primer set are used in the reaction so that consensus between them can be checked. This method helps to maintain the true ratio of nucleotide bases in heterogeneous mixtures, which may otherwise become skewed as a result of bias during amplification and sequencing (Schmitt et al., 2012). Two overlapping amplicons are used, spanning the HVSI (nucleotide position 15997-16401) and HVSII (29–408) regions, with each set targeting a region of between 236-256 bp. While this has been demonstrated as a successful approach for forensic sample types meeting the mtDNA input requirement of 1 ng (Davis et al., 2015), the minimum fragment size of 236 bp may be an optimistic target for highly degraded DNA.

It is essential that the capabilities and limitations of this generation of newly developed, commercially available assays are explored, as well as those ‘in-house’ developed methods for forensic application, as these data will help inform the forensic community as they discuss the potential for the introduction of MPS into casework.

1.4 Aims and objectives

1.4.1 Aim

The overall aim of this thesis is to explore the use of enhanced experimental strategies for the analysis of sub-optimal biological material that is encountered during forensic investigations. This aim will be addressed as two areas of research; the first of which will be an investigation into underexploited low-template DNA (ltDNA) sources in forensic casework. The second will focus on sample types that are expected to yield severely degraded DNA. Novel, existing and emerging laboratory strategies will be utilised to fulfil these aims.

1.4.2 Objectives

Historically, the use of single-cell analysis within a forensic context is rarely carried out due to challenges associated with DNA amplification, contamination prevention and data interpretation. Over the last decade, in addition to the availability of increasingly sensitive DNA profiling kits our knowledge and understanding of the factors affecting amplification in low DNA template samples has developed. The first part of this research aims to investigate the analysis and interpretation of DNA derived from single-cells in order to determine their efficacy for use in forensic casework.

The second part of this research will investigate an application of single-cell DNA analysis within a forensic context through testing of the hypothesis that when textile fibres are transferred from one surface to another, they could also be acting as a vector for the wearer's own DNA, via cells that have adhered to the fibre surfaces. If proven, this would offer forensic scientists a method by which transferred fibres could be targeted to obtain a DNA profile, with the two evidence types being utilised alongside one another in forensic casework.

The focus of the final part of this thesis is directed towards teeth and bone that is expected to contain severely degraded DNA due to either age or exposure to a hostile environment. The value of massively parallel sequencing techniques for degraded DNA, through the targeted sequencing of both nuclear DNA markers and regions of the mitochondrial genome will be tested, using traditional PCR-capillary electrophoresis techniques as a point of comparison.

2. Materials & methods

2.1 Anti-contamination measures

Personal protective equipment was worn at all times during sample preparation, DNA extraction and PCR set-up; this included a hairnet, facemask, gloves, oversleeves and a dedicated labcoat. All pre-PCR sample manipulations were carried out inside a UV cabinet within a restricted access laboratory that post-PCR material did not enter. The laboratory was deep cleaned approximately every 4-6 weeks using a 10% (w/v) Sodium hypochlorite solution. As a minimum, all bench surfaces were cleaned using 10% (w/v) Sodium hypochlorite and the preparation hood was treated with UV light for 10 min before and after every use. DNA free disposable equipment was used where possible, including the use of pre-sterilised, aerosol-resistant filter tips to prevent contamination of the pipettes. Plates and tubes were treated with UV light for a minimum of 10 min in a UV Crosslinker box at an exposure output of approximately 999.9 J/cm^2. All other equipment was wiped using 10% (w/v) Sodium hypochlorite solution and then, where possible, treated in a UV Crosslinker box for a minimum of 10 min (999.9 J/cm^2). Where possible, solutions were prepared immediately prior to use in order to reduce the risk of contamination and, if not DNA-free were then treated with UV light for a minimum of 10 min (999.9 J/cm^2).

Negative controls were introduced during DNA extraction, DNA quantification and pre-PCR set-up and processed exactly as the samples, in order to detect any DNA contamination. All post-PCR manipulations were carried out in a dedicated area, wearing a dedicated labcoat and disposable gloves.

Use of the Upright Leica DM5000 B Fluorescent Microscope System was in an isolated cubicle within a multi-user room. This area was cleaned down using 10 % (w/v) Sodium hypochlorite and PPE, including a disposable labcoat, was worn at all times. All equipment was cleaned and samples prepared in the dedicated pre-PCR lab as previously described.

2.2 Additional measures for ancient DNA

In addition to the above the pre-PCR laboratory was deep cleaned using a 10% (w/v) Sodium Hypochlorite solution immediately prior to the start of work and then after DNA extraction was completed. The lab was also deep cleaned between the processing of different samples. Equipment that was not DNA-free was soaked in a 10% (w/v) Sodium Hypochlorite solution, rinsed with DNA-free water and then treated in a UV Crosslinker box for a minimum of 20 min (999.9 J/cm^2). In addition to the previously described PPE, two pairs of gloves were worn at all times, with the outer set changed at regular intervals.

During the period when ancient DNA samples were being extracted, no other samples were processed in the pre-PCR lab and lab access was restricted to only the operator.

2.3 Equipment, consumables & suppliers

2.3.1 Company information

Company	Company information
BioRad	Hemel Hempstead, United Kingdom
CS Cleaver Scientific Ltd	Rugby, UK
DREMEL®	Mount Prospect, IL, USA
Edge BioSystems	Gaithersburg, MD, USA
Eppendorf	Hamburg, Germany
Faraday Scientific Ltd	Sheffield, UK
GE Healthcare	Chalfont, UK
Geospiza, Inc.	Seattle, WA, USA
Gilson, Inc.	Middleton, WI, USA
Grant	Shepreth, UK
Integrated DNA Technologies	Coralville, IA, USA
Illumina Inc.	San Diego, CA, USA
J&M Analytik AG	Essingen, Germany
KDP Tools Ltd	Devon, UK
Killer Ink Ltd	Liverpool, UK
Kimberly-Clark Professional	Kent, UK
Labnet International Inc.	Woodbridge, NJ, USA
Leica Microsystems	Wetzlar, Germany
Linde Group	Munich, Germany
Orange-Sol Companies	Gilber, AZ, USA
Pal International Ltd	Lutterworth, UK
Promega Corporation	Madison, WI, USA
Qiagen	Venlo, Limburg, Netherlands
Main Man supplies Ltd	Cradley Heath, UK
Merck Millipore	Billerica, MA, USA
Microsoft	Redmond, WA, USA
Molecular Bio-Products, Inc.	San Diego, CA, USA
Sarstedt	Nümbrecht, Germany
Scenesafe Ltd	Essex, UK
Scientific Laboratory Supplies	Hessel, Yorkshire, UK
Sellotape®	Cheshire, UK
Shurtape®	Hickory, NC, USA

Sigma Aldrich	St. Louis, MO, USA
Sigma Laborzentrifugen GmbH	Osterode am Harz, Germany
Silverline Tools Ltd	Somerset, UK
Spex CertiPrep	Metuchen, NJ, USA
Swann-Morton®	Sheffield, UK
Syngene	Cambridge, UK
TCS Biosciences Ltd	Buckingham, UK
Thermo Fisher Scientific	Oyster Point, CA, USA
TWD TradeWinds, Inc.	Pleasant Prairie, WI, USA
UVP	Upland, CA, USA
VELP Scientifica	Usmate, Italy
WA Products (UK) Ltd	Essex, UK
Waldemar Knittel Glasbearbeitungs	Braunschweig, Germany
3M™	Bracknell, UK

2.3.2 Commercial kits

Kit name	Company	Catalogue number
Alexa Fluor 488 Goat Anti-mouse SFX Kit	Thermo Fisher Scientific	A31620
AmpFℓSTR® SGM Plus® PCR Amplification Kit	Thermo Fisher Scientific	4307133
AmpFℓSTR® NGM™ PCR Amplification Kit	Thermo Fisher Scientific	4415020
AmpFℓSTR® Identifiler® Direct PCR Amplification Kit	Thermo Fisher Scientific	4467831
AmpFℓSTR® Yfiler® PCR Amplification Kit	Thermo Fisher Scientific	4359513
AmpFℓSTR® MiniFiler® PCR Amplification Kit	Thermo Fisher Scientific	4373872
BigDye® Terminator v1.1 Cycle Sequencing Kit	Thermo Fisher Scientific	4337449
CC5 Internal Lane Standard 500	Promega	DG1521
ForensicGEM® Tissue	ZyGEM	FT10200
GeneScan™ 500 ROX™ dye Size standard	Thermo Fisher Scientific	401734
GeneScan™ 500 LIZ™ dye Size Standard	Thermo Fisher Scientific	4322682
Hyperladder™ 50bp	Bioline	BIO-

		33054
PERFORMA® DTR Gel Filtration Cartridges	Edge BioSytems	42453
PowerPlex® ESX 16 System	Promega	DC6711
PowerPlex® ESI 16 System	Promega	DC6771
QIAamp DNA Mini Kit	Qiagen	51304
QIAGEN Multiplex PCR Kit	Qiagen	206143
QIAquick PCR Purification Kit	Qiagen	28104
Quantifiler® Duo DNA Quantification Kit	Thermo Fisher Scientific	4387746
Quantifiler® Trio DNA Quantification Kit	Thermo Fisher Scientific	4482910
Qubit® dsDNA HS Assay Kit	Thermo Fisher Scientific	Q32854
QuickStep™2 PCR Purification Kit (Single cartridges)	Edge BioSystems	33617
SYBR® Green PCR Master Mix	Thermo Fisher Scientific	4309155

2.3.3 Reagents

Reagent	Company	Catalogue number
Agarose for electrophoresis	Sigma Aldrich	A5304
Ammonium acetate molecular biology ≥ 98%	Sigma Aldrich	A1542
AmpFlSTR® Control DNA 007	Thermo Fisher Scientific	-
AmpFlSTR® Control DNA 9947A	Thermo Fisher Scientific	-
Boric acid for electrophoresis	Sigma Aldrich	B7901
Calcium chloride Anhydrous, BioReagent	Sigma Aldrich	C5670
Chloroform:isoamyl alcohol (24:1)	Sigma Aldrich	C0549
Cytokeratin 10 Monoclonal Antibody, Mouse (VIK-10) 100 µg/ 200 µl	Thermo Fisher Scientific	395300
DAPI, dilactate ≥ 98% (HPLC)	Sigma Aldrich	D9564
5X DNA Loading Buffer Blue	Bioline	BIO-37045
DPBS, no calcium, no magnesium	Gibco	14190
DPX Mountant for histology	Sigma Aldrich	06522
Eosin Y (1% aqueous)	TCS Biosciences	HS250
Ethanol	Sigma Aldrich	E7023

Ethylenediaminetetraacetic acid BioULTRA, anhydrous, ≥ 99% (titration)	Sigma Aldrich	EDS
Genetic Analyzer 10X Buffer with EDTA	Thermo Fisher Scientific	402824
Guanidine Thiocyanate	Thermo Fisher Scientific	AM9422
Haematoxylin Harris Modified (no acetic acid)	TCS Biosciences	HS351
Hi-Di™ Formamide	Thermo Fisher Scientific	4311320
Hydrochloric Acid ACS reagent, 37%	Sigma Aldrich	258148
Liquid Nitrogen	Linde Group	-
Paraformaldehyde powder 95%	Sigma Aldrich	158127
Phenol:chloroform:isoamyl alcohol (25:24:1)	Sigma Aldrich	77617
POP-4® Polymer for 3130/3130xl Genetic Analyzers	Thermo Fisher Scientific	4363752
2-Propanol (isopropanol) BioReagent for molecular biology ≥ 99.5%	Sigma Aldrich	I9516
Proteinase K powder	Thermo Fisher Scientific	AM2544
NucBlue® Fixed Cell ReadyProbes® Reagent	Thermo Fisher Scientific	R37606
Scenesafe FAST™ Box	Scenesafe	K545
Silicon dioxide ~ 99%, 0.5-10 µm (approx. 80% between 1-5 µm)	Sigma Aldrich	S5631
Sodium chloride for molecular biology ≥ 98% (titration)	Sigma Aldrich	S3014
Sodium hydroxide ACS reagent, ≥ 97.0%, pellets	Sigma Aldrich	221465
Sodium hydroxide solution BioUltra, for molecular biology, 10 M in H_2O	Sigma Aldrich	72068
Sodium dodecyl sulfate	Sigma Aldrich	436143
Sodium hypochlorite	Fisher Scientific	10553864
SYBR® Safe DNA Gel Stain	Thermo Fisher Scientific	S33102
Sticky Stuff Remover	Orange-Sol Companies	-
Triton™ X-100	Sigma Aldrich	X100
Trizma base for molecular biology, ≥ 99.8% (T)	Sigma Aldrich	93362
Water Molecular Biology Grade Reagent	Sigma Aldrich	W4502

2.3.4 Equipment & consumables

Equipment	Company	Catalogue number
ART Pre-Sterilised Aerosol Resistant Tips	Molecular BioProducts	-
Aura™ Particulate Respirator 9320	3M™	GT500073165
Capillary Array 36 cm for 3130	Thermo Fisher Scientific	4333464
CFX96 Touch™ Real-Time PCR Detection System	BioRad	185-5196
CL-1000 Ultraviolet Crosslinker box	UVP	95-0174-02
Cobex Sheet 175 mm 4" x 6"	WA Products	B20912-100
Deburring and Engraving Set 30Pc	KDP	KDPHB265
Deckglaser crystal clear glass coverslip	Waldemar Knittel Glasbearbeitungs	GG-1212
DNA Free Swab	Scenesafe Ltd	K650
Dremel® 8200 (8200-2/45)	DREMEL®	F0138200JE
Dremel® Flexible Shaft (225) length 1070 mm	DREMEL®	26150225JA
Drum Sanding Kit 10 Piece	Silverline Tools	675272
Gilson Pipetman Classic™	Gilson	F167500/300
Hard-Shell® Low-Profile Thin-Wall 96-Well Skirted PCR Plate	BioRad	HSP-9641
Heater block	Grant	-
Sellotape Double-sided Tape 15 mm x 5 m with Dispenser	Sellotape®	-
Falcon Conical 50 ml PP Centrifuge Tubes Bagged	Scientific Laboratory Supplies	352070
Falcon Petri Dish 60 X 15 mm Standard Style Petri dishes	Scientific Laboratory Supplies	351007
Ink Precision Bug Pin Sterile Stainless Steel Tattoo Needles Round Liners	Killer Ink	50-08RL
Innova®44 incubator	Eppendorf	M1282-

		0002
Kimtech Science® A7 Disposable Lab Coats	Kimberly-Clark Professional	1137935
Kimtech Science® Purple Nitrile 24 cm gloves	Kimberly-Clark Professional	90627
KIMCARE® Medical Wipes - Interfolded / White	KIMCARE® Medical	3020
MicroAmp® Fast Optical 96-Well Reaction Plates	Thermo Fisher Scientific	4358297
MicroAmp® Fast Reaction Tube with Cap, 0.1 ml	Thermo Fisher Scientific	4358297
MicroAmp® Fast-Tube Strip tube 0.1 ml	Thermo Fisher Scientific	4358293
MicroAmp® Optical 8-Cap Strip	Thermo Fisher Scientific	4323032
MicroAmp® Optical Adhesive Film	Thermo Fisher Scientific	4311971
Microbox Precision Drills 20Pc	KDP	KDPDR074
Microcon YM-30 centrifugal filter device Ultracel-30 membrane	Merck Millipore	MRCF0R030
Microgard™ 2500 disposable oversleeves	Sigma Aldrich	Z679283-50PR
Microscope slides plain, size 25 mm x 75 mm	Sigma Aldrich	S8902
Microseal® 'C' Optical Seals	BioRad	MSC-1001
Spectrafuge™ Mini Laboratory Centrifuge	Labnet International Inc.	C1301
MiSeq Desktop Sequencer	Illumia	-
Mob Cap	Pal® International	D93
MSP400 UV-Vis Range Microspectrophotometer	J&M Analytik AG	-
NanoDrop® Spectrophotometer ND-1000	Thermo Fisher Scientific	-
Plate Septa, 96-well	Thermo Fisher Scientific	4315933
PowerPac™ Basic Power supply	BioRad	164-5050
Qubit® 2.0 Fluorometer	Thermo Fisher Scientific	Q32866
Qubit® Assay Tubes, 0.5 ml	Thermo Fisher Scientific	Q32856
Shurtape JLAR Clear to the core Tape 2'' x 72 yds	Shurtape®	-
SIGMA 1-14 Microfuge	Sigma L.GmbH	10014

Spex CertiPrep 6750 Freezermill	Spex CertiPrep	-
Sterile disposable forceps	TWD TradeWinds	DF8988P-S
Surgical scalpel Sterile Stainless Steel Blade No.10A	Swann-Morton®	0302
Syngene G:BOX Gel Doc	Syngene	-
Upright Leica DM5000 B Fluorescent Microscope System	Leica Microsystems	-
UV Sterilisation Cabinet	CS Cleaver Scientific	CSL-UVCAB
Vented Diamond Cutting Disc Kit 6-Piece 22 mm	Silverline Tools	719813
Veriti® 96-Well Fast Thermal Cycler	Thermo Fisher Scientifc	4375305
Wizard Vortex Mixer	VELP Scientifica	F202A0175
Warrior Blue Powderfree Nitrile Glove	Main Man Supplies	0117SNDPF
Whatman® Grade 1 filter paper	GE Healthcare	1001-0155
Micro tube 1.5 ml SafeSeal	Sarstedt	72.706.400
Microcentrifuge Tubes 2.0 ml, graduated, flat cap, pre-sterilised	Molecular BioProducts	3453
3130 Genetic Analyzer	Thermo Fisher Scientific	-
7500 FAST Real Time PCR System	Thermo Fisher Scientific	4351106

2.3.5 Software

Software	Company
Finch TV Version 1.4.0	Geospiza Inc.
Genemapper® *ID* version 3.2	Thermo Fisher Scientific
GeneSys.Ink software	Syngene
Genetic Analyzer Data Collection Software version 3.0	Thermo Fisher Scientific
Leica Application Suite Software Version 4.3.0	Leica Microsystems
Microsoft Excel 2010	Microsoft
ND-1000 Version 3.8.1	Thermo Fisher Scientific
ONYX Software	Faraday Scientific
SeqScape Version 5.2	Thermo Fisher Scientific
7500 Software version 2.0.6.	Thermo Fisher Scientific

2.4 Sample preparation

2.4.1 Reagent preparation

All chemicals and reagents were stored at room temperature for a maximum of 2 years unless otherwise stated. Any purchased reagents that were kept frozen were first separated into aliquots of an appropriate volume to avoid freeze-thaw cycles.

2.4.1.1 DAPI stock solution

To prepare a 1 mg/ml stock of DAPI solution, 10 mg of DAPI dilactate powder was added to 10 ml of Molecular Biology Grade (MBG) water and inverted several times to mix. 500 µl of stock solution was aliquoted into 0.5 ml tubes and stored at -20 °C until required.

2.4.1.2 DAPI working solution

A 1 in 10,000 working solution of DAPI dilactate was prepared by adding 4 µl of 1 mg/ml stock solution to 39.996 ml of sterile DPBS in a 50 ml Falcon tube. The tube was covered in aluminium foil and inverted several times to mix.

2.4.2 Reference buccal swab samples

Reference samples for operators and volunteers involved throughout this project were taken by swabbing the inside surface of both cheeks using a single DNA-free swab. The swab was then stored at -20 ºC until required for further work.

2.4.3 Buccal cell harvesting

The head of the buccal swab was placed into a 1.5 ml tube with 500 µl of MBG water and vortexed for 15 sec. The swab head was then carefully removed and discarded using forceps and the sample centrifuged for 3 min at 14,000 x *g* in a SIGMA 1-14 Microfuge. All of the supernatant was removed and another 500 µl of MBG water was added to the pellet, which was resuspended by vortexing for 15 sec. The sample was centrifuged as described previously and the supernatant removed. The cell pellet was then resuspended in 30 µl of MBG water.

2.4.3.1 Sample applied directly onto microscope slide

5 µl of the cell suspension was pipetted onto a microscope slide and allowed to air dry prior to staining. 5 µl of MBG water was used in place of the cell suspension to act as a negative control for the staining process. All prepared samples were placed into a DNA clean Petri dish ready for staining.

2.4.3.2 Sample applied onto J-LAR® tape

A microscope slide was prepared by placing one strip of double-sided tape at both ends and removing the protective covering. A strip of J-LAR tape was then secured to the slide adhesive side up, using the double-sided tape strips. 5 µl of the cell suspension was pipetted onto the J-LAR tape and allowed to air dry prior to staining. 5 µl of MBG water was used in place of the cell suspension to act as a negative control for the staining process. Prepared samples were placed into a DNA clean Petri dish ready for staining.

2.4.4 Fibre tapes

Five microscope slides were prepared by placing one strip of double-sided tape at both ends and removing the protective covering. The fibre tape to be stained was carefully peeled off its acetate and then positioned horizontally across the slides, secured in place using the double-sided tape. A disposable 10A scalpel was then used to separate the slides, which were each placed into a DNA clean Petri dish ready for staining with the appropriate dye.

2.4.5 Bone samples

2.4.5.1 Bone

The Vindolanda femur bone sample (see Chapter 5) was prepared by removing a section of approximately 3 cm x 2 cm. This was accomplished by first using a junior hacksaw to make two parallel cuts perpendicular to the bone surface and then using a Dremel® 8200 series hand tool with a Dremel® Flexible shaft attached, fitted with a Silverline Vented Diamond Cutting Disc to cut along the length as shown in Figure 2.1. The section was then removed using sterile forceps and placed in a container whilst the area was cleaned down. Approximately 1-2 mm was removed from every bone surface using the Dremel® hand tool with a Silverline sanding band attachment and the area was then cleaned down again. Once completed the bone section was cleaned thoroughly using two DNA-free swabs, the first with a 5% (w/v) sodium hypochlorite solution and the second with MBG water. The bone section was cut into pieces of less than 1 cm x 1cm, using the Dremel® hand tool with a vented diamond cutting disc attachment. Bone pieces were then soaked in a 5% (w/v) sodium hypochlorite solution and rinsed with MBG water. The sections were allowed to air dry for approximately 15 min and were then placed into a UV Crosslinker box at an exposure output of 999.9 J/cm^2. Samples were treated with UV light for 20 min and then turned over 180° and treated for an additional 20 min. The cleaned sections were then placed into a DNA clean 50 ml Falcon tube and stored at room temperature until required. This cleaning process is an amalgamation of a number of published protocols (Rohland and Hofreiter, 2007, Kemp et al., 2007, Bouwman and Brown, 2005).

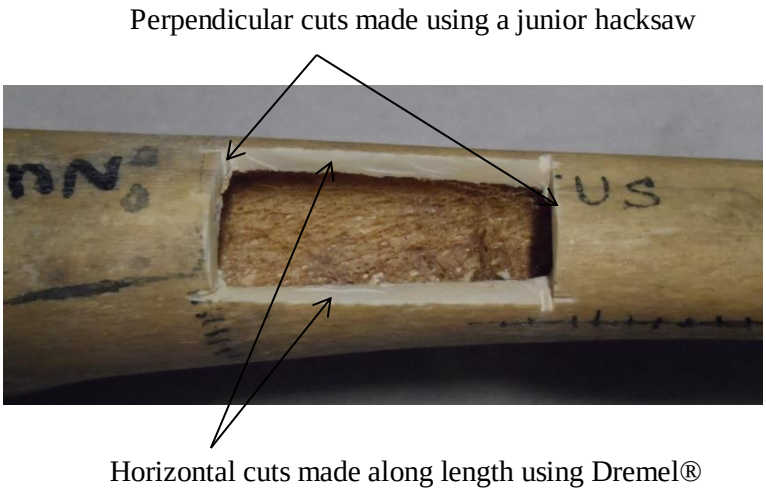


Figure 2.1 Image to illustrate removal of section from an intact bone sample.

2.4.5.2 Teeth

The entire outside surface of a tooth was cleaned using a DNA-free swab soaked in a 5% (w/v) sodium hypochlorite solution and then with a second swab using MBG water. The tooth was then soaked in a 5% (w/v) sodium hypochlorite solution for 10 minutes, rinsed using MBG water and then allowed to air dry. The tooth was then placed into a UV Crosslinker box set at an exposure output of 999.9 J/cm^2 for 20 min on one surface and then turned over 180º for an additional 20 min. The cleaned tooth was then placed into a DNA clean 50 ml Falcon tube and stored at room temperature until required.

2.4.5.3 Removal of dentin and cementum from teeth

In order to preserve the outside casing so that the sample could be returned for display in the museum, some samples were obtained by removing the dentin and cementum from the inside surface of the tooth. In order to access the dentin/cementum the top section was first removed by slicing the crown away using a Dremel® 8200 series hand tool with a Dremel® Flexible shaft attached, along with a Silverline Vented Diamond Cutting Disc attachment. The tooth root section was placed on to a DNA clean sheet of acetate and a variety of Dremel® accessories were then utilised to remove as much of the inside surface material as possible from the tooth without damaging the surface (Figure 2.2). Both drill and engraving bits were used with the size and shape being determined by the size and shape of the tooth/roots. The powder generated through drilling was collected into a DNA clean 50 ml Falcon tube using a sterile disposable 10A scalpel blade and stored at room temperature until required. This cleaning process is an amalgamation of a number of published protocols (Rohland and Hofreiter, 2007, Kemp et al., 2007, Bouwman and Brown, 2005).

Figure 2.2 Example of tooth casing with all inner material removed. Use of this method allowed for the preservation of the casing for return to the museum so that it could be put back on display.

2.4.5.4 Freezermill 6750

Either a bone or tooth sample was placed into a grinding vial and sealed with a magnetic rod in the centre. The grinding vial was then placed into liquid nitrogen inside a Spex CertiPrep 6750 Freezermill, which was used to grind the entire sample to a fine powder using a program of a 15 min pre-cool, followed by three cycles of 2 mins of grinding and 2 mins of cooling. The sample powder was then transferred from the grinding vial to a DNA clean 50 ml Falcon tube and stored at room temperature until required for further processing.

2.5 Staining of cells

2.5.1 Histological staining

The sample of cells was removed from its Petri dish and placed onto a staining rack. Enough Haematoxylin (Harris modified) was applied to cover the sample and allowed to incubate for 2 min at room temperature. Sterile DPBS was used to rinse the sample three times and Eosin Y (1% v/v) was then used to cover the sample and left in place for 30 sec. Sterile DPBS was again used to rinse the sample. Samples were stained immediately prior to imaging with a high-powered light microscope using x 100 and x 400 magnification.

2.5.2 Fluorescent staining

All fibre/cell samples stained with a fluorescent dye were imaged using an Automated Upright Leica DM5000 B Fluorescent Microscope System with Leica Application Suite Software Version 4.3.0.

2.5.2.1 NucBlue® Fixed Cell ReadyProbes® Reagent

NucBlue® Fixed Cell ReadyProbes® Reagent was applied directly from the bottle until the sample was covered. The sample Petri dish was covered with aluminium foil for a 30 min incubation at room temperature. Samples were then removed from the Petri dish and rinsed three times with sterile DPBS followed by MBG water. Samples were then placed back into the Petri dish, which was again covered in aluminium foil. Staining was carried out immediately prior to microscope imaging.

2.5.2.2 DAPI

A P1000 pipette was used to cover the sample with approximately 300 µl of a working solution of DAPI dilactate. The lid was replaced and the Petri dish covered with aluminium foil and incubated at room temperature for 1 hour. Samples were removed from the Petri dish and rinsed three times with sterile DPBS followed by MBG water. Samples were then placed back into the Petri dish, a coverslip applied using DPX if required, and then wrapped in aluminium foil until use. Staining was carried out immediately prior to microscope imaging.

2.6 Single-cell work

Double-sided tape was placed in two strips at either end of a pre-cleaned microscope slide and the protective covering removed. Strips of J-LAR tape were then secured to the slide adhesive side up, using the double-sided tape strips. A disposable 10A scalpel was used to cut squares of J-LAR tape/ Scenesafe FAST™ minitapes approximately 1 mm x 1 mm which were secured to the slide, again adhesive side up, by placing onto the J-LAR strips as shown in Figure 2.3.

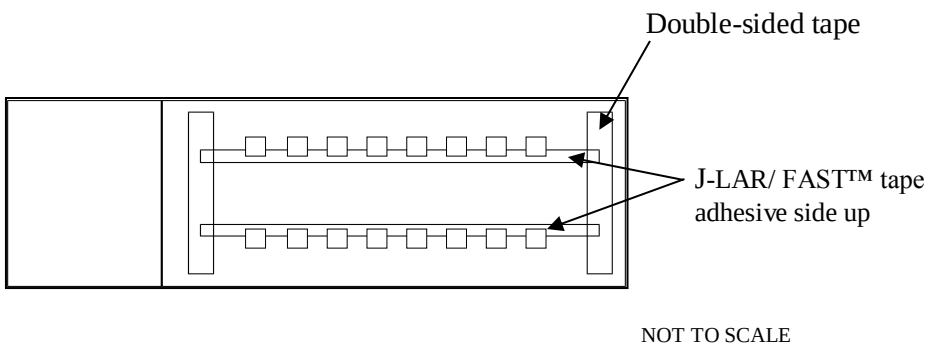


Figure 2.3 Preparation of J-LAR/ FAST™ minitapes to be used for the transfer of single-cells. This set up allowed for easy access to the adhesive squares whilst working, preventing the occurrence of contamination.

The required number of 0.1 ml individual capped MicroAmp® Fast Reaction Tubes were prepared for cell transfer by placing 1 µl of sticky stuff adhesive remover (SSR) mixed with MBG water in the bottom centre of the tube. Both the prepared tubes and microscope/tape slides were then treated with UV light for 10 min in a CL-1000 Ultraviolet Crosslinker box at an exposure output of 999.9 J/cm^2.

The single-cell of interest was carefully removed from its surface using a micro-needle with a tip diameter of 0.25 mm, and transferred directly onto the adhesive surface of an upturned square of J-LAR tape. The tape section was immediately viewed under the fluorescent microscope to confirm transfer and an image captured if required. A pair of DNA clean forceps was used to then remove the tape section from the slide and transfer it directly into the 1 µl of MBG water in a 0.1 ml capped tube.

2.7 DNA extraction

2.7.1 Reagent preparation

2.7.1.1 0.45/0.5 M EDTA

13.15/14.61 g of EDTA powder was added to 80 ml of MBG water and inverted several times to mix. Sodium hydroxide pellets were added to dissolve the solid EDTA and adjust the pH to 8.0, MBG water was then added to a final volume of 100 ml.

2.7.1.2 Proteinase K reconstitution buffer

To prepare a 50 mM Tris-HCl pH 8.0, 10 mM $CaCl_2$ reconstitution buffer, 1.21 g of Trizma base was added to 150 ml of MBG water and adjusted to pH 8.0 using 0.1 M HCl. 0.22 g of $CaCl_2$ was then added and the solution was inverted several times to mix before topping up to a total volume of 200 ml with MBG water.

2.7.1.3 Proteinase K 10 mg/ml

0.1 g of Proteinase K (Pro K) powder was added to 10 ml of Pro K reconstitution buffer and inverted several times to mix.

2.7.1.4 8M HCl

To prepare a 8M HCl solution, 0.68 ml 37% HCl was placed in a DNA clean 1.5 ml tube and made up to a final volume of 1 ml using MBG water.

2.7.1.5 5 M Ammonium acetate

38.54 g of ammonium acetate was added to 80 ml of MBG water and inverted several times to mix. Once dissolved, additional MBG water was added to a final volume of 100 ml to provide a 5 M solution.

2.7.1.6 Extraction solution

To prepare 10 ml of extraction solution, 0.25 ml of Pro K (10 mg/ml) was added to 9.75 ml of 0.45 M EDTA and inverted to mix (Rohland and Hofreiter, 2007).

2.7.1.7 Binding buffer

To prepare 400 ml of binding buffer (5 M GuSCN, 25 mM NaCl, 50 mM Tris) 0.58 g NaCl, 2.42 g Trizma and 236.32 g GuSCN were added to 300 ml of MBG water and inverted to mix. The flask was placed in warm water, with periodic mixing, until fully dissolved and then made up to a final volume of 400 ml using MBG water (Rohland and Hofreiter, 2007).

2.7.1.8 Silica suspension

4.8 g of silicon dioxide was suspended in 40 ml of MBG water and allowed to settle for 1 hour in a DNA clean 50 ml Falcon. 39 ml of the suspension supernatant was then transferred to a fresh tube and the remaining settled silica discarded. The suspension was allowed to sediment for a further 4 hours before 35 ml of the supernatant was removed and discarded. 48 µl of 8M HCl was added to the pellet, which was then vortexed to mix (Rohland and Hofreiter, 2007).

2.7.1.9 Washing buffer

100 ml of washing buffer (125 mM NaCl, 10 mM Tris, 1 mM EDTA, 50% (v/v) EtOH) was prepared by adding 0.73 g NaCl, 0.12 g Trizma and 222 µl EDTA [0.45 M] to 50 ml MBG water, which was then topped up with EtOH to a final volume of 100 ml and inverted to mix until fully dissolved (Rohland and Hofreiter, 2007).

2.7.1.10 Extraction Buffer

50 ml of extraction buffer (0.5 M EDTA pH 8.0, 0.5% (w/v) SDS, 0.1 mg/ml Pro K) was prepared by adding 0.25 g SDS and 0.5 ml Pro K (10 mg/ml) to 35 ml of EDTA [0.5 M] and inverting several times to mix. The solution was heated gently until fully dissolved and then MBG water was added to a final volume of 50 ml (Bouwman and Brown, 2005).

2.7.1.11 TE buffer

To prepare 100 ml TE buffer (10 mM Tris, 1 mM EDTA, pH 8.0) 0.12 g Trizma and 222 µl EDTA [0.45 M] was dissolved in 100 ml of MBG water.

2.7.2 Reference samples

2.7.2.1 Buccal swab using QIAamp® DNA mini kit

Following the manufacturers instructions (Qiagen, 2015b), either the sample swab head, or a DNA-free swab for use as an extraction negative control, was removed from the shaft and placed into a 1.5 ml plastic tube containing 400 µl of PBS and 400 µl of AL buffer. 20 µl of Pro K (10 mg/ml) was then added and the sample was vortexed for 15 sec to mix. The sample was incubated for 10 min 56 °C in a dry heater block. The tube was very briefly vortexed and the swab head removed and discarded. 400 µl of absolute ethanol was added to the tube, which was again vortexed for 15 sec. 675 µl of the sample was transferred to the centre of a QIAamp Mini spin column and centrifuged in a SIGMA 1-14 Microfuge at 4722 x *g* for 1 min. The collection tube was then emptied and the step repeated using the remainder of the sample mixture. The collection tube was replaced and 500 µl AW1 buffer was added to the centre of the silica column. The column was centrifuged for 1 min at 4722 x *g* and the collection tube containing the filtrate was discarded and replaced. 500 µl AW2 buffer was then added to the centre of the spin column which was centrifuged at 14462 x *g* for 3 min. The filtrate was discarded and the spin column centrifuged for an additional 1 min to ensure that any remaining AW2 buffer was removed and the column was then transferred into a 1.5 ml tube. 60 µl of AE buffer was added to the centre of the silica membrane and the tube was then left to incubate with the cap closed for 3 min. The sample was centrifuged for 1 min at 4722 x *g* and the column was discarded; the eluate was retained and stored at -20 °C until required for further use.

2.7.3 Single-cells

2.7.3.1 ForensicGEM® Tissue

A mastermix of ForensicGEM® Tissue extraction reagents was prepared in a 1.5 ml tube for the number of samples to be processed, containing 0.6 µl 10X Gold Buffer, 0.3 µl prepGEM enyme and 4.1 µl of MBG water per sample. For ltDNA samples (<0.1ng/µl) the volumes were as follows: 0.5 µl 10X Gold Buffer, 0.25 µl prepGEM enyme and 3.25 µl of MBG water per sample. The mastermix was vortexed for 15 sec and 5 µl (4 µl for the low template DNA assay) was dispensed into each sample tube, pipetting up and down several times to mix. An equivalent amount of mastermix was added to an unused sample tube for use as an extraction negative control. All reagents and samples were kept on ice throughout the preparation period. Samples were incubated for 10 min at 75 °C, followed by 5 min at 95 °C on a Veriti® 96-Well Fast Thermal Cycler to deactivate the prepGEM enzyme. Samples were placed on ice and immediately prepared for PCR as described in Section 2.9.

2.7.4 Bone samples

2.7.4.1 Phenol chloroform

Protocol based on Kemp et al. (Kemp et al., 2007). 2 ml of 0.5 M EDTA (pH 8.0) was added to 0.5 g of powdered bone/tooth sample in a 50 ml Falcon tube and inverted several times to mix. EDTA only, was transferred to an additional 50 ml Falcon tube to act as a negative extraction control throughout processing. Samples were then placed into sealed bags and incubated in an Innova®44 incubator at 20 °C for 24-36 hours, rotating at 100 rpm to in order to agitate gently. Throughout the incubation, samples were periodically checked and inverted to remove any sediment from the bottom of the tube. 300 µl Pro K (10 mg/ml) was then added and samples were inverted to mix and placed back into sealed bags in the incubator at the increased temperature of 65 °C. Samples were removed after a four-hour period and then centrifuged at 5000 x *g* for 2 minutes in a Sigma 3-18KS Centrifuge. The supernatant was transferred to a fresh 50 ml falcon tube and an equal volume of phenol:chloroform:isoamyl alcohol (25:24:1) was added and inverted several times to mix. Samples were centrifuged at 1665 x *g* for 5 min and the aqueous phase was transferred to a fresh tube. An equal volume of phenol:chloroform:isoamyl alcohol (25:24:1) was added to each sample and the previous step repeated. An equal volume of chloroform:isoamyl alcohol (24:1) was then added to each of the samples, which were vortexted to mix and then centrifuged again for 3 min at 1665 x *g*. The aqueous phase was again transferred to a fresh tube and a one half volume of 5 M ammonium acetate was added to each sample and vortexed briefly to mix. An equal volume of room temperature isopropanol was then added to each sample, vortexed and left to incubate at room temperature for ten hours.

The following day the DNA was pelleted by centrifuging the samples at 1665 x *g* for 30 min and the supernatant was then removed and discarded. The sample pellet was allowed to air dry at room temperature for approximately 15 min prior to resuspension in 50 µl of MBG water. The extracted DNA was stored at -20 °C until required for further work.

2.7.4.2 Silica column

Protocol based on Bouwman and Brown (Bouwman and Brown, 2005) using reagents from the PCR Purification Kit (Qiagen). 1.5 ml of extraction buffer was added to 0.5 g of powdered bone/tooth sample in a 50 ml Falcon tube and inverted several times to mix. An equivalent amount of extraction buffer was added to an additional Falcon tube for use as an extraction negative control throughout processing. Samples were placed in sealed bags and put into an Innova®44 incubator at 56 °C for 24 – 36 hours, rotating at 100 rpm to agitate gently. Throughout the incubation, samples were periodically checked and inverted to remove any sediment from the bottom of the tube. Samples were centrifuged in a Sigma 3-

18KS Centrifuge at 693 x *g* for 5 min and the supernatant transferred into a fresh Falcon tube containing 3.75 ml PB buffer (Qiagen) and the samples were then vortexed to mix. 750 µl of sample/PB buffer mix was added to a Qiaquick column, which was then centrifuged in a SIGMA 1-14 Microfuge for 1 min at 14,462 x *g* and the filtrate discarded. This step was repeated until the entire sample had passed through the column and the collection tube was replaced. 750 µl of buffer PE (Qiagen) was then added to the sample column, centrifuged for 1 min at 14,462 x *g* and the collection tube replaced. The column was centrifuged for an additional 2 min to ensure that the PE buffer was completely removed from the sample. The column section was transferred to a 1.5 ml tube and 50 µl of buffer EB (Qiagen) added to the centre of the silica membrane. Samples were left to incubate for 10 min at room temperature and were then centrifuged at 14,462 x *g* for 1 min. The eluate was retained and stored at -20 °C until required for further analysis.

2.7.4.3 Silica suspension

Protocol based on Rohland and Hofreiter (Rohland and Hofreiter, 2007). 10 ml of extraction solution was added to 0.5 g of powered bone/tooth sample in a 50 ml Falcon tube and inverted several times to mix. 10 ml of extraction solution was also added to an additional 50 ml Falcon tube to act as a negative control throughout the extraction process. Samples were placed inside sealed bags and placed in an Innova®44 incubator at room temperature for 24 – 36 hours, rotating at 100 rpm to agitate gently. Throughout the incubation, samples were periodically inverted to remove any sediment from the bottom of the tube. The temperature of the incubator was increased to 56 °C for one hour prior to the removal of the samples which were then centrifuged for 2 min at 5000 x *g* in a Sigma 3-18KS Centrifuge. The supernatant was transferred to a fresh 50 ml Falcon tube containing 40 ml binding buffer and inverted several times to mix. The silica suspension was briefly vortexed to resuspend and 100 µl was added to the sample/binding buffer mix. The sample was inverted to mix and then the pH was adjusted using 8M HCl to pH 4.0. Samples were then sealed into bags and placed back into the incubator for an additional 3 hours at room temperature with a rotation of 100 rpm. Once the incubation period had completed, samples were centrifuged for 2 min at 5000 x *g* and the supernatant removed but retained in a fresh tube until after quantification. 1 ml binding buffer was added to the sample and the silica pellet resuspended by pipetting up and down several times. The silica/binding buffer mix was then transferred to a 2 ml tube and centrifuged at 16,000 x *g* for 30 sec in a SIGMA 1-14 Microfuge, the supernatant was removed and 1 ml of washing buffer added to the sample. The silica pellet was resuspended and then the sample was centrifuged at 16,000 x *g* for 30 sec. The supernatant was again removed and the wash step repeated a second time. The supernatant was removed as completely as possible and the pellet allowed to air dry with the lid open for

10 min. 100 µl TE buffer was added to resuspend the pellet, which was then left to incubate for 10 min. The sample was centrifuged for 3 min at 16,000 x *g* and the supernatant transferred to a fresh tube, being careful not to disturb the silica pellet. An additional 100 µl TE buffer was added and the step repeated once more to give a total extract volume of 200 µl which was stored at -20 °C until required.

2.7.4.4 Microcon® Centrifugal filter device

A Microcon YM-30 centrifugal filter device with an Ultracel-30 membrane was utilised to combine and concentrate multiple DNA extracts from the same sample into approximately 60-80 µl. This was done by adding up to 0.5 ml of DNA extract to the membrane in the sample reservoir and centrifuging the vial for 7 min at 10,000 x *g*. The sample reservoir was then reversed and placed into a fresh vial and centrifuged for 3 min at 1000 x *g*. The sample reservoir was discarded and the retenate retained. Samples were then stored at -20 °C until required.

2.8 Quantification

2.8.1 Reagent preparation

2.8.1.1 Primer resuspension buffer

To prepare 100 ml of primer resuspension buffer (10 mM Tris, 0.1 mM EDTA pH 8.0) 0.12 g Trizma and 22.2 µl of EDTA [0.45 M] was added to 80 ml of MBG water and inverted several times to dissolve. MBG water was then added to a final volume of 100 ml

2.8.1.2 Primer preparation

All lyophilised primers were resuspended to a 100 µM or 10 µM stock solution by adding the required volume primer suspension buffer directly to the primer tube. This was then vortexed for 15 sec to completely resuspend the pellet. Primer stocks were stored at either -20 °C or -80 °C.

10 µl of 100 µM primer stock was added to 90 µl MBG water in a 1.5 ml tube to obtain a working solution at 10 µM. Diluted primers were vortexed for 15 sec to mix thoroughly.

2.8.2 NanoDrop® Spectrophotometer ND-1000

The NanoDrop® Spectrophotometer ND-1000 with software ND-1000 Version 3.8.1 was used for the quantification of DNA in reference samples, which had an expected concentration exceeding 5 ng/µl. The sampling arm and lower measurement pedestal were both cleaned using water and then 1.5 µl of water was placed onto the lower pedestal to initialise the instrument. The pedestal was then wiped and 1.5 µl of AE buffer (Qiagen) was

used as a blank sample in order for the level of background absorbance to be recorded. After cleaning, 1.5 µl of DNA extract was transferred to the lower pedestal and the sample absorbance measured and recorded. Quantification was carried out in triplicate and an average of the result taken for use in dilution calculations prior to PCR.

2.8.3 Quantifiler® Duo and Trio DNA Quantification Kits

Nuclear DNA in low template or ancient DNA samples was quantified using either the Quantifiler® Duo (Quant Duo) or Trio (Quant Trio) kits. Both kits were set-up in MicroAmp® Fast Optical 96-Well Reaction plate and processed on a 7500 FAST Real Time PCR System with 7500 Software version 2.0.6.

2.8.3.1 Template DNA preparation

A serial dilution was prepared using (Quant Duo) DNA Standard (male genomic DNA, 200 ng/µl) and (Quant Duo) Dilution Buffer as shown in Table 2-1 to generate a standard curve for absolute quantification of the unknown samples. In a similar manner, a serial dilution for Quant Trio was prepared using Quantifiler® THP DNA Standard (male genomic DNA, 100 ng/µl) and Quantifiler® THP Dilution Buffer as shown in Table 2-2 In a slight modification of the manufacturer's instructions, six standards were prepared as opposed to the recommended five, in order to try and capture any samples with a DNA concentration of less than 0.005 ng/µl. A fresh set of standards was prepared prior to each run.

Table 2-1 Preparation of DNA standards for the Quantifiler® DUO DNA Quantification Kit

Standard	Final concentration (ng/µl)	Volume of previous sample (µl)	Volume of dilution buffer (µl)	Dilution factor
1	50.00	*10	30	4X
2	16.70	10	20	3X
3	5.60	10	20	3X
4	1.80	10	20	3X
5	0.60	10	20	3X
6	0.21	10	20	3X
7	0.07	10	20	3X
8	0.02	10	20	3X

*****10 µl of Quantifiler® Duo DNA Standard stock

Table 2-2 Preparation of DNA standards for the Quantifiler® Trio DNA Quantification Kit

Standard	Final concentration (ng/µl)	Volume of previous sample (µl)	Volume of dilution buffer (µl)	Dilution factor
1	50.000	*10	10	2X
2	5.0000	10	90	10X
3	0.5000	10	90	10X
4	0.0500	10	90	10X
5	0.0050	10	90	10X
6	0.0005	10	90	10X

*****10 µl of Quantifiler® THP DNA Standard stock

All DNA extracts to be quantified were thoroughly mixed by vortex for 15 sec and then 2 µl of sample was transferred to the reaction plate in either duplicate or triplicate. 2 µl of MBG water was added in duplicate to the reaction plate to act as the no template control (NTC) for the real-time PCR.

2.8.3.2 Reaction set-up

A mastermix of the relevant PCR mix and primer mix was prepared for the number of samples to be analysed as shown in Table 2-3. 23 µl (Quant Duo) or 18 µl (Quant Trio) of mastermix was then dispensed into each sample well to give a total reaction volume of 25 µl or 20 µl respectively, and the plate was sealed with a MicroAmp® Optical Adhesive Film.

Table 2-3 Quantifiler® Duo/Trio DNA Quantification Kit reaction set-up

Kit	Reagent	Volume per sample (µl)
Quant Duo	Quantifiler® Duo Primer Mix	10.5
	Quantifiler® Duo Reaction Mix	12.5
	Total	23.0
Quant Trio	Quantifiler® Trio Primer Mix	8.0
	Quantifiler® Trio Reaction Mix	10.0
	Total	18.0

2.8.3.3 PCR run conditions

The real-time PCR was programmed to run for 40-cycles as per the manufacturer's instructions for each kit as shown in Figure 2.4 and Figure 2.5.

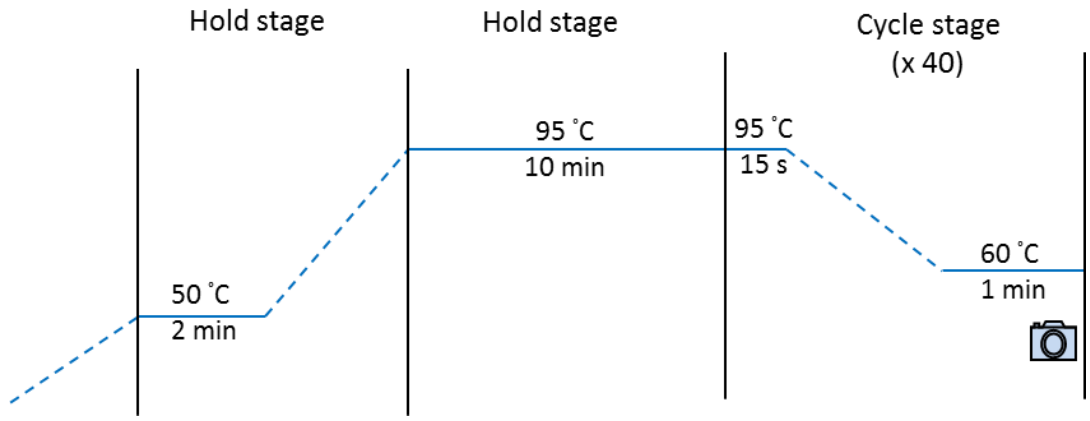


Figure 2.4 Quantifiler® Duo DNA Quantification Kit amplification conditions. An initial incubation was carried out at 50 ºC for 2 min followed by a second incubation for 10 min at 95 ºC. The DNA then underwent 40-cycles consisting of a DNA denaturation step for 15 sec at 95 ºC followed by an anneal/extend step for 1 min at 60 °C. At the end of cycle an image was captured in order monitor the level of fluorescence in real time.

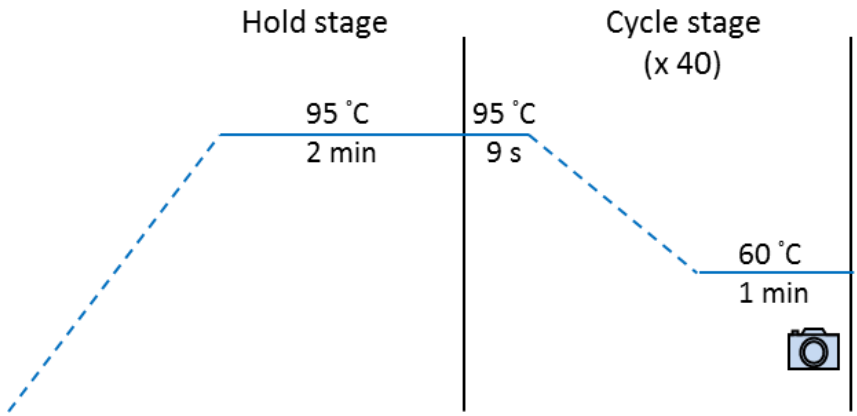


Figure 2.5 Quantifiler® Trio DNA Quantification Kit amplification conditions. An initial incubation was carried out for 2 min at 95 ºC. The DNA then underwent 40-cycles consisting of a DNA denaturation step for 9 sec at 95 ºC followed by an anneal/extend step for 1 min at 60 ºC. At the end of cycle an image was captured in order monitor the level of fluorescence in real time.

2.8.3.4 Analysis

Data analysis for both kits was carried out using the integrated 7500 Software v2.0.6 and Microsoft Excel 2010. Results were accepted if the standard curve had a R^2 value >0.98.

2.8.4 In-house mtDNA real-time PCR assay

An in-house SYBR green assay was designed as described in Section 5.4 to quantify the amount of mitochondrial DNA present in sample DNA extracts.

2.8.4.1 Template DNA preparation

A synthetic DNA fragment (Integrated DNA Technologies, Coralville, IA, USA) was used in lieu of extracted DNA as described in Section 5.4 to prepare samples for the standard curve, which were set-up using the volumes outlined in Table 2-4 immediately prior to each run. A 20 µl aliquot of dsDNA fragment stock was taken from -80 °C storage, allowed to defrost on ice and then mixed by vortex for 15 sec; this tube formed standard 1. 990 µl of MBG water was added to each of four 1.5 ml tubes and labelled accordingly. 10 µl of standard 1 was transferred into standard 2, which was then vortexed for 15 sec to mix thoroughly and this step was repeated for subsequent standards. 2 µl of each standard was transferred into a Hard-Shell® Low-Profile Thin-Wall 96-Well Skirted PCR Plate in duplicate.

All DNA extracts to be quantified were thoroughly mixed by vortex for 15 sec and 2 µl of each sample was then transferred to the reaction plate in either duplicate or triplicate. 2µl of MBG water was added in duplicate to the reaction plate to act as the NTC for the real-time PCR.

Table 2-4 Preparation of DNA standards for in-house real time mtDNA assay

Standard	Final copy number/µl	Volume of previous sample (µl)	Volume of MBG water (µl)	Dilution factor
1	$5*10^9$	#20	-	1X
2	$5*10^7$	10	990	100X
3	$5*10^5$	10	990	100X
4	$5*10^3$	10	990	100X

#20 µl of synthetic dsDNA fragment stock [$5*10^9$ copies/µl]

2.8.4.2 Reaction set- up

SYBR® Green PCR Master Mix reagent kit along with the primers (Integrated DNA Technologies, Coralville, IA, USA) listed in Table 2-5 were used to set-up the reactions in a total volume of 10 µl as shown in Table 2-6. The reaction plate was sealed using Microseal® 'C' Optical Seals and samples were amplified using a CFX96 Touch™ Real-Time PCR Detection System with the program shown in Figure 2.6.

Table 2-5 Primer sequences used for in-house real time PCR mtDNA assay

Primer	Sequence 5' – 3'
F9638	CACCTGAGCTCACCATAGTCT
R9772	GAAATGGTGAAGGGAGACTCG

Table 2-6 Reaction set-up for in-house real time PCR mtDNA assay

Reagent	[Initial]	[Final]	Volume per sample (µl)
SYBR® Green PCR Master Mix	2X	1X	5
Primer F9638	10 µM	0.3 µM	0.3
Primer R9772	10 µM	0.3 µM	0.3
MGB water	-	-	2.4
		Total	10

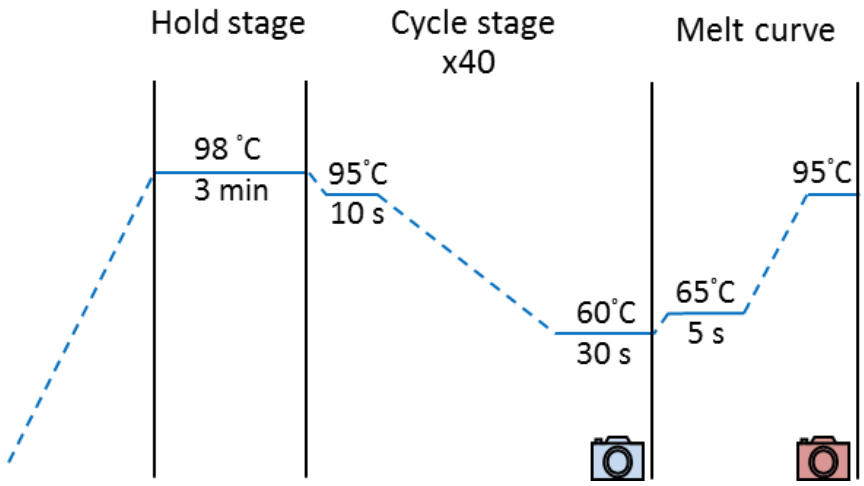


Figure 2.6 Amplification conditions for in-house real time PCR mtDNA assay PCR. An initial incubation for 3 min at 98 ºC was followed by 40-cycles, consisting of a DNA denaturation step for 10 sec at 95 ºC, followed by an anneal/extend step for 30 sec at 60 ºC. At the end of cycle an imaged was captured in order monitor the level of fluorescence in real time. The temperature was then increased in increments of 5 °C every 5 sec and an image take to produce the melt curve.

2.8.5 Qubit™ protocol

A Qubit® 2.0 Fluorometer was used to measure the quantity of double-stranded DNA in a DNA extract using a Qubit™ dsDNA HS Assay Kit.

A working solution was prepared by adding *X* µl of Qubit™ dsDNA HS reagent to *X**198 µl of Qubit™ dsDNA HS buffer, where *X* is the number of samples plus standards to be analysed.

The standards were prepared by pipetting 190 µl into two 0.5 ml Qubit® Assay Tubes and adding 10 µl of either Qubit™ dsDNA HS standard 1 (0 ng/µl) or 2 (100 ng/µl). The standards were vortexed to mix and left to incubate for 2 min in the dark, at room temperature. The dsDNA High Sensitivity option was selected on the Qubit® Fluorometer, followed by the option to calibrate and standard 1 was inserted into the sample holder for measurement. When complete, this step was repeated for standard 2 and the standard curve accepted or rejected.

198 µl of the prepared working solution was added to a 0.5 ml clear-walled tube and 2 µl of DNA extract was then added and vortexed to mix. Samples were incubated for 2 min at room temperature in a dark environment and then quantified in turn.

2.9 DNA amplification

2.9.1 Commercial kits

2.9.1.1 DNA template preparation

Each sample to be amplified was vortexed for 15 sec to mix and the required volume of DNA extract was transferred to a 0.1 ml MicroAmp® Fast-Tube Strip tube and sealed with a MicroAmp® Optical 8-Cap Strip. 007 or 9947A Control DNA was added, either directly or pre-diluted with MBG water, to give a total DNA input of 0.5 ng for a 28-cycle program or 0.05 ng for a 34-cycle program. An equivalent volume of MBG water was added to one tube for the PCR negative control.

2.9.1.2 Reaction set-up

The required volume of each PCR reagent summarised in Table 2-7 for the number of samples to be processed was transferred to a DNA clean 1.5 ml tube and vortexed for 15 sec to mix. The appropriate volume of mastermix was then dispensed into each of the sample tubes, including the PCR positive and negative samples, pipetting up and down to mix. Samples were briefly vortexed and placed in a mini-microcentrifuge for 5 sec to ensure all liquid was removed from the lid.

The sealed PCR tubes were then placed on a Veriti® 96-Well Fast Thermal Cycler and amplified using the manufacturer's instructions for PCR run conditions as described in Table 2-8.

Table 2-7 Summary of the PCR set-up, cycle number and size standards required for commercial amplification kits

Amplification kit	Application	Cycle number	Reaction mix (µl)	Primer mix (µl)	AmpliTaq Gold (µl)	DNA template/ MGB water (µl)	Reaction volume (µl)	Control profile	Size standard
AmpFℓSTR® SGM Plus® PCR Amplification Kit	Buccal samples Low template samples	28 34	5.25	2.75	0.25	5	12.5	007	500 ROX
AmpFℓSTR® SGM Plus® PCR Amplification Kit	Single-cells	34	6.3	3.3	0.3	*6	15	007	500 ROX
AmpFℓSTR® NGM™ PCR Amplification Kit	Buccal samples	28	5.0	2.5	N/A	5	12.5	007	500 LIZ
AmpFℓSTR® NGM™ PCR Amplification Kit	Low template samples Single-cells	30/34	10.0	5.0	N/A	10	25	007	500 LIZ
AmpFℓSTR® Identifiler® Direct PCR Amplification Kit	Single-cells	34	12.5	12.5	N/A	#N/A	25	9947A	500 LIZ
AmpFℓSTR® Yfiler® PCR Amplification Kit	Low template samples	34	9.2	5.0	0.8	10	25	007	500 LIZ
PowerPlex® ESX 16 System	Low template samples Single-cells	30/34	5.0	2.5	N/A	17.5	25	007	CC5
PowerPlex® ESI 16 System	Low template samples Single-cells	30/34	5.0	2.5	N/A	17.5	25	007	CC5

*PrepGEM extraction reagents added

#1µl of control DNA added for PCR positive sample

Table 2-8 Summary of PCR run conditions required for commercial kits

Amplification kit	Initial denaturation	Denature	Anneal	Extend	Final extension	Hold	Cycle number
AmpFℓSTR® SGM Plus® PCR Amplification Kit	95 °C 11 min	94 °C 1 min	59 °C 1 min	72 °C 1 min	60 °C 45 min	4 °C ∞	28/34
AmpFℓSTR® NGM™ PCR Amplification Kit	95 °C 11 min	94 °C 20 sec	59 °C 3 min		60 °C 10 min	4 °C ∞	28/30/34
AmpFℓSTR® Identifiler® Direct PCR Amplification Kit	95 °C 11 min	94 °C 20 sec	59 °C 2 min	72 °C 1 min	60 °C 25 min	4 °C ∞	28/34
AmpFℓSTR® Yfiler® PCR Amplification Kit	95 °C 11 min	94 °C 1 min	61 °C 1 min	72 °C 1 min	60 °C 80 min	4 °C ∞	28/34
PowerPlex® ESX/ESI 16 System	96 °C 2 min	94 °C 30 sec	59 °C 2 min	72 °C 90 sec	60 °C 45 min	4 °C ∞	30/34

2.9.2 Low-template DNA assay

2.9.2.1 Template DNA preparation

DNA extracts were vortexed for 15 sec to mix and 5 µl of each sample was transferred to a 0.1 ml MicroAmp® Fast-Tube Strip tube, which was sealed with a MicroAmp® Optical 8-Cap Strip. 007 Control DNA was pre-diluted to give a total DNA input of 0.05 ng and an equivalent volume of MBG water was added to one tube for the PCR negative control.

Single-cell samples were briefly vortexed and then centrifuged in a mini-microcentrifuge for 5 sec to remove any liquid from the tube cap.

2.9.2.2 Reaction set-up

Qiagen's Multiplex PCR Kit was used to set-up reactions in a total volume of 15 µl using the fluorescently tagged primers (Integrated DNA Technologies, Coralville, IA, USA) listed in Table 2-9. Primers were resuspended and diluted as described in section 2.8.1.1. The mastermix was prepared by transferring the required volume for the number of samples to be processed into a DNA clean 1.5 ml tube, as described in Table 2-10 and vortexing for 15 sec to mix. 10 µl of mastermix was then dispensed into each of the sample tubes, including the PCR positive and negative samples, pipetting up and down to mix. Samples were vortexed and placed in the mini-microcentrifuge for 5 sec to ensure all liquid was removed from the lid.

The sealed PCR tubes were then placed on a Veriti® 96-Well Fast Thermal Cycler and amplified using the run conditions shown in Table 2-11.

Table 2-9 Primer sequences used for in-house ltDNA amplification assay

Primer	Sequence 5' – 3'	Reference
D10S1248_F	NED-TTA ATG AAT TGA ACA AAT GAG TGA G	(Coble and Butler, 2005)
D10S1248_R	GCA ACT CTG GTT GTA TTG TCT TCA T	
PENTA_E_F*	6FAM-TGG GCG ACT GAG CAA GAC TC	(Butler et al., 2003)
PENTA_E_R	GGT TAT TAA TTG AGA AAA CTC CTT ACA	
Amelo_F	PET-CCC TGG GCT CTG TAA AGA ATA GTG	(Sullivan et al., 1993)
Amelo_R	ATC AGA GCT TAA ACT GGG AAG CTG	
D12S391_F	VIC-TCT CCA GAG AGA AAG AAT CAA CAG GA	(Guo et al., 2014)
D12S391_R*	CCT CCA TAT CAC TTG AGC TAA TTC C	

* Primer modified

Table 2-10 PCR set-up for in-house ltDNA amplification assay

Reagent	[Initial]	[Final]	Volume per sample (µl)
Qiagen 2X Buffer	2X	1X	7.5
Primer mix	4.5 µM (each)	0.3 µM (each)	1
Q solution	5X	0.5X	1.5
		Total volume	10

Table 2-11 PCR amplification conditions for in-house ltDNA assay

Initial denaturation	Denature	Anneal	Extend	Final extension	Hold	Cycle number
95 °C 15 min	94 °C 30 sec	59 °C 90 sec	72 °C 1 min	60 °C 30 min	4 °C ∞	34

2.9.3 Mitochondrial DNA amplification

2.9.3.1 Template DNA preparation

Samples were vortexed for 15 sec to mix and 5 µl was then transferred to a 0.1 ml MicroAmp® Fast-Tube Strip tube, which was sealed with a MicroAmp® Optical 8-Cap Strip. 5 µl of pre-diluted 007 Control DNA was added to one tube, to give a total DNA input of 0.05 ng, as a PCR positive control and an equivalent volume of MBG water was added a tube for the PCR negative control.

2.9.3.2 Reaction set-up

Individual mastermixes were prepared for each primer set by transferring the required volume of each reaction component into a DNA clean 1.5 ml tube as summarised in Table 2-12 to Table 2-17. All primers (Integrated DNA Technologies, Coralville, IA, USA) listed in Table 2-18 to Table 2-20, were resuspended and diluted as previously described in Section 2.8.1.1. The positions of the Edson primer sets listed in Table 2-18 are shown in Figure 2.7. Each mastermix was vortexed and 15 µl transferred to the appropriate tubes to give a total reaction volume of 20 µl.

Table 2-12 PCR set-up for mtDNA HVSI primer sets A-C and HVSII sets E, H & J

Reagent	[Initial]	[Final]	Volume per sample (µl)
Buffer II	10X	1X	2
$MgCl_2$	25 mM	3 mM	2.4
Forward Primer	10 µM	1 µM	2
Reverse Primer	10 µM	1 µM	2
dNTP mix	25 mM (each)	0.2 mM (each)	0.16
BSA	10 mg/ml	1 mg/ml	2
Taq	5 U/µl	1 U	0.2
MBG water	-	-	4.24
		Total mastermix volume	15

Table 2-13 PCR set-up for mtDNA HVSI primer set D

Reagent	[Initial]	[Final]	Volume per sample (µl)
Buffer II	10X	1X	2
$MgCl_2$	25 mM	2 mM	1.6
Forward Primer	10 µM	1.5 µM	3
Reverse Primer	10 µM	1.5 µM	3
dNTP mix	25 mM (each)	0.2 mM (each)	0.16
BSA	10 mg/ml	1 mg/ml	2
Taq	5 U/µl	1 U	0.2
MBG water	-	-	3.04
		Total mastermix volume	15

Table 2-14 PCR set-up for Illumina D-loop protocol primer mix MI

Reagent	[Initial]	[Final]	Volume per sample (µl)
10X buffer II	10X	1X	2
Primer mix	0.25µM (each)	0.04	3.2
dNTPs	25mM (each)	0.4mM (each)	0.32
Taq	5U/µl	1U	0.2
$MgCl_2$	25mM	6mM	4.8
MBG water	-	-	2.48
Q solution	5X	0.5X	2
		Total mastermix volume	15

Table 2-15 PCR set-up for Illumina D-loop protocol primer mix MII

Reagent	[Initial]	[Final]	Volume per sample (µl)
10X buffer II	10X	1X	2
Primer mix	0.25µM (each)	0.05	4
dNTPs	25mM (each)	0.6mM (each)	0.48
Taq	5U/µl	1U	0.2
$MgCl_2$	25mM	6mM	4.8
MBG water			1.52
Q solution	5X	0.5X	2
		Total mastermix volume	15

Table 2-16 PCR set-up for Illumina D-loop protocol primer mix MIII

Reagent	[Initial]	[Final]	Volume per sample (µl)
10X buffer II	10X	1X	2
Primer mix	0.25µM (each)	0.05	4
dNTPs	25mM (each)	0.6mM (each)	0.48
Taq	5U/µl	1U	0.2
$MgCl_2$	25mM	7mM	5.6
MBG water			2.72
		Total mastermix volume	15

Table 2-17 PCR set-up for Illumina D-loop protocol primer mix MIV

Reagent	[Initial]	[Final]	Volume per sample (µl)
10X buffer II	10X	1X	2
Primer mix	0.25µM (each)	0.05	4
dNTPs	25mM (each)	0.4mM (each)	0.32
Taq	5U/µl	1U	0.2
$MgCl_2$	25mM	5mM	4
MBG water			2.48
Q solution	5X	0.5X	2
		Total mastermix volume	15

Table 2-18 Primer sequences for HVSI, II & III amplicons using Edson primer sets (Edson, 2004) for samples to be analysed using the 3130 Genetic Analyzer

HVSI	Primer ID	Primer Sequence 5' - 3'
Set A	F15989	CCC AAA GCT AAG ATT CTA AT
	R16158	TAC TAC AGG TGG TCA AGT AT
Set B	F16112	CAC CAT GAA TAT TGT ACG GT
	R16251	GGA GTT GCA GTT GAT GT
Set C	F16190	CCC CAT GCT TAC AAG CAA GT
	R16322	TGG CTT TAT GTA CTA TGT AC
Set D	F16268	CAC TAG GAT ACC AAC AAA CC
	R16410	GAG GAT GGT GGT CAA GGG AC
Set E	F034	GGG AGC TCT CCA TGC ATT TGG TA
	R159	AAA TAA TAG GAT GAG GCA GGA ATC
Set H	F220	TGC TTG TAG GAC ATA ATA AT
	R389	CTG GTT AGG CTG GTG TTA GG
Set J	F403	TCT TTT GGC GGT ATG CAC TTT
	R569	GGT GTA TTT GGG GTT TGG TTG

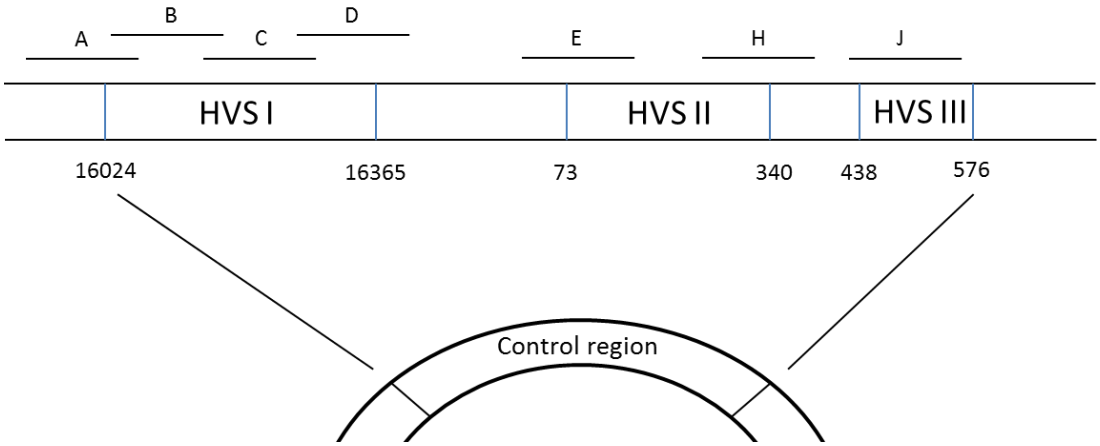


Figure 2.7 Image to show position of Edson primer sets used. This primer set has been designed to generate overlapping amplicons which are suitable for amplification of degraded DNA. The exact positions of primer sets A-E, H and J and the primer sequences are listed in Table 2-18.

Table 2-19 Primer sequences for HVSI amplicons A-D [4] for samples to be analysed on the MiSeq Desktop Sequencer

HVSI	Primer Id	Primer Sequence 5' - 3'
Set A	AF1	AATGATACGGCGACCACCGAGATCTACACCGCTGATTCATGCATTGCCCAAAGCTAAGATTCTAAT
	AF2	AATGATACGGCGACCACCGAGATCTACACAGGCGATTCATGCATTGCCCAAAGCTAAGATTCTAAT
	AF3	AATGATACGGCGACCACCGAGATCTACACGCGGTATTCATGCATTGCCCAAAGCTAAGATTCTAAT
	AR1	CAAGCAGAAGACGGCATACGAGATAATACACCTGAGTCGTGTACTACAGGTGGTCAAGTAT
	AR2	CAAGCAGAAGACGGCATACGAGATTTTTCACCTGAGTCGTGTACTACAGGTGGTCAAGTAT
Set B	BF1	AATGATACGGCGACCACCGAGATCTACACGTGACATTCATGCATATCACCATGAATATTGTACGGT
	BF2	AATGATACGGCGACCACCGAGATCTACACCCCTAATTCATGCATATCACCATGAATATTGTACGGT
	BF3	AATGATACGGCGACCACCGAGATCTACACAGGTCATTCATGCATATCACCATGAATATTGTACGGT
	BR1	CAAGCAGAAGACGGCATACGAGATCACTTACTTGAATATCCGGAGTTGCAGTTGATGT

	BR2	CAAGCAGAAGACGGCATACGAGATATGCTACTTGAATATCCG GAGTTGCAGTTGATGT
Set C	CF1	AATGATACGGCGACCACCGAGATCTACACCCTGTATTCATAC ATAGCCCCATGCTTACAAGCAAGT
	CF2	AATGATACGGCGACCACCGAGATCTACACCGATCATTCATAC ATAGCCCCATGCTTACAAGCAAGT
	CF3	AATGATACGGCGACCACCGAGATCTACACGTCCAATTCATAC ATAGCCCCATGCTTACAAGCAAGT
	CR1	CAAGCAGAAGACGGCATACGAGATGGTAAACCTGAGTCGCCT GGCTTTATGTACTATGTAC
	CR2	CAAGCAGAAGACGGCATACGAGATGATTCACCTGAGTCGCCT GGCTTTATGTACTATGTAC
Set D	DF1	AATGATACGGCGACCACCGAGATCTACACCTGGTATTCATGC ATATCACTAGGATACCAACAAACC
	DF2	AATGATACGGCGACCACCGAGATCTACACGCATCATTCATGC ATATCACTAGGATACCAACAAACC
	DF3	AATGATACGGCGACCACCGAGATCTACACGGAACATTCATGC ATATCACTAGGATACCAACAAACC
	DR1	CAAGCAGAAGACGGCATACGAGATATTCCACTTGAATATATG AGGATGGTGGTCAAGGGAC
	DR2	CAAGCAGAAGACGGCATACGAGATAAGGAACTTGAATATATG AGGATGGTGGTCAAGGGAC

Table 2-20 Primer sequences for Illumina protocol D-loop amplification

HVSI	Primer Id	Primer Sequence 5' - 3'
Set MI	F15997-A	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCACCATTAGCAC CCAAAGCT
	F15997-B	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCACCATTAGCA CCCAAAGCT
	R16236-A	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCTTTGGAGTTGCA GTTGATG
	R16236-B	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCTTTGGAGTTGC AGTTGATG
Set MII	F16159-A	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTACTTGACCACC TGTAGTAC
	F16159-B	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTACTTGACCAC CTGTAGTAC
	R16401-A	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTGATTTCACGGA GGATGGTG
	R16401-B	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTGATTTCACGGA GGATGGTG
Set MIII	F29-A	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGGTCTATCACCC TATTAACCAC
	F29-B	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGGTCTATCACC CTATTAACCAC
	R285-A	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGGGGTTTGGTGG AAATTTTTTG
	R285-B	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGGGGTTTGGTGG AAATTTTTTG
Set MIV	F172-A	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGATTATTTATCGCA CCTACGT
	F172-B	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGATTATTTATCGC ACCTACGT
	R408-A	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCTGTTAAAAGTG CATACCGCCA
	R408-B	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCTGTTAAAAGTG CATACCGCCA

The sealed PCR tubes were then placed on a Veriti® 96-Well Fast Thermal Cycler and amplified using the PCR run conditions shown in Table 2-21 with the annealing temperatures specified in Table 2-22.

This set-up was used for samples going on to be sequenced by both the 3130 Genetic Analyzer and the MiSeq Desktop Sequencer. Reaction set-up for samples to be sequenced by the MiSeq using the Edson set of primers modified with flowcell adaptors, as explained in more detail in Section 5.3.6.7.

Table 2-21 PCR run conditions for mtDNA HVS amplification 'X ºC' denotes the annealing temperature for each of the HVS primer sets as described in Table 2-22

Initial denaturation	Denature	Anneal	Extend	Final extension	Hold	Cycle no.
95 ºC 10 min	94 ºC 30 sec	X ºC 1 min	72 ºC 1 min	60 ºC 7 min	4 ºC ∞	35

Table 2-22 Target fragment size and annealing temperatures for amplifications of mtDNA primer sets HVSI A-D (both Edson and modified Edson sets), HVSII E, H & J (Edson) and Optimised Illumina D-loop protocol sets MI-MIV

Primer Set	Target size (bp)	Annealing temperature °C
HVSI Set A	169	52
HVSI Set B	139	53
HVSI Set C	132	55
HVSI Set D	142	59
HVSII Set E	125	56
HVSII Set H	169	46
HVSIII Set J	166	58
MI	239	59
MII	242	60
MIII	256	59
MIV	236	59

2.10 PCR clean-up

2.10.1 QuickStep™2 PCR Purification Kit

To prepare PCR products for clean-up using the QuickStep™2 PCR Purification Kit, MBG water was added to each sample so to a final total volume of 20 µl. The SOPE resin was vortexed briefly to mix and 4 µl added to each sample, pipetting up and down several times to mix. The samples were left to incubate at room temperature for 5 min whilst the Performa Gel Filtration Cartridges were placed in a SIGMA 1-14 and centrifuged at 850 x *g* for 3 min. The columns were then transferred to fresh 1.5 ml tubes and the SOPE/sample mixture transferred to the centre of the column gel bed. Samples were then centrifuged at 850 x *g* for 3 min, the columns discarded and the eluates retained. Samples were stored at -20 °C until required for further processing.

2.11 Sequencing reaction

2.11.1 BigDye® Terminator v1.1 Cycle Sequencing Kit

2 µl of purified PCR product from each sample was transferred into a 0.1 ml MicroAmp® Fast-Tube Strip tube and sealed with a MicroAmp® Optical 8-Cap Strip. 2 µl of MBG water was added to one tube for a sequencing negative control.

The appropriate primers (Integrated DNA Technologies, Coralville, IA, USA) and PCR reagents included in the BigDye® Terminator v1.1 Cycle Sequencing Kit were briefly mixed by vortex and a mastermix was prepared in a 1.5 ml tube for each individual primer as shown in Table 2-23. The mastermix was mixed for 15 sec by vortex and 8 µl was then dispensed into each sample tube. Samples were placed on a Veriti® 96-Well Fast Thermal Cycler and amplified for 25-cycles using the program shown in Table 2-24.

Table 2-23 Sequencing reaction set-up for BigDye® Teminator kits

Reagent	[Initial]	[Final]	Volume per sample (µl)
Ready Reaction Premix	2.5X	0.5X	2
BigDye Sequencing Buffer	5X	0.5X	1
Primer (either forward or reverse)	10 µM	1 µM	1
MBG water	-	-	4
	Total mastermix volume		8

Table 2-24 Sequencing reaction run conditions for BigDye® Terminator kits

Initial denaturation	Denature	Anneal	Extend	Hold	Cycle no.
96 °C 1 min	96 °C 10 sec	50 °C 5 sec	60 °C 75 sec	4 °C ∞	25

2.12 Sequencing reaction clean-up

2.12.1 PERFORMA® DTR Gel Filtration Cartridges

The required number of PERFORMA® DTR Gel Filtration Cartridges were placed in a SIGMA 1-14 Microfuge centrifuge at 750 x *g* for 2 min (850 x *g* for 3 min for BigDye® Teminator v 3.1 Cycle Sequencing Kit) and the columns transferred to a fresh kit tube. The entire sample was transferred to the centre of the gel bed and the columns centrifuged at 750 x *g* for 2 min (850 x *g* for 3 min for BigDye® Teminator v 3.1 Cycle Sequencing Kit). The columns were then discarded and the eluates retained. Samples were stored at 4 °C until required for further processing.

2.13 DNA profile visualisation

2.13.1 Reagent preparation

2.13.1.1 TBE buffer

To prepare 1 litre of a 5X TBE buffer stock solution 54 g of Trizma, 27.5 g boric acid and 20 ml of 0.5 M EDTA pH 8.0 was added to 900 ml of deionised water and inverted several times to dissolve. Deionised water was added to a final volume of 1 L and the solution was stored at room temperature for a maximum of 6 months.

A 1 in 5 dilution in deionised water was carried out to prepare a 1X working solution of TBE buffer. The solution was stored at room temperature for a maximum of 6 months.

2.13.1.2 3% agarose gel

To prepare a 3% (w/v) agarose gel, 1.5 g of agarose powder was added to 50 ml of 1X TBE buffer and heated in a microwave for approximately 90 sec until melted. The gel was allowed to cool at room temperature for 2 min and 5 µl of SYBR®Safe and gently swirled to mix. The gel was poured into a cassette mould holding either an 8 or 15 well comb and left

to set under aluminium foil for 20 min. Once set, the comb was removed from the gel and the cassette transferred to a gel tank, which was then covered in 1X TBE buffer.

Samples were prepared by adding 5 µl of PCR product into a 1.5 ml tube and adding 1 µl of 5X Loading Buffer Blue (Bioline), pipetting up and down to mix. The sample/loading buffer was then transferred into one of the gel sample wells; this was repeated for samples. 5 µl of Hyperladder™ 50 bp was added to the wells on the left and right of the samples. The tank was sealed and the electric current applied for 40 min at 120 V.

Products were visualised and captured using a Syngene G:BOX Gel Doc with integrated GeneSys.Ink software.

2.13.2 Applied Biosystems 3130 Genetic Analyzer

2.13.2.1 Fragment analysis

To process samples for fragment analysis on the 3130 Genetic Analyzer, a size standard mix was prepared in a 1.5 ml tube containing either 8.5 µl or 8.7 µl of Hi-Di™ Formamide and 0.5 µl or 0.3 µl of GeneScan™ 500 ROX™ or 500 LIZ™ dye Size Standard respectively. The size standard mix was briefly vortexed and 9 µl was then dispensed into the number of required wells of a 96-well plate. 1 µl of the sample PCR product was then added to each well. For Promega ESI/ESX 16 kits, 10 µl of Hi-Di™ Formamide was mixed with 1 µl of CC5 Internal Lane Standard 500 and 11 µl transferred to the reaction plate per well.

The appropriate allelic ladder for the amplification kit used in PCR was vortexed to mix and 1 µl was added to a well at a ratio of one ladder for every two runs.

A plate septum was used to seal the plate and was then placed on a Veriti® 96-Well Fast Thermal Cycler for 3 min at 95 °C, prior to being put on ice to snap cool for 5 min. The plate was then secured in the plate adaptor and placed onto the 3130 machine for processing using a 36-cm length capillary array with POP-4.

Genetic Analyzer Data Collection Software v 3.0 was used to set-up the plate sample sheet using the appropriate run module with parameters as described in Table 2-25. The sample sheet was then linked to the plate and the run initiated.

Table 2-25 3130 Genetic Analyzer run module capillary electrophoresis parameters

Run module	Oven temp (ºC)	Injection voltage (kV)	Injection time (sec)	Run voltage (kV)
HIDFragmentAnalysis36_POP4_1	60	3	5	15
HIDFragmentAnalysis36_POP4_1_LI	60	3	15	15
UltraSeq36_POP4_1	55	1.2	10	15

2.13.2.2 Sequencing analysis

To prepare for sequencing analysis of samples on the 3130 Genetic Analyzer, 8 µl of Hi-Di™ Formamide was dispensed into the required number of wells on a MicroAmp® Optical Reaction Plate. Samples were briefly vortexed to mix and 2 µl was then transferred to the reaction plate, pipetting up and down several times to mix. The reaction plate was then sealed with a plate septum, secured in the plate adaptor and placed onto the 3130 machine.

A new plate was created, named and 'sequencing analysis' selected. Sample names were inputted onto the sample sheet and the run module UltraSeq36_POP4_1, with the parameters as shown in Table 2-25, was selected. The KB basecaller mobility file KB_3130_POP4_BDTv1.1/3.1 was used and set to remove bases from the end until fewer than 4 bases out of 20 have QVs < 20. The sample sheet was then linked to the plate and the run commenced.

2.13.3 Data analysis

2.13.3.1 Fragment analysis

Files produced by the 3130 run collection software were imported into Genemapper® *ID* version 3.2 for analysis. DNA fragments were first sized relative to the appropriate size standard for the commercial kit as shown in Table 2-7, and then alleles designated by comparison with the appropriate ladder. Table 2-26 outlines the analysis settings used for all amplification kits. As no ladder was available for the in-house developed ltDNA assay, all alleles were sized by comparison to GeneScan™ 500 LIZ™ dye Size Standard and the alleles designated manually as described in Section 3.3.6

2.13.3.2 Sequencing analysis

Files created by the 3130 run collection software and KB base caller were imported into SeqScape Version 5.2 software and then the resulting ABI format chromatogram files were opened and analysed using Finch TV Version 1.4.0.

Table 2-26 GeneMapper ID v 3.2 Analysis settings

Amplification kit used		Panel set	Bin set	Peak amplitude threshold	Size standard
AmpFℓSTR® SGM Plus®	Reference Low-template	AmpFLSTR_Panels_v1	AmpFLSTR_Bins_v1	50 25	CE_F_HID_GS500
AmpFℓSTR® NGM™	Reference Low-template	AmpFLSTR_NGM_v1	NGM_bins_v1	50 25	CE_G5_HID_GS500
AmpFℓSTR® Identifiler® Direct	Reference Low-template	IdentifilerDirect_GS500_Panels_v1	IdentifilerDirect_GS500_Bins_v1	50 25	CE_G5_HID_GS500
AmpFℓSTR® Yfiler®	Reference Low-template	AmpFLSTR_YFiler_Panel_v2	YFiler_v2	50 25	CE_G5_HID_GS500
PowerPlex® ESI 16 System	Reference Low-template	Promega_ESI_16	Promega_ESI_16_bins	50 25	CC5_ILS_GS500
PowerPlex® ESX16 System	Reference Low-template	Promega_ESX_16	Promega_ESX_16_bins	50 25	CC5_ILS_GS500
ltDNA Assay	Reference Low-template	ltDNA v1.1	NA	50 25	CE_G5_HID_GS500

2.14 Fibre spectral profile visualisation

2.14.1 Microspectrophotometer

The fibre of interest was mounted on a microscope slide using DPX with a Deckglaser crystal clear glass coverslip and left for at least 1 hour to set before analysis. For fibres examined in the UV range, glycerol was used to mount the sample under quartz coverslips.

The slide was placed onto the microscope stage of a TIDAS MSP800 UV-Vis Range Microspectrophotometer, which was set with the following parameters for acquisition: scan type – Absorbance, start wavelength – 380 nm, end wavelength – 700 nm, integration time – 600 ms, no. of averages – 10, binning – 1, dark current – selected. The measurement window was aligned alongside the fibre and a background spectrum recorded. The measurement window was then positioned over the fibre of interest, which was orientated in a vertical direction and ten measurements were recorded along the fibre length, which were then averaged prior to comparison between spectra from different fibres. Data was analysed and visualised using integrated ONYX software.

3. Single-cell DNA analysis

3.1 Introduction

LtDNA analysis is carried out routinely in forensic casework, particularly on so-called 'touch DNA' samples; however the sensitivity required for such work often results in the amplification of background DNA in addition to the target DNA, thus resulting in mixed DNA profiles. Whilst under some circumstances, i.e. with clear major/minor contributors, a mixture can be resolved; often the complexity of these profiles renders them uninterpretable. The use of individual single-cells as a DNA source would minimise the potential for DNA mixtures; however this type of analysis is seldom carried out in forensic casework.

The analysis of DNA extracted from single nucleated cells was first reported by Li et al. in 1988 (Li et al., 1988) who used the then relatively new technique of PCR to amplify two genetic loci from both a single sperm cell and a diploid cell. It was not until almost a decade later that the potential for use of forensic analysis of single-cell DNA was first explored and reported by Findlay et al. (Findlay et al., 1997), who amplified at least four out of six STR markers in approximately two thirds of the two hundred and twenty six buccal cells tested. Whilst these results looked promising for application in forensic science the authors noted the requirement for a robust interpretation strategy due to both the stochastic effects that are associated with processing ltDNA extract and allele drop-in/contamination events, which were observed in 17% of their results. Allele drop-in is a sporadic contamination events thought to be a result of amplification of atmospheric DNA or potentially cell-free DNA that is adhered to the surface of the cell.

In order to interpret any DNA profile, a peak amplitude threshold (PAT) describing the minimum peak heights in relative fluorescent units (RFU) will be used, which for the identification of allelic peaks against the background noise signal (Budowle et al., 2009b). Minimum peak heights of 50 RFU and 150 RFU for heterozygous and homozygous peaks respectively have been demonstrated as suitable PAT thresholds for standard DNA profiles generated using 28-cycle PCR (Gill et al., 1997b). For ltDNA samples a nominal minimum threshold of 25 RFU is frequently utilised (Cowen et al., 2011) in order to obtain the maximum amount of information from each sample.

Once all peaks of interest have been identified, characteristics such as stutter peaks, heterozygous balance between peaks within a locus, balance across all the loci, allelic drop-out and allelic drop-in must all be considered (Gill et al., 1997b).

Stutter peaks are a common and unavoidable artifact that results in the appearance of additional peaks exactly one or more repeat units less than the true allele peak or, less frequently, in the plus-one repeat unit position (Walsh et al., 1996). It is thought stutter

peaks arise during DNA synthesis as a consequence of the polymerase ‘slipping’ off the template strand, thus causing the two DNA strands to separate slightly, which due to misalignment upon reannealing results in the newly synthesised strand consisting of one repeat unit shorter than the original template (Levinson and Gutman, 1987, Strand et al., 1993). To account for those stutter peaks that are one repeat unit longer than the template DNA strand, it is thought that the extending strand may loop out during synthesis, which will then act as a template in the next round of amplification (Walsh et al., 1996). A number of factors have been shown to effect the occurrence of stutter peaks including repeat unit length (Walsh et al., 1996, Klintschar and Wiegand, 2003), the longest stretch of uninterrupted tandem repeats within the allele (Bright et al., 2014, Klintschar and Wiegand, 2003) and the primary DNA sequence (Aponte, 2015, Walsh et al., 1996, Schlotterer and Tautz, 1992). The degree to which the stutter peak is present will be determined predominantly by the cycle number and frequency with which the slippage event occurs. As demonstrated recently by Aponte et al. (Aponte, 2015), the frequency and range of stutter increased by up to approximately 3% and 5% respectively when using MPS as compared with CE, most likely due to the additional amplification during sequencing.

Stutter peaks must be considered with care as they can be interpreted as true alleles, therefore indicating a mixed profile, or incorrectly discounted from the profile thereby implying results from a single DNA source, when in fact it is mixed (Walsh et al., 1996, Clayton et al., 1998). Stutter peaks are identified through calculation of the stutter ratio (SR) in peak height/area (RFU) between the true allele peak and the suspected stutter peak using the following equation according to Gill et al. (Gill et al., 2000a): SR = stutter peak φ / allele φ (‘n’) + φ (‘n+1’), where ‘φ’ is peak area in RFU and ‘n’ is the allele peak minus the additional adenine base. The proportion of stutter (SP) observed at each locus is calculated using the equation SP = number of loci which stutter/total number of loci. Using the original SGM assay it was shown that stutter peaks were under 15% (0.15) of the true allele height (Gill et al., 1997b), and this threshold is often still used for the designation of stutter peaks if more precise, locus specific figures are not available through extensive validation work (Gibb et al., 2009). For ltDNA samples however it has been demonstrated that the stutter ratio value is increased due to stochastic variation of PCR; Whitaker et al. (Whitaker et al., 2001), for example, observed stutter peaks that were as high as 0.40 for alleles with a peak area of < 10 000 RFU.

The ratio in peak height or area (RFU) between two peaks at heterozygous loci is referred to as the heterozygous balance (Hb). If the starting quantity of DNA is optimal for the amplification kit used, then the amplification rate of both alleles at a locus should be consistent and therefore Hb = 1. When the starting quantity of DNA is sub-optimal however,

random selection will increase the probability that one allele is amplified to a greater extent than the other during the early cycles of PCR (Garvin et al., 1998, Ray et al., 1996). As the number of copies of this allele increases with each cycle, there are relatively fewer copies of the second allele, which is then visualised as imbalance in the peak height at the locus, becoming more severe as the Hb value draws towards zero (allelic drop-out). An increase in the number of PCR cycles, a method often used to enhance the amplification of ltDNA samples (Section 3.1), will exacerbate this effect, resulting in a higher frequency of observed heterozygous imbalance (Gill et al., 2000b). Hb by peak area can be calculated using either Hb = lower φ allele/ higher φ allele or Hb = LMW allele φ / HMW allele φ (Gill et al., 2000a). The latter of these equations provides additional information in that a result over Hb > 1, indicates preferential amplification of the LMW allele and H < 1 of the HMW allele.

An association between the mean peak area (RFU) and the variation in peak asymmetry at a locus has been demonstrated by Whitaker et al. (Whitaker et al., 2001) using loci that have peaks with an area >10,000 RFU, Hb min = 0.63 Hb max = 1.59, whilst loci with peaks <10,000 are expected to have more variation, Hb min = 0.58 Hb max = 1.89.

The difficulty that arises more frequently in ltDNA profiles is the potential to incorrectly assign a locus as homozygous, when in actuality there may only be a single peak present as a result of allele drop-out. The consequences of this in casework are potentially serious as if the reference profile is heterozygous for a particular locus, and the crimestain is mistakenly designated homozygous, then the true offender might be excluded as the DNA source. To address this issue, a ltDNA threshold can be determined from a known source of DNA as the minimum peak height at which the drop-out of a second allele can be excluded, thus identifying a single peak as a homozygote (Gill et al., 2009).

Allelic drop-in is the term used to describe the presence of one or two alleles, which have occurred independently as a result of the fragmented chromosomal DNA that is ubiquitous in the environment (Gill and Buckleton, 2010). Allelic drop-in is inescapable when working with methods and technologies that have been designed to detect even minute traces of DNA. The presence of more than two alleles is indicative of a likely contamination event rather than random allelic drop-in and it is important not to confuse the two concepts.

It was first theorised by Navidi et al. (Navidi et al., 1992) and later tested by Taberlet et al. (Taberlet et al., 1996) that the use of a 'multiple tubes approach' would enable a more complete set of DNA results, a process widely known as consensus profiling. Consensus profiling requires a minimum of two amplifications from the same DNA extract, the profiles for which can then be compared and only alleles that are observed at least twice are recorded (Cowen et al., 2011). This approach somewhat alleviates the interpretation of artifacts such

as high stutter peaks and allele drop-in, as the stochastic effects of PCR are random by nature and so any two profiles, even from the same DNA extract, will exhibit differences. The conservative nature of this method results in a higher degree of confidence when analysing ltDNA profiles as demonstrated by Cowen et al. (Cowen et al., 2011), who found that the number of drop-in alleles was reduced by at least 95% when all alleles had to be observed twice.

It is not possible to generate a consensus profile in the traditional sense from a single-cell DNA source as in having only two copies of the genome (diploid cells), multiple amplifications will result in differing regions of the DNA template being randomly amplified. It is however possible to amplify multiple single-cells in independent reactions, and then to generate consensus across the samples. Allele dropout has been shown to effect single-cell DNA amplification at a rate of approximately 10% per locus (Findlay et al., 1998, Findlay et al., 1997). Due to the random nature of the preferential amplification of one allele over another in ltDNA samples (Garvin et al., 1998, Ray et al., 1996), misinterpretation of a heterozygous locus as homozygous due to allele dropout can be reduced by the amplification and DNA profiling of multiple individual single-cells to provide consensus profiles. For example Garvin et al. (Garvin et al., 1998) demonstrated that when using a total of eight single-cells (four loci DNA profiles), a locus would be incorrectly designated as a homozygote in just 0.017% of cases where allelic dropout was observed at 62% per cell.

In order to try and assuage some of the stochastic effects of PCR, thereby improving the quality of the resulting DNA profile, varying modifications to standard PCR and CE protocols have been researched.

A number of modifications to the stage of DNA amplification have been shown to increase PCR sensitivity and thus the ability to detect and interpret DNA profiles generated from minimal quantities of starting DNA template. The most effective of these is to increase the number of cycles in the PCR program from the standard number of 28-cycles up to a total of 34-cycles, either in one program (Findlay et al., 1997) or as two sequential amplifications referred to as 28 + 6 cycles (Kloosterman and Kersbergen, 2003). Due to the increased presence of stochastic effects resulting from PCR using ltDNA, increasing the number of PCR cycles above 34 has been shown offer no additional advantages (Gill et al., 2000b).

A reduction in the volume of reagents used for PCR has shown to increase the sensitivity with which low level template DNA can be detected as it increases the ratio of DNA to reaction components (Gaines et al., 2002, Proff, 2006, Li, 2009, Leclair et al., 2003). For example, a volume of 5 µl or 10 µl were shown to be most effective in increasing PCR sensitivity when using the AmpFℓSTR® Profiler Plus™ Kit (Gaines et al., 2002) with 22%

and 32% of expected alleles recovered respectively, whilst no amplification was observed using reaction volumes of 15, 20 and 50 µl.

Direct PCR allows for the addition of the forensic material straight to the amplification reaction (Mercier et al., 1990); whilst there are associated reductions in time and cost using this method, importantly for ltDNA samples the potential for loss of DNA or introduction of contamination during the extraction process is eliminated. The addition of extraction reagents either to the PCR components (Goldenberger et al., 1995) or as a step prior to amplification (carried out in the same reaction tube), can facilitate cell lysis and therefore the release of DNA to increase PCR efficiency. A disadvantage of this method is the potential for PCR inhibitors to be added into the amplification reaction along with the sample. The effect of this is alleviated somewhat due to the inclusion of buffers and enzymes with a higher tolerance for inhibitors in kits such as Applied Biosytems AmpFℓSTR® Identifiler® Direct PCR Amplification Kit (Wang et al., 2012, Wang et al., 2011) and Promega's PowerPlex® 18D System (Oostdik et al., 2013), as well as the direct-PCR compatible extraction reagents (Park et al., 2008). It has also been shown that some standard amplification kits, designed for use with an extraction step, may also be used effectively for direct PCR (Hall and Roy, 2014).

PCR enhancers are components added to the PCR reagents that may improve the efficiency of amplification, particularly for DNA templates with a high GC content (Henke et al., 1997, Jensen et al., 2010), a high number of homopolymeric regions (Fazekas et al., 2010) or localised repeat sequences (Haqqi et al., 2002), and can also increase the stringency with which the reaction occurs, reducing the presence of stochastic effects such as stutter (Marshall et al., 2015). The effectiveness of betaine, a commonly used PCR enhancer, for ltDNA samples has been demonstrated by Marshall et al. (Marshall et al., 2015) who observed an increase in PCR yield for samples with a starting DNA quantity of 100 pg or 25 pg, using both AmpFℓSTR® Identifiler and Identifiler Direct amplification kits and also an 'in-house' two-locus assay.

In addition to modifications during PCR, optimisation of post-PCR processes has also shown effective in increasing the sensitivity with which DNA can be detected.

During electrokinetic injection, negatively charged primers and ions will compete with DNA for uptake by the electrodes prior to capillary electrophoresis, thereby limiting the quantity of DNA available for visualisation. The use of post-PCR clean-up methods to remove these additional PCR components from the sample, have shown to result in an increase in peak height (RFU) (Smith and Ballantyne, 2007, Forster et al., 2008), thus assisting interpretation of present alleles on the generated electropherogram. The advantage in utilising post-PCR

methods over modifications during pre-PCR stages, is the reduced risk of sample contamination using amplified DNA (Smith and Ballantyne, 2007); however, the reaction volume used for amplification must be considered as there is potential for sample loss during this stage. Additionally the use of a highly deionised form of formamide for sample injection will reduce competition from ions during electrokinetic injection and increase DNA uptake into the capillaries (Butler et al., 1995).

Adjusting the electrokinetic injection voltage or length of time for which the voltage is applied will facilitate the uptake of DNA for capillary electrophoresis and so an associated increase in peak height RFU is observed (Westen et al., 2009, Forster et al., 2008, Cowen et al., 2011). This can be particularly useful to increase the height of potential allelic peaks above the analytical threshold, which under standard conditions would not have been recorded in the results. Cowen et al. (Cowen et al., 2011) were able to demonstrate that amplified DNA may be present but left undetected, as shown when up to 50% of dropped-out alleles were detected following CE at 4 kV for 30 s as compared with the standard CE settings of 3 kV for 10 s.

In some circumstances however, the nuclear DNA within the sample will have degraded to such an extent that any amplification and subsequent information obtained is limited. Mitochondrial DNA is known to persist better than nuclear DNA and so offers an alternative for forensic investigation as previously discussed in Section 1.2.4.3.

Despite the interpretation issues discussed above, there are two substantial advantages to single-cell profiling; the reduced risk of obtaining a mixed DNA profile and the ability to obtain forensic DNA information from any biological material.

3.2 Aims and objectives

The use of individual single-cells in forensic casework is limited due to concerns regarding difficulties associated with reliable DNA amplification, potential for contamination and interpretation challenges due to both the expected exacerbation of stochastic effects associated with any ltDNA sample and inability for a consensus profile to be generated. The aim of this part of this thesis is to investigate the analysis and interpretation of DNA derived from individual single-cells to determine their efficacy for use in forensic casework.

The objectives of this study are to first develop an effective method for the visualisation and recovery of single-cells for DNA analysis that does not cause inhibition for downstream processing, and which can be carried out efficiently and under DNA anti-contamination conditions. The use of the AmpFℓSTR® SGM PCR Amplification Kit (SGM Plus), as the commercial kit utilised for loading to the NDNAD at the commencement of this study, will be optimised in order to maximise the information obtained from the DNA template available in a single diploid cell, approximately 6 pg. Once an optimal process has been determined, data will be generated to evaluate parameters for interpretation, i.e. stutter percentages etc. in order to assess challenges relative to published work on other ltDNA samples. In light of the recent transition to DNA 17 amplification kits for loading to the NDNAD, a preliminary look at the performance of three commercially available kits, AmpFℓSTR® NGM PCR Amplification Kit and the PowerPlex® ESX 16 and ESI System, will be carried out to assess their performance against SGM Plus. Finally, in addition to this, an ‘in-house’ assay will be developed targeted for amplification of ltDNA samples to determine if the use of fewer low molecular weight loci will improve amplification reliability in single-cell levels of DNA.

3.3 Materials and methods

All single-cell DNA and assay development work for this project was approved by the Northumbria University Ethics Committee under the reference number RE25-10-121294.

3.3.1 Inhibition tests

50 pg of 007 DNA was added to twenty-eight 0.1 ml PCR tubes and an equivalent volume of MBG water was transferred to an additional six PCR tubes for negative controls. A square of UV-treated J-LAR tape or Sellotape, approximately 0.2 x 0.2 mm, was placed into each of three DNA-containing tubes and two water-containing tubes. In order to facilitate the release of the single-cell from the tape, two different adhesive removers, Sticky Stuff Remover (SSR) and Brava® Adhesive Remover (AO), were tested at volumes of 1 µl, 2 µl or 3 µl; added in triplicate to PCR tubes containing DNA. SSR is currently used by the FBI as an adhesive remover to recover textile fibres from under fibre tapes for further analysis (R. Palmer 2013 pers. comm. 8th February). Data on the components in the SSR solution is limited to hydrocarbons, n-alkanes, isoalkanes, cyclics and aromatics (De-Solve-it, 2012). Brava® Adhesive Remover is used by individuals who have a stoma, to remove any adhesive from the surface of the skin without causing damage (Coloplast, 2013); no information is available on the contents. The remaining four DNA samples and two water samples were used as control samples and PCR negatives respectively. In all cases the volume was made up to a total of 6 µl using MBG water. SGM Plus reagents were added to all samples and amplified for 34-cycles as described in Section 2.9.1. All SGM Plus PCR products were then visualised and analysed as described in Sections 2.13.2.1 and 2.13.3.1.

3.3.2 PCR volume trial

Single-cells and a tape negative control were prepared and transferred into a 0.1 ml PCR tube containing 1 µl of SSR and SGM Plus reagents were then added to each tube as described in Table 3-1 to give final reaction volumes of 5, 10, 15, 20 & 25 µl. Samples were amplified for 34-cycles according to manufacturer's instructions as detailed in Section 2.9.1. All SGM Plus PCR products were then visualised and analysed as described in Sections 2.13.2.1 and 2.13.3.1.

Table 3-1 SGM Plus component volumes for each reaction volume tested

Reaction Volume (µl)	SGM reaction mix (µl)	SGM Plus primer set (µl)	Taq polymerase (µl)	MBG water (µl)
5	2.1	1.1	0.1	1.7
10	4.2	2.2	0.2	3.4
15	6.3	3.3	0.3	5.1
20	8.4	4.4	0.4	6.8
25	10.5	5.5	0.5	8.5

3.3.3 Initial ForensicGEM extraction trial

Six individual single-cells and a tape negative control were prepared and transferred into a 0.1 ml PCR tube containing 1 µl of SSR. A mastermix prepared on ice, containing 0.6 µl 10X Gold Buffer, 0.3 µl prepGEM enzyme and 4.1 µl of MBG water was then added to each sample to give a total extraction volume of 6 µl. Samples were incubated on a Veriti® 96-Well Fast Thermal Cycler for 10 min at 75 °C, followed by 5 min at 95 °C to deactivate the prepGEM enzyme. Samples were then placed on ice and SGM Plus reagents added to each sample as described in Section 2.9.1. These results were then compared with three single-cells which had been amplified without the use of the ForensicGEM reagents as a control. All SGM Plus PCR products were visualised and analysed as described in Sections 2.13.2.1and 2.13.3.1.

3.3.4 Single-cell DNA profiles

The donor volunteer for buccal cells used in this experiment was chosen based on his SGM Plus DNA profile as he was heterozygous at nine of the ten STR loci and also shared minimal alleles with the operator, facilitating detection of sample contamination.

A buccal sample was cell harvested as described in Section 2.4.3, with the additional wash step added to the protocol in order to remove all salivary fluid from the cell samples, and prepared for staining with haematoxylin and eosin as described in Sections 2.4.3.2 and 2.5.1. Fifty single-cells were then recovered and prepared as described in Section 3.3.3. All SGM Plus PCR products were visualised and analysed as described in Sections 2.13.2.1 and 2.13.3.1, using both a standard and extended capillary electrophoresis injection time.

3.3.5 Assay comparison

Buccal samples were prepared as described in Section 2.4.3 and transferred directly to a microscope slide for staining with haematoxylin and eosin as explained in Sections 2.4.3.2 and 2.5.1. Forty individual cells were collected and transferred into 0.1 ml PCR tubes for

treatment using ForensicGEM reagents as detailed above (Section 3.3.3). Each of the amplification kits to be tested, AmpFℓSTR® NGM PCR Amplification Kit (NGM), PowerPlex® ESX 16 System (ESX), and PowerPlex® ESI 16 System (ESI), were added to ten cells and amplified according to the manufacturer's instructions as summarised in Section 2.9.1.

The final ten cells were amplified using an 'in-house' developed assay (Section 3.3.6) for ltDNA (ltDNA assay) as described in Section 2.9.2. All PCR products were visualised and analysed using the kit relevant reagents and protocols as described in Sections 2.13.2.1 and 2.13.3.1.

3.3.6 Development of an in-house low-template DNA amplification assay

In order to try and improve the reliability of the amplification results for single-cell DNA analysis it was thought that a miniSTR kit with shorter target fragment sizes might result in a more consistent result. In the absence of funding for a commercially available kit, an in-house ltDNA assay was designed and optimised. The requirements of the assay were to multiplex loci that have a relatively small fragment size, as this facilitates amplification as discussed in Section 1.2.1, but with a high degree of variability to enable identification of the donor source wherever possible.

It was decided to include just four target STR loci in this assay, one in each of the 6-FAM, VIC, NED and PET dye lanes to allow for flexibility in primer design whilst keeping all loci at a minimal size. The fifth dye lane was then used to accommodate the LIZ-labelled size standard in order to size DNA fragments and identify alleles following PCR.

A number of resources were utilised to determine which STR loci would be ideal candidates for inclusion in the multiplex, including published literature, particularly (Butler et al., 2003, Coble and Butler, 2005, Hill et al., 2008, Guo et al., 2014), and databases such as the ALlele FREquency Database (ALFRED) (Osier et al., 2001) and the ALLST*R autosomal STR database (Götz, 2015). Parameters such as allele range, fragment size, power of discrimination and heterozygosity value amongst the British population were considered during selection. Loci D10S1248, D12S391 and Penta E were chosen as suitable STR markers, in addition to which Amelogenin was included for sex determination.

D10S1248 (Coble and Butler, 2005) is an ideal locus for inclusion in this multiplex for the purposes of sensitivity as the allele range spans only 44 bp and by using primer binding sites immediately adjacent to the repeating region, the total target fragment size can be kept minimal (Coble and Butler, 2005), thus facilitating locus amplification in both low-template and degraded DNA samples. For this reason, as well as a relatively high power of

discrimination (PD) value of 0.87 (sample size of 2949 individuals) amongst twenty-six countries from the European Union (Welch et al., 2012), D10S1248 was proposed as an addition to the European standard set of loci (ESS) in 2006 (Gill et al., 2006a). D12S391 is a tetranucleotide STR (Lareu et al., 1996) that has over fifty different alleles (Butler, 2012) and was also recommended as an addition to the ESS of loci in 2006 due to a high PD value of 0.97 (sample size of 2949 individuals) amongst the same twenty-six countries from the European Union (Gill et al., 2006b, Welch et al., 2012). Finally Penta E (Schumm, 1998), which is located on chromosome 15, was selected for inclusion in the multiplex as this locus has a wide range of alleles and therefore a high power of discrimination. As Penta E is not widely utilised in Europe, there is little data available; however, the PD value for the UK is given as 0.98 (sample size of 156 individuals) (Götz, 2015). Partially due to its pentanucleotide nature, Penta E exhibits very low levels of stutter <2% (Schumm, 1998) relative to tetranucleotides, which will facilitate interpretation in the presence of the stochastic effects associated with low-template and degraded DNA material. The high heterozygosity values for D10S1248, D12S391 and Penta E of 0.76, 0.89 and 0.90 respectively (Götz, 2015) will also aid identification of source DNA, even in the presence of allele dropout. Calculating the match probability using the product rule as previously described in section 1.3.1, the probability of an unrelated individual having the same DNA profile as that generated using the ltDNA assay for the 007 control, was calculated as approximately 1 in 51,000 (allele frequencies obtained from (Götz, 2015).

Primer sequences were initially obtained from published literature (see Table 3-2) and checked for multiplex compatibility using both NetPrimer (http://www.premierbiosoft.com/netprimer/index.html, last accessed 24/05/2015) and the NIST Autodimer tool (Vallone and Butler, 2004). Due to a difference in melting temperature (Tm) across the primer sets of 11.5 °C, the Penta E forward and D12S391 reverse primers were redesigned using Primer-BLAST (Ye et al., 2012) resulting in a decrease in the Tm range to 6.5 °C.

The fluorescent intensity of the NED and PET dyes is low, approximately 2.5x and 4x less respectively, relative to the 6-FAM and VIC dyes and so, to balance the peak heights across the profile, these two dyes were coupled with the two smallest fragment targets of Amelogenin and D10S1248, as the LMW targets are preferentially amplified, therefore accounting for some of the intensity defect. Each dye was attached to the forward primer of each locus set as shown in Table 3-2. The primer concentration was also increased for the two smaller-sized loci during the initial testing; however, this did not affect the peak height balance between the loci. A diagram depicting the final assay layout is shown in Figure 3.1.

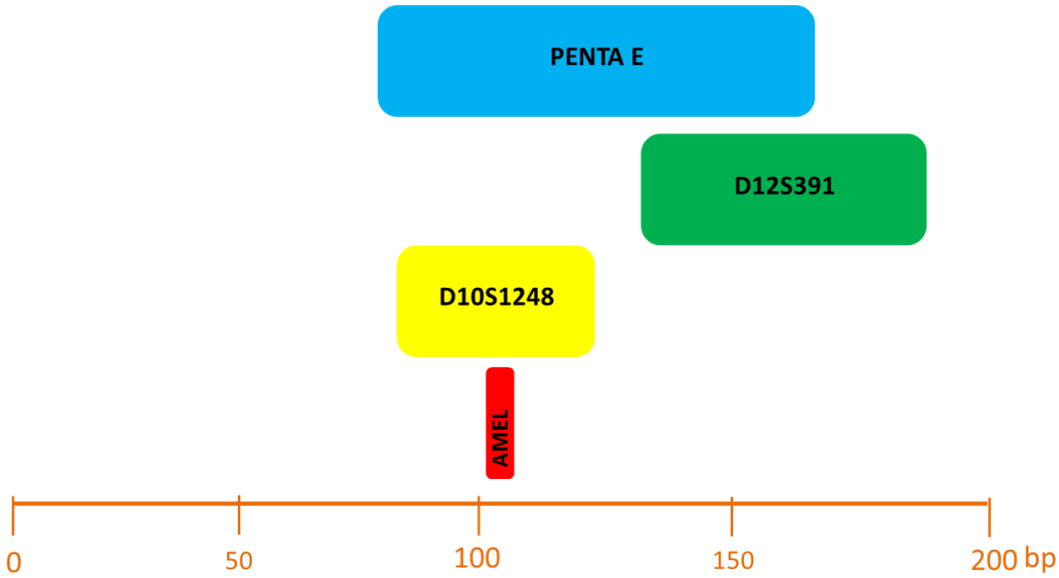


Figure 3.1 Overview of positions of loci included in the ltDNA amplification assay relative to the size standard in base pairs (bp) in orange colour. Each set of primers was tagged with a different fluorescent tag to allow for detection on the 3130 Genetic Analyzer; loci are shown in the colour related to the attached dye.

Primers were set-up for DNA amplification using the Qiagen Multiplex PCR kit that includes a buffer containing K^+ and $NH4^+$, which together facilitate the specific and stable hybridisation of primers to the DNA template, enabling an efficient extension of the DNA strand (Qiagen, 2010). A series of experiments were conducted in order to optimise the assay to ensure maximum sensitivity, including the annealing temperature tested at 58-60 °C, primer concentrations 0.1–0.5 µM (each) and inclusion of an additional unit of HotStar Taq DNA Polymerase. The addition of the Q-solution additive included in the Qiagen Multiplex PCR Kit, which is designed to increase assay tolerance to the annealing temperature range across the primer sets, was also tested. Results were determined using relative peak heights and the overall quality of the EPG, i.e. presence of PCR artefacts, using 007 control DNA at a total input of 50 pg. The following conditions were determined as optimal for this assay; a primer concentration of 0.3 µM (each), the addition of Q solution to the reaction mixture and an annealing temperature of 59 °C. Alleles were designated using their fragment size (see appendix 8.2) relative to the included size standard. An example of a ltDNA assay profile is shown in Figure 3.2.

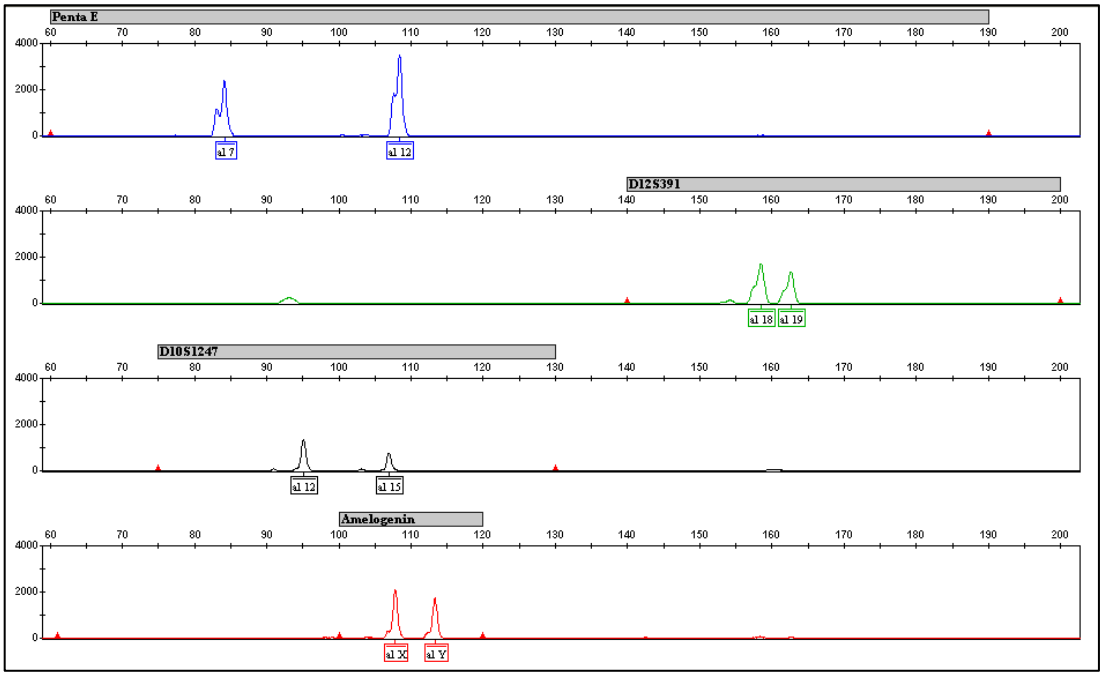


Figure 3.2 Example of EPG for ltDNA profile generated from a total DNA input of 50 pg using 007 control DNA. Alleles at STR loci Penta E, D12S391 and D10S1247 have been sized relative to a size standard (not shown) and designated using expected allele sizes (see appendix 8.2). Alleles at Amelogenin were also sized relative to the size standard and designated X or Y based on expected fragment size.

A serial dilution was set-up using 007 control DNA at total DNA inputs of approximately 100 pg, 50 pg, 25 pg, 13 pg, 6 pg and 3 pg, in order to test the sensitivity of the assay under optimised conditions; reactions were set-up in triplicate as described in Section 2.9.2.1.

An minimum average of 96% of the expected alleles were amplified down to a total DNA input of 25 pg, shown in Figure 3.3. Successful amplification of alleles was also observed in all samples at DNA inputs of approximately 13 pg, 6 pg and 3 pg, with averages of 79%, 63% and 38% of expected alleles observed respectively.

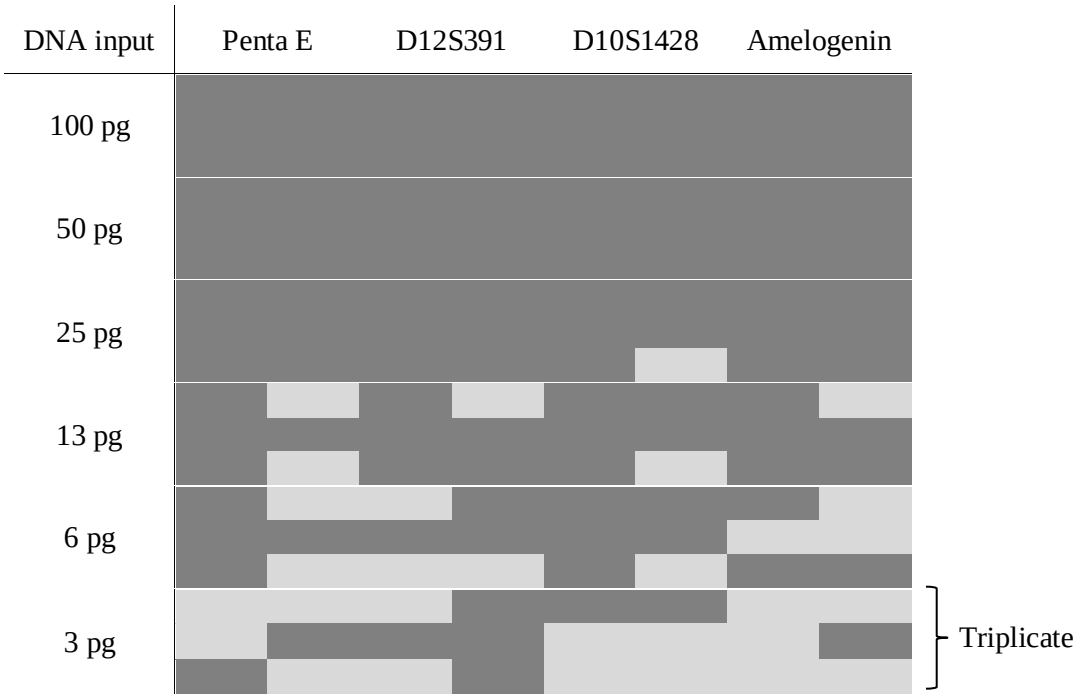


Figure 3.3 Summary of sensitivity test results for the ltDNA assay showing successful amplification of alleles (grey) and allele drop out (pale grey) at each individual locus for decreasing total DNA inputs, from 100 pg – 3 pg. Tests were carried out in triplicate with all results shown.

Table 3-2 Summary of primer information for the four loci included in the ltDNA amplification assay. Two of the primers, D12S391 (reverse) and Penta E – (forward) were redesigned for this multiplex (denoted by *)

Locus	Genbank accession	Primer set	Repeat motif	Fragment size range (bp)	Allele range	Reference
D10S1248	AL391869	NED-TTAATGAATTGAACAAATGAGTGAG GCAACTCTGGTTGTATTGTCTTCAT	GGAA (simple)	79-123	8-19	(Coble and Butler, 2005)
D12S391	G08921	PET-TCTCCAGAGAGAAAGAATCAACAGGA GCCATATCACTTGAGCTAATTCCTCTAATA	AGAT/AGAC (compound)	145-195	15-27.2	(Guo et al., 2014)*
Penta E	AC027004	6FAM-GGCGACTGAGCAAGACTC GGTTATTAATTGAGAAAACTCCTTACA	AAAGA (simple)	73-168	5-24	(Butler et al., 2003)*
Amelogenin	M55418 M55419	VIC-CCCTGGGCTCTGTAAAGAATAGTG ATCAGAGCTTAAACTGGGAAGCTG	NA	106 112	X Y	(Sullivan et al., 1993)

3.4 Results

3.4.1 Inhibition tests

Allele peak area (RFU) was recorded at each locus and compared against a control sample to determine if the addition of either adhesive tape or adhesive remover to the sample tube resulted in PCR inhibition. Of the two tapes tested, J-LAR tape did not appear to affect peak area (AV 24471 RFU, SD 12397 RFU) when compared with the control sample (AV 20241 RFU, SD 9729 RFU), whilst a decrease in peak height relative to the control was observed across the DNA profile for Sellotape (AV 9051 RFU, SD 4927) as shown in Figure 3.4.

In order to facilitate release of the single-cell from the tape two adhesive removers were tested for indication that PCR would be inhibited if added to the sample tube. No PCR inhibition was observed in the presence of 1 µl SSR (AV 15482 RFU, SD 7203 RFU) relative to the control sample (Figure 3.5) however, the addition of 2 µl and 3 µl resulted in complete inhibition of the PCR. All volumes of AO resulted in complete PCR inhibition with no DNA amplification observed. Data for charts included in appendix 8.3.

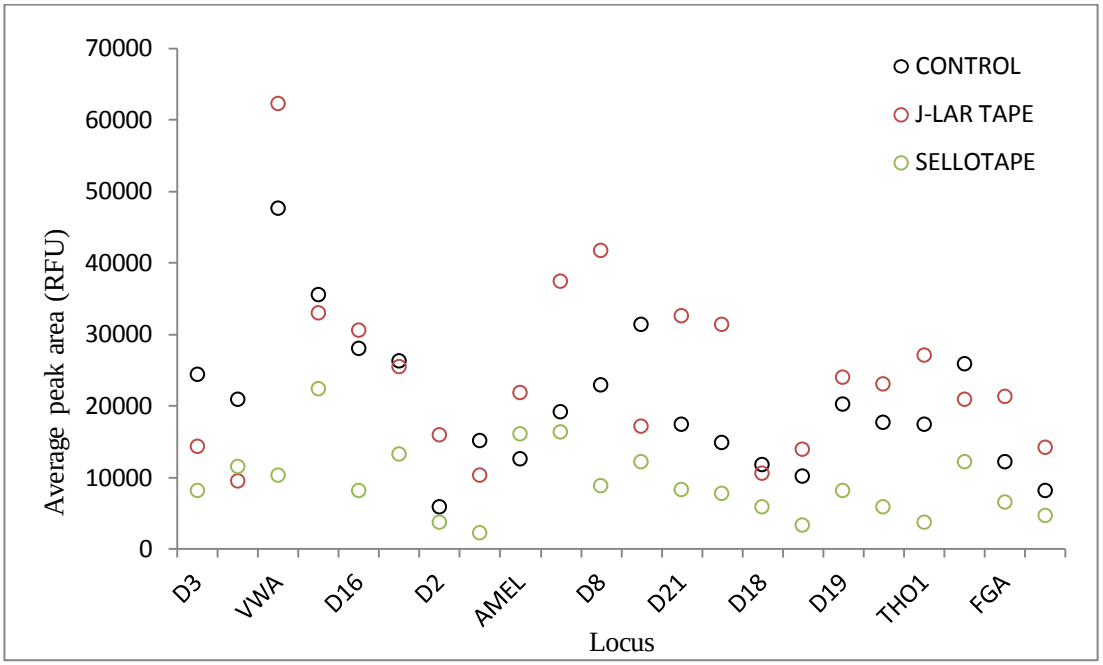


Figure 3.4 Peak area of alleles at each locus following addition of adhesive tape into sample tube during PCR as compared with a control sample. Tests were carried out in triplicate and the average peak area (RFU) for each amplified allele in the DNA profile is shown for Sellotape (red) J-LAR tape (green). The average results for a control sample (black) tested without the addition of adhesive tape is also shown for comparison. A decrease in the average peak area across all loci, relative to the control sample is seen for the Sellotape indicating PCR inhibition, whilst the peak area of alleles J-LAR tape was comparable with the control sample.

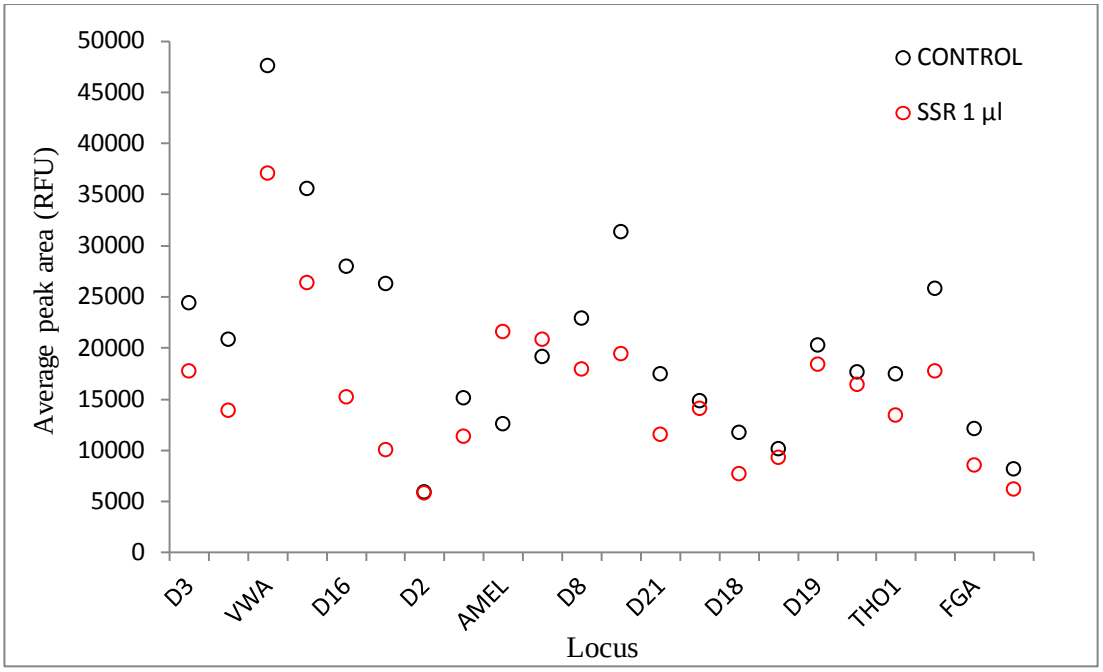


Figure 3.5 Peak area of alleles at each locus following addition of 1 µl SSR into sample tube during PCR as compared with a control sample. Tests were carried out in triplicate and the average peak area (RFU) for each amplified allele in the DNA profile is shown for 1 µl SSR (red) and a control sample (black), which was amplified without the addition of an adhesive remover for comparison. Although a slight decrease in peak area can be seen at some loci, no significant difference in the results was observed.

3.4.2 Amplification reaction volume trial

Single-cells were prepared in triplicate and amplified using SGM Plus reagents, in reaction volumes of between 5 – 25 µl on a 34-cycle PCR program. The average DNA results summarised in the chart below (Figure 3.6) indicate that 15 µl is the most effective reaction volume for single-cell DNA amplification, whilst 5 µl was least effective in these experiments. Data for chart included in appendix 8.4.

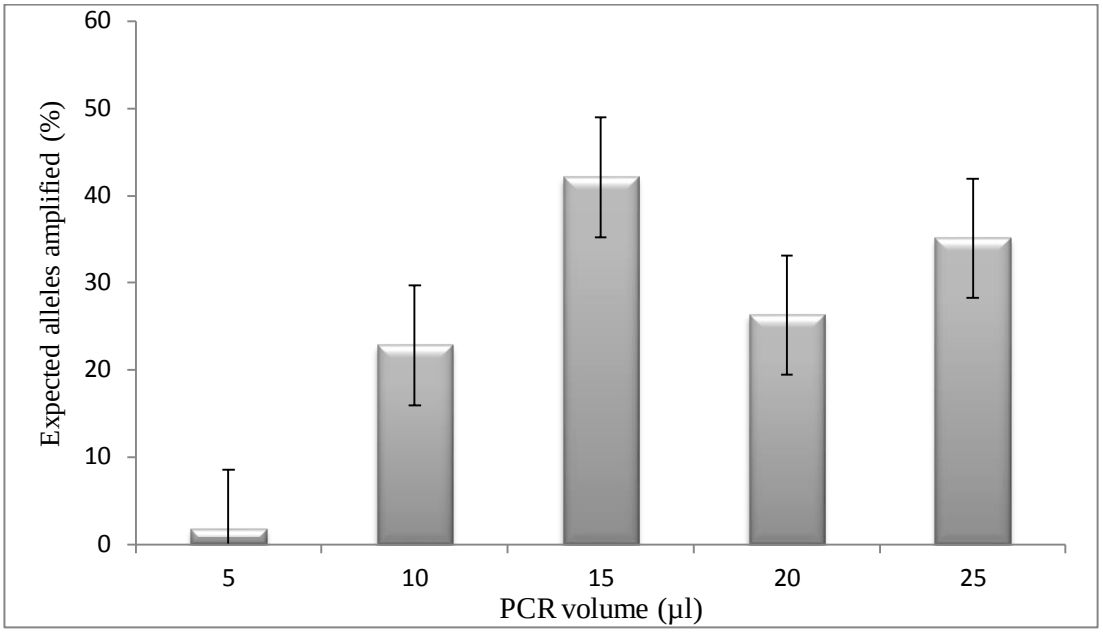


Figure 3.6 Effect of PCR volume on the average number of expected alleles amplified from individual single-cells using a 34-cycle PCR. The fewest expected amplified alleles was observed in 5 µl (2%), followed by 10 µ (23%), 20 µl (26%), 25 µl (35%) and finally the highest number of expected alleles were amplified in 15 µl (43%).

3.4.3 Initial ForensicGem extraction trial

The use of the ForensicGem reagents, including the incubation period prior to the addition of the SGM Plus reagents was shown to increase the average number of expected alleles amplified by 19%.

3.4.4 Single-cell overall results

Fifty individual single-cells were prepared and amplified from the same buccal source for 34-cycles and then visualised using capillary electrophoresis at both standard and extended injection times. Of the fifty resulting DNA profiles using a standard injection time, thirty-eight partial profiles were obtained, of which 61% had at least five alleles present (Figure 3.7). No amplification was observed for eleven samples and one result was excluded from the data set due to evidence of contamination as more than two unexpected alleles were present as set out in Gill and Buckleton (Gill and Buckleton, 2010). Increasing the CE injection time resulting in the amplification of an additional DNA profile contain at least five alleles. Data for chart included in appendix 8.5.

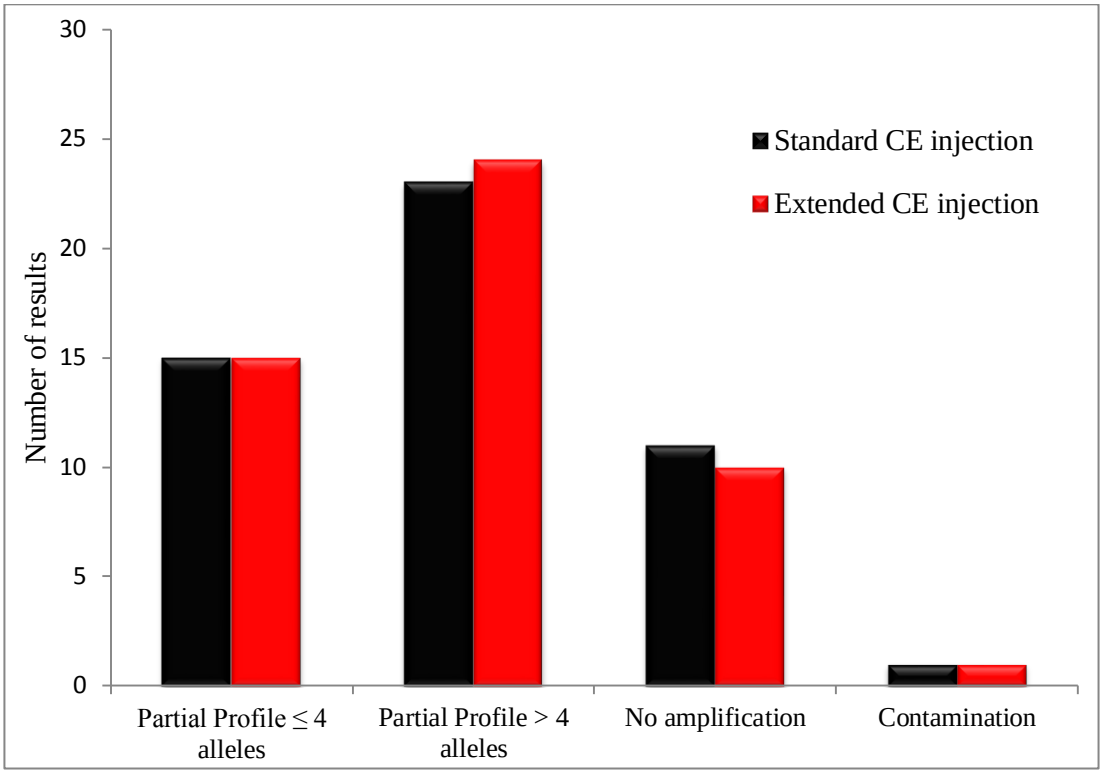


Figure 3.7 Summary of DNA profile results; using a standard CE injection the following results were observed: 23 partial profiles with more than four alleles, 15 partial profiles with four or less alleles, 11 of samples with no amplification observed and 1 sample in which contamination was evident. For the extended CE injection the number of samples where no amplification was present was reduced to 10, and the number of partial profiles with more than four alleles was increased to 24.

3.4.5 Amplified alleles

The number of alleles amplified at each locus was plotted as a percentage of the maximum number of alleles across all forty-nine DNA profiles, with loci arranged in ascending order of molecular weight. As shown in Figure 3.8 as the amplicon size increases, the frequency with which the allele is amplified decreases. The exception to this is vWA, excluded from chart, as it is the only homozygous locus in the donor's DNA profile and was observed in 51% and 53% of single-cell DNA profiles for standard and enhanced CE conditions respectively.

The amplicon size (bp) was plotted against peak height (RFU) for each of the observed alleles amplified in the forty-nine single-cell DNA profiles (Figure 3.9), for both standard and extended CE injection times. When the injection time was increased by a period of ten seconds, the peak heights increased by an average factor of 3 (SD 0.3) relative to the standard injection time. A slight decline can be seen in the peak height as the size of the amplicon increased, which is more pronounced when using the extended injection time. Data for charts included in appendix 8.6 and 8.7.

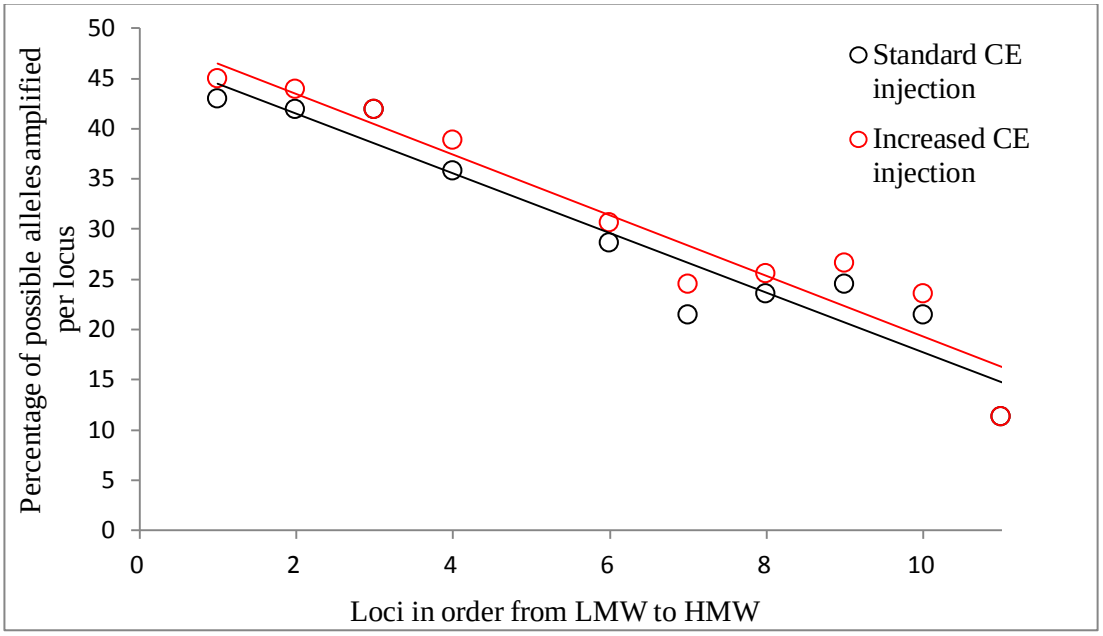


Figure 3.8 Relationship between the molecular weight of each locus with the number of times an allele was observed for both standard and increased CE injection times. As the molecular weight of the amplicon size at each locus increased, the percentage of alleles amplified can be seen to decrease.

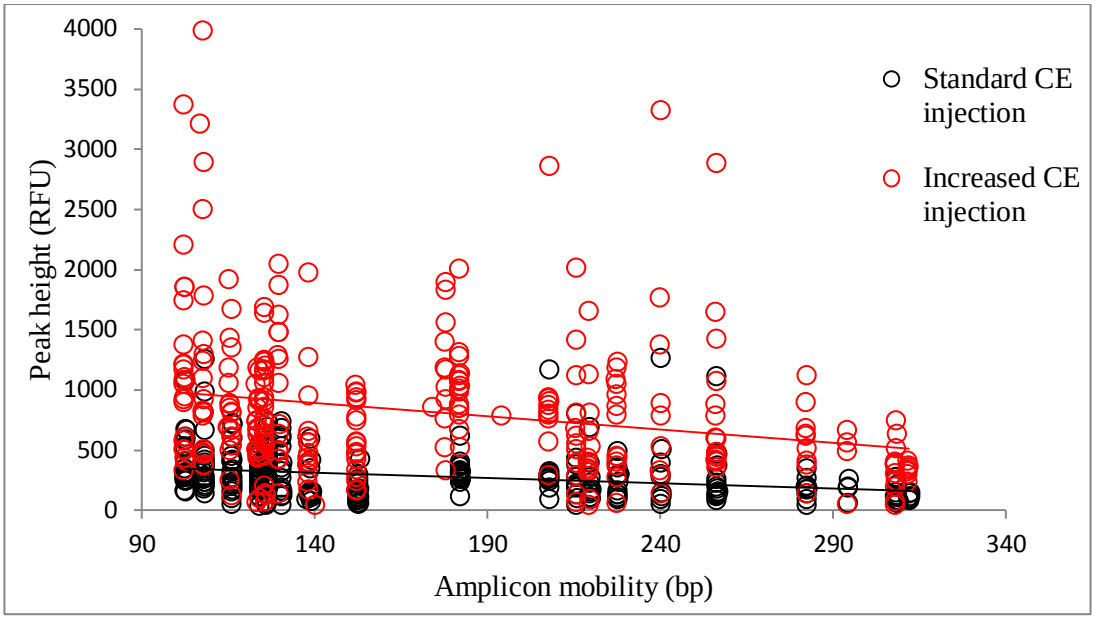


Figure 3.9 Relationship between amplicon size (bp) and peak height (RFU) for both standard and increased capillary electrophoresis injection time. As the amplicon size increases a slight reduction in peak height is observed.

3.4.6 Heterozygous balance

Heterozygous balance (Hb) was assessed using the allelic peak areas (φ) (combining the 'n' and 'n+1' peaks where present) of 83/90 loci, for standard/increased CE injection time, according to the following equation as described by Gill et al. (Gill et al., 2000a), Hb = LMW allele φ / HMW allele φ. The results were plotted as two scatterplots shown in Figure 3.10 and Figure 3.11, for those loci with an average peak height ≤10,000 RFU and >10,000 RFU respectively. Under standard CE conditions 51% of loci had a value of Hb ≥0.6, average Hb = 0.6 (SD 0.22), and a Hb range of 0.84, whilst using the enhanced CE settings resulted in values of 49% Hb ≥0.6, average Hb = 0.59 (SD 0.26), with the range Hb = 0.91. The increased injection time did not result in a significant decrease in heterozygous imbalance (P = 0.99). In cases where the maximum peak area was below 10,000 RFU, a total of 44% of loci had a value of Hb ≥0.6 (both standard and increased settings) whereas 57% of loci for which the maximum peak area exceeded 10,000 RFU had values of Hb ≥0.6. Irrespective of peak area, all loci except D3 and D16 had a median value of Hb > 1, thus indicating a slight tendency towards preferential amplification of the LMW allele. Amplification of D2 did not result in the presence of both alleles in any sample. Data for charts included in appendix 8.8, and statistics calculations are included in appendix 8.12.

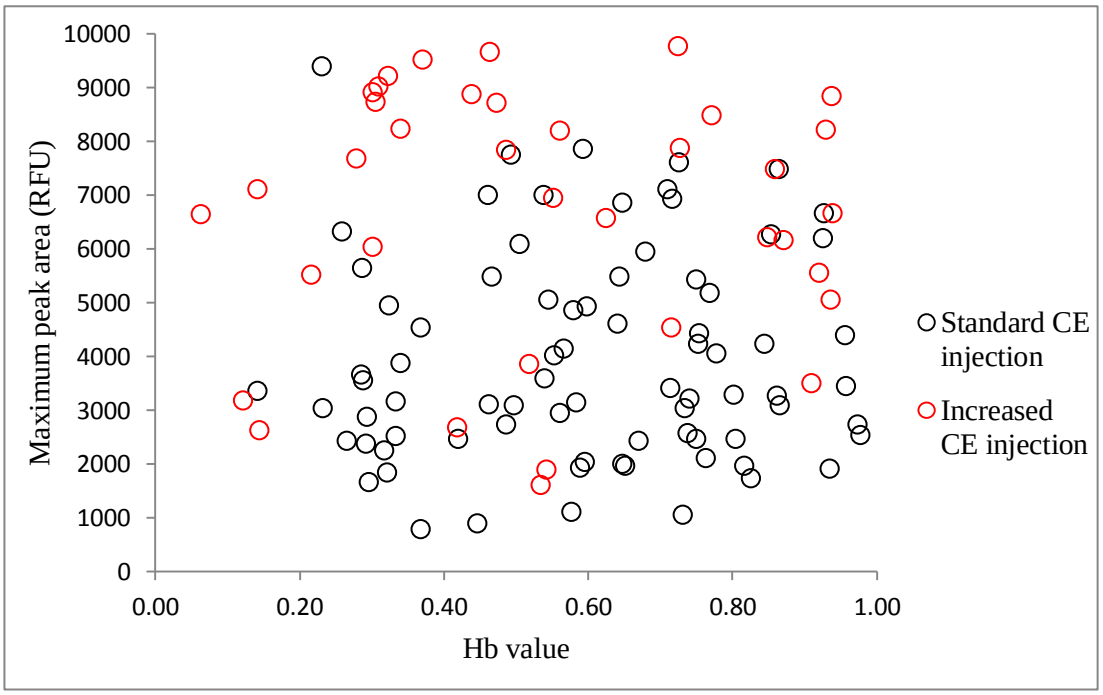


Figure 3.10 Relationship between heterozygous balance and maximum peak area, where area is under 10,000 RFU for both standard and increased injection times. No relationship is observed between the heterozygous balance for alleles with a peak area of <10,000 RFU for either injection time.

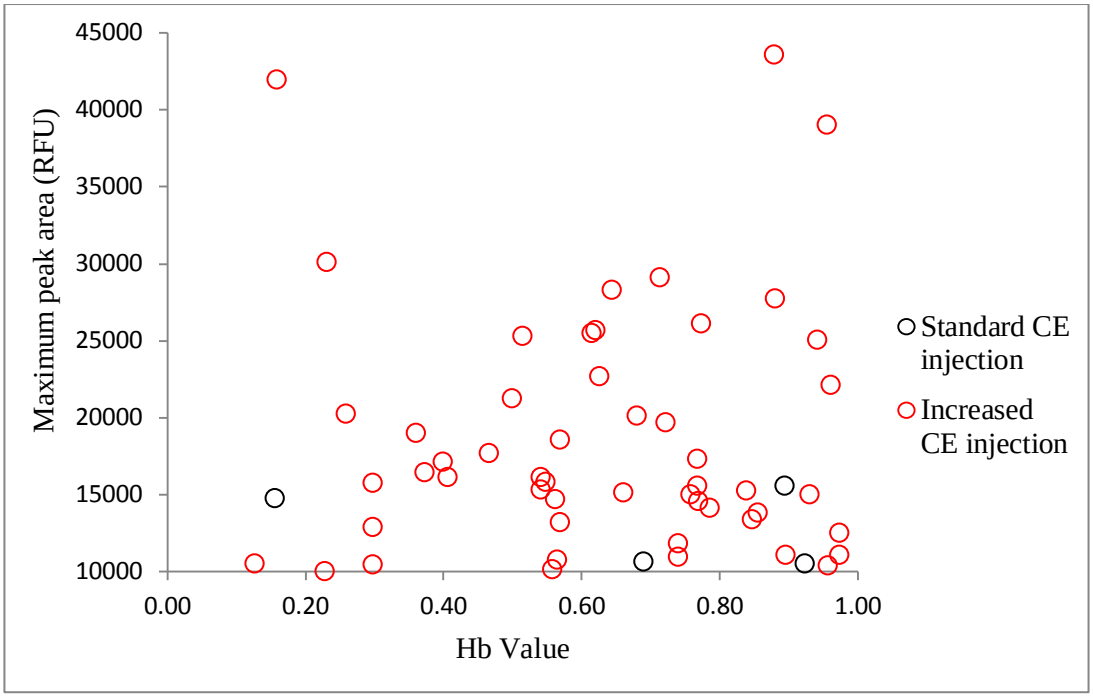


Figure 3.11 Relationship between heterozygous balance and maximum peak area, where area exceeds 10,000 RFU for both standard and increased injection times. Only four data points for loci where the peak area >10,000 RFU were recorded for a standard CE injection.

3.4.7 Locus dropout

Of a possible 429 loci, complete drop-out was observed at 46% of loci using a standard CE injection and 47% when using an increased injection time. The increased injection time did not result in a significant decrease in the level of locus dropout observed, with a P value of 0.95 (see appendix 8.12 for details). The loci with the highest proportion of locus drop-out are D2 and D16, whilst D19 has the lowest as shown in Figure 3.12. Two instances were recorded following an increased injection time during CE, for which alleles were detected at loci that had previously been recorded as dropped out using the standard setting. In both cases the allele could not be designated in the first instance due to poor morphology/peak height, which was improved following the enhanced injection time, see

Figure 3.13. Data for charts included in appendix 8.9.

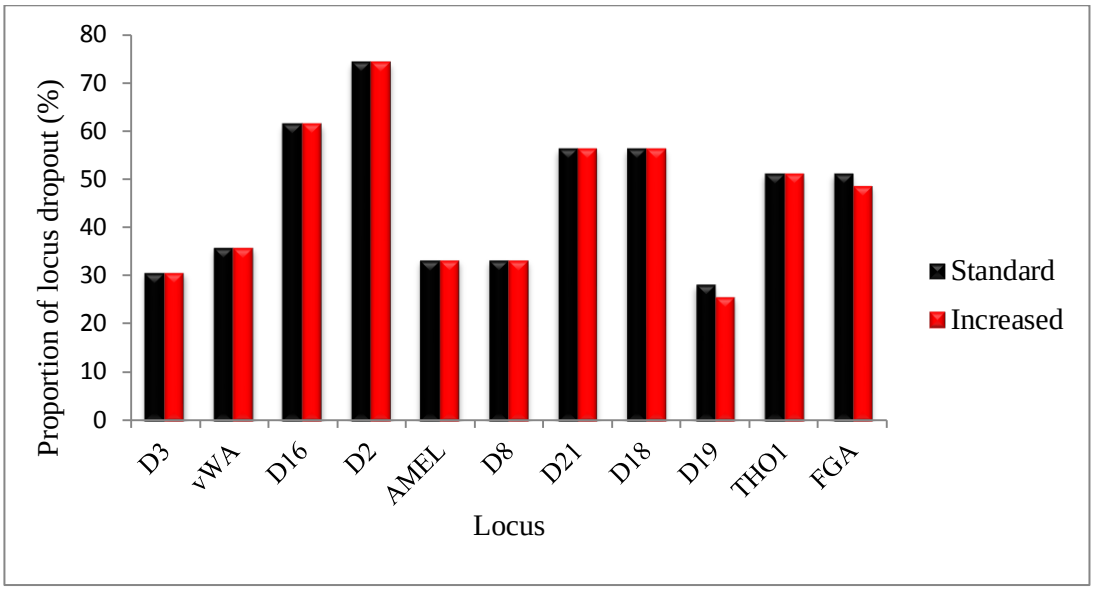


Figure 3.12 Frequency at which complete locus dropout was observed at each locus using both standard and increased CE injection times. Values at loci were the same, except for at D19 and FGA where the percentage is slightly reduced.

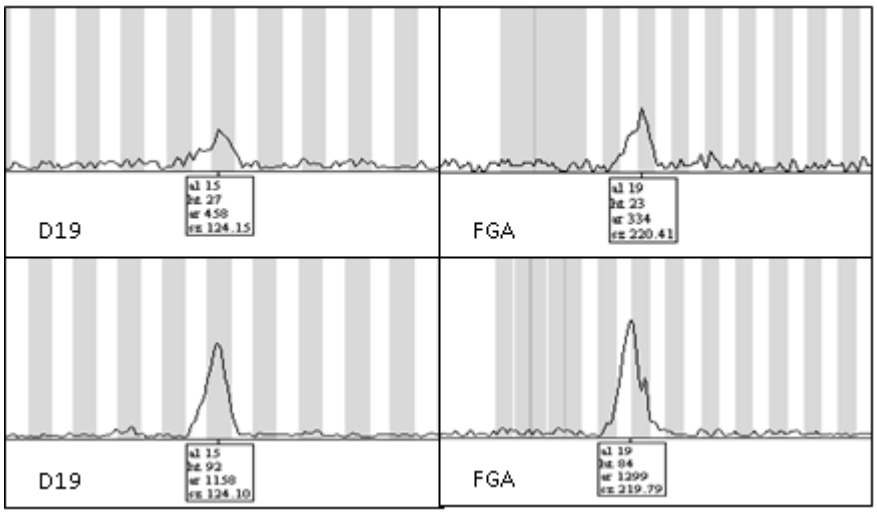


Figure 3.13 Images showing EPG of alleles for D19 (left) and FGA (right) for both standard (top) and increased (below) injection times during CE. It can be seen that under the standard injection times the peaks could have been mistaken for baseline noise on the electropherogram, however with the increase in injection time the morphology of the peak in each case becomes more distinct.

3.4.8 Allele dropout

The frequency of allelic drop-out was recorded for each locus in addition to which, if the allele was the lower or higher molecular weight allele at that locus was also noted (Table 3-3). Allelic drop-out was observed at 31% of all heterozygous loci; this was reduced to 29% using the increased injection during CE. The use of the increased injection time at CE did not result in a significant decrease in the level of allele dropout observed, P = 0.61 (see appendix 8.12). There was no bias towards the higher or lower molecular weight alleles using the standard injection, 50% for each, whilst the increased injection time resulted in a slight bias towards drop-out of the lower molecular weight (LMW) allele at 53%, than the higher molecular weight (HMW) allele at 47%.

The increase in injection time resulted in a reduction of 2% in allelic drop-out. In five out of six of the observed alleles the low peak area, relative to the previously present first allele, lead to severe imbalance (see Figure 3.14) as reflected with Hb values as follows: D2 0.12, D8 0.14, D21 15.43 and D18 15.37.

Figure 3.15 shows the peak heights (RFU) of the homozygous locus vWA and the remaining allele for loci at which allelic drop-out has occurred. The maximum peak height of the present allele, where the other has dropped out is 708 RFU and 1827 RFU for standard and increased injection times respectively. Data for charts included in appendix 8.9.

Table 3-3 Summary of allelic drop-out results at all heterozygous loci for both standard and increased injection times

Locus	Number of times allele dropped out					
	Standard			Enhanced		
	Total	LMW	HMW	Total	LMW	HMW
D3	13	1	12	13	1	12
D16	6	6	0	6	6	0
D2	9	3	6	8	5	3
D8	17	10	7	16	9	7
D21	11	5	6	10	5	5
D18	13	8	5	11	8	3
D19	15	8	7	15	8	7
THO1	10	7	3	10	7	3
FGA	15	8	7	14	7	7
AMEL	10	4	6	10	4	6
Total	119	60	59	113	60	53

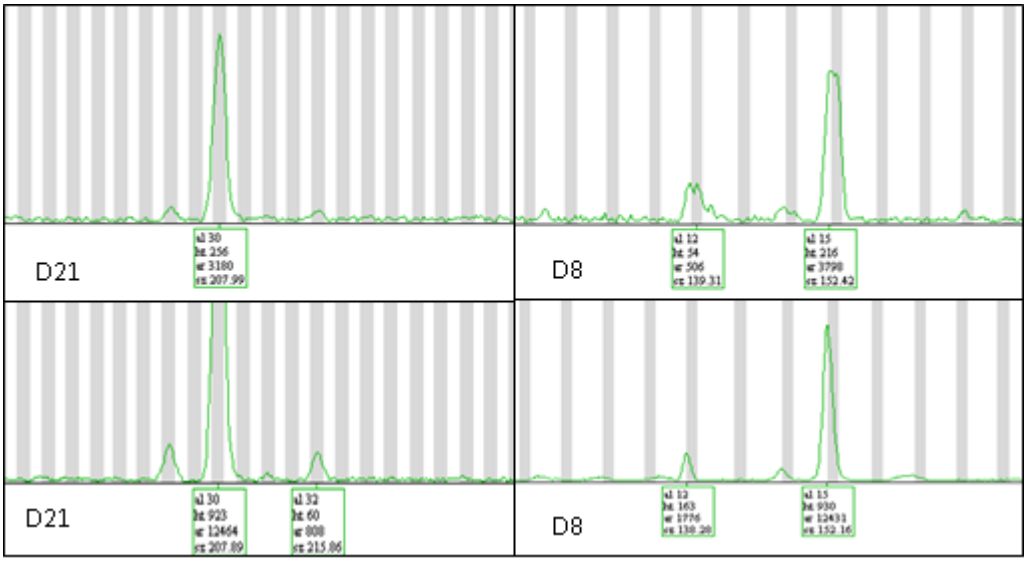


Figure 3.14 Images showing EPG of imbalanced alleles at D21 (left) and D8 (right) for both standard (top) and increased (below) injection times during CE. Due to the level of preferential amplification during PCR, the imbalance between allele peaks is extreme with dropout occurring under a standard injection time (D21). With the increase in injection time however, the smaller allele is detected.

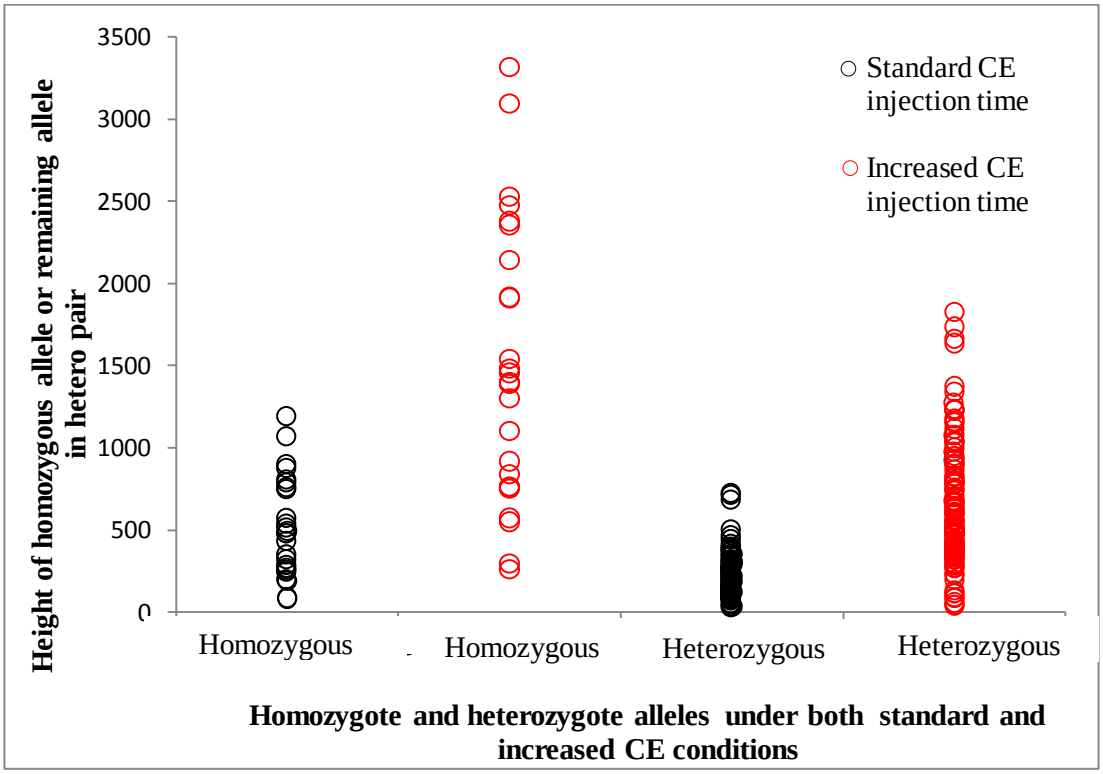


Figure 3.15 Peak height (RFU) of both homozygous alleles and, in cases where allelic drop-out has occurred, the peak height of the remaining allele, under both standard and increased CE injection times.

3.4.9 Stutter peaks

The relative area of stutter peaks (SR) was assessed using the allelic peak areas (φ) (combining the 'n' and 'n+1' peaks where present) according to the following equation as described by Gill et al. (Gill et al., 2000a), SR = stutter peak φ / allele φ ('n') + φ ('n+1'). Stutter peaks (-4 bp) were observed for 27% of the 272 alleles observed using a standard CE injection time (Figure 3.16), which increased to 65% (of 281 alleles) when the increased CE injection time was applied (Figure 3.17). The increase in injection time did not result in a significant increase in the level of stutter observed (P = 0.86). 64% of the stutter peaks observed using the standard injection times were seen to decrease in stutter ratio following the use of enhanced CE settings. 88% and 92% of stutter peaks had a stutter ratio value of ≤0.15 for standard and increased injection times respectively (P = 0.15), whilst the 95th percentile value for stutter ratio in each case was SR ≤0.20 and ≤0.19. As the peak area of the allele peak decreased, the stutter ratio was seen to increase as shown in Figure 3.18. The maximum stutter ratio observed where the true allele peak area < 10 000 RFU was at 0.33/0.32, and values at >10 000 RFU were 0.13/0.29, for standard/increased CE settings respectively. No stutter peaks greater in area than the true allele were observed in this dataset. Data for charts included in appendix 8.10 and statistics results are included in appendix 8.12.

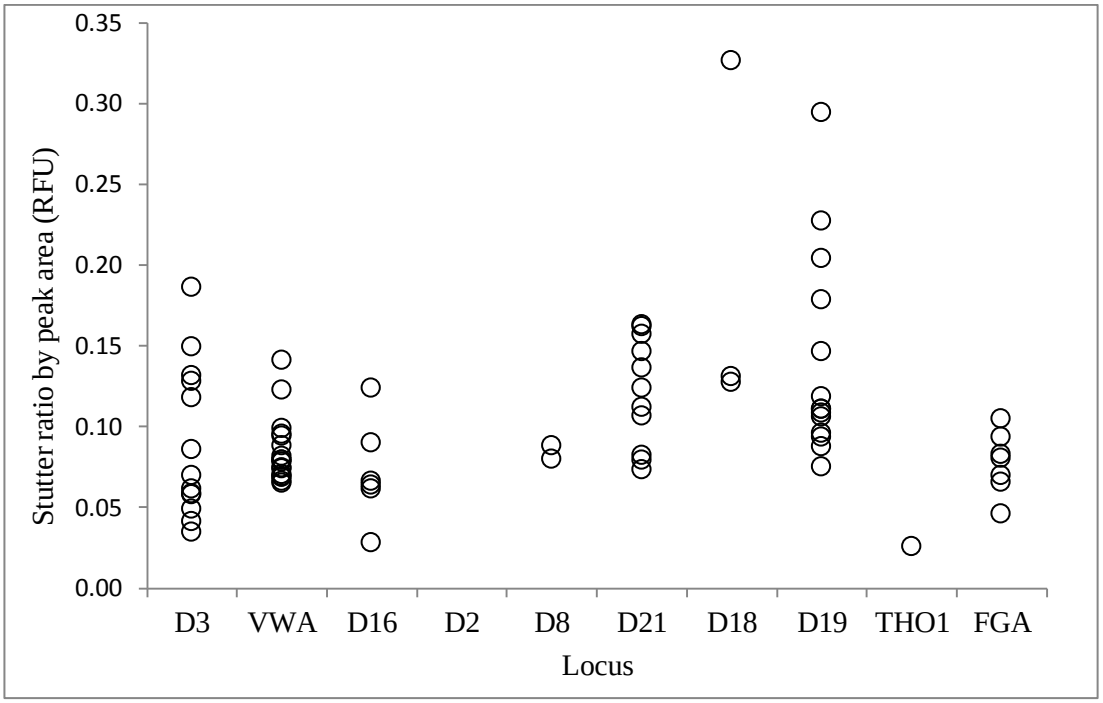


Figure 3.16 Stutter ratio by peak area for each STR locus using a standard CE injection time

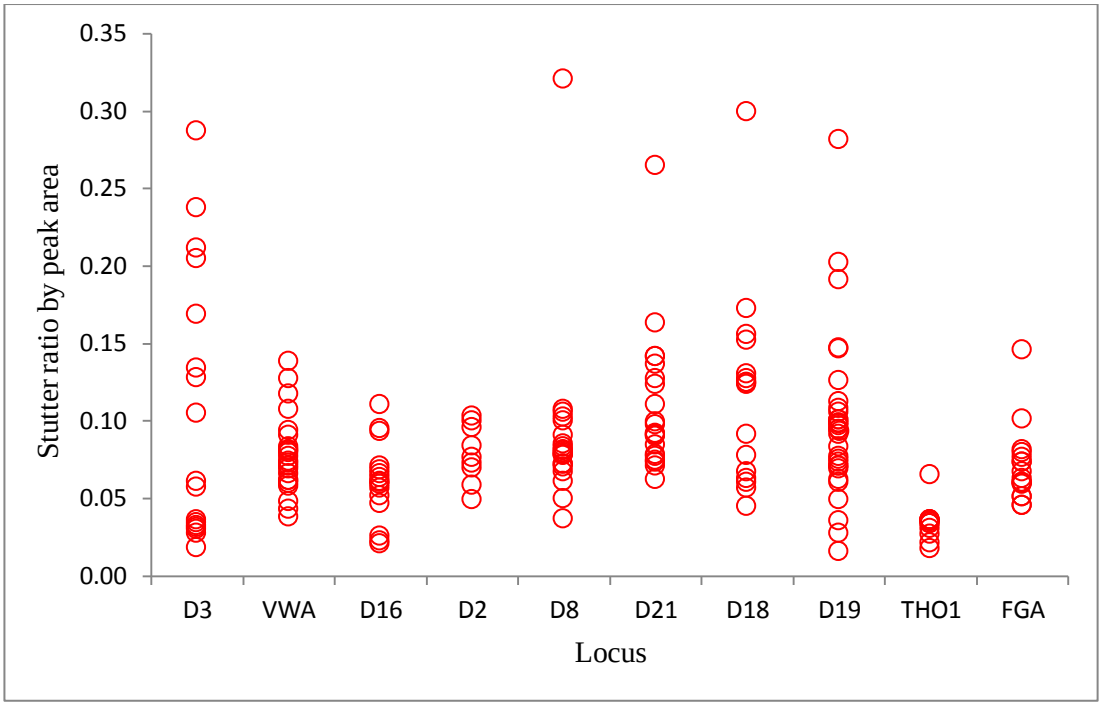


Figure 3.17 Stutter ratio by peak area for each STR locus using an increased CE injection time

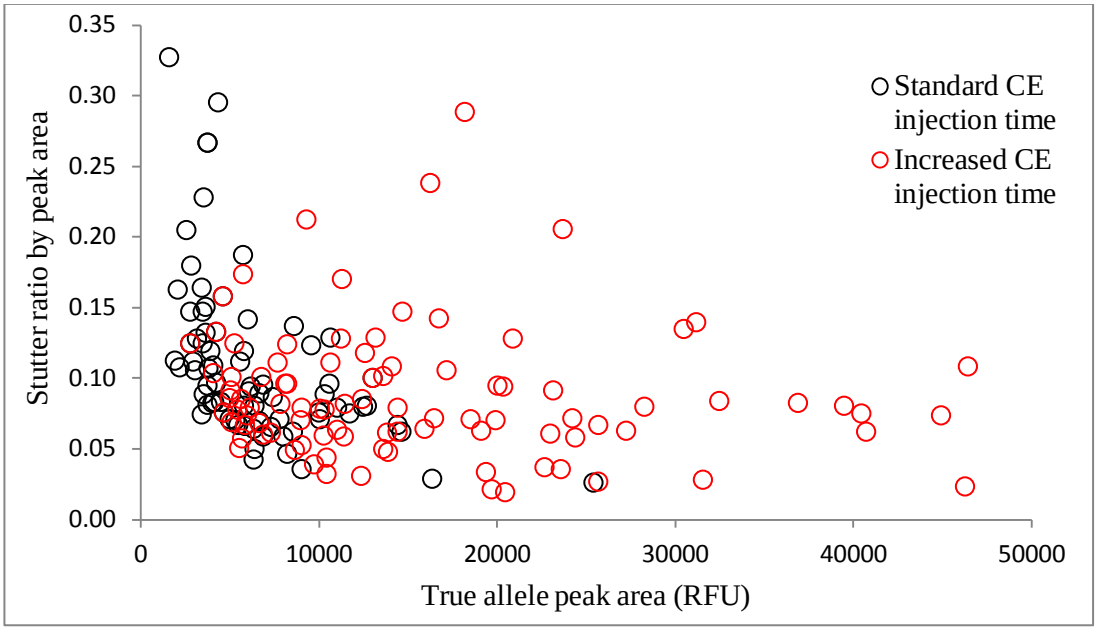


Figure 3.18 Stutter ratio by area relative to peak area of true allele for standard and increased CE injection times – outlier removed from THO1 with peak area of >70,000 RFU

3.4.10 Assay comparison

A comparison was carried out to assess the performance of three commercial kits; AmpFℓSTR® NGM PCR Amplification Kit (NGM), PowerPlex® ESX 16 System (ESX) and PowerPlex® ESI 16 System (ESI), and an 'in-house' developed assay for ltDNA was also evaluated.

Two samples were excluded from the data set; firstly one from the ESI data on the basis that more than two unexpected alleles were observed in the DNA profile thereby indicating the presence of contamination, and secondly from the ESX dataset as a full profile was obtained from one sample, the peak areas for which were at least double that of all other samples, thus suggesting that a higher starting quantity of DNA was present in the PCR tube. No amplification was observed for one single-cell using the ESX assay and for two cells using all other assays.

The average result for the number of expected alleles amplified per cell was taken (excluding those results showing no amplification) with the following results: ESX 75%, ltDNA assay 51%, NGM 49%, SGM Plus 46% and ESI 28% (Figure 3.19). Using the ESX assay, over 85% of the expected alleles were obtained from each of four single-cells, an example of the EPG for which is shown in Figure 3.20.

The random match probabilities (RMP), that is the frequencies with which the generated DNA profile would be expected to occur in the population, were calculated using the online European STRbASE (Parson, 2003, Gill et al., 2003, Welch et al., 2012) with the results summarised in Table 3-4. Both the minimum and average RMP generated from the single-cell results for all assays exceeded the 1 in a billion number which is used as the ceiling figure in forensic casework. The ltDNA assay was not included at this point as the database used does not currently hold the allele frequencies for locus Penta E. Data for charts included in appendix 8.11.

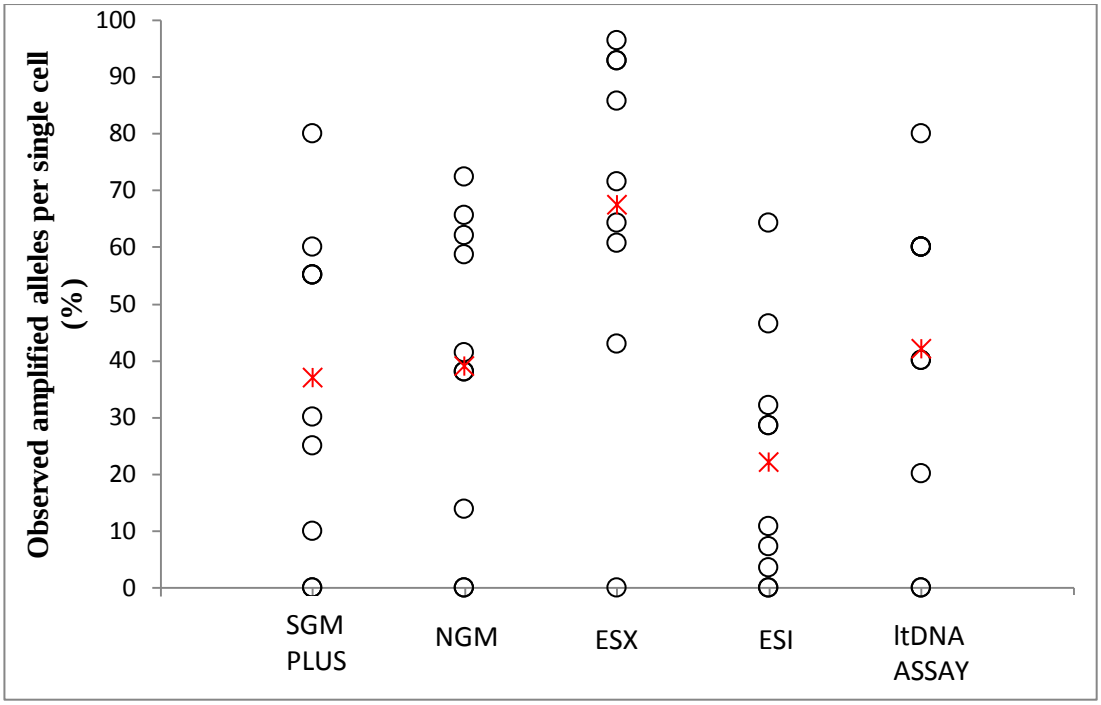


Figure 3.19 Percentage of alleles amplified from each individual single-cell tested as a percentage of the total number of alleles possible for the kit utilised. The mean value for each kit is represented by the red coloured marker with the values as follows: SGM 37%, NGM 39%, ESX 67%, ESI 22% and the ltDNA assay 42%.

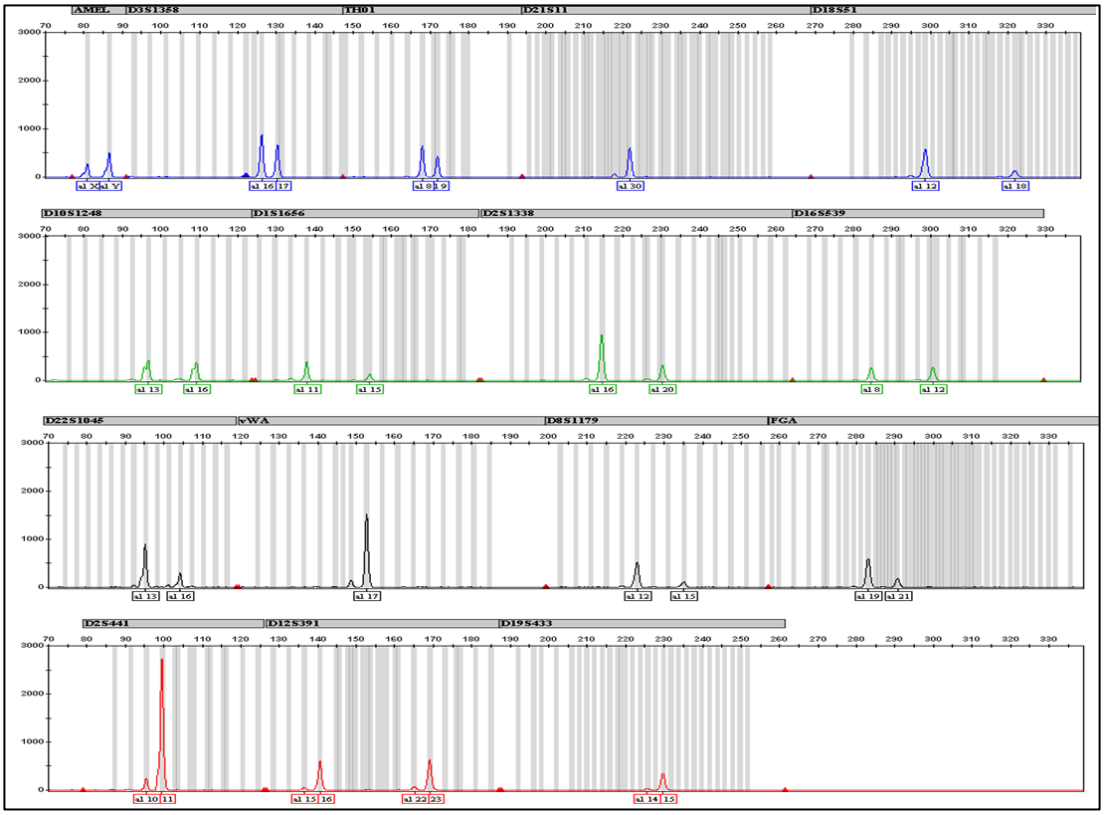


Figure 3.20 Example EPG of partial profile generated from a single-cell using ESX-16

Table 3-4 Average and minimum random match probabilities (RMP) for assays as calculated using STRbASE (Parson, 2003)

Assay	Average RMP	Minimum RMP
SGM Plus	2.81E-15	2.81E-16
NGM	4.95E-17	5.18E-18
ESI 16	2.64E-13	2.64E-14
ESX 16	1.01E-21	1.34E-22

3.5 Discussion

In order to evaluate the efficacy of individual single-cells as a source for forensic DNA analysis, a method by which each cell could be recovered from a surface and transferred into a PCR tube was developed. In the absence of an automated system such as laser microdissection (Di Martino et al., 2004a, Di Martino et al., 2004b), it was found most effective to stain the cells *in situ* using haematoxylin and eosin, and then under a high-powered microscope at x100 magnification to manually recover the individual cell using ultra-fine micro needles. The length of the needle allowed for its manoeuvring under the objective, whilst a tip width of just 0.25 mm facilitated the precision manipulation of each single-cell, in addition to which each disposable needle arrives pre-sterilised using ethylene oxide and so can be used with minimal risk of DNA contamination. Once recovered from the surface of interest, the cell was transferred to a UV-treated square of adhesive tape approximately 2 x 2 mm, for examination under the microscope at x200 magnification in order to confirm both the transfer of the cell from the needle point and also to ensure that only one cell was present to make certain that any DNA amplification observed could be attributed to just one cell. It was determined that placing the tape/cell into 1 µl of sticky stuff remover (SSR) that had been pre-pipetted into the PCR tube, was able to facilitate release of the single-cell from the adhesive during amplification, without causing inhibition during PCR.

It has been demonstrated previously that reducing the reaction volume can increase the PCR sensitivity thereby generating in a better quality DNA profile. The results of this study suggest that 15 µl is the most effective reaction volume for single-cell DNA amplification, contradicting previous research which has suggested that use of a smaller volume is more effective (Gaines et al., 2002). This is most likely due to the requirement to include SSR in the PCR tube in order to ensure that the cell is effectively released from the adhesive tape; in the 10 µl and 5 µl reactions the concentration of SSR may have inhibited the PCR thus resulting in the amplification of relatively fewer of the expected alleles. The use of a volume of 15 µl also allowed for the ForensicGEM extraction reagents to be utilised, occupying 6 µl of the total volume, with the SGM Plus amplification reagents making up the remaining 9 µl. Discovered in the 1980's by Morgan and Daniel (Bergquist et al., 1987) the neutral proteinase, derived from *Bacillus sp.* is included in the ForensicGEM kit, facilitates cell lysis through the breakdown of proteins into soluble peptides, therefore making the DNA template more accessible to the amplification reagents during PCR. An increase of approximately 20% was observed in the average number of expected alleles amplified when ForensicGEM was used as compared with single-cells amplified without the use of the extraction enzyme. As discussed in Section 3.1, the use of direct PCR decreases the number

of transfer steps necessary during extraction, thus reducing the risk of contamination and, particularly pertinent for single-cell work, the potential for sample loss.

Across all forty-nine generated DNA profiles, a total of 314 alleles were observed under a standard CE injection setting, and 324 alleles using the increased injection time.

The frequency with which alleles were observed at each locus was shown to correlate with the molecular weight; as the molecular weight of the loci increased, the number of times an allele was observed at a particular locus was shown to decrease. This can be explained through preferential amplification, the effects of which are known to be exacerbated in ltDNA samples. With the exception of vWA as the only homozygous locus, Amelogenin, D19 and D3 were found to amplify most frequently observed in over 40% of the DNA profiles, whilst D2 was present in just 11%.

As the size of the amplicon fragment increased, a decline in the peak height could be observed. Although this trend is one of the hallmarks of degraded DNA, in this study all single-cells were prepared from a fresh buccal source and so is likely related to the reduced frequency of amplification observed of the LMW loci. The use of an increased injection time during CE resulted in an average increase in peak height (RFU) by a factor of three, which is expected given that the injection time was increased by a factor of three from the standard setting of, 3 kV for 5 s (15 kV), to the enhanced setting of 3 kV for 15 s (45 kV).

It has been demonstrated that as the area of peak decreases below 10,000 RFU, the heterozygous balance (Hb) at each locus is subject to more variation (Whitaker et al., 2001). As 95% of all amplified alleles had a peak area <10,000 RFU when using the standard CE injection time, the results of this study reflect this, with no correlation observed between the average peak area and Hb value; this is also observed when the peak height remains at <10,000 RFU following an increased CE injection time. Further analysis of those loci for which Hb < 0.2 found that in all bar one of these cases an allele that had previously been recorded as dropped out under standard injection times was detected with the increased injection time, as the previously present peak area was also enhanced by the additional injection time, the difference between the two allele peaks resulted in severe imbalance at the locus (see Section 3.1). As the peak height exceeds 10,000 RFU, there was an increase of over 10% in the number of loci with Hb> 0.6 and almost 15% in those with Hb> 0.5, relative to those with a peak area <10,000 RFU.

In this study, using standard CE injection settings, it was found that 27% of observed alleles had an associated stutter peak present, which is consistent with the observations of Whitaker et al. (Whitaker et al., 2001) of 30%. The stutter ratios calculated for both standard and

enhanced CE injection times showed that 88% and 92% of stutter peaks had a stutter ratio value of ≤0.15, which is often the standard figure used for interpretation (Gill et al., 2000a), whilst the 95th percentile value for stutter ratio in each case was SR ≤0.20 and ≤0.19. The maximum stutter ratio observed for this dataset was 0.33, which falls within figures observed by both Whitaker et al. at 0.40 (Whitaker et al., 2001), and Petricevic et al. at 0.57 (Petricevic et al., 2010) as part of their validation studies. Also consistent with these studies is the observation that with a decrease in allele peak area, the stutter ratio was seen to increase. This is demonstrated when the peak area is divided; the maximum stutter ratio observed where the true allele peak area <10 000 RFU was at 0.33/0.32, and values at >10 000 RFU were 0.13/0.29, for standard/increased CE settings respectively.

Locus drop-out was observed at just under half of the possible loci across all samples. This is markedly higher than results published using low concentrations of extracted DNA, for which less than 8% locus dropout was recorded (Petricevic et al., 2010, Whitaker et al., 2001). This is possibly due to the higher quantities of DNA utilised in these papers and included in their calculations, for example 6.25/12.5 pg (Petricevic et al., 2010) and 12.5/25 pg (Whitaker et al., 2001) thus reducing the frequency of locus drop-out observed. The locus exhibiting the highest frequency of drop-out was D2, which was also observed by Petricevic et al. (Petricevic et al., 2010) and Whitaker et al. (Whitaker et al., 2001), who also noted that the increased injection time had minimal impact on the proportion of locus dropout. As the SGM Plus assay is a multiplex PCR kit, it is probable that the efficiencies with which the different STR loci amplify will vary and may account for the relatively higher amount of drop-out exhibited by locus D2 (Petricevic et al., 2010).

Allelic drop-out was observed in 31% and 29% of all heterozygous loci for the standard and increased CE injection times respectively. No bias with regard to the drop-out of either the LMW or HMW alleles was observed using standard CE settings, whilst a slight tendency towards drop-out of the LMW alleles was noted using the extended injection time. These results concur with those of Whitaker et al. (Whitaker et al., 2001) who observed a slight bias towards LMW drop-out and Petricevic et al. (Petricevic et al., 2010) who found a slight tendency towards HMW drop-out but noted no significant difference from the Whitaker data set.

The use of an extended CE injection time resulted in a 2% decrease in the proportion of allelic drop-out observed, however the additional alleles that were designated (D2, D8, D21, and D18) resulted in peak imbalance with values of Hb < 0.15 and Hb >15. The usefulness of these additional alleles is therefore limited, as in the absence of prior DNA profile

knowledge such a severe imbalance could be interpreted as the presence of a low level mixture.

In order to determine an approximate value for the ltDNA threshold as defined by Gill et al. (Gill et al., 2009) the peak heights of the remaining allele at all heterozygous loci for which the second allele fell below the lower limit of detection (25 RFU), were analysed and determined to be 750 RFU and 1850 RFU for standard and increased injection times respectively. If the peak height of a single allele at a locus exceeds this value then the peak can be designated as homozygous. Using these thresholds, 28% of the alleles amplified for vWA using the standard injection, and 32% using the increased injection, would have been designated as homozygous without prior knowledge of the expected result.

Both the ESX and ESI assays have been shown to perform with equivalent success during validation tests with approximately 27% of expected alleles amplified using 7.8 pg extracted DNA template (Tucker et al., 2012a, Tucker et al., 2011). The results for the ESI assay with an average of 28% of the expected alleles amplified from a single-cell in this study are consistent with those observed using the extracted DNA, however the ESX results far exceeded the validation results with an average of 75% of alleles observed. Given that these samples were prepared at the same time and using the same source, the discrepancy between the two kits cannot be accounted for, unless perhaps the ForensicGEM reagents provided more favourable conditions for the ESX chemistry than the ESI assay chemistry.

It has been demonstrated through direct comparison that the more recently developed ESI amplification kit is more sensitive than the SGM Plus (Tucker et al., 2012b), which at first glance would appear inconsistent with the results of this study with averages of 28% and 46% expected alleles amplified for the two kits respectively. However, aside from the additional number of loci targeted by the ESI assay, as part of this study the SGM Plus kit has been optimised for use in terms of the PCR volume utilised, use of ForensicGEM extraction reagents and inhibition tests, which the ESI kit has not been subjected to. This in itself demonstrates the value of optimising even a commercially available kit when it is to be used for a specific purpose; this may also explain why the NGM kit performed little better than its predecessor.

A full DNA profile, which was generated using the ESX amplification kit, was excluded from analysis as, whilst almost complete DNA profiles (>85%) were observed in four other samples, the peak area (RFU) was at least double that observed on all other EPGs, indicating the presence of a higher quantity of starting DNA material. As all samples were prepared under the microscope using stained cells and cell transfer checked, it is possible that more than one cell was present, for example if cells were directly situated on top of one another,

and so multiple transfer was not detected, or the cell transferred may have been at the end stage of mitosis, prior to cytokinesis, therefore containing twice the DNA template. The capability to generate a full profile from such a small starting quantity of DNA demonstrates the impressive sensitivity of this assay.

The ‘in-house’ ltDNA assay was developed using just three STR markers and Amelogenin to determine if reducing the targets for amplification would improve the reliability of PCR for such a low level of template DNA. Whilst the results produced were comparable with those of the commercial assays (excluding ESX), the reduced number of loci did not improve the reliability of the assay thus, given the loss of discrimination, does not carry any advantage over assays that contain more, HMW loci.

3.6 Conclusions

The described method for the recovery of individual single-cells provides an effective and efficient means by which a cell can be detected and transferred from its surface to a PCR tube without sample loss. The use of the ForensicGEM extraction reagents facilitates cell lysis prior to amplification and allows for the use of a one-tube step with SGM Plus reagents added directly to the tube for PCR.

Overall, the results for the amplification of DNA from individual single-cells demonstrate that the frequency and intensity of artifacts resulting from the stochastic effects of PCR are comparable with those observed from using extracted ltDNA samples and can, therefore, be interpreted using the same parameters. In the absence of a consensus profile, the use of an increased injection time during capillary electrophoresis can aid interpretation of peaks and artifacts that are around the minimum peak amplitude threshold. Although the use of an increased injection can help in some instances such as resulting in a reduction in the level of locus and allele dropout observed; this was not found to be significant with all P values exceeding 0.1.

The use of single-cells as a DNA source could be of crucial importance in cases where very little biological evidence is identified, providing a source of intelligence for investigators, or as an alternative technique to identify a single contributor in cases where a complex mixtures is obtained.

Following initial results, Promega's ESX-16 amplification assay, used in conjunction with ForensicGEM extraction reagents, appears to be particularly effective for amplification of DNA from individual single-cells.

4. Cell transfer via textile fibres

4.1 Introduction

The concept that the transfer of textile fibres between two (or more) surfaces as possible means to link the victim(s), suspect(s) and scene(s) involved in a crime, has been described in literature for almost a century (Locard, 1920). Through mapping the fibre transfers that have occurred during a criminal incident, movement and interaction between individuals and scenes can be established, which can help investigators by supporting or refuting the version of events given by those involved.

Despite the demonstrated utility of fibre evidence in casework, there is still a general perception that analysis is time consuming, costly and without much return in terms of evidentiary value, often leading to fibre evidence being under-utilised by investigators. Extensive research across the field, including the use of block colour (Grieve et al., 2005, Palmer et al., 2009, Biermann, 2007), population (Roux and Margot, 1997, Palmer and Oliver, 2004, Watt et al., 2005, Grieve et al., 2003), target (Cook et al., 1997, Jones and Coyle, 2011, Kelly and Griffin, 1998, Wiggins et al., 2004), transfer and persistence studies (Akulova et al., 2002, Grieve and Biermann, 1997b, Palmer and Polwarth, 2011, Roux et al., 1999, Salter and Cook, 1996, Kidd and Robertson, 1982, Palmer and Burch, 2009), is helping to demonstrate and reinforce the value and reliability of fibre evidence for both intelligence purposes and for use in court; dismissing, for example, the notion that two fibres of the same type and colour cannot be distinguished in the majority of cases.

A landmark three-part study in 1975 by Pounds & Small, although preliminary, is the first detailed account of the factors influencing the transfer and persistence of textile fibres (Pounds and Smalldon, 1975c, Pounds and Smalldon, 1975a). They proposed that the textile fibres are transferred to a second surface through three distinct mechanisms; fibres that are loose on the outside surface of the garment, those that are drawn out of the fabric as a result of friction, and those that are fragmented during the contact (Pounds and Smalldon, 1975b). It has been demonstrated that the transfer of fibres between two surfaces is dependent upon a number of factors such as the nature of both the donor and the recipient surfaces, the length of contact time and pressure applied during the event (Pounds and Smalldon, 1975a, Kidd and Robertson, 1982, Grieve and Biermann, 1997b, Roux et al., 1999, Palmer and Burch, 2009). Following a transfer event however, it has been shown that the percentage of fibres lost from the recipient surface is consistent irrespective of the surface type, displaying an exponential decay with increasing time (Pounds and Smalldon, 1975c, Akulova et al., 2002, Palmer and Polwarth, 2011, Roux et al., 1999). Understanding the characteristics of transfer and persistence for fibre evidence facilitates the forensic scientist in forming their expectations regarding the likelihood of evidence recovery given a set of circumstances.

Several techniques are available for the recovery of textile fibres and other debris from the recipient surface to which they have been transferred, including the use of forceps (Schotman, 2015), shaking/scraping techniques (Pounds, 1975, Lowrie and Jackson, 1991) and polystyrene rods (Keutenius et al., 2013). Whilst these techniques have been shown to be effective methods for the recovery of fibres in some circumstances, they can be time consuming, may present a contamination risk (Roux et al., 2001) and additional information on the location of fibre populations over the surface of the item is lost.

Care must be taken during recovery as fibre contamination can occur between items, through direct contact or indirectly via the examiner, as well as from unrelated background fibres in the examination area. Roux et al. (Roux et al., 2001) suggested the use of separate examination rooms where possible for related items, the use of a disposable covering over the benchtop during examination and that a disposable lab coat be worn by the examiner, in order to minimise the risk of contamination.

Tape lifting has been the method of choice for fibre recovery in many laboratories (Wiggins, 2001) since it was first proposed by Frei-Sulzer in 1951 (Keutenius et al., 2013). Following the recovery of transferred fibres from an item, an initial examination of the fibre tapes is conducted under a low-powered microscope. If no source material has been identified, then fibre populations of possible interest are searched for and used to inform the investigation. If a potential source material has been identified, then tapes are searched for target fibres that are possible matches based on similar colour and morphology. Following detection, fibres of interest are removed from the fibre tape and mounted under a coverslip using a medium such as Phytohistol (Grieve and Deck, 1995) or Entellen (Wiggins and Drummond, 2007) on a glass microscope slide for further examination.

Comparison microscopy allows for the observation of two fibres immediately adjacent to one another in the same field of view. This enables a direct comparison of colour and morphological detail under white, polarised and fluorescent light sources. The use of comparison microscopy is in itself a highly discriminating tool for fibre examination and can be utilised to quickly exclude fibres from the investigation (Palmer et al., 2009). Fibres that cannot be distinguished from one another using microscopy are then analysed using microspectrophotometry (MSP), which provides the forensic analyst with the ability to objectively discriminate colour between two fibres based on their absorption spectra (Robertson and Grieve, 1999). It has been demonstrated that by extending the analysis into the UV range, further discrimination between fibres is possible, for example, the discriminatory power for non-denim blue cotton fibres was found to increase by 7% (Palmer

et al., 2009) and 9% (De Wael, 2015) to a value of 0.98 following additional analysis in the UV range.

The recovery of DNA from fibre tapes is usually attempted only if visible cellular material such as skin flakes are present and a specific connection is sought or, most commonly, in cold casework where the available evidence is limited. The chemical structure of differing fibre types is expected to affect the strength of the interaction between the fibre and carbohydrates on the cellular membrane, or directly with cell-free DNA, depending on their functional groups (Seah et al., 2004). For example, the hydroxyl groups on the repeating cellulose units of cotton will form strong hydrogen bonds that enable a relatively stable binding with DNA/cell membrane, whereas weaker dipole-dipole interactions are formed with polyester carbonyl and cyano groups (Linacre et al., 2010, Mulligan et al., 2011). It is has been suggested that the varying strengths of these interactions may account for the observed ability to amplify DNA from some fabrics such as cotton and nylon with more success than from others such as polyester (Seah et al., 2004, Linacre et al., 2010). The results for one study did not show any significant difference between cotton and polyester (Mulligan et al., 2011), however this study differed from the previous two in that the focus was on fibres as a method for DNA retrieval, i.e. for swabbing. Sethi et al. (Sethi, 2013) used cotton and polyester fabrics to carry out simulated 'grabbing' incidents, but were unable to amplify DNA from the 'assailant' due to the low quantities of DNA (< 18 pg/µl) recovered.

The current methods used in casework to obtain a transferred DNA profile involve either the cutting or swabbing of fibre tapes for DNA extraction (May, 2009). Having tested methods for the recovery of DNA from fibre tapes using five different swabbing solvents and a total of seven extraction methods, May and Thomson (May, 2009) found that the use of xylene to swab the fibre tape followed by a Chelex™ DNA extraction was the most effective method in recovering DNA. Although it has been demonstrated that DNA can be successfully recovered in this manner, these methods do have two significant disadvantages; the potential for the loss of fibre evidence, and that it is likely that a DNA mixture will be obtained, with the donor's profile at risk of being completely obscured in a very high wearer DNA background. The loss of both of these evidence types will be detrimental to the investigation.

4.2 Aims and objectives

It is hypothesised that when textile fibres are transferred from a donor to a recipient, the fibres themselves may be acting as a vector for the transfer of nucleated cells from the donor garment and could therefore be targeted to obtain a DNA profile. If possible, this would provide a more refined technique for the recovery of DNA from fibre tapes, increasing the successful detection and amplification of the donor's DNA, even in the presence of a high level of background recipient (wearer) DNA. The aim of this chapter is to test this hypothesis, whilst ensuring that all methods have no detrimental effect on either the subsequent fibre analysis or DNA processing, thus allowing the two evidence types to be investigated alongside one another.

The objectives of this study are to first demonstrate that nucleated cells can be seen *in situ* on the surface of textile fibres and that once located, can be recovered for single-cell DNA analysis utilising the methods developed in the previous chapter. Any techniques used for visualisation of the nucleated cells will require testing for compatibility with downstream DNA and fibre analysis processes. The extent to which nucleated cells are present on the outside surface of clothing items will then be investigated in order to determine the potential for transfer via the fibre surface. Finally, fibre transfer experiments will be used to demonstrate the transfer of nucleated cells on the surface of textile fibres.

4.3 Method development and results

This project had been conducted with ethical approval from the Northumbria University Ethics committee under reference number RE25-10-121294. Volunteers were provided with information sheets and requested to sign consent forms prior to participation in the study; copies of which are included in appendix 8.12.

The methods and results sections for this chapter are presented together in order to effectively reflect and describe the development of the methods involved, with sections clearly marked as either 'experiment' or 'method development' as appropriate.

4.3.1 Visualisation of cells *in situ* on the fibre surface (method development)

In order to demonstrate the transfer of single-cells on the surface of textile fibres, a method by which the cells could be visualised *in situ* was first developed.

4.3.1.1 Testing of fluorescent dyes for staining of cells *in situ* on fibre tapes

For the initial proof of concept work, it was decided that a DAPI-based dye would be an ideal candidate stain, as it will fluoresce nuclear DNA without the requirement of additional fixing reagents, thus limiting manipulations and therefore potential loss of sample, as well as impact on downstream DNA/Fibre analyses. Two DAPI solutions were considered; DAPI dilactate is supplied as a powder which was added to MBG water to obtain a concentrated stock at 1 mg/ml (see Section 2.4.1.1), which was then diluted further in sterile DPBS prior to use, and NucBlue® ReadyProbes® Reagent which is a ready to use DAPI solution.

4.3.1.2 Determining optimum DAPI dilactate dilution for staining cells

Buccal cells were prepared as described in Section 2.4.2 and 2.4.3 and then a diluted sample was transferred to a microscope slide and allowed to air dry as outlined in Section 2.4.3.1. A 1 in 2 serial dilution of DAPI dilactate (hereafter referred to as DAPI) was carried out using sterile DPBS to prepare 1 in 2500, 1 in 5000, 1 in 10000 and 1 in 20000 solutions. Samples, including a negative MBG water control, were stained *in situ* on the microscope slide as described in Section 2.5.2.2. Cells were then examined under a Leica DM5000 B Fluorescent Microscope System using filter A (excitation filter BP 340-380, emission filter LP 425) and also using differential interference contract microscopy.

As shown in Figure 4.1, the nuclear DNA is visible at even the lowest DAPI concentration; however, it was noted that at this 1 in 20000 dilution there were cells present that had not stained as intensely. The 1 in 10000 DAPI dilution (Figure 4.1C) was determined as the most effective concentration to use for this study as the nuclear DNA was clearly visible in

all cells as compared with the DIC images, and this low concentration is less likely to effect the downstream fibre/DNA analyses then the higher concentrations (Figure 4.1A and B). No fluorescence was observed for any of the negative controls.

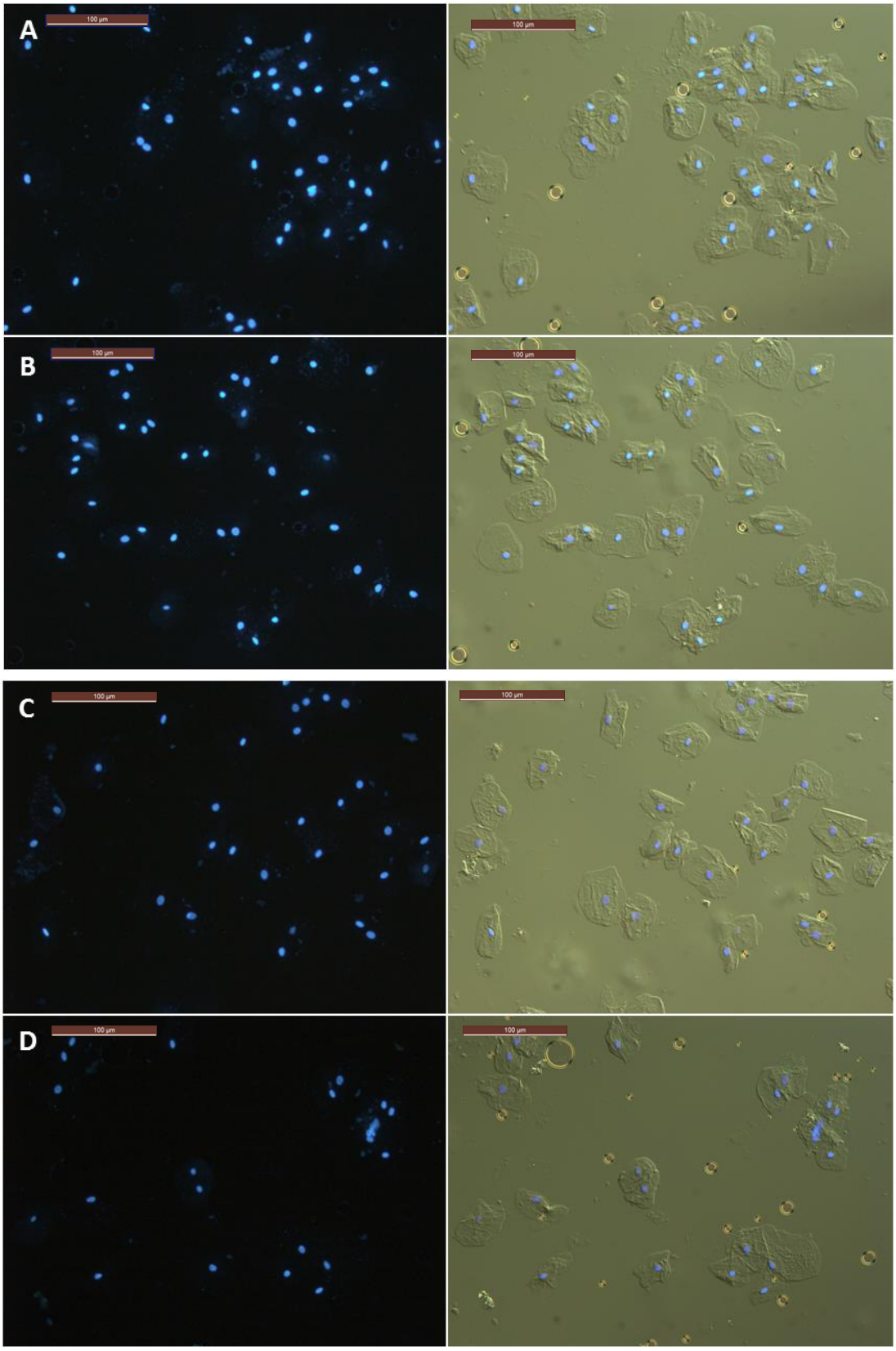


Figure 4.1 Images of buccal cells stained with a DAPI solution at 1 in 2500 (A) 1 in 5000 (B) 1 in 10000 (C) and 1 in 20000 (D) taken under the Leica DM5000 B fluorescent microscope at x200 magnification with filter A (left) and as overlaid with DIC images taken (right). All scale bars show 100 µm.

4.3.1.3 Testing DAPI and NucBlue® for PCR inhibition

A buccal swab was obtained from a volunteer, which was extracted using the QIAamp® DNA mini kit as described in Section 2.7.2.1 and then quantified using the NanoDrop® Spectrophotometer ND-1000 as outlined in Section 2.8.2. The DNA extract was diluted to a concentration of 0.1 ng/µl with MBG water and 5 µl was added to each of the required PCR tubes for a total DNA input of 0.5 ng, whilst 5 µl of MBG water was added to a tube for the PCR negative control. A 1 in 2 serial dilution of DAPI was prepared in sterile DPBS for testing of 1 in 1250, 1 in 2500, 1 in 5000, 1 in 10000 and 1 in 20000 DAPI solutions. 1 µl of sterile DPBS, NucBlue® or diluted DAPI was then added to PCR tubes in duplicate and 1 µl of water was added to both the positive and negative control samples. A mastermix of AmpFℓSTR® SGM Plus® reagents was then prepared as described in Section 2.9.1.2 and 9 µl added to each tube for a total reaction volume of 15 µl. Amplification was carried out using a Veriti® thermal cycler following the 28-cycle program outlined in Section 2.9.1.2.

The electropherograms (EPGs) for each reaction were examined for indications that inhibition had occurred during the PCR, such as locus or allele dropout and a reduction in the peak heights, particularly in the high molecular weight (HMW) loci (Opel et al., 2010).

The NucBlue® solution was tested as it is a ready-to-use, DAPI based solution, which can be stored at room temperature and can be directly applied to the sample. Complete dropout at locus D21S11 was observed in both duplicate reactions indicating specific inhibition at this locus as shown in Figure 4.2. The specificity of this inhibition suggests an interaction between the DAPI and the DNA sequence of this STR itself or its associated primers (Opel et al., 2010), because if it were as a result of interactions with the Taq polymerase directly or other components of the reaction mix, then indications of inhibition would be expected at more, or all loci in the multiplex. Inhibition at this locus was also observed for the most concentrated DAPI at a 1 in 1250 dilution. In this case peak heights were reduced but present, suggesting that there may be a component in the NucBlue® reagent that is exacerbating the interaction between the DAPI dye itself and the DNA sequence which is preventing PCR at this location.

Inhibition was only observed in the most concentrated DAPI solution tested (1 in 1250 dilution) where, aside from the aforementioned inhibition at D21S11, a reduction in the HMW alleles relative to the control sample was observed as shown in Figure 4.3. No indication of PCR inhibition was observed however for the 1 in 5000 dilution and as a 1 in 10000 solution will be used in this study, it has been demonstrated that the DAPI dye preparation is compatible with downstream DNA processing.

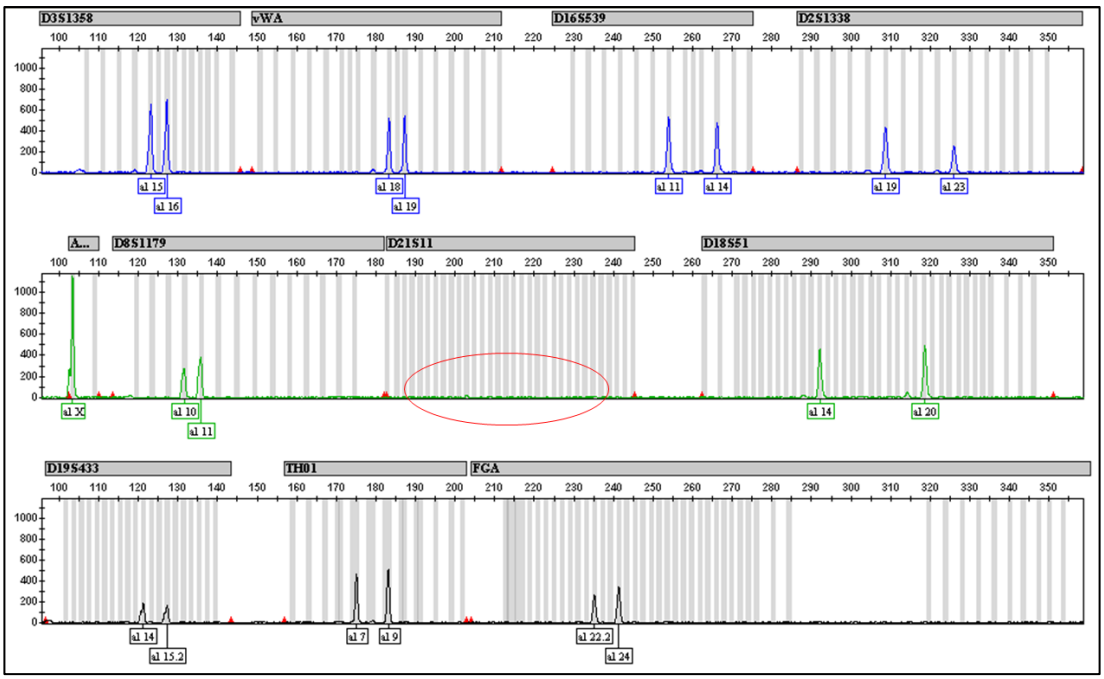


Figure 4.2 Example of EPG for NucBlue® solution showing complete locus dropout at D21S11, circled in red. Peaks are observed at all other loci, which are all clearly visible with good peak heights.

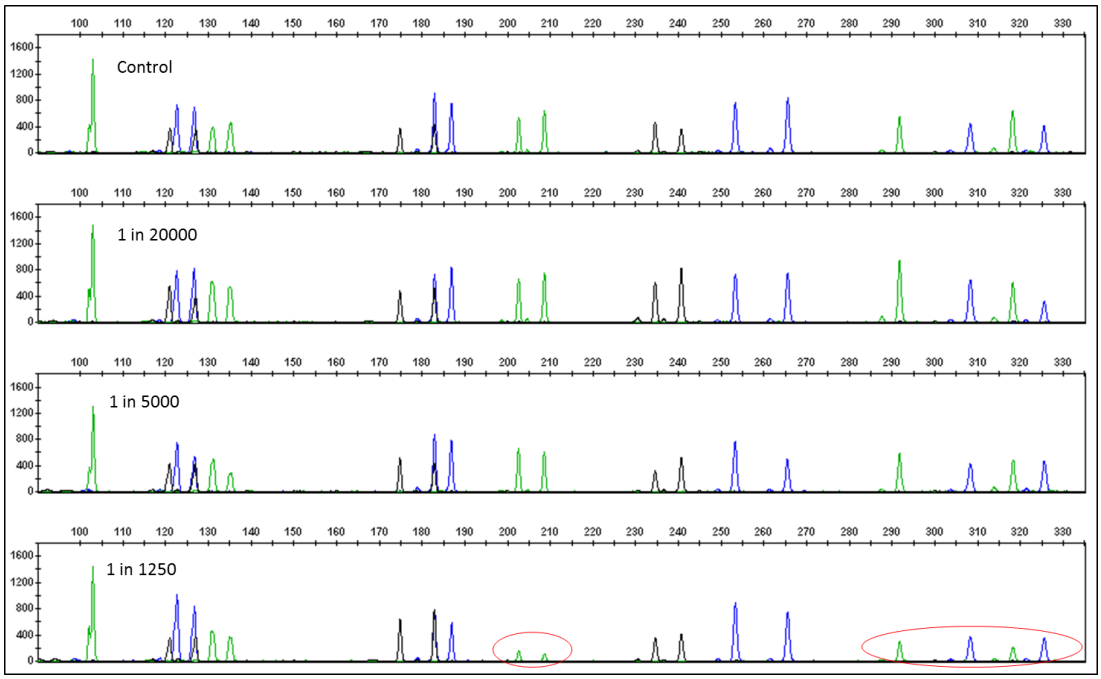


Figure 4.3 EPG examples for PCR inhibition tests with DAPI at 1 in 20000, 1 in 5000 and 1 in 1250 dilutions. A reduction in peak height relative to the control was observed at locus D21S11 and HMW loci D2S1338 and D18S51, shown circled in red.

4.3.1.4 Effect of DAPI staining on MSP

Ten fibres were soaked in PBS or a DAPI solution of 1 in 1000, 1 in 5000 or 1 in 10000, and mounted on a microscope slide using glycerol and a quartz coverslip as described in Section 2.14. In addition to this, ten unstained control fibres were also prepared in the same way for comparison of absorption spectra generated using the J & M TIDAS MSP 800™ (UV–vis) microspectrophotometer (MSP) in order to determine if staining the fibres with a DAPI solution would affect the downstream fibre MSP analysis. Three readings in the UV-vis range were taken along the length of each fibre, which were then averaged. The stained fibres were then compared with the unstained control fibre spectra as shown in Figure 4.4. As these spectra were taken using cotton fibres, some natural variation would be expected; it appears that the staining has not adversely affected the spectral profile of the fibre.

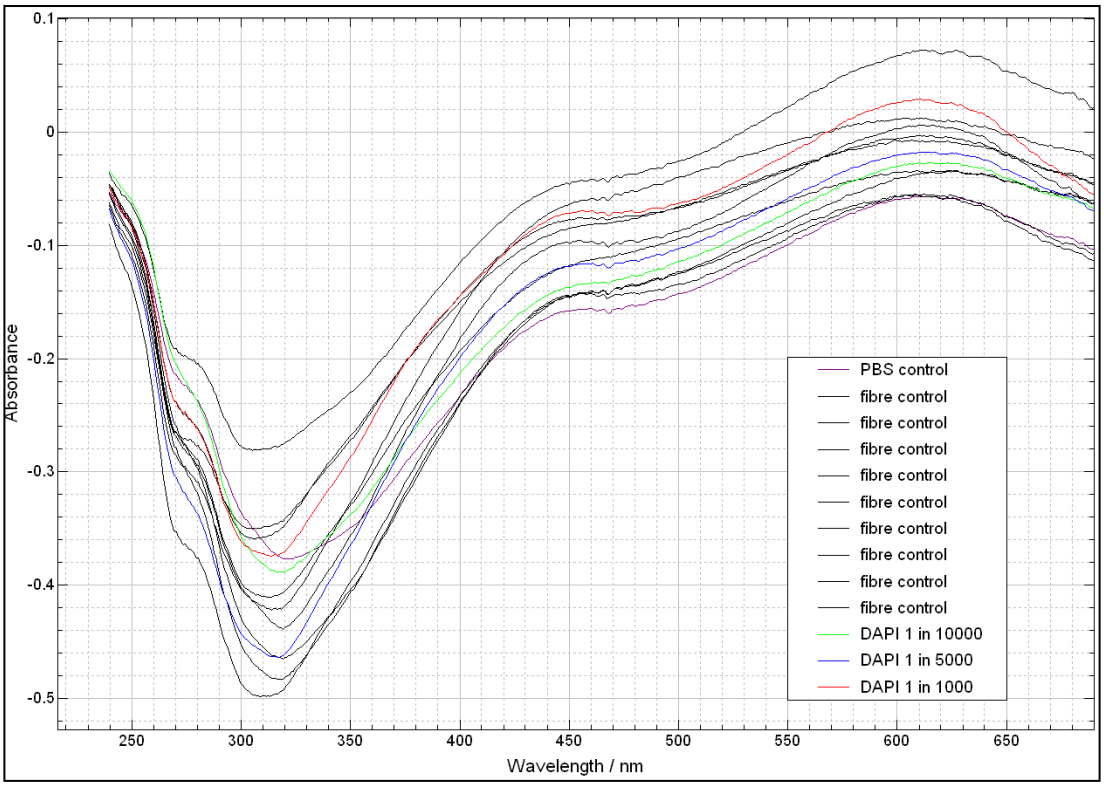


Figure 4.4 Spectra generated for fibres stained with DAPI at 1 in 1000 (red), 1 in 5000 (blue) and 1 in 10000 (green) dilutions, and control fibres (shown in black) using MSP

4.3.2 Demonstration of cells on outer surface available for transfer (experiment)

4.3.2.1 Demonstration of cells on outer surface available for transfer – materials and methods

Five volunteers were recruited and asked to wear an item of upper washed clothing, i.e. cardigan or jumper, of their own. Following at least six hours of wear, the garment was removed by the volunteer and placed into a bag, which was then sealed.

The outside front surface of each garment was then fibre taped in three sections, upper right front/right sleeve, upper left front/left sleeve and lower right/left using one strip of J-LAR tape per section, which was secured on an acetate sheet. The garment details were recorded as shown in Table 4-1.

Table 4-1 Summary of garment information for each of the five participants

Volunteer	Garment type	Fibre type
1	Orange/peach coloured wool style jumper	Unknown
2	Blue/green coloured zip up cardigan	100% acrylic
3	Blue coloured cardigan without fastenings	96% viscose, 4% elastane
4	Navy coloured cardigan with ¾ length sleeves	70% viscose, 30% nylon
5	Burgundy coloured long sleeve top	100% cotton

Fibre tapes were prepared and stained for examination as described in Section 2.4.4 and 2.5.2.2. The prepared microscope slides were then examined using an upright Leica DM5000 fluorescent microscope under x100 - x200 magnification. Each tape section was searched and the number of nucleated cells observed on each tape was recorded as the number of cell 'events', i.e. a single-cell or a cluster of cells equates to one event each, and the total number of cells in each event; an approximation was recorded for clusters in which not all cells could not be counted individually.

4.3.2.2 Demonstration of cells on outer surface available for transfer - results

In order to demonstrate the presence of cells on the outer garment surface and thus the availability of cells for transfer, the outside front surface of the tops of five volunteers were fibre taped. The tapes were stained with DAPI for examination under the high-power fluorescent microscope and the number of nucleated cells was counted. Both single-cells and cell clusters were observed on tapes taken from all volunteers, examples of which are shown in Figure 4.5 to Figure 4.7. The number of cell events and an approximate value for the total

cell count, i.e. taking into account all cells within a cluster, were recorded as shown in Figure 4.8. The number of cell events observed was between 672-2198 and the total number of cells within the range 1748-4410. The average number of cells observed per cell event for each volunteer is two, except for one volunteer with an average of four. The mode and median for all five volunteers is one, demonstrating that although clusters of cells increase the cell count, single-cells were observed more frequently.

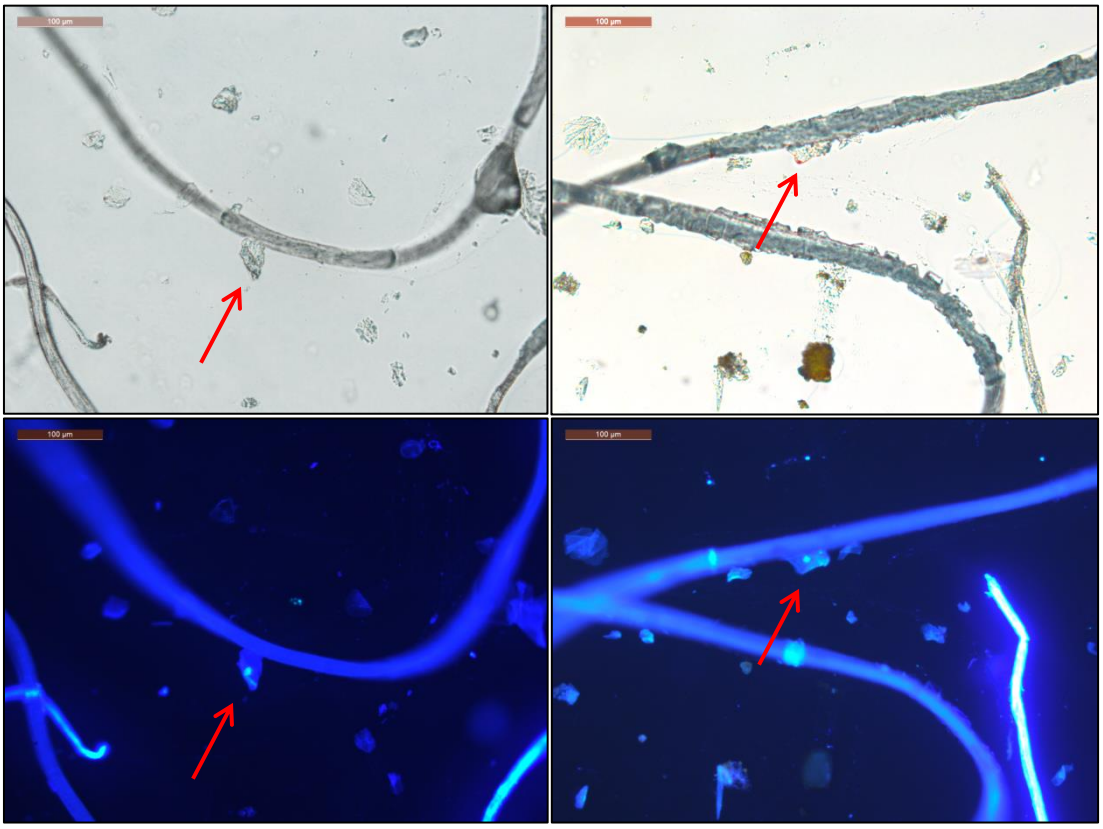


Figure 4.5 Examples of stained cells observed *in situ* on fibre tapes using a high-powered fluorescent microscope. All scale bars show 100 µm.

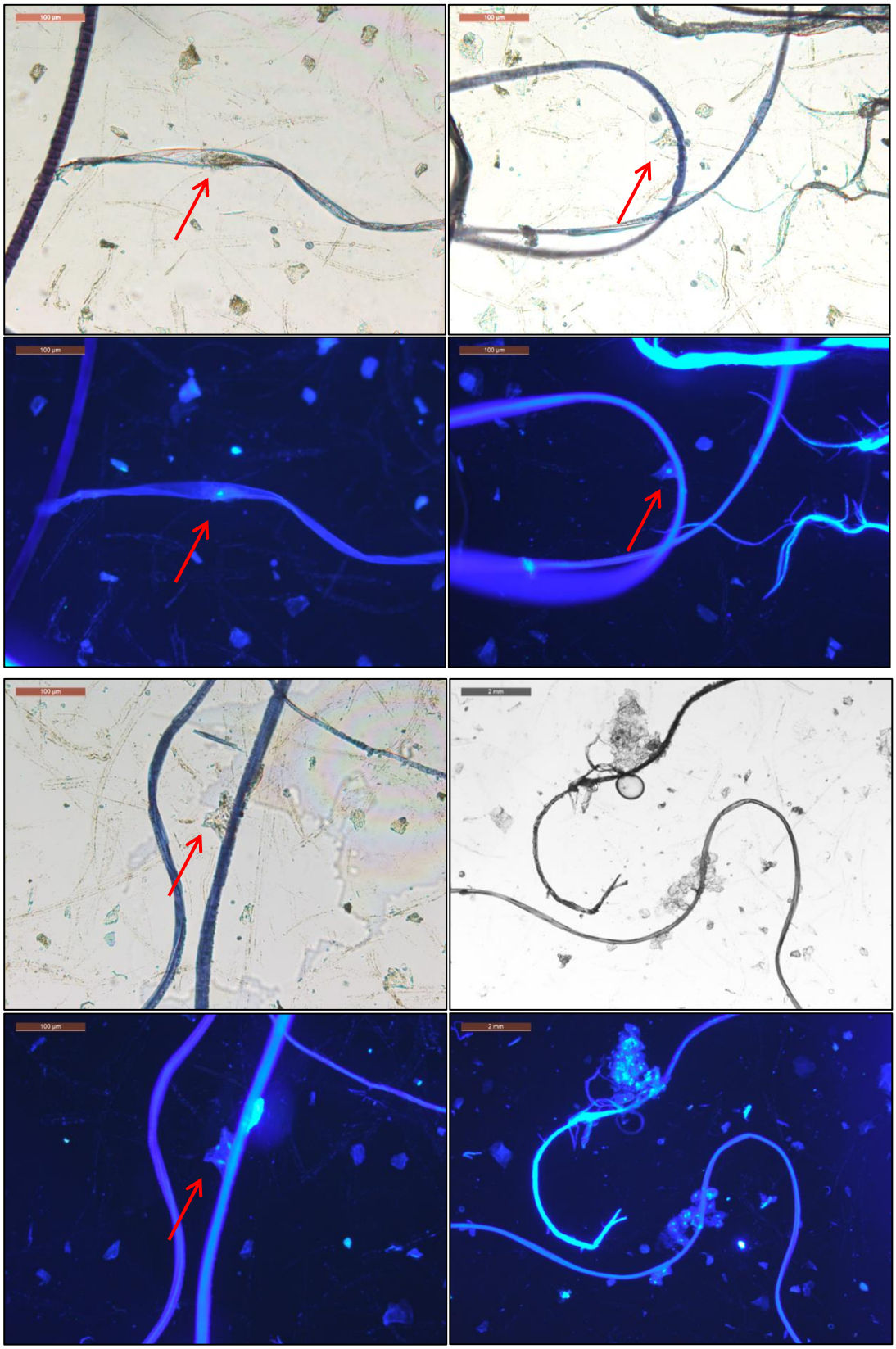


Figure 4.6 Examples of stained cells observed *in situ* on fibre tapes using a high-powered fluorescent microscope. All scale bars show 100 µm aside from the bottom two right hand panels for which the scale bar shows 2 mm.

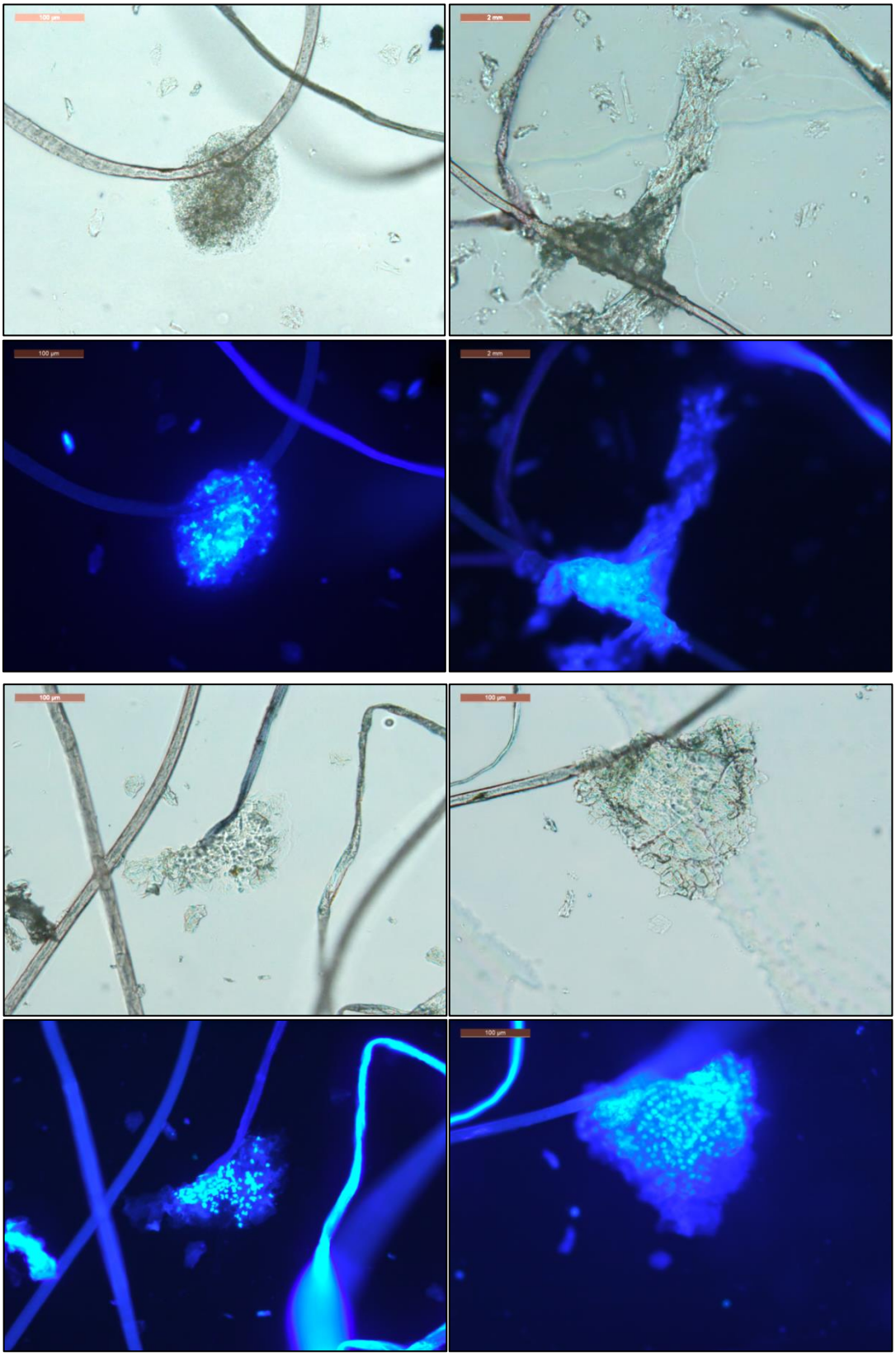


Figure 4.7 Examples of stained cells observed *in situ* on fibre tapes using a high-powered fluorescent microscope. All scale bars show 100 µm aside from the top two right hand panels for which the scale bar shows 2 mm.

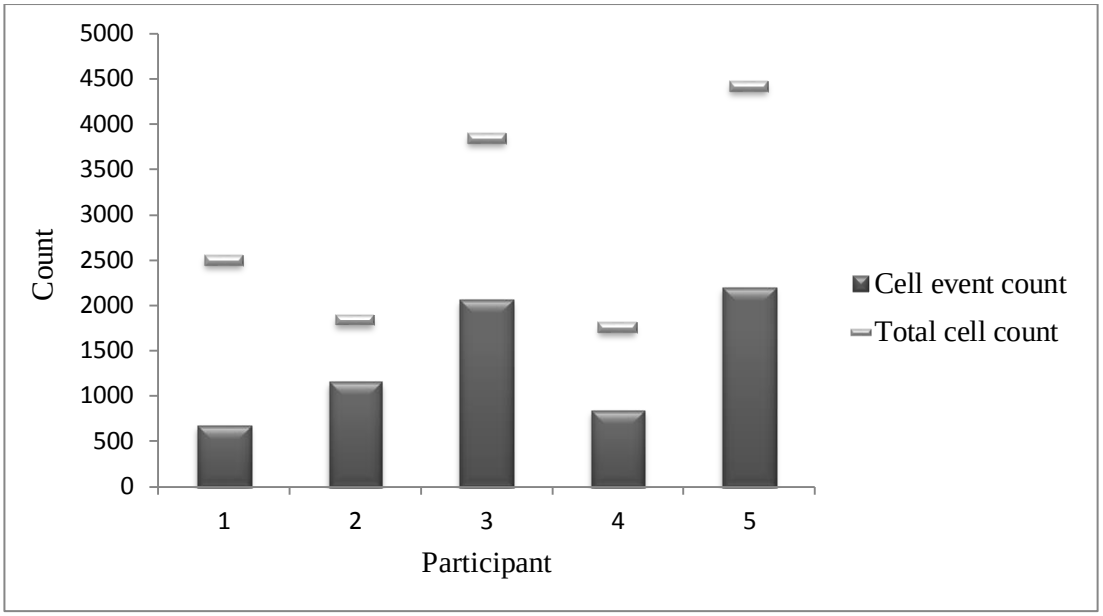


Figure 4.8 Bar chart showing both the number of cell events and the total cell count for each participant. The total cell count includes those clusters of cells that were frequently observed.

4.3.3 Demonstrating cell transfer – two-way transfer (experiment)

4.3.3.1 Demonstrating cell transfer – two-way transfer - materials and methods

Six, 100% cotton long sleeved-tops, three purple and three green, were purchased on separate days. The tops were grouped by colour and washed in separate machines, using a biological washing powder in an attempt to remove exogenous DNA from the surface. The tops were allowed to air dry in a DNA clean environment and then all outside surfaces were taped to remove foreign fibres and adhering cellular material, before being sealed in bags prior to use. Six participants were recruited and split into three pairs, each individual within the pair was provided with either a green or purple coloured top. Participants were requested to wear the provided top for a minimum of eight hours whilst carrying out their normal daily activities. Immediately prior to meeting each individual was asked to wear a hairnet, facemask, disposable gloves and over-sleeves to prevent the transfer of DNA through means other than contact between garments. The participants were then asked to hug for a two minute period, one minute with arms around the shoulders and one minute with arms around the waist, following which the tops were immediately placed in separate bags and were sealed until examination.

Fibre taping was carried out on separate, non-subsequent days for each of the tops and inside different examination hoods, which had been cleaned using 10% (v/v) sodium hypochlorite

solution and then lined with brown paper. The front outside surface of each top was fibre taped using J-LAR tape with a 1:1 taping method and each tape was then secured and sealed on an acetate sheet.

All tapes were searched for transferred fibres using a low powered microscope on x25 magnification. Fibres of interest were then recovered by cutting around the fibre using a scalpel and removing the J-LAR tape section using forceps. Microscope slides were prepared by securing a strip of J-LAR tape adhesive side up using double-sided sticky tape and treating with UV light for ten minutes. The J-LAR tape section containing the fibre of interest was then secured adhesive side up on a microscope slide for staining with DAPI.

The removed sections of fibre tape were prepared and stained with DAPI for examination as described in Section 2.4.4 and 2.5.2.2. The prepared microscope slides were then examined using an upright Leica DM5000 fluorescent microscope under x100-x200 magnification. The surface of each recovered fibre was searched for nucleated cells and once located, an image of the cell or cluster of cells *in situ* was captured using the microscope camera. The cell(s) was then removed from the fibre surface utilising a micro-needle and transferred under the microscope onto a fresh section of J-LAR tape, and then into an individual 0.1 ml PCR tube containing 1 µl of sticky stuff remover (SSR) as described in Section 2.6. The associated fibre was then recovered from the J-LAR tape section using SSR and mounted on a microscope slide using DPX and a coverslip as outlined in Section 2.14.1, for further fibre analysis should a positive PCR result be obtained.

Samples were prepared using ForensicGEM® Tissue extraction reagents and then set-up for direct, 34-cycle PCR using the AmpFℓSTR® SGM Plus® PCR Amplification Kit as described in Sections 2.7.3.1 and 2.9.1.2 respectively. PCR products were visualised and analysed as outlined in Sections 2.13.2.1 and 2.13.3.1.

4.3.3.2 Demonstrating cell transfer – two-way transfer - results

Between 640-1300 transferred fibres were recovered from each of the six individuals who participated in the two-way transfer experiments (Figure 4.9). A total of forty-five cell events were observed in which a nucleated cell or cluster of cells was associated with the surface of the transferred fibre. These were imaged *in situ* and then recovered for DNA profiling, some examples of which are shown in Figure 4.11 and Figure 4.12. As shown in Figure 4.10, seventeen of the forty-five samples resulted in successful amplification of DNA, with sixteen of these being attributed to background wearer DNA, one of which was a full profile obtained from a cluster of six nucleated cells shown in Figure 4.13. The remaining DNA profile could not be attributed to either of the participants involved in the

transfer, or to any person involved with this study. No DNA amplification was observed for twenty-eight of the samples. Data for chart included in appendix 8.14.

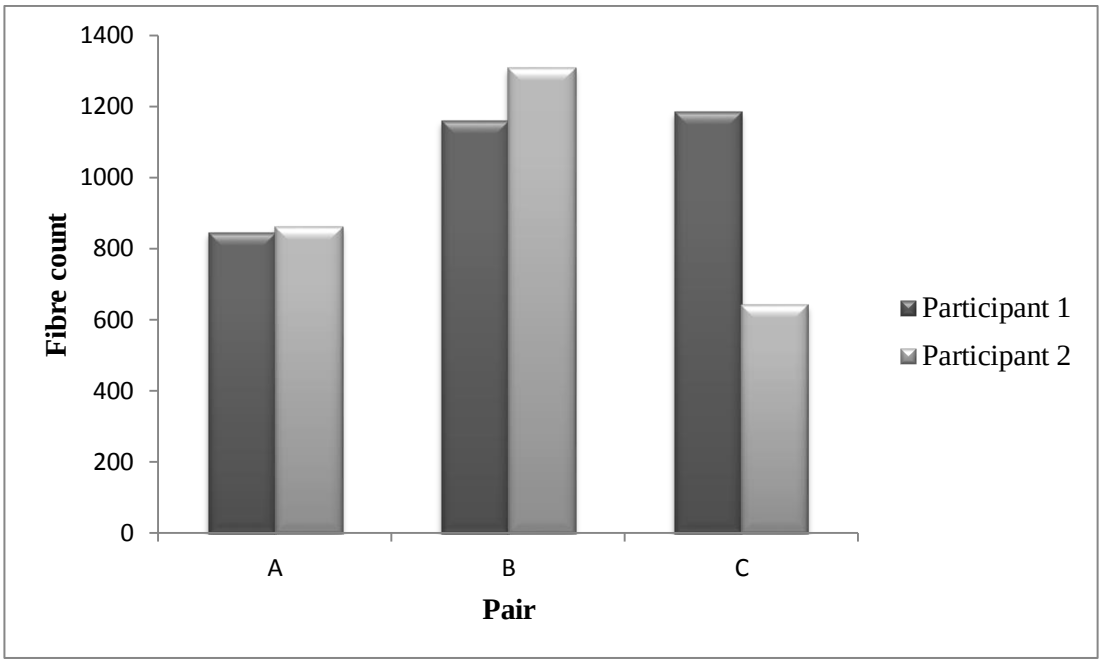


Figure 4.9 Number of transferred fibres recovered from each participant's garment

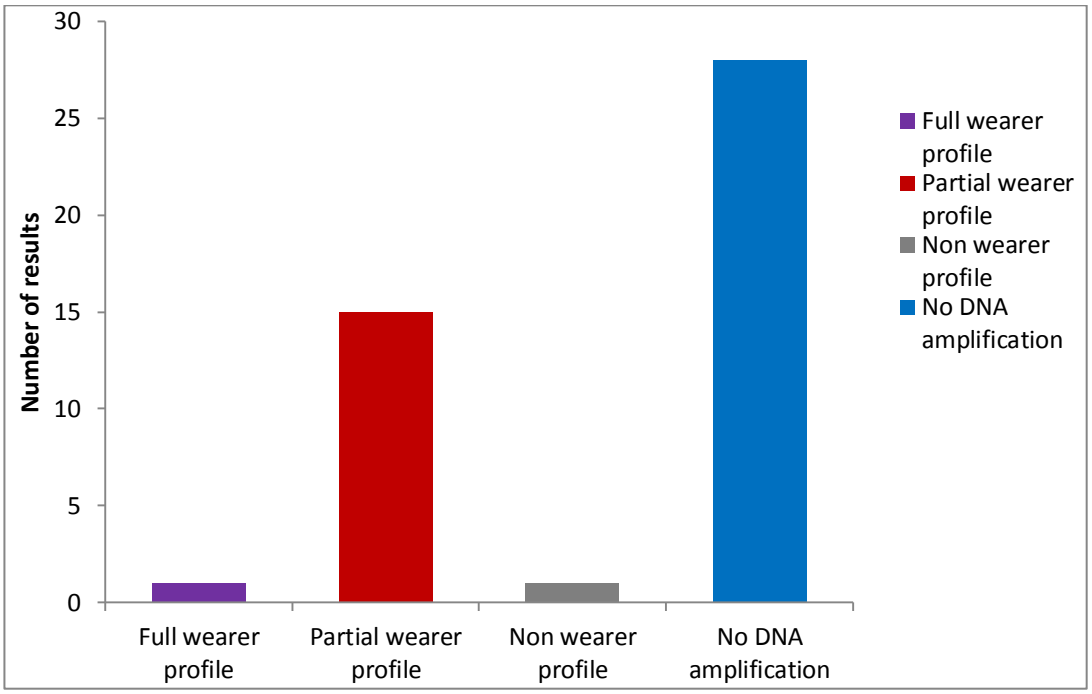


Figure 4.10 Amplification results obtained for a total of forty-five cells recovered from the surface of textile fibres

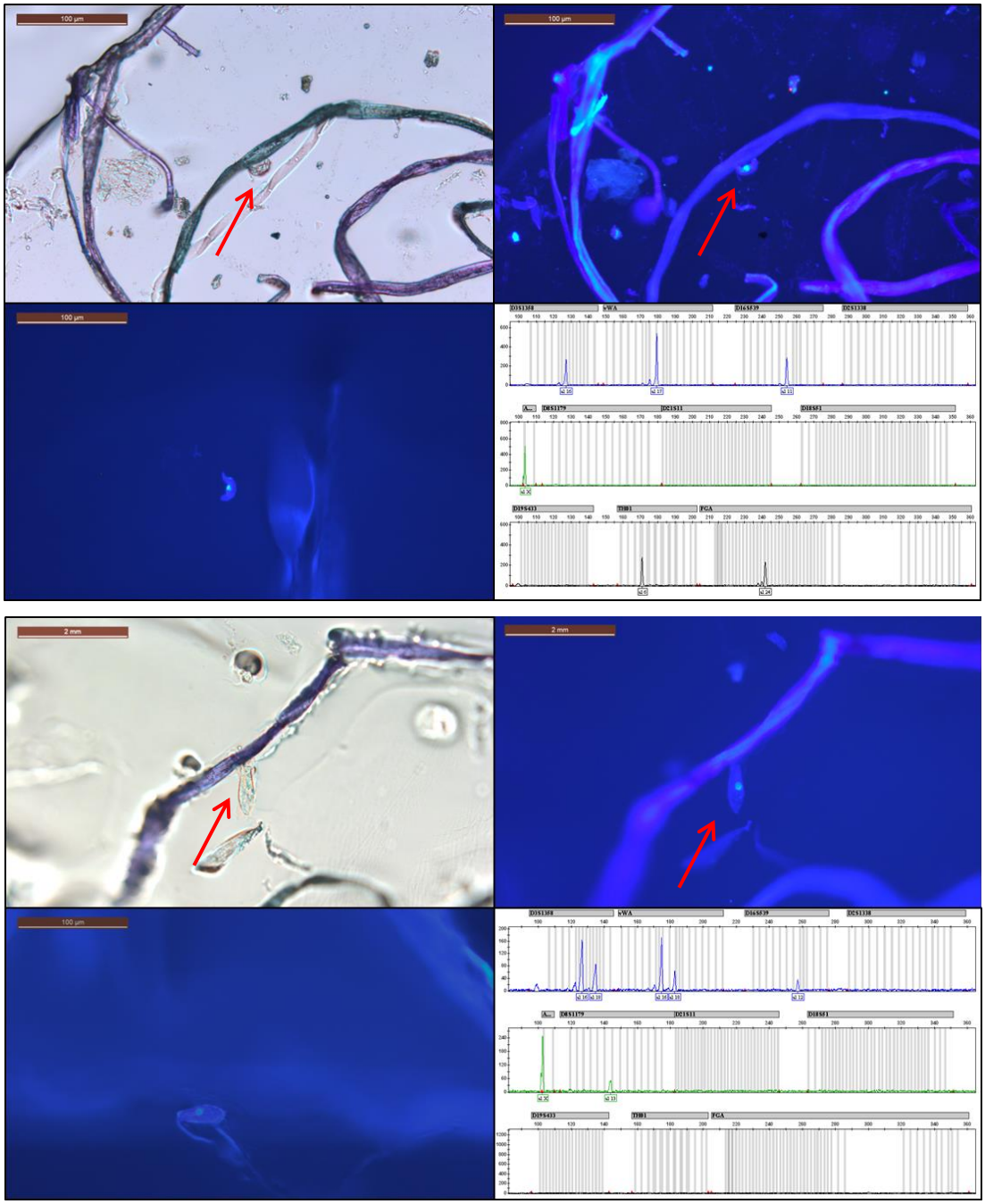


Figure 4.11 Examples of partial SGM Plus DNA profiles obtained from a single-cell recovered from the surface of a textile fibre. Each four image panel shows the cell *in situ* on the fibre surface under white (top left) and fluorescent (top right)t light, an image of the cell on a section of tape prior to transfer into a PCR tube (bottom left) and the resulting SGM Plus EPG. All scale bars in the top panel show 100 µm, and those in the bottom panel show 2 mm.

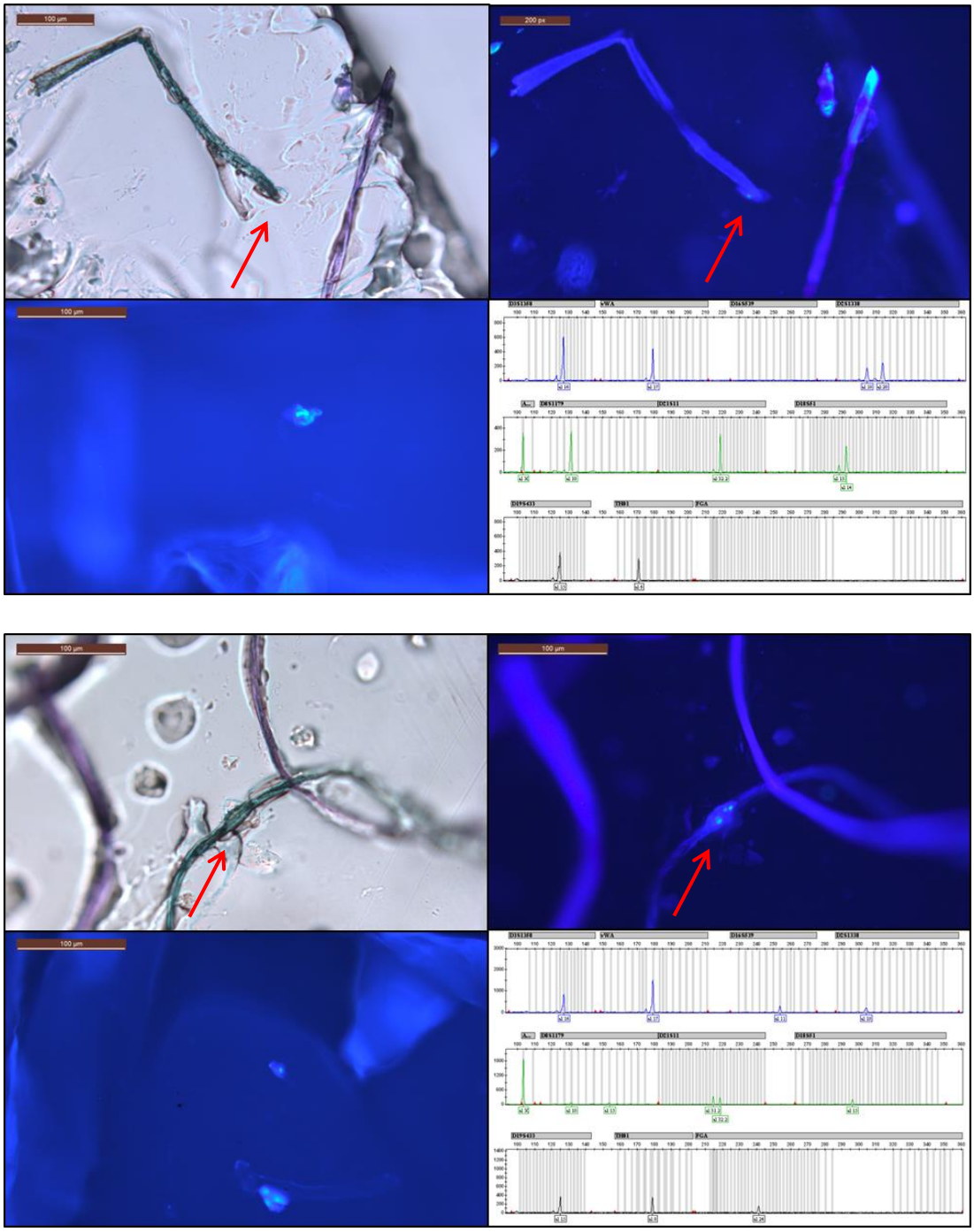


Figure 4.12 Examples of partial SGM Plus DNA profiles obtained from a single-cell (top panel) and a cluster of cells (bottom panel) recovered from the surface of a textile fibre. Each four image panel shows the cell(s) *in situ* on the fibre surface under white (top left) and fluorescent (top right)t light, an image of the cell(s) on a section of tape prior to transfer into a PCR tube (bottom left) and the resulting SGM Plus EPG. All scale bars show 100 µm.

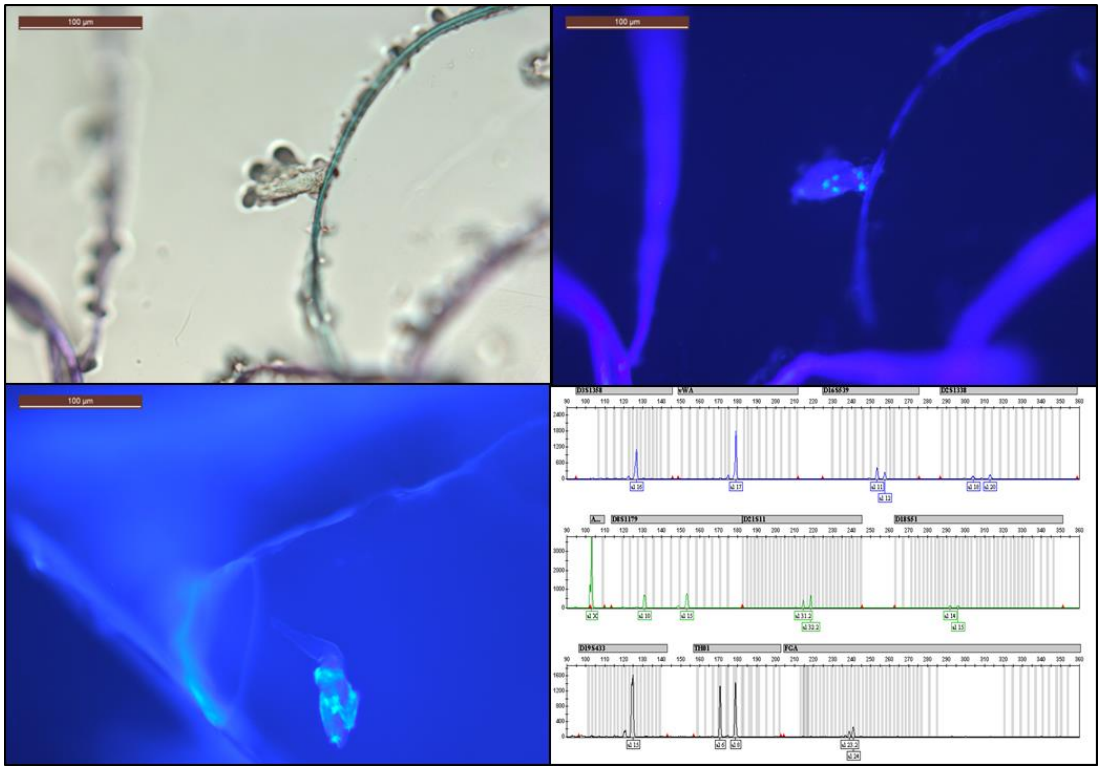


Figure 4.13 A full SGM Plus DNA profile obtained from a cluster of six cells recovered from the surface of a textile fibre. The cells ares shown *in situ* on the fibre surface under white (top left) and fluorescent (top right) light, an image of the cells on a section of tape prior to transfer into a PCR tube (bottom left) and the resulting SGM Plus EPG. All scale bars show 100 µm.

4.3.4 Further method development

Following analysis of the results from the two-way transfer experiment, see Section 4.3.3.2, it was determined that modifications to the DNA analysis protocol would be required prior to the commencement of the next transfer experiment; specifically on the medium used for transfer of the single-cell into the PCR tube.

4.3.4.1 Effect of increased time of cell exposure to SSR

In order to ascertain the effect of prolonged SSR exposure on the amplification of DNA in ltDNA samples, buccal cells were prepared as described in Sections 2.4.2 and 2.4.3. Cells were stained using haematoxylin and eosin as described in Section 2.5.1. Ten buccal cells were transferred in duplicate to pre-prepared J-LAR tape sections as described in Section 2.4.3.2. It was decided that ten buccal cells would allow the effect of SSR exposure on a low DNA template to be assessed whilst ensuring successful amplification for analysis. The J-LAR tape sections were transferred to a 0.1 ml PCR tube with 1 µl of SSR and left to incubate at room temperature. At hourly intervals, an additional two samples were prepared and transferred into SSR until the final sample, which was prepared for PCR immediately

following transfer into the SSR. All samples were then prepared and amplified using 34-cycle SGM Plus® amplification as described in Sections 2.7.3.1 and 2.9.1.2.

Shown in Figure 4.14, as the length of cell exposure time to the SSR increased over one hour a decrease in the average peak heights at all loci was observed. To determine if SSR could be removed from the amplification tube entirely, an experiment was set-up to test the effectiveness of FAST™ minitapes rather than J-LAR as the vector for transfer of the single-cell into the PCR tube. As the layer of adhesive on the Scenesafe FAST™ minitapes is relatively thin, it was tested to see if water could replace SSR in the PCR tube without affecting the amplification reliability. Data for chart included in appendix 8.15.

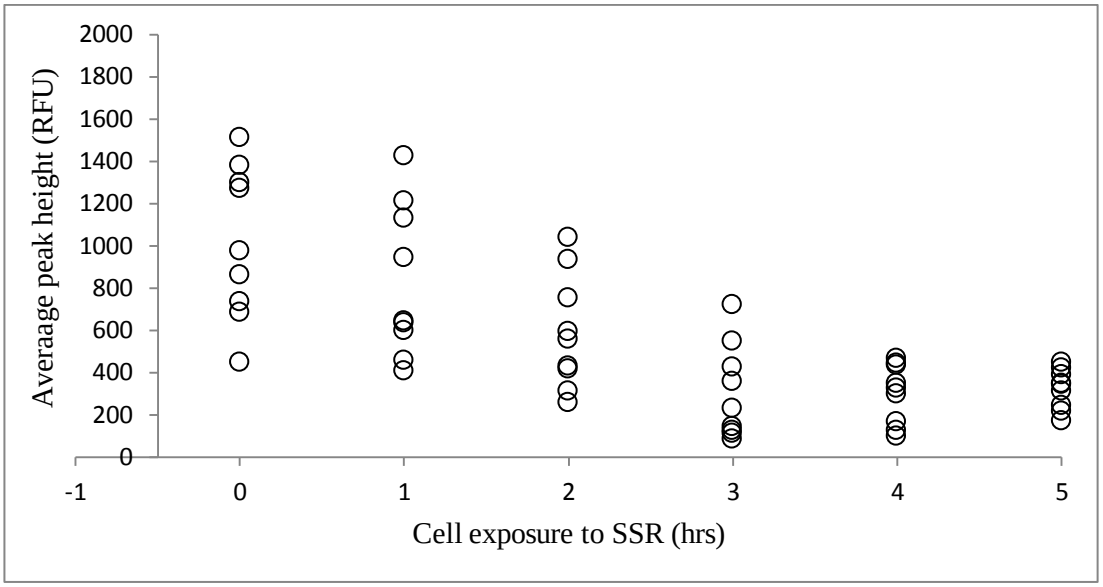


Figure 4.14 Results of DNA profiles generated from cell samples exposed to SSR for an increasing length of time. Average peak heights for all ten SGM Plus STR loci are shown at hourly time increments.

4.3.4.2 Effectiveness of FAST™ minitapes with MBG water

A buccal sample was prepared as described in Sections 2.4.2, 2.4.3 and 2.4.3.2, and then stained using a 1 in 10000 diluted DAPI solution as outlined in Section 2.5.2.2. Under the fluorescent microscope a single-cell was placed on a pre-prepared section of FAST™ minitapes (Chapter 2.5) and then transferred into a 0.1 ml PCR tube containing 1 µl of MBG water. This was carried out in triplicate and then repeated again using 1 µl of SSR. The same steps were taken using J-LAR tape for the cell transfer in order for the results to be compared. All samples were then prepared and amplified using 34-cycle SGM Plus® amplification as described in Sections 2.7.3.1and 2.9.1.2.

As determined in the previous chapter, SSR is required to facilitate the release of the single-cell from the layer of adhesive on the J-LAR tape, for amplification of the DNA. Figure 4.15 shows that the use of sections of FAST™ minitapes for cell transfer into the PCR tube, rather than J-LAR tape results in a more consistent amplification of DNA. Furthermore SSR does not appear to be required for release of the cell from the thinner layer of the FAST™ minitapes adhesive. Data for chart included in appendix 8.16.

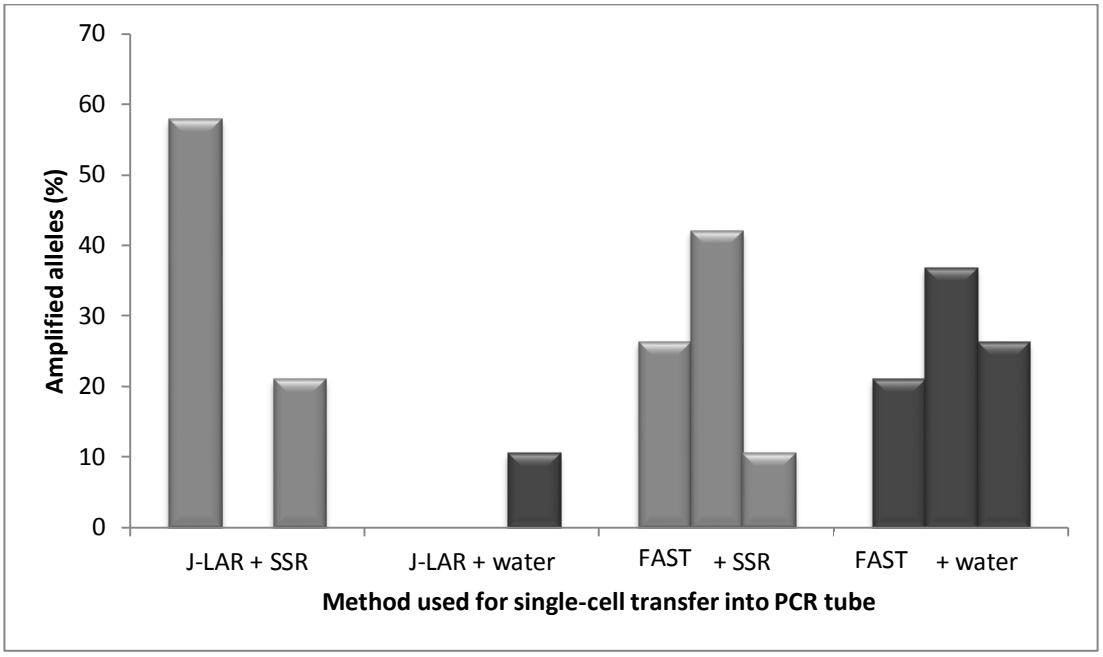


Figure 4.15 Methods for single-cell transfer to a PCR tube results showing the number of observed alleles amplified from a single-cell (in triplicate) following use either J-LAR tape or FAST™ minitape with SSR (grey) or water (dark grey).

The subsequent one-way transfer experiments were carried out using the modified protocols.

4.3.5 Demonstrating cell transfer – one-way; part A (experiment)

4.3.5.1 Demonstrating cell transfer – one-way; part A – materials and methods

Three green coloured, long sleeved tops (95% cotton, 5% elastane) and three black coloured long sleeved tops, 100% cotton, were purchased on separate days. As described previously, the tops were grouped by colour, washed in separate machines using a biological washing powder and then fibre taped to remove any exogenous DNA and fibres respectively. A strip of J-LAR tape was applied five times across the front surface of each of the three black coloured recipient tops, which was then stained with a 1 in 10000 dilution DAPI solution and examined under the high powered fluorescent microscope for nucleated cells as described in Sections 2.4.4 and 2.5.2.2. No nucleated cells were observed on any of these control tapes.

Three volunteers were recruited at random to act as the fibres/DNA donors whilst one individual was recruited to act as the recipient in all three transfer experiments. The donor participant was asked to wear the provided green top for at least eight hours whilst carrying out their normal daily activities, and then given a hairnet, face mask, disposable gloves and oversleeves to wear immediately prior to contact. The recipient was provided with a hairnet, facemask, disposable gloves and a crime scene suit, which was worn underneath the black cotton recipient garment as shown in Figure 4.18. The two participants were then asked to hug for two minutes, in the same manner as previously described. At the end of the contact period, the recipient garment was immediately taken off and sealed in a bag until examination.

4.3.5.2 One-way transfer study; part A - results

The number of transferred fibres recovered from the recipient garments following contact was approximetly 630-720 as shown in Figure 4.16.

In once instance, a cluster of three cells was observed to be associated with a transferred fibre shown in Figure 4.17. The cells were recovered from the fibre surface for DNA analysis, however no amplification was observed.

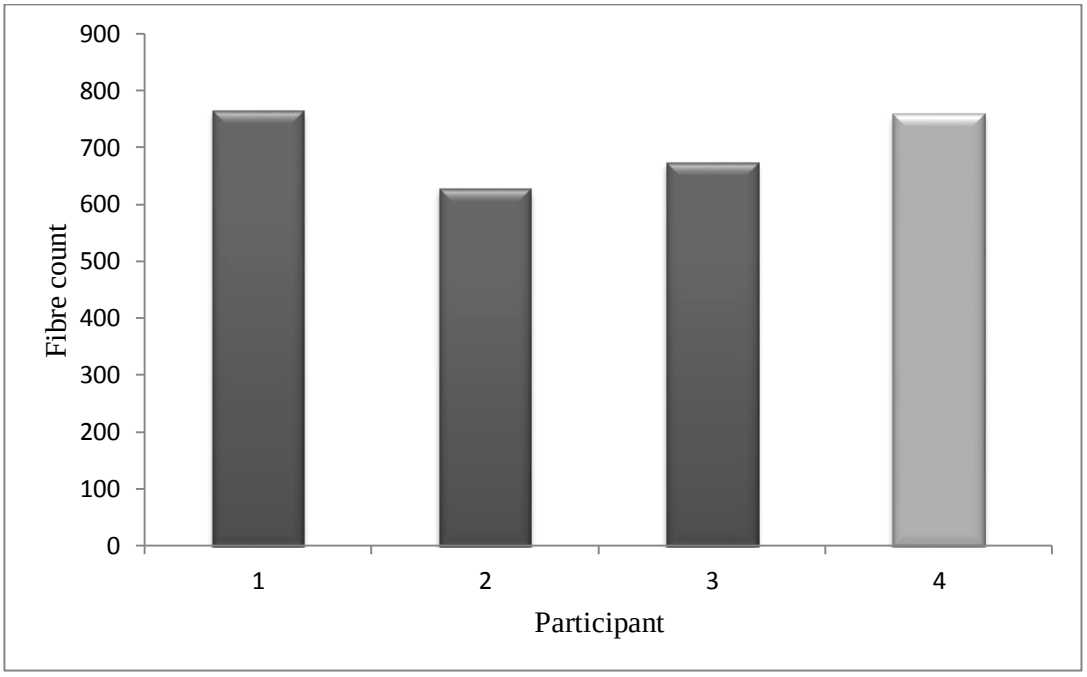


Figure 4.16 Number of transferred fibres recovered from the recipient garment following contact with the donor for part A (particpants 1-3) and part B (participant 4)

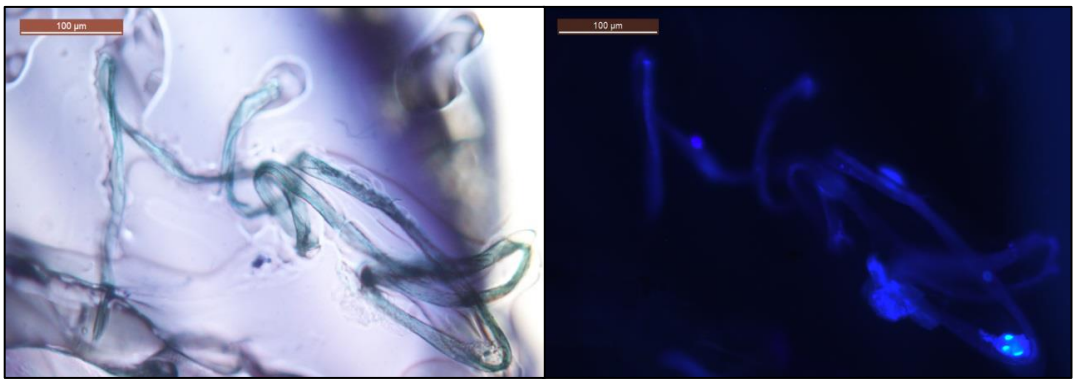


Figure 4.17 Images of nucleated cells *in situ* on the surface of a transferred fibre as viewed under both white (left) and fluorescent (right) light sources. Both scale bars show 100 µm.

4.3.6 Demonstrating cell transfer – one-way; part B (experiment)

4.3.6.1 Demonstrating cell transfer – one-way; part B – materials and methods

This part of the experiment was set-up exactly as described for part A, however a male individual who had taken part in the initial part of this study was selected to act as the fibre/DNA donor using the same burgundy coloured, long sleeved, 100% cotton top that he had previously worn. The participant was asked to wear this top for a minimum of sixteen hours over a two day period prior to transfer, shown in Figure 4.18.

All work was carried out exactly as described for the two-way transfer experiment up until the point at which a cell sample was to be recovered from the surface of a fibre.

The cell(s) was then removed from the fibre surface utilising a micro-needle and transferred under the microscope onto a fresh section of FAST™ minitapes, and then into an individual 0.1 ml PCR tube containing 1 µl of MBG water as described in Section 2.6. The associated fibre was then recovered from the J-LAR tape section using SSR and mounted on a microscope slide using DPX and a coverslip as outlined in Section 2.14.1, for further fibre analysis should a positive PCR result be obtained.

Samples were prepared using ForensicGEM® Tissue extraction reagents and then set-up for direct, 34-cycle SGM Plus® amplification as described in Section 2.7.3.1 and 2.9.1.2. PCR products were visualised and analysed as outlined in Sections 2.13.2.1 and 2.13.3.1.

Figure 4.18 Photographs taken during one-way transfer experiment of the recipient prior to transfer (right) and during contact (left).

4.3.6.2 One-way transfer study; part B - results

A total of 760 transferred fibres were recovered from the fibre tapings taken from the receipient's gament following contact. There were no cell events observed to be associated with any of these transferred fibres. Following this, four fibre tapes corresponding to the upper front chest area of the recipient top were selected and swabbed for DNA. The results for the positive and negative controls for these samples were as expected, however no DNA amplification was observed for any of the sample fibre tapes.

4.3.7 DNA extraction directly from tapes (experiment)

4.3.7.1 DNA extraction directly from tapes – materials and methods

Four fibre tapes were selected for further DNA analysis, which had been taken from the upper front sections of the recipient garment. Each tape was carefully peeled off its acetate and swabbed using xylene to remove the adhesive and concentrate the sample onto the swab head as described in (May, 2009). A positive control was set-up by pipetting 10 x 1 µl of

buccal cells which had been cell harvested from a buccal swab as described in Section 2.4.3, onto a strip of J-LAR tape, which was allowed to air dry before being sealed on acetate. The negative control was a strip of plain J-LAR tape secured on a sheet of acetate. Both controls were then treated as the sample fibre tapes. 200 µl of xylene was added to each of the tubes, which were vortexed and then left with open lids overnight in a fume hood until the xylene had evaporated. The four tapes, alongside the positive and negative controls, were then extracted using a Chelex™ extraction and were set-up for PCR amplification using the AmpFℓSTR® NGM™ PCR Amplification Kit as outlined in Section 2.9.1.2. PCR products were visualised and interpreted as described in Sections 2.13.2.1 and 2.13.3.1.

4.3.7.2 DNA extraction directly from tapes – results

Although positive and negative results were as expected for this method, no DNA amplification was observed for any of the four tested fibre tapes as described above.

4.3.8 Persistence of single-cells in fibre tapes (experiment)

4.3.8.1 Persistence of single-cells in fibre tapes - material and methods

Using the same cell harvested sample as described in Section 3.3.4, 1 µl droplets of cells were transferred to a strip of UV-treated J-LAR tape, at intervals of approximately 5 cm and allowed to air-dry in a DNA clean environment. The strip of J-LAR tape was then secured to a sheet of UV-treated acetate and left at room temperature for a period of 24 months. Samples were then prepared for staining with haematoxylin and eosin as described in Sections 2.4.3.2 and 2.5.1, amplified using SGM Plus reagents, see Section 2.9.1.2 and all PCR products were visualised and analysed as described in Sections 2.13.2.1 and 2.13.3.1.

4.3.8.2 Persistence of single-cells in fibre tapes - results

In order to demonstrate the ability to amplify DNA from individual single-cells following a prolonged storage at room temperature, buccal cells were recovered for DNA analysis after a period of 24 months. Of the ten cells recovered, nine resulted in partial DNA profiles with the no amplification observed for the remaining sample. The number of alleles observed per single-cell ranged between three and seventeen alleles; the results were compared with the number of alleles observed from the previous fifty-cell experiment as these cells, from the same donor, were prepared on the day of sampling (day 0), shown in Figure 4.19. The range of alleles observed from the single-cells following the 24 month period of storage, was the same as that observed for those cells prepared on the same day as sampling (both datasets having a mean value of 8 alleles) indicating that the two-year storage of cells on a fibre tape does not cause a consistent reduction in the number of alleles amplified. To determine if a reduction in HMW loci was observed following the 24 month period, results were compared

with those of the fifty-cell DNA results carried out in Section 3.4.4. As shown in Figure 4.20, although the percentage of alleles is seen to decrease as the molecular weight of the allele increases, the results follow the same trend as observed for the freshly prepared cell samples, supporting the persistence of the DNA in these single-cells without consistent signs of degradation of even the HMW loci. Data for chart included in appendix 8.17.

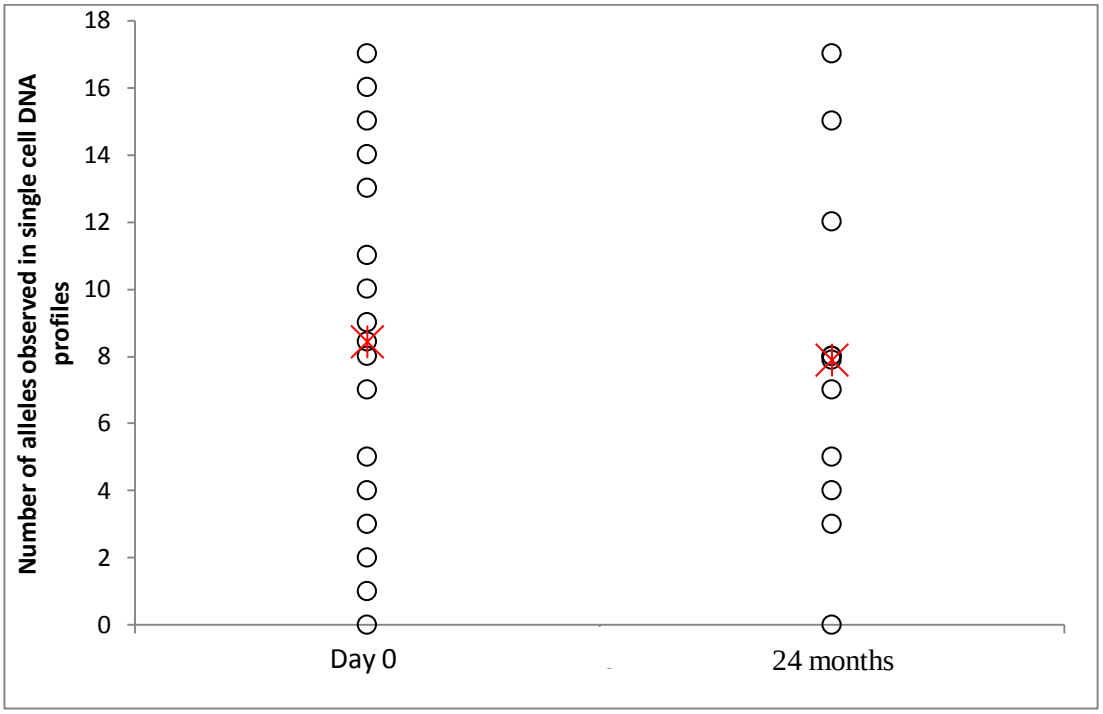


Figure 4.19 Comparison of alleles amplified at each locus in single-cell DNA profiles for both control cells (day 0) and following a 24 month storage period. Mean values for both data sets are shown with a red marker.

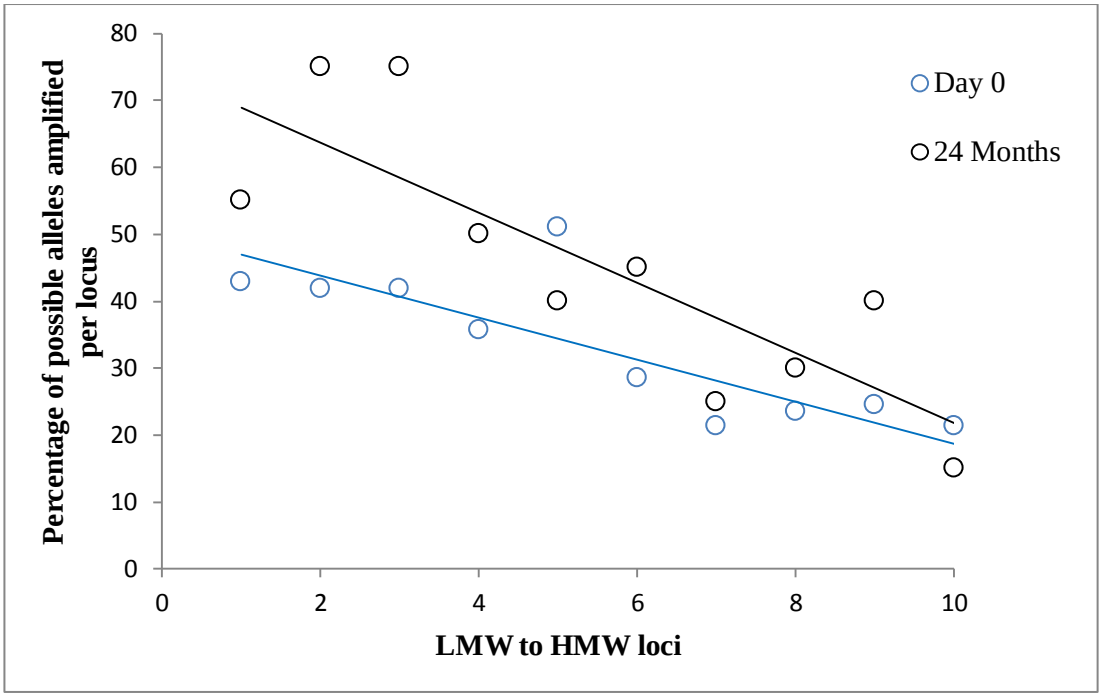


Figure 4.20 Comparison of alleles amplified at each locus in single-cell DNA profiles for both control cells and following a 24 month storage period. A decrease in the amplification of HMW alleles can be seen for both time periods.

4.4 Discussion

It is thought that during the period over which an item of clothing is worn, nucleated cells are being continuously deposited on the outside surface, as it is known that DNA can be transferred onto fabric through contact with the skin (Linacre et al., 2010, Daly et al., 2012), i.e. when adjusting sleeves or wiping hands, or via saliva aerosols whist talking (Rutty et al., 2003, Port et al., 2006). The results of the first part of this study confirmed the presence of nucleated cells on the outer layer of upper garments with an average of approximately 1300 cell events, being observed on fibre tapings for each volunteer. Given that the entire outside front surface of each top was taped, this number of cells is relatively low but does demonstrate that nucleated cells are present and therefore available for transfer to a recipient surface.

The presence of individual single-cells on the surface of the textile fibres was prevalent, however numerous clusters of cells were also observed for each participant. These clusters, particularly those made up of over thirty cells, were observed wrapped around the fibre surface, which is expected to facilitate their transfer with the fibre to a recipient surface. In addition to this, the increased quantity of template DNA within the group of cells will aid DNA amplification during PCR. It is probable that the cellular clusters are present as a result of saliva aerosols rather than from contact with skin, as very few nucleated cells and no clusters of more than eight cells were observed during previous examination of touch DNA samples under a high-powered microscope (data not included in this study), which is also supported by observations in (Alessandrini et al., 2003) and (Kita et al., 2008). Although outside of the scope of this study, RNA profiling could be used to determine if saliva is in fact the source of these cellular clusters.

As cotton is the most frequently encountered relevant fibre type (Watt et al., 2005, Grieve and Biermann, 1997a, Houck, 2003, Palmer and Oliver, 2004, Marnane et al., 2006, Roux and Margot, 1997, Cantrell et al., 2001, Fong and Inami, 1986, Palmer and Burch, 2009), all items purchased for use in the transfer experiments were at least 95% cotton, and selected for their ability to shed well in order to maximise fibre transfer. In addition to this, all donor tops were selected to be either green or purple in colour, as both are less frequently observed in the population (Houck, 2003, Grieve et al., 2003), thus reducing the chance that fibres unrelated to the study were misidentified during the initial examination under the low powered microscope.

The transfer experiments themselves used were set-up to reflect those used in fibre transfer and persistence studies (K. Sheridan 2014 pers. comm. 4th, June), with the addition of DNA anti-contamination precautions integrated into the methodology. A total of ten transfer

experiments were conducted over the course of this study, which resulted in the recovery and examination of just fewer than nine thousand individual fibres. Less variability in the number of transferred fibres recovered between experiments was observed following the one-way experiments than the two-way transfers. This may be due to the fact that the same individual acted as the recipient in each of the four transfer events, thereby partially standardising the pressure applied during the contact period.

Just over a third of the cell samples tested resulted in partial wearer DNA profiles, with between one to sixteen alleles (average of six) amplified per sample. One full profile was obtained from a cluster of six cells, approximately 37 pg DNA, which is consistent with the earlier sensitivity test results (Chapter 3) carried out using pristine 007 control DNA.

Just under half of the cell samples processed following the two-way transfer study resulted in no amplification, which may be as a result of the SGM Plus kit's limit of sensitivity for single-cell analysis if the DNA is sub-optimal, however it was also noticed that the time taken to examine the fibre surfaces under the fluorescent microscope, image, recover and transfer cells to the PCR tube, was by far in excess of that taken during the initial single-cell studies carried out in Chapter 3. On further investigation it was demonstrated that the SSR was in fact having an increasingly detrimental effect on the ability to amplify the DNA as the length of time of cell exposure was increased and therefore may have contributed to these results. As it has been shown in initial studies that the SSR does not cause PCR inhibition at the low volume of 1 µl used in each reaction (see Section 3.4.1), it is likely that the SSR is interacting with the DNA template over the increased timeframe, either through binding with the DNA or causing degradation to occur. The SSR was originally added into the PCR tube to facilitate the release of the single-cell off the J-LAR tape section and so a several methods for transfer of the cell(s) to the PCR tube were investigated. The use of FAST™ minitapes, which are typically employed for collecting touch DNA samples from exhibits, was shown to be an effective substitute for the J-LAR tape, due to the relatively thin layer of adhesive on the tape, which has been shown to release cells easily for DNA extraction (Hansson, 2009, Verdon et al., 2014). As a result of this, all further testing was conducted using MBG water in place of the SSR, which was shown to be effective when used in conjunction with the FAST™ minitapes.

Following the number of wearer DNA profiles obtained from the two-way transfer study, it was decided that in order to demonstrate the transfer of cells from the donor garment, a one-way transfer experiment would be utilised, whereby the recipient garment was prepared to minimise the presence of background cells and fibres. The following three transfer experiments resulted in only one cell event, which did not result in DNA amplification. In a

final attempt to demonstrate the transfer of cells via the fibre surface a volunteer, who took part in the earlier potential for cell transfer experiment and had a top that shed fibres well and also had the highest cell count, was asked to wear his top again for a minimum of sixteen hours over a two day period. This was part of the experiment was designed with the intention of maximising the potential for nucleated cells to be present on the outside surface of the garment and therefore increase the chance of transfer to the recipient. However, no cell events associated with transferred fibres were observed.

As only one cell event was observed in the one-way transfer studies, two possible explanations were considered that may account for the absence of observed cells; firstly that as only the sections of J-LAR tape containing the fibre of interest were stained and examined under the fluorescent microscope, cells may be present over the remainder of the fibre tape, or secondly that there is little or no transfer of DNA to the fibre tapes as a direct result of contact between two items of clothing. In order to further explore this, fibre tapes were selected from the recipient garment used in the final experiment and swabbed for DNA according to current practice in operational casework (May, 2009).

The contact scenario used for the latter part of this study was set-up to maximise the potential for DNA transfer, and in the absence of any background DNA to obscure the results, no amplifiable DNA was recovered from any of the swabbed fibre tapes. This could be a result of the precautions taken during this study to ensure that DNA transfer could only take place as a direct result of contact between the donor and recipient garments which, although preliminary, suggests that when DNA is recovered from fibre tapes, it may in fact have been transferred to the recipient garment through means other than surface-surface contact, i.e. via direct contact with the skin during an altercation or even from deposits of saliva following a verbal exchange. As such, recovering DNA from the surface of a garment should not be taken as proof of direct contact in the absence of other evidence.

4.5 Conclusions

The use of DAPI as a DNA stain, together with fluorescence microscopy, has been shown to be an effective method by which single-cells can be visualised *in situ* on the surface of textile fibres. Using a microneedle, it is possible to manually recover detected cells to a section of adhesive FAST tape for transfer directly into a PCR tube, thus ensuring that the sample is not lost. It has been demonstrated that the inclusion of DAPI and FAST™ minitapes in the PCR does not inhibit DNA amplification, nor does the DAPI adversely affect the MSP spectral profile, allowing for further analysis of the fibre if required.

It has been demonstrated that nucleated cells are present on the outside surface of an individual's upper clothing and as such represent a source of DNA for possible transfer to a recipient surface.

Whilst it has been demonstrated that nucleated cells can be visualised *in situ* on fibre tapes and recovered for DNA analysis without adversely affecting subsequent fibre analysis, the results of this study suggest that if fibres are in fact acting as a vector for the transfer of DNA then they do so rarely and as such, specifically targeting transferred fibres for an associated DNA profile is unlikely to yield a result.

From a forensic casework point of view, the use of staining, examining and recovering cells from the whole fibre tape may prove a favourable alternative to the current methods employed in forensic casework of swabbing or cutting the fibre tape for DNA extraction, which would have the dual advantages of preserving the fibre evidence as well as reducing the background debris and adhesive that would allow for a more efficient DNA extraction or a direct PCR.

5. Massively parallel sequencing approaches to degraded DNA analysis

5.1 Introduction

Forensic samples are no longer 'living' and so, having lost their innate DNA repair capabilities an accumulation of DNA damage in the cells occurs (Lindahl, 1993). Regardless of its localised or ubiquitous nature, the degradation of DNA in any forensic sample has the potential to impede efforts to generate reliable data. The successful amplification of targeted regions of DNA is dependent upon the ability of the polymerase enzyme to access the DNA template and synthesise the new strand; damage to the DNA template can partially or completely inhibit this process. If amplified product is obtained, then consideration must be given to the veracity of any data obtained as base modifications and polymerase chain reaction (PCR) artefacts may have become incorporated into the DNA sequences.

The first generation Sanger sequencing instruments are rapidly becoming superseded by the use of massively parallel sequencing (MPS) platforms across the biological sciences. The development of MPS was originally driven by the need to reduce both the time and costs associated with Sanger sequencing (Sanger et al., 1977) for large scale genomic study (Collins et al., 2003), however this multifaceted technology has found numerous additional and diverse applications across the spectrum of genetics, including in clinical diagnostics (Jones, 2015), molecular ecology (Ekblom and Galindo, 2011) and functional genomics (Morozova and Marra, 2008). The fields of forensic science and bioarchaeology are no exception to this as the potential is explored to develop both novel approaches to their work or for opportunities to make improvements on traditional techniques in order to generate a better quality and quantity of results from sub-optimal samples.

The investigation-governed nature of forensic DNA analysis requires a focused approach, exploiting highly informative markers in order to answer questions posed during the inquiry. Sequencing of large regions, or indeed the whole human genome, is therefore a redundant endeavour in terms of the added value of the additional information, as well as the time and monetary considerations. There are two notable exceptions to this type of approach; firstly, the recently demonstrated distinguishing capability in haplotype when the whole mtGenome is sequenced (Just et al., 2015, King et al., 2014a) as discussed in Chapter 1, and secondly is the use of whole nuclear genome mapping by MPS to differentiate between monozygotic twins based on rare polymorphisms, who would otherwise be indistinguishable from one another (Weber-Lehmann et al., 2014).

As discussed in detail in Section 1.3.3, the markers traditionally utilised in casework are currently undergoing development and evaluation for use on the MPS platforms, in addition to the rigorous validation processes that accompany any new technology for application in a

forensic context, with the long-term potential to entirely replace the current PCR-CE techniques in forensic casework.

One of the significant advantages to using MPS for forensic analysis is that rather than a profile based on fragment size analysis, the actual sequence of the STR is obtained. Numerous isoalleles (Heinrich et al., 2005, Rockenbauer et al., 2014, Gelardi et al., 2014, Dalsgaard et al., 2014) as well as SNP sites in the flanking regions of the repeat structures (Divne et al., 2010) have already been documented, which provide the capability to distinguish between two individuals with the same allele call for a locus, thus increasing the discriminatory value of the allele and also aiding the deconvolution of mixtures. The allele sequence may also prove particularly useful for distinguishing stutter from the true allele in situations where one allele is a single repeat shorter than the other (Zeng et al., 2015).

The ability to multiplex a variety of markers of the same size is a significant advantage of MPS over the traditional PCR-CE methodologies for the processing of degraded forensic samples. The extent of DNA fragmentation due to cell death mechanisms or strand breaks caused by DNA hydrolysis, will ultimately determine the quantity and quality of DNA template that is available for DNA extraction and the potential for successful amplification. Targeting markers suitable for degraded DNA, i.e. SNPs, mini-STRs and mtDNA (see Chapter 1), will facilitate DNA amplification, enabling the maximum amount of information to be obtained from the remaining DNA.

Bioarchaeological interest in a particular set of skeletal remains, with queries around sex, identity, ancestry and phenotypic characteristics have shown to benefit from a forensic approach to the DNA analysis.

The first successful project involving this type of historical DNA work include the identification of five members of the Romanov family in 1994 (Gill et al., 1994) followed by the remaining two children in 2008 (Coble et al., 2009). The initial work was carried out using Amelogenin, five tetrameric STR markers and mtDNA analysis of the HVSI region, which enabled the identification of the Tsar, Tsarina and three of their children from the nine present skeletons. Further work was then able to confirm two skeletons found in an additional grave as the missing daughter and son of the Tsar and Tsarina. Updated DNA analyses were carried out for all members of the family, using the AmpFℓSTR® Identifiler™, MiniFiler™ and YFiler™ PCR Amplification Kits to confirm the identity of Tsar Nicholas II as well as the results from the previous study.

The investigation into skeletal remains excavated in 2012, which were alleged and subsequently confirmed to be those of King Richard III (King et al., 2014b) utilised a variety

of DNA markers and following protocols for ancient DNA analysis, carried out in two laboratories for independent verification. Whole genome sequencing of mtDNA from the skeletal sample and also two maternal relatives (separated by 19/21 generations) for comparison, was carried out on both the Illumina HiSeq2000 and Ion Torrent PGM MPS platforms. Phenotypic characteristics were also investigated by targeting HlrisPlex SNPs, the results of which inferred that the subject had a 96% chance of having blue-coloured eyes and a 77% chance of blonde-coloured hair. Both of the above examples demonstrate the value of using a variety of marker types in order to provide a comprehensive DNA evaluation, thus increasing confidence in the results obtained. Although this information has clear application in field of forensics, it must be noted that many phenotypic characteristics, such as hair, eye and skin colour, can to varying degrees be altered by the individual, with their appearance no longer reflecting their genetics.

It has been hypothesised that the reason behind the preservation of DNA in skeletal remains is due to its affinity for binding with calcium phosphate apatite minerals (Lindahl, 1993) and collagen I (Mrevlishvili and Svintradze, 2005b), which form the inorganic and organic components of the structural matrix respectively. Corroboration for the adsorption of DNA on apatite has been provided in 2001 by Okazaki (Okazaki et al., 2001) and more recently (2014) by Grunenwald (Grunenwald et al., 2014), who both concluded that there is a strong affinity between the phosphate groups of the DNA backbone and the calcium ions in the apatite. Collagen molecules cluster together to form fibrils, which are required to provide structural support in bone materials (Rho et al., 1998). It is thought that DNA is incorporated into the centre of these fibrils, forming a stable nanostructure with the collagen molecules (Pidaparti et al., 2009, Mrevlishvili and Svintradze, 2005a).

Within the bone matrix, a network of cross-linked, fully hydrated collagen fibrils is formed, the alignment of which results a series of regularly interspersed gap zones (Rho et al., 1998). The mineralisation process is initiated in these gaps through the formation of hydroxyapatite platelets, which then progresses throughout the fibril network via the formation of hydroxyapatite crystals that encapsulate the fibrils, including the DNA component, and develop in size to fill the interfibrillar space (Campos et al., 2012). Extracellular DNA that is present within the interfibrillar space as a result of cell death processes, chemical degradation or microbial activity, may become trapped through adsorption with the hydroxyapatite crystallites, effectively facilitating their preservation (Campos et al., 2012).

Although DNA can be extracted from both the apatite and collagen fractions independently (Campos et al., 2012), the relationship between the two appears to be complex, with collagen potentially acting to shield the apatite-adsorbed DNA (Götherström, 2002, Roberts

et al., 2002), as the loss of DNA in the collagen faction also correlates with a loss in the mineral fraction (Campos et al., 2012). As proposed by Campos (Campos et al., 2012), if the collagen/DNA fibril molecules were to degrade, i.e. though microbial activity (Child, 1995), then apatite crystals located in the interfibrillar gap zones would also then be exposed to degradation. It appears therefore probable that the preservation of DNA is principally dependent upon the integrity of the bone matrix.

Each tooth is made up of pulp, a soft tissue, and three structural components, enamel, dentin and cementum, which are made up of 97%, 70% and 45% mineral respectively (Brudevold et al., 1960). All four of these components contain both DNA and mtDNA, which have been shown to degrade differentially *post mortem* (Higgins et al., 2015). Due to the mineral encapsulated cementocytes in the cementum and odontocytes in the dentin, the DNA in these two components invariably persists for a longer period of time than the pulp tissue (Malaver and Yunis, 2003, Higgins et al., 2015).

There are numerous examples in which the ability of DNA to persist for longer periods of time in environments that have a cooler climate has been demonstrated (Burger et al., 1999, Hoss et al., 1996, Higgins et al., 2015). Whilst it has been shown that the structure of the bone itself does not undergo any significant structural changes under 250 ˚C (Etok et al., 2007), the rate of depurination, one of the predominant forms of DNA degradation in *post mortem* samples, has been shown to be temperature dependent (Lindahl and Nyberg, 1972). This knowledge, along with localised modern and historical temperature data, allows the thermal age of a sample, i.e. the extent of DNA degradation expected if the temperature is at 10 ˚C, to be calculated. The estimated thermal age has been shown to be a better indicator of the state of DNA preservation in a sample than the actual age (Smith et al., 2003).

The availability of water within the localised environment is known to increase the rate of DNA degradation, as damage from hydrolytic activity, particularly depurination and deamination is frequently observed in degraded samples (Hedges and Millard, 1995, Nielsen-Marsh and Hedges, 2000, Bollongino et al., 2008).

Although the release of hydroxyapatite crystals into the surrounding environment provides bone material with an effective buffering capacity within the range pH 3-9 (Bada et al., 1980), if the acidity of the burial environment exceeds this level, the bone matrix has been shown to break down (High, 2015). Whereas it has been demonstrated that slightly alkaline conditions, such as limestone environments (Burger et al., 1999) favour DNA preservation.

The exposure to sunlight or more specifically UVA and UVB light, will vary depending on the individual circumstances surrounding death, e.g. if the skeletal remains were in the open

environment, as well as excavation and storage. The absorption of light photons by nucleotides, particularly pyrimidines, can lead to a variety of oxidation reactions, including the formation of 8-OxoG and other hydantoins (Peak and Peak, 1989, Zhang et al., 1997, Ravanat et al., 2001, Hoss et al., 1996).

Finally, bioerosion of bone material as a result of microbial activity, particularly the breakdown of collagen by collagenase/protease-producing bacteria (Child, 1995) will also detrimentally effect the preservation of DNA within bone material (Campos et al., 2012).

The greatest challenge faced for the analysis of degraded samples, especially for ancient specimens, is the potential for any extracted DNA to have come from a source other than the target. Aside from a couple of examples of exceptional DNA preservation (Meyer et al., 2012, Reich et al., 2010), it has been shown in a number of studies that on average a degraded DNA extract consists of 95-99.9% exogenous DNA and just 0.1-5% target DNA (Burbano et al., 2010, Meyer et al., 2014, Fu et al., 2013).

Nine criteria for ensuring the authenticity of ancient DNA results were proposed by Cooper and Poinar (Cooper and Poinar, 2000) in an effort to standardise data evaluation, these include that pre-amplification work is carried out in a physically isolated area, that appropriate negative controls are utilised throughout processing and a minimum of two independent and sequential extractions of each sample should be carried out, by two different laboratories; although with archaeological specimens this may not always be feasible due to limited sample availability. At least two DNA amplifications should also be carried out for each extract in order to generate consensus sequences that will help identify PCR artefacts arising from stochastic variation, and that an inverse relationship should be observed between the amplification strength and the PCR product size.

5.2 Aims and objectives

DNA material contained within human skeletal remains that has severely degraded as a result of temporal or environmental factors can represent a significant challenge for forensic DNA analysis. MPS techniques are proposed as a significant a step forward in terms of the quantity and reliability of information that can be obtained from samples and as such, should offer advantages for analysis of these sample types. The aim of this chapter is to investigate the efficacy of MPS techniques for degraded DNA through the targeted sequencing of both DNA markers and regions of the mtDNA genome, using traditional PCR-CE techniques as a point of comparison.

The initial objective of this study is to identify the most effective method of extracting DNA from the skeletal remains, assessing both the DNA quantity and quality through the use of qPCR techniques. In the absence of a commercially available kit, an 'in-house' assay for the quantification of mtDNA will be developed and tested for use with these samples. DNA extracts will then be subjected to both PCR-CE and two PCR-MPS techniques, using Illumina's suggested protocol for amplification of the control region of the mtGenome and their recently launched ForenSeq™ Signature Prep Kit for DNA, in order to assess the capability of the latter to provide a more comprehensive evaluation of the degraded DNA samples.

5.3 Materials and methods

5.3.1 Samples

Two different sample sets were used as part of this study in order to look at both the effect of age and of environment in the analysis of DNA using MPS techniques.

5.3.1.1 Brazil

To investigate the use of MPS for the sequencing of highly degraded DNA, twenty samples (labelled LAFR01–LAFR20) were obtained through a collaborative project with the Department of Pathology and Legal Medicine, Ribeirão Preto Medical School, University of São Paulo, Brazil. These samples were taken from the skeletal remains of twenty individuals excavated from the Cemetery of Bom Pastor in Ribeirão Preto. This cemetery was used as a burial site for individuals of either unknown identity or those who did not have means for other arrangements. Due to the site been sold, the graves were excavated and the remains donated to the Ribeirão Preto Medical School.

The weather conditions in Ribeirão Preto, for the time period 2000-2012 had an average high temperature of 26 ºC, with an average of 124 mm rainfall over 10.8 days a month (www.worldweatheronline.com, accessed 18/07/2015). These samples were of particular interest as the heat and humidity in Brazil creates a hostile environment for DNA, therefore it was expected that any DNA remaining in the skeletal samples would be severely degraded. As such, mtDNA sequencing was the primary target of this study although some nDNA analysis was also carried out.

These samples included a section of the femur from twenty different individuals that had been buried 11-14 years previously. In addition to this, a tooth sample was also provided for samples LAFR02, 07, 09, 11 and 20. Ethical permission for this project has been granted by both the University of São Paulo Ribeirão Preto Medical School and Northumbria University Ethics Committee (copy included in appendix 8.18).

5.3.1.2 Vindolanda Trust

Positioned on the old Stanegate Road, approximately a mile south of Hadrian's Wall, Vindolanda is an archaeological dig site dating back to the early Roman occupation (Birley, 2009). Throughout this period, the infrastructure of Vindolanda was subjected to frequent re-building projects, resulting in the construction of at least nine distinct forts. As one fort was demolished, the ground was prepared for new structures by filling and covering with up to 40 cm of turf or clay, creating layers in which any debris was protected from both oxygen and surface water (Birley, 2009). These unique conditions have consequently enabled

Vindolanda to yield some of the best preserved exhibits from across the entire Roman Empire (Birley, 2007).

Two samples have been obtained from the Vindolanda Trust for DNA analysis; sample 'V12-35N 29180' Femur and sample 'SF8658 – Tooth from Human Skull'. The Femur bone was recovered during excavation of a fort ditch, which has been dated to circa 110CE. The tooth sample, a molar removed from the jaw bone, was taken from a skull that has been dated to 208-211CE. Permission for the destruction of the samples for DNA analysis was obtained from the Vindolanda Trustees in December 2014 (copy included in appendix 8.19).

5.3.2 Sample preparation

5.3.2.1 Brazil bones & teeth

The Brazil samples were prepared by cleaning and pulverising the entire pre-cut section of bone sample or tooth as described in Section 2.4.5. The samples were prepared sequentially, in two batches as shown in Table 5-1, with a full clean-down of the preparation area in between each sample.

5.3.2.2 Vindolanda femur

A 3 cm x 2.3 cm section of the femur was removed immediately adjacent to the area that had previously been removed by York University as shown in Figure 5.1. All surfaces were sanded, cleaned and treated with UV light to remove any exogenous DNA prior to pulverisation using a freezer mill as described in Section 2.4.5. A total of 6.9 g of powdered sample was collected.

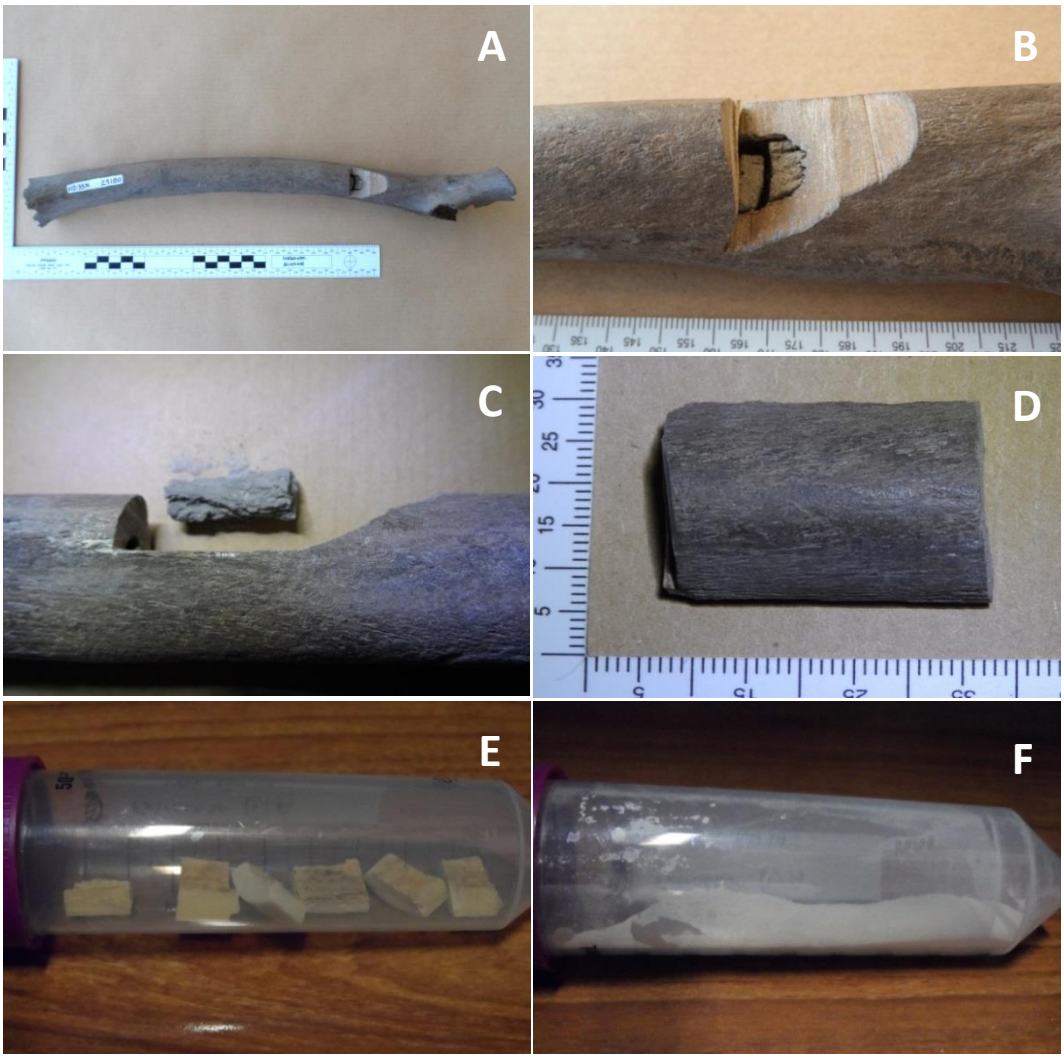


Figure 5.1 Images illustrating preparation of femur sample. 5.1A – image of Vindolanda femur sample, 5.1B – close up of section removed previously by York University, 5.1C – section removed, 5.1D – removed section cleaned and sanded to remove exogenous DNA, 5.1E – section cut into small pieces prior to pulverisation in freezermill, 5.1F – bone sample powdered.

5.3.2.3 Vindolanda tooth

The tooth was cleaned and treated with UV light in order to remove any foreign DNA from the outside surface. In an effort to preserve the tooth casing, the tooth was drilled as described in Section 2.4.5.3; 0.26 g of powdered sample was collected (Figure 5.2).

In a second attempt to obtain nuclear DNA, the entire tooth surface was re-cleaned and pulverised as detailed in Section 2.4.5, resulting in 1.12 g of sample powder

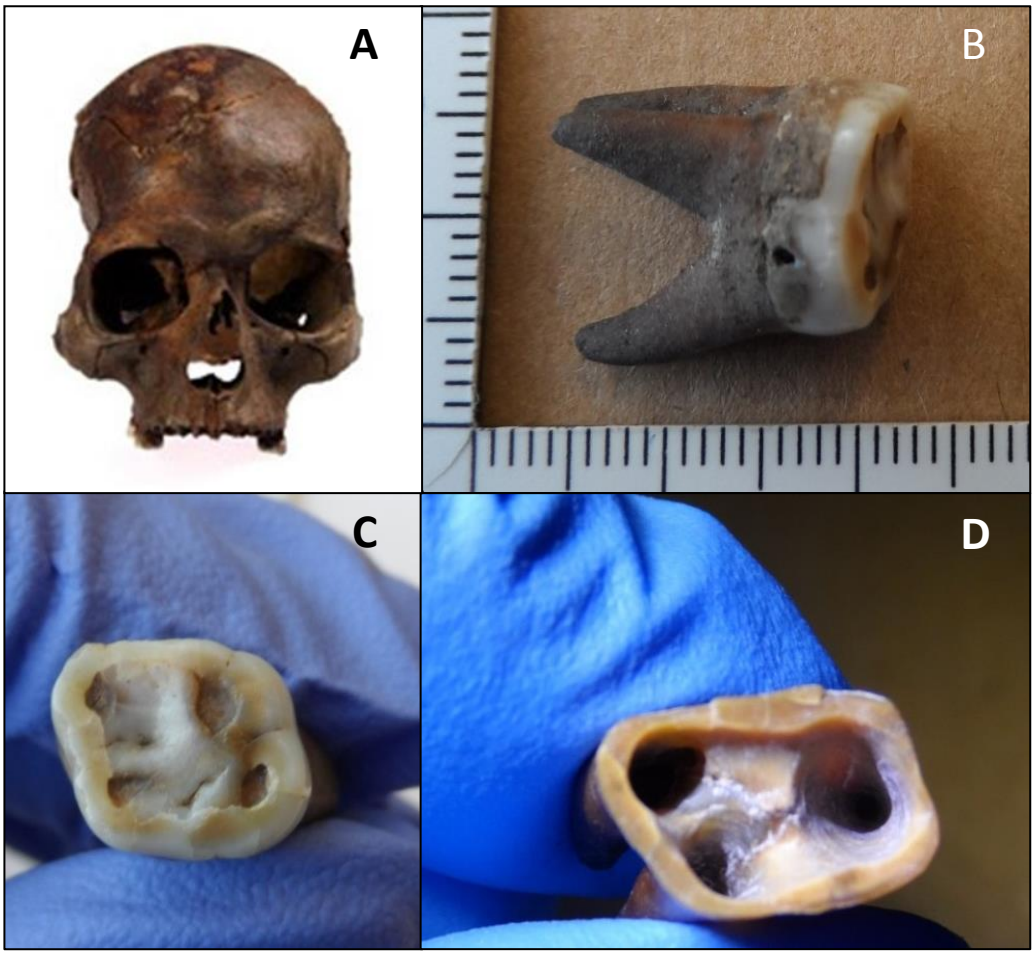


Figure 5.2 Images of the Vindolanda tooth sample. 5.2A – image of excavated skull prior to removal of the tooth sample (taken from http://www.vindolanda.com/educate/artefacts), 5.2B-side profile 5.2C – top section of crown, 5.2D – tooth case section post removal of inner material.

5.3.2.4 Control sample - *Cervus elaphus*

Tooth and bone samples from a Red deer (*Cervus elaphus*) were cleaned and prepared in exactly the same way as the human samples, one for each preparation batch and labelled as 'Negative A'. This sample was used as a process control and to monitor for the presence of any human contamination throughout processing.

5.3.3 DNA extraction

5.3.3.1 Reference buccal swabs

The reference buccal samples required for this work were collected as described in Section 2.4.2 and extracted using the QIAamp® DNA mini kit as described in Section 2.7.2.1.

5.3.3.2 Brazil bones and teeth

Samples 005B, 007B, 011B, 013B and 020B were extracted using phenol chloroform (method 1), silica column (method 2) and silica suspension (method 3) approaches, following the protocols detailed in Section 2.7.4. Samples 002T, 007T, 009B, 009T, 012B, 017B were extracted using method 1 only, samples 011T and 020T were extracted using only method 2, samples 001B-004B, 006B, 008B, 010B, 014B-016B, 018B and 019B were extracted using only method 3 as summarised in Table 5-1.

5.3.3.3 Vindolanda samples

Two sequential extractions of the pulverised femur were carried out, each one using a total of 3 g of bone powder, labelled 'Extraction 1' and 'Extraction 2'. The 3 g of bone powder was further split into six samples of 0.5 g, each of which was extracted as described in Section 2.7.4.3. Extracts were each eluted in 200 µl of TE buffer and were then combined and concentrated into approximately 60-80 µl, using a Microcon YM-30 centrifugal filter device, described in Section 2.7.4.4.

The first tooth extraction, consisting of 0.26 g of material removed from the inside surface of the casing, was extracted as outlined in Section 2.7.4.3. Volumes of the extraction solution and binding buffer were reduced to 6 ml and 24 ml respectively; all other volumes remained the same.

The pulverised whole tooth sample was split into two samples of 0.55 g and extracted in the same manner as the femur sample. Due to the limited quantity of sample, it was not possible to carry out a second independent extraction of this sample.

5.3.3.4 Negative control

A reagent negative control, labelled as 'Negative B' was set-up using an equivalent volume of digestion buffer as that added to the sample powder, and was then treated in exactly the same way as the samples throughout processing. This control was used to monitor any contamination of the DNA extraction reagents or occurring as a result of sample manipulation during processing.

Table 5-1 Summary of the sample details and preparation, extraction methods and amplification kit used for samples received from Ribeirão Preto, Brazil

Internal Sample reference	*Post mortem* interval (years)	Sex	Batch	Phenol chloroform (method 1)	Silica column (method 2)	Silica suspension (method 3)	AmpFℓSTR® NGM™ PCR Kit	mtDNA HVSI 'in-house' MiSeq assay	ForenSeq DNA Signature Kit
001B	14	(Male)	2			✓	✓		
002B	14	Male	2			✓	✓		
002T	14	Male	1	✓			✓		✓
003B	14	Male	2			✓	✓		
004B	14	Male	2			✓	✓		
005B	12	Male	1	✓	✓	✓*	✓	✓	✓
006B	12	Male	2			✓	✓		
007B	12	Female	1	✓	✓	✓*	✓	✓	✓
007T	12	Female	1	✓					
008B	11	Male	2			✓	✓		
009B	11	Male	1	✓			✓		✓
009T	11	Male	1	✓					
010B	11	Male	2			✓			
011B	11	Male	1	✓	✓*	✓	✓		✓
011T	11	Male	1		✓				

012B	8	(Female)	1	✓			✓		✓
013B	12	Male	1	✓	✓*	✓	✓	✓	✓
014B	12	Male	2			✓	✓		
015B	12	Female	2			✓	✓		
016B	12	Male	2			✓	✓		
017B	12	Male	1	✓			✓	✓	✓
018B	11	Male	2			✓	✓		
019B	10	Female	2			✓	✓		
020B	10	Male	1	✓	✓*	✓	✓	✓	✓
020T	10	Male	1		✓				

*denotes which DNA extract was used for NGM profile generation where multiple extractions were carried out

Sex listed in brackets reflects a conflict in assigning biological sex between the *ante mortem* and *post mortem* examinations with the *ante mortem* result provided

5.3.4 DNA quantification

5.3.4.1 Nuclear DNA

The nuclear DNA recovered from the Brazil bone and tooth samples 001-020, along with the two Vindolanda femur extracts and the Vindolanda tooth extract were each quantified in triplicate, using the Quantifiler® Trio DNA Quantification Kit as detailed in Section 2.8.3.

The degradation index (DI) was used to assess the level of DNA degradation in each of the samples. This was calculated using the formula:

$$DI = \frac{Small\ autosomal\ target\ [DNA]\ (ng\ /\ \mu l)}{Large\ autosomal\ target\ [DNA]\ (ng\ /\ \mu l)}$$

The level of degradation for each sample was then categorised according to Vernarecci et al. as indicated by the DI value; < 1-1.5 no degradation, 1.5-4 mildly degraded, 4-10 degraded and >10 severely degraded (Vernarecci et al., 2015).

An internal positive control (IPC) was run with each sample, which is a synthetic sequence of DNA that is not found in nature. If the expected Ct value of the IPC is delayed, this is indicative of the presence of inhibitors within the sample that are affecting DNA amplification. Samples were flagged as showing signs of PCR inhibition if their IPC value exceeded the mean Ct value for all of the standards in the run plus a variance of two, as suggested in the manufacturer's guidelines.

Sex was only assigned in samples where the total DNA concentration (short target) was determined to be >2 pg/µl. The sex was assigned as male when the average DNA concentration of the Y target exceeded that of the total DNA (short target) concentration. In cases where the average total DNA concentration (short target) is greater than the average value for male DNA (Y target), then the sample was determined to be single source only if the mean ± the standard deviation, for the two targets overlapped. If the two values did not overlap, i.e. the total DNA concentration exceeds the male DNA concentration; a mixed source of male/female DNA was then recorded. Where there was no amplification of the Y target observed, the sample was recorded as female.

The red deer negative controls were quantified using both the Quantifiler® Trio Kit, in order to detect any present human DNA and also the Qubit® 2.0 Fluorometer, along with the Qubit™ dsDNA HS Assay Kit as detailed in Section 2.8.5, to confirm a positive extraction result. The Nanodrop-1000 was used to quantify the DNA concentration of all reference samples required for this work in this section as described in Section 2.8.2 .

5.3.5 Mitochondrial qPCR assay

5.3.5.1 Assay design

In the absence of a commercially available kit, an in-house quantitative-PCR (qPCR) SYBR-Green assay for the quantification of mitochondrial DNA was designed and optimised for use. Amongst chordates, the arrangement of the thirty-seven mitochondrial genes is identical and so, as much of the mtGenome is conserved between species (Boore, 1999), the primary challenge when designing a qPCR assay was ensuring that it is human-specific. It is also preferable to amplify a target outside of the control region, which is required for sequencing, in order to minimalise the risk of laboratory contamination (Kavlick et al., 2011). Finally, as previously discussed, mtDNA is more usually targeted for DNA analysis in samples that are expected to contain degraded DNA and as such, the target fragment size should be less than 200bp to aid mtDNA detection in these sample types.

Cytochrome C oxidase is a transmembrane enzyme involved in the electron transport chain activity (Capaldi, 1990). It is a protein made up of thirteen subunits, the three largest of which COX1-3, are encoded for by the mtGenome (Kadenbach et al., 1983). The inter- and intra-species variability in the sequences of these subunits, predominately COX1 (Hebert et al., 2003) but also COX3 (Voigt, 2012), has been exploited for the DNA barcoding of species allowing rapid identification.

Sequences for the COX1-3 regions of the mtGenome were obtained from Genbank and aligned with the *Homo sapiens* sequence using the web-based version of Clustal Omega (v1.2.1) (McWilliam et al., 2013). A comparison was first carried out between the sequence for *Homo sapiens* and the genetically closely related higher primates (accession numbers listed in Table 5-2), looking for two clusters of human-specific SNPs separated by no more than 150 bases, over which human-specific primers could be designed. One such example was identified in the COX3 sequence as a possible target for amplification, shown in Figure 5.3.

Forward and reverse primers were then designed to incorporate three of the human-specific SNPs and checked, first manually, against eleven species, Figure 5.4, and then across all organisms using Primer-BLAST (Ye et al., 2012) for specificity. Only four results were returned from the Primer-BLAST search, three of which were extinct species closely related to *Homo sapiens* (see Figure 5.5).

Table 5-2 Accession numbers for sequences of species used in assay design

Species	Genbank accession number
Homo sapiens	NC_012920.1
Gorilla gorilla	NC_011120.1
Pan troglodytes	NC_001643.1
Pan paniscus	NC_001644.1
Cervus elaphus	NC_007704.2
Lemur catta	NC_004025.1
Canis lupus	AY656745.1
Felis catus	NC_001700.1
Mus musculus	NC_005089.1
Scomberomorus semifasciatus	NC_021391.1
Xenopus laevis	NC_001573.1
Bos taurus	AF492351.1

```
Gorilla_gorilla     GGTGTCTCAATTACCTGAGCCCACCATAGCCTAATAGAAAATAACCGTAACCAAATAATT
Human_Cyt_COX3      GGAGTATCAATCACCTGAGCTCACCATAGTCTAATAGAAAACAACCGAAACCAAATAATT
Pan_troglodytes     GGAGTATCAATTACTTGAGCCCATCACAGCTTAATAGAAAATAACCGAAACCAAATAATT
Pan_paniscus        GGAGTATCAATTACTTGAGCCCATCACAGCTTAATAGAAAATAACCGAAACCAAATAATT
                    ** ** ***** ** ***** ** ** **  ********** ***** ************

Gorilla_gorilla     CAAGCACTACTTATCACAATTTTACTGGGCCTCTACTTCACCCTCCTACAAGCCTCAGAG
Human_Cyt_COX3      CAAGCACTGCTTATTACAATTTTACTGGGTCTCTATTTTACCCTCCTACAAGCCTCAGAG
Pan_troglodytes     CAAGCACTGCTTATTACGATTCTACTAGGTCTTTATTTTACCCTCCTACAAGCCTCAGAA
Pan_paniscus        CAAGCACTACTCATTACAATTTTACTAGGTCTCTATTTCACCCTCCTACAAGCCTCAGAA
                    ******** ** ** ** *** **** ** ** ** ** ********************

Gorilla_gorilla     TACTTTGAAGCCCCCTTTACCATTTCCGATGGTATCTATGGCTCAACATTTTTTGTAGCC
Human_Cyt_COX3      TACTTCGAGTCTCCCTTCACCATTTCCGACGGCATCTACGGCTCAACATTTTTTGTAGCC
Pan_troglodytes     TATTTCGAATCCCCTTTTACCATTTCCGATGGCATCTACGGCTCAACATTCTTTGTAGCC
Pan_paniscus        TACTTCGAATCCCCTTTTACCATTTCCGATGGCATCTACGGCTCAACATTCTTTGTAGCC
                    ** ** **  * ** ** *********** ** ***** *********** *********
```

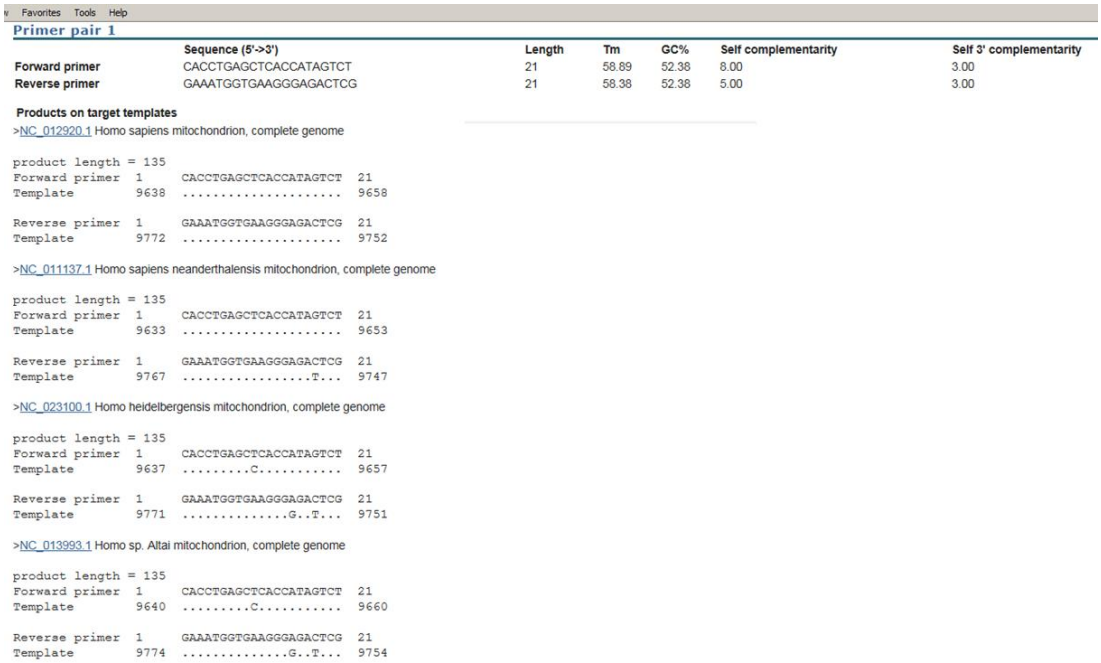
Figure 5.3 Clustal Omega (v1.2.1) alignment of the *Homo sapiens* mtGenome COX3 region with three higher primates; human-specific SNPs are highlighted in grey with a distance of 139 bases between the two furthest positions

```
Scom_semifasciatus  GGAGTTACAGTTACCTGAGCTCACCACAGCATTATAGAAGGCAATCGAAAAGAAGCAATC
Xenopus_laevis      GGAGTTACAGTTACCTGAGCTCACCACAGCATTATAGAAGGCAATCGAAAAGAAGCAATC
Mus_musculus        GGTGTTTCAATTACATGAGCTCATCATAGCCTTATAGAAGGTAAACGAAACCACATAAAT
Lemur_catta         GGCGTATCCATCACCTGAGCTCACCATAGCCTAATAGAAGGAAACCGAGTACAAATACTC
Gorilla_gorilla     GGTGTCTCAATTACCTGAGCCCACCATAGCCTAATAGAAAATAACCGTAACCAAATAATT
Human_Cyt_COX3      GGAGTATCAATCACCTGAGCTCACCATAGTCTAATAGAAAACAACCGAAACCAAATAATT
Pan_troglodytes     GGAGTATCAATTACTTGAGCCCATCACAGCTTAATAGAAAATAACCGAAACCAAATAATT
Pan_paniscus        GGAGTATCAATTACTTGAGCCCATCACAGCTTAATAGAAAATAACCGAAACCAAATAATT
Canis_familiaris    GGAGTATCTATTACTTGAGCCCATCATAGTTTAATAGAAGGTAATCGCAAACATATACTT
Felis_catus         GGAGTATCAATCACCTGAGCTCACCACAGTTTGATGGAGGGAAATCGAAAACACATGCTT
Cervus_elaphus      GGAGTTTCTATCACCTGAGCCCATCATAGCCTTATAGAAGGAAACCGCAACCATATACTA
Bos_taurus          GGAGTTTCTATTACCTGAGCCCATCATAGTTTAATAGAAGGGGACCGAAAGCATATATTA
                    ** **  *  * ** ***** ** ** **  * ** **     * **     *

Scom_semifasciatus  CAGTCTCTAGCACTGACAATTCTCCTCGGTTTCTACTTCACCTTCCTTCAAGCTATAGAA
Xenopus_laevis      CAGTCTCTAGCACTGACAATTCTCCTCGGTTTCTACTTCACCTTCCTTCAAGCTATAGAA
Mus_musculus        CAAGCCCTACTAATTACCATTATACTAGGACTTTACTTCACCATCCTCCAAGCTTCAGAA
Lemur_catta         CAAGCTCTACTCATCACTATTACCCTAGGTCTCTATTTCACACTCCTCCAAGCCTCAGAA
Gorilla_gorilla     CAAGCACTACTTATCACAATTTTACTGGGCCTCTACTTCACCCTCCTACAAGCCTCAGAG
Human_Cyt_COX3      CAAGCACTGCTTATTACAATTTTACTGGGTCTCTATTTTACCCTCCTACAAGCCTCAGAG
Pan_troglodytes     CAAGCACTGCTTATTACGATTCTACTAGGTCTTTATTTTACCCTCCTACAAGCCTCAGAA
Pan_paniscus        CAAGCACTACTCATTACAATTTTACTAGGTCTCTATTTCACCCTCCTACAAGCCTCAGAA
Canis_familiaris    CAAGCCTTATTCATTACAATCTCCTTAGGCGTATATTTTACGCTATTACAGGCCTCCGAA
Felis_catus         CAAGCACTATTTATTACAATCTCTTTAGGGGTCTACTTTACACTCCTCCAAGCCTCCGAA
Cervus_elaphus      CAAGCCTTATTTATTACTATCGCACTAGGCGTCTATTTTACACTACTACAAGCCTCAGAA
Bos_taurus          CAAGCCCTATTTATCACCATCACATTAGGAGTCTACTTCACACTACTACAAGCCTCAGAA
                    **  *  *     * ** **     * **  * ** ** **  *  * ** **    **

Scom_semifasciatus  TACTATGAAGCTCCCTTTACAATCGCAGACGGAGTCTACGGTTCCACATTCTTTGTAGCC
Xenopus_laevis      TACTATGAAGCTCCCTTTACAATCGCAGACGGAGTCTACGGTTCCACATTCTTTGTAGCC
Mus_musculus        TACTTTGAAACATCATTCTCCATTTCAGATGGTATCTATGGTTCTACATTCTTCATGGCT
Lemur_catta         TATTTTGAAACATCCTTTACAATCTCAGACGGCGTGTATGGTTCTACATTCTTTATGGCA
Gorilla_gorilla     TACTTTGAAGCCCCCTTTACCATTTCCGATGGTATCTATGGCTCAACATTTTTTGTAGCC
Human_Cyt_COX3      TACTTCGAGTCTCCCTTCACCATTTCCGACGGCATCTACGGCTCAACATTTTTTGTAGCC
Pan_troglodytes     TATTTCGAATCCCCTTTTACCATTTCCGATGGCATCTACGGCTCAACATTCTTTGTAGCC
Pan_paniscus        TACTTCGAATCCCCTTTTACCATTTCCGATGGCATCTACGGCTCAACATTCTTTGTAGCC
Canis_familiaris    TACTATGAGACATCTTTTACAATCTCCGATGGGGTGTACGGATCTACCTTTTTTATAGCC
Felis_catus         TACTATGAAACATCATTCACGATCTCGGACGGAGTATACGGATCTACCTTCTTCATGGCC
Cervus_elaphus      TACTATGAAGCACCCTTTACCATTTCAGACGGAGTTTATGGCTCAACCTTCTTTGTAGCT
Bos_taurus          TACTATGAAGCACCTTTTACTATCTCCGACGGAGTTTACGGCTCAACTTTTTTTGTAGCC
                    ** *  **  *  * **  * **  * ** **  * ** ** ** ** ** **  * **
```

Figure 5.4 Clustal Omega (v1.2.1) alignments of the target fragment of the mtGenome COX3 region for a variety of species. Sequences were obtained from Genbank for comparison, the accession numbers for which are listed in Table 5-2. Proposed human-specific primer hybridisation sites are highlighted in grey and nucleotides which differ from this sequence are highlighted in pale grey.



Favorites Tools Help

Primer pair 1

	Sequence (5'->3')	Length	Tm	GC%	Self complementarity	Self 3' complementarity
Forward primer	CACCTGAGCTCACCATAGTCT	21	58.89	52.38	8.00	3.00
Reverse primer	GAAATGGTGAAGGGAGACTCG	21	58.38	52.38	5.00	3.00

Products on target templates

```
>NC_012920.1 Homo sapiens mitochondrion, complete genome

product length = 135
Forward primer  1     CACCTGAGCTCACCATAGTCT  21
Template        9638  .....................  9658

Reverse primer  1     GAAATGGTGAAGGGAGACTCG  21
Template        9772  .....................  9752

>NC_011137.1 Homo sapiens neanderthalensis mitochondrion, complete genome

product length = 135
Forward primer  1     CACCTGAGCTCACCATAGTCT  21
Template        9633  .....................  9653

Reverse primer  1     GAAATGGTGAAGGGAGACTCG  21
Template        9767  ................T...  9747

>NC_023100.1 Homo heidelbergensis mitochondrion, complete genome

product length = 135
Forward primer  1     CACCTGAGCTCACCATAGTCT  21
Template        9637  .........C...........  9657

Reverse primer  1     GAAATGGTGAAGGGAGACTCG  21
Template        9771  .............G..T...  9751

>NC_013993.1 Homo sp. Altai mitochondrion, complete genome

product length = 135
Forward primer  1     CACCTGAGCTCACCATAGTCT  21
Template        9640  .........C...........  9660

Reverse primer  1     GAAATGGTGAAGGGAGACTCG  21
Template        9774  .............G..T...  9754
```

Figure 5.5 Screenshot of Primer-BLAST (Ye et al., 2012) page showing four results returned from a search against all organisms. *Homo sapiens neanderthalensis*, *Homo heidelbergensis* and *Homo sapiens Altai* are now extinct, and so this assay is highly modern human-specific, allowing for an accurate quantification of the human mtDNA within a sample. Both primers are 21 nucleotides in length with a GC content of approximately 52% and a Tm of 58-59 ºC.

5.3.5.2 Standards

A 200 bp synthetic gene fragment was used to generate qPCR standards containing a known copy number of the COX3 target so that the mtDNA copy number in each sample could be determined. The DNA fragment was ordered as two single-stranded complementary oligonucleotides (Integrated DNA Technologies, Coralville, IA, USA), which were each resuspended to a concentration of 10 µM in primer resuspension buffer as described in Section 2.8.1.1 .

The following calculations were utilised to determine the mtDNA copy number of synthetic standard.

Forward strand

1 copy MW = 61095.6 g/mol (provided by IDT)

1 copy = $\frac{61095.6\ g/mol}{6.022\times 10^{23}\ mol}$ (divide by mol)

1 copy = $1.01*10^{-19}$ g ($1.01*10^{-10}$ ng)

Total amount of oligo = 90,000 ng (provided by IDT)

Total number of copies = $\frac{90000\ ng}{1.01\times10^{-10}\ ng}$ (mass of total amount of oligo/mass of 1 copy)

Total number of copies = 8.91 *10^{14}

Resuspended to 10 µM in 153.1 µl of primer resuspension buffer

1 µl = $\frac{8.91\times10^{14}\ copies}{153.1\ \mu l}$

1 µl = 5.82*10^{12} copies in a 10 µM stock

A working solution was prepared with a concentration of 2 µM, by adding 10 µl of 10 µM stock, containing total of 5.82*10^{13} copies of the DNA fragment, to 40 µl of MBG water and mixed briefly by vortex.

Theoretical no. of copies in 1 µl = $\frac{5.82\times10^{13}\ copies}{50\ \mu l}$

Theoretical no. of copies in 1 µl of a 2 µM solution = 1.16*10^{12}

The 2 µM oligonucleotide solution was quantified using the NanoDrop® Spectrophotometer ND-1000 with software ND-1000 V3.8.1 as described in Section 2.7.2, in order to determine the actual number of copies of the fragment in this diluted solution.

Average DNA concentration = 103.7 ng/µl

Actual no. of copies in 1µl = $\frac{103.7\ ng}{1.01\times10^{-10}\ ng}$

Actual no. of copies in 1 µl of 2 µM solution = 1.03*10^{12}

Complimentary strand

1 copy MW = 62345.4 g/mol

1 copy = $\frac{62345.4\ g/mol}{6.022\times 10^{23}\ mol}$

1 copy = 1.04*10^{-19} g (1.04*10^{-10} ng)

Total amount of oligo = 30,000 ng

Total number of copies = $\frac{30000\ ng}{1.04\times10^{-10}\ ng}$

Total number of copies = 2.90 *10^{14}

Resuspended to 10 µM in 50.1 µl of primer resuspension buffer

1 µl = $\frac{2.90\times10^{14}\ copies}{50.1\ \mu l}$

1 µl = 5.79*10^{12} copies in a 10 µM stock

A working solution of the oligonucleotide was again prepared with a concentration of 2 µM, by adding 10 µl of 10 µM stock, containing total of 5.79*10^{13} copies of the DNA fragment, to 40 µl of MBG water and mixing briefly by vortex.

Theoretical no. of copies in 1 µl = $\frac{5.79\times10^{13}\ copies}{50\ \mu l}$

Theoretical no. of copies in 1 µl of a 2 µM solution = 1.16*10^{12}

The 2µM oligonucleotide solution was quantified using the NanoDrop® Spectrophotometer as previously described in order to determine the actual number of copies of the fragment in this diluted solution.

Average DNA concentration = 116.0 ng/µl

Actual no. of copies in 1µl = $\frac{116.0\ ng}{1.04\times10^{-10}\ ng}$

Actual no. of copies in 1 µl of 2 µM solution = 1.12*10^{12}

Both oligonucleotides were then diluted to a concentration of 1*10^{10} as shown below:

Forward strand – stock = 1.03*10^{12} copies/µl

Dilution factor = $\frac{(1.03\times10^{12}\ copies/\mu l)\times 1\mu l}{1\times10^{10}\ c\ opies}$

Dilution factor = 103

1 µl of 2 µM oligonucleotide solution was added to 102µl of MBG water and centrifuged briefly to mix.

Complimentary strand – stock = 1.12*10^{12} copies/µl

Dilution factor = $\frac{(1.12\times10^{12}\ copies/\mu l)\times 1\ \mu l}{1\times10^{10}\ copies}$

Dilution factor = 112

1 µl of 2 µM oligonucleotide solution was added to 111µl of MBG water and centrifuged briefly to mix.

Both oligonucleotides were transferred to a fresh tube in a 1:1 ratio to give a concentration of $1*10^9$ copies/µl. The tube was heated to 95 °C for 3 minutes and then left to cool slowly at room temperature.

All stock and working solutions of both oligonucleotides were stored at -80 °C.

5.3.5.3 Optimisation

The primer annealing temperature was tested between 55-60 ºC using the thermal gradient feature on the CFX96 Touch™ Real-Time PCR Detection System and although it was marginal, 60 ºC was determined to be the optimum. The specificity of the primer set was demonstrated by the single melt peak observed for all samples tested indicating the presence of one PCR product only as shown in Figure 5.6.

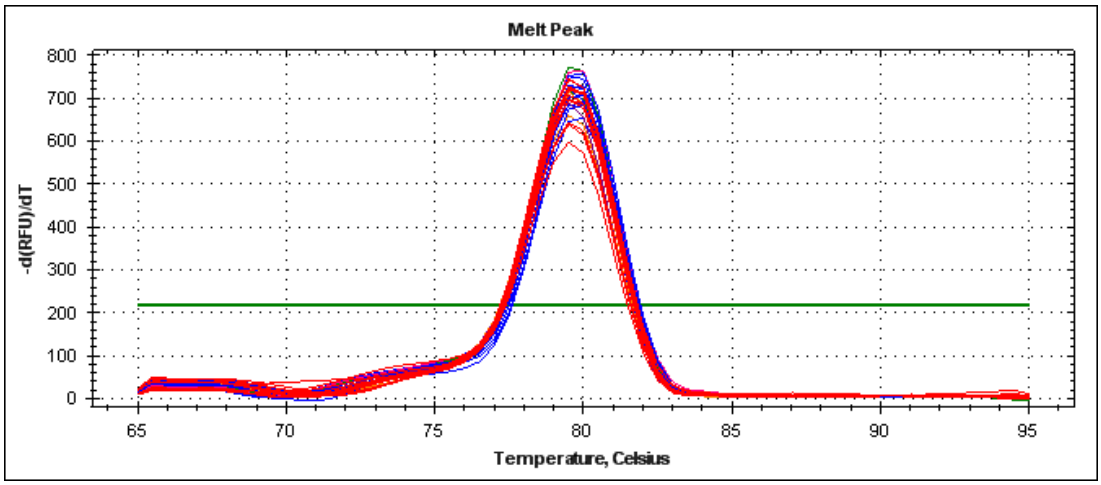


Figure 5.6 Single melt peak generated post PCR cycling steps, demonstrating the presence of just one PCR product. Extremely low levels of primer dimer were also observed, falling well below the Ct detection threshold, thus having no impact upon the Ct values.

As can be seen in Figure 5.7, although mtDNA was amplified down to the lowest DNA input of 50 copies/µl, the linear range of the assay was determined to be $5*10^3$ – $5*10^9$ copies/µl as the linear relationship between the quantity of DNA and the Ct value was lost below this value. Whilst samples with an mtDNA DNA copy number below 5000 will be detected, the copy number cannot be accurately determined.

The assay performance can be measured through comparison with the theoretical number of cycles (n) between Ct values for adjacent standards, expected for a perfect doubling reaction using the equation $2^n = X$, where X is the dilution factor used in generating the standard curve. The theoretical value for this assay was calculated as 3.32 cycles and the actual value as an average of 3.34 cycles with a standard deviation of 0.7, indicating that the assay is highly optimised. This is supported by both the amplification efficiency value of 96%, i.e.

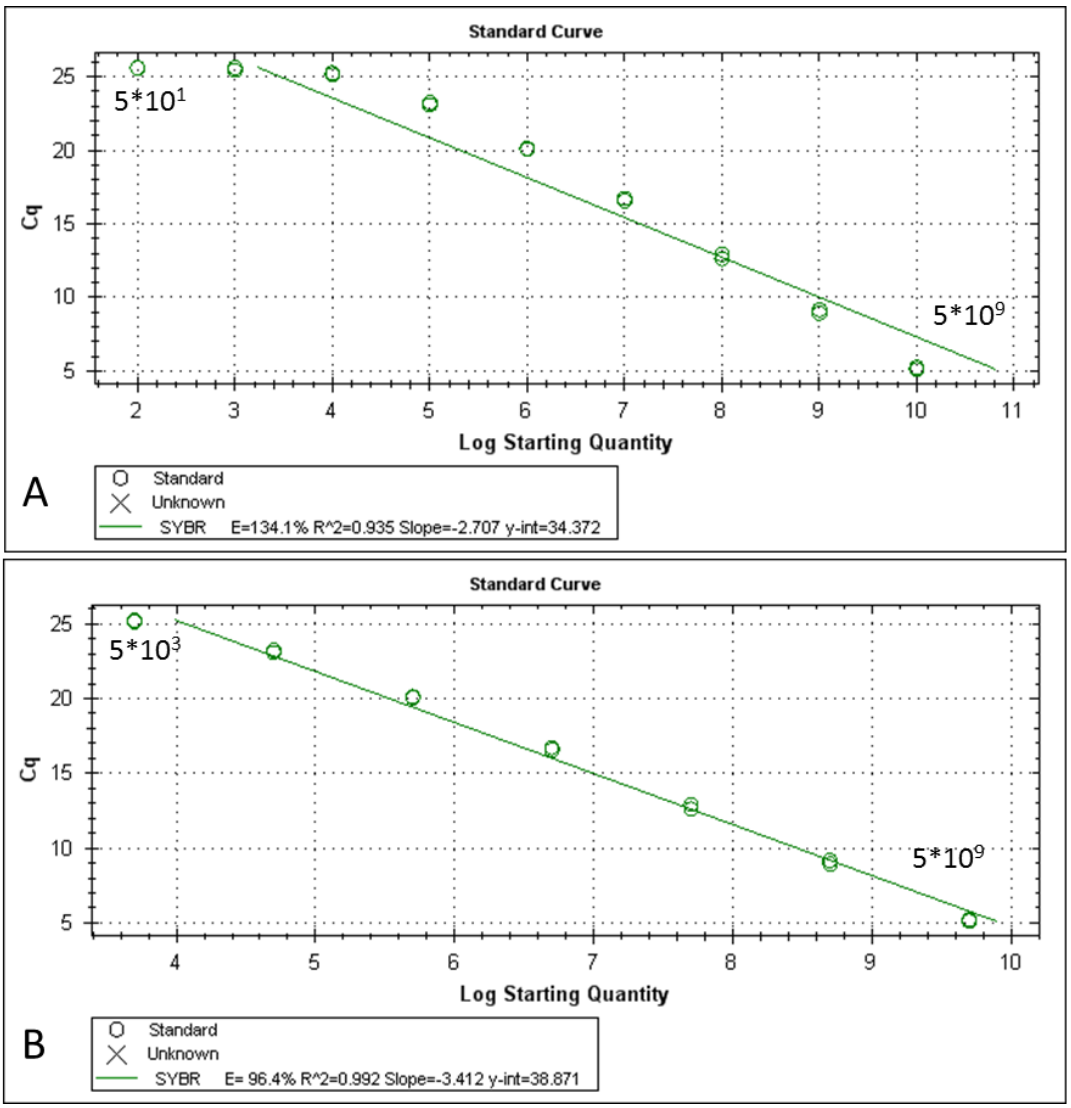


Figure 5.7 Standard curve generated for the mtDNA synthetic standard using a 1 in 10 serial dilution of the stock concentration for the range of $5*10^9$–$5*10^1$ copies/µl (3XA) or $5*10^9$–$5*10^3$ copies/µl (3XB). The linear relationship is lost below $5*10^3$ copies/µl as demonstrated by the plateau in the Ct (Cq) detection threshold for lower values and high amplification efficiency value of 134%. Copy number values calculated using the standard curve equation shown in 3XA would be inaccurate. The linear relationship is demonstrated for the standards $5*10^9$ – $5*10^3$ copies/µl in 3XB.

96% of the DNA template was amplified at the end of each cycle, and the R^2 value of 0.99.

The sensitivity of the assay was tested using samples with a known DNA concentration by preparing a serial dilution (1 in 10) using Quantifiler Trio standard DNA with DNA input concentrations of 5000 pg/µl, 500 pg/µl, 50 pg/µl and 5 pg/µl. As shown in Figure 5.8, amplified mtDNA was detected for all samples within the linear range of the assay thus demonstrating its robustness at low levels of DNA.

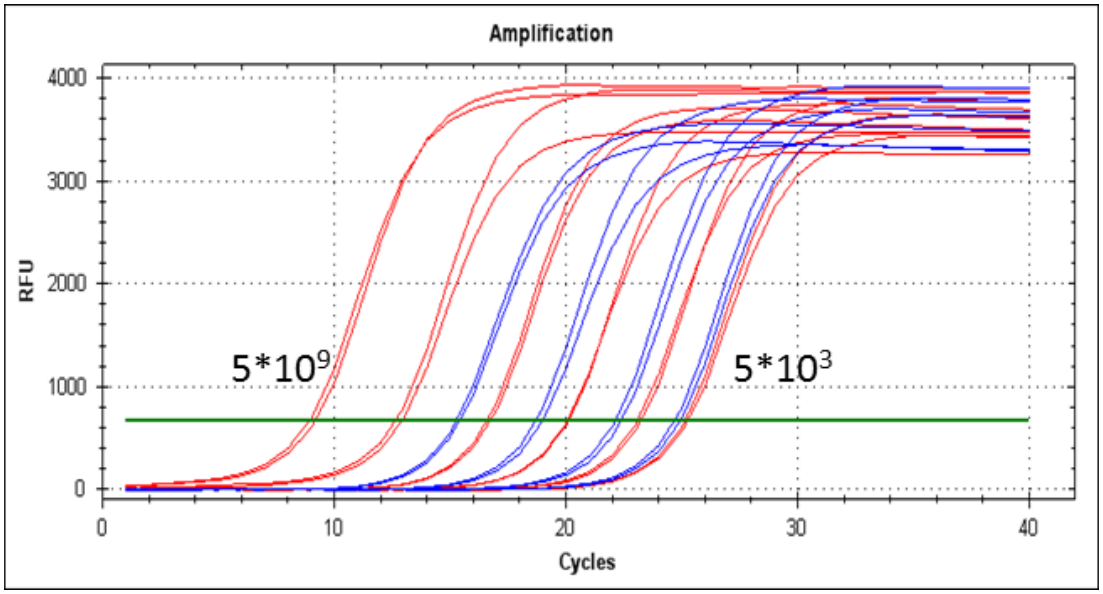


Figure 5.8 Sensitivity test results for mtDNA qPCR assay showing the amplification curves for a 1 in 10 serial dilution at DNA inputs of approximately 5000 pg/µl, 500 pg/µl, 50 pg/µl and 5 pg/µl (blue). All samples were quantified within the range of the mtDNA standards (red).

5.3.5.4 Species specificity test

In order to check for non-human mtDNA amplification, DNA extracts from eight species; fox, pig, hedgehog, deer, dog, cat *pseudomonas aeruginosa*, mc and fungus, were tested using the qPCR assay. 007 Control DNA was used at an input of 50 pg/µl as a point of comparison against the species tested. All species had an average Ct value of < 28, and so low level amplified product was detected. However the extraction negative control had a Ct value of 29, suggesting very low level human contamination may be present in all samples; if compared with the human control of just 50 pg/µl, this must be equivalent to only several pg/µl DNA. As can be seen in Figure 5.9 the Ct values for all species tested falls below that of even the last standard, equating to only a few copies of mtDNA/µl. As all species DNA was freshly extracted, if amplification was of a non-specific product/target then a greater Ct value would have been observed. The results therefore demonstrate the human-specific nature of this assay.

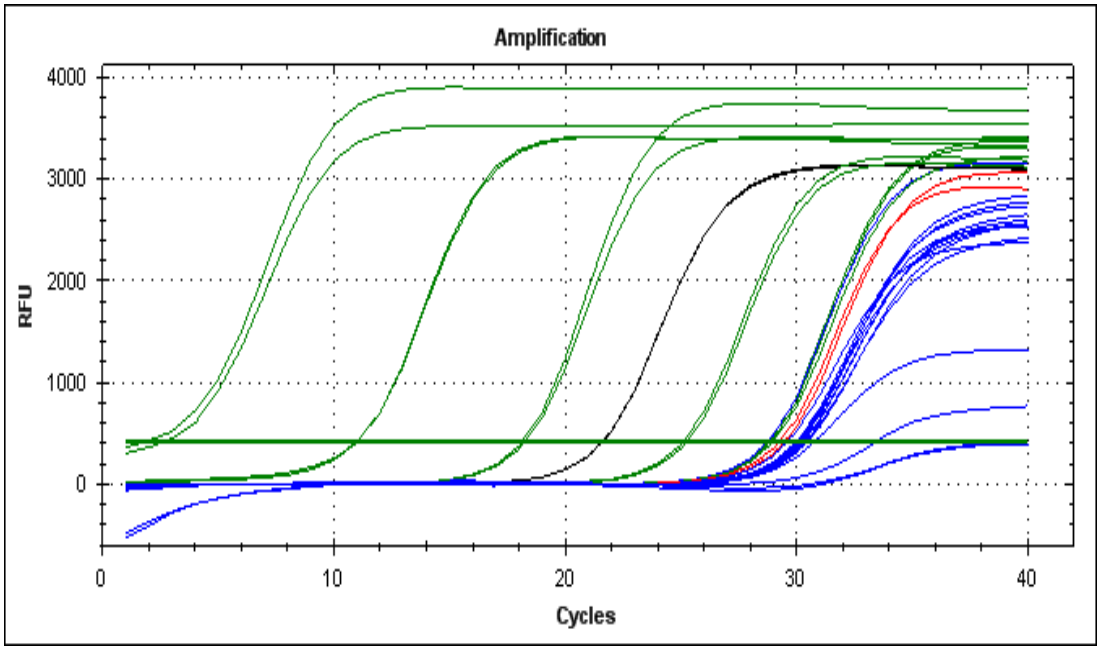


Figure 5.9 Human specificity test results. Standards at 10^9- 10^1 copies (green). Human control 007 DNA used as a positive control at an input of 0.5 pg/µl (black), extraction negative control (red) and all species samples (blue)

5.3.5.5 Reaction set-up for mtDNA assay

Each of the Brazil bone and tooth samples, alongside a fresh set of mtDNA standards, were set-up for qPCR using the in-house mtDNA assay in a 10 µl volume and amplified using a CFX96 Touch™ Real-Time PCR Detection System as described in Section 2.8.4.

5.3.6 DNA amplification, visualisation and data analysis

5.3.6.1 AmpFℓSTR® NGM™ PCR Amplification Kit

The AmpFℓSTR® NGM™ PCR Amplification Kit was used to amplify each of the Vindolanda DNA extracts, Brazil samples 001B–020B and 002T and all control samples. Reactions were set-up using 10 µl of DNA extract in a 25 µl reaction volume as described in Table 2-7 and amplified on a Veriti thermal cycler for 30 cycles according to manufacturer's instructions as shown in Table 2-8. Two amplification reactions were set-up for each DNA extract to enable consensus DNA profile interpretation. The exception to this was with the Vindolanda femur extracts, which were compared with one another to generate a consensus profile. Reference samples were set-up using 5 µl of DNA extract in a 12.5 µl reaction as described in Table 2-7, and amplified using the 28-cycle program described in Table 2-8.

1 µl of PCR product was used, together with Hi-Di™ Formamide and GeneScan™ 500 LIZ™ dye Size Standard for visualisation using the 3130 Genetic Analyzer as described in Section 2.13.2.1. Results were analysed as described in Section 2.13.3.1, using GeneMapper ID v3.2, with either standard or ltDNA guidelines followed for reference or bone/tooth samples respectively.

5.3.6.2 AmpFℓSTR® Yfiler® PCR Amplification Kit

10 µl of each of the Vindolanda extracts was used to set-up amplification AmpFℓSTR® Yfiler® PCR Amplification Kit reactions for a 34-cycle PCR. The tooth extract was set-up in duplicate so that a consensus profile could be generated, whilst the femur results were compared with one another to generate the consensus profile. Reaction volumes of 25 µl were used for each reaction as shown in Table 2-7 and amplification was carried out on a Veriti thermal cycler as instructed by the manufacturer, outlined in Table 2-8.

1 µl of PCR product was used, together with Hi-Di™ Formamide and GeneScan™ 500 LIZ™ dye Size Standard for visualisation using the 3130 Genetic Analyzer as described in Section 2.13.2.1. Results were analysed as described in Section 2.13.3.1, using GeneMapper ID v3.2, with either standard or ltDNA guidelines followed for reference or bone/tooth samples respectively.

5.3.6.3 ForenSeq™ DNA Signature Kit

20 µl of DNA extract from each of the Vindolanda femur and tooth samples, along with Brazil samples 002T, 005B, 007B, 009B, 011B, 012B, 013B, 017B & 020B and an operator reference sample, were sent to the Illumina Inc. laboratory (Little Chesterford, Essex, UK) for amplification using the ForenSeq™ DNA Signature Prep Kit and sequencing on the MiSeq FGx Forensic Genomics System (see appendix 8.20 for workflow details). The Brazil samples were amplified using the database compatible primer set A, whilst primer set B was used for the Vindolanda samples, which includes additional phenotypic and ancestry markers.

5.3.6.4 mtDNA amplification for Sanger sequencing for York results comparison

In order to support the authenticity of the DNA obtained from the Vindolanda femur sample, 5 µl of the two DNA extracts was used for each of the two primer sets listed in Table 2-18 covering regions of HVSII (primer set H) and HVSIII (primer set J) of the mtGenome, which had previously been sequenced by York University. Following the extraction of DNA from a section of the femur sample using a silica based method, as described in

Rohland et al. (Rohland et al., 2010), researchers at York University then amplified the control region of the mtGenome using nine overlapping primer sets (Gonzalez Fortes, 2014)

Reactions were set-up as described in Table 2-12 in a total volume of 20 µl and amplified on a Veriti thermal cycler following the 35-cycle program outlined in Table 2-21 and Table 2-22. Reference profiles for all mtDNA work were set-up using all of the primer sets listed in Table 2-18 in the same way as above, for sequencing on the 3130 Genetic Analyzer.

To prepare amplicons for sequencing by the 3130 Genetic Analyzer, PCR product was first cleaned using the QuickStep™ 2 PCR Purification Kit as described in Section 2.10.1. 2 µl of the purified PCR product was then used to set-up the sequencing reaction with the BigDye® Terminator v1.1 Cycle Sequencing Kit as explained in Section 2.11.1. Reaction product was then cleaned prior to sequencing using PERFORMA® DTR Gel Filtration Cartridges as outlined in Section 2.12.1.

5.3.6.5 Illumina D-loop protocol amplification of HVSI/HVSII regions of the mtGenome

Samples 004B, 008B, 009T, 019B, 013B and 014B were each set-up for amplification of four overlapping DNA fragments (two for each HVSI/II region) using the primer set mixes listed in Table 2-20, as described by the Illumina D-loop protocol. Reactions were set-up in a 20 µl volume using optimised volumes as described in Table 2-14 to Table 2-17 using 5 µl of DNA extract per primer mix. PCR was carried out using a Veriti thermal cycler using the parameters outlined in Table 2-21 and Table 2-22. Control DNA 007 was used as a PCR-positive control for each primer set with a total DNA input of 50 pg. PCR product was visualised on a 3% (w/v) agarose gel as described in Section 2.13.1.2.

5.3.6.6 Amplification of HVSII set E of the mtGenome

Samples 004B, 008B, 009T, 019B, 013B and 014B were each set-up for amplification of one fragment of the mtGenome HVSII region using Edson primer set E, the sequence for which is listed in Table 2-18. The amplification of this 125 bp target was used to demonstrate the presence of mtDNA in samples that had failed PCR using the Illumina suggested primer sets. Reactions were set-up using 5 µl of DNA extract per primer set, in a total reaction volume of 20 µl as outlined in Table 2-12 and amplified on a Veriti thermal cycler using the parameters outlined in Table 2-21 and Table 2-22. AmpFlSTR® Control DNA 007 was used as a PCR positive control for each primer set with a total DNA input of 50 pg. PCR products were visualised on a 3% (w/v) agarose gel as described in Section 2.13.1.2

Data produced by sequencing using the 3130 Genetic Analyzer were analysed as described in Section 2.13.3.2.

5.3.6.7 mtDNA in-house HVSI amplification for MiSeq sequencing

To allow for the MPS of DNA from extremely degraded samples using a one-step PCR process, the protocol developed by Kozich et al. (Kozich et al., 2013) has been adapted for use amplifying fragments of mtDNA. Rather than fragmenting the DNA library and adding the platform-specific adaptors and sample indices in a second PCR step, all the required components are added to the primers, see Figure 5.10, for incorporation into the amplicon during target amplification. This has the dual advantage of decreasing the number of sample manipulations, and therefore time taken for sample preparation, and should also simplify sequence alignment during data analysis.

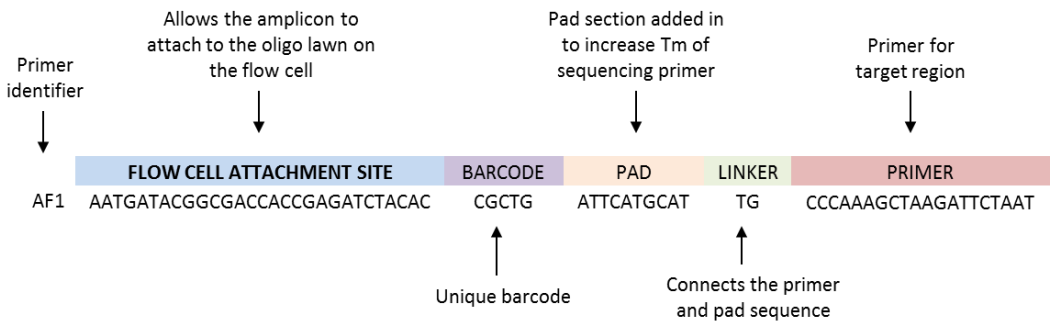
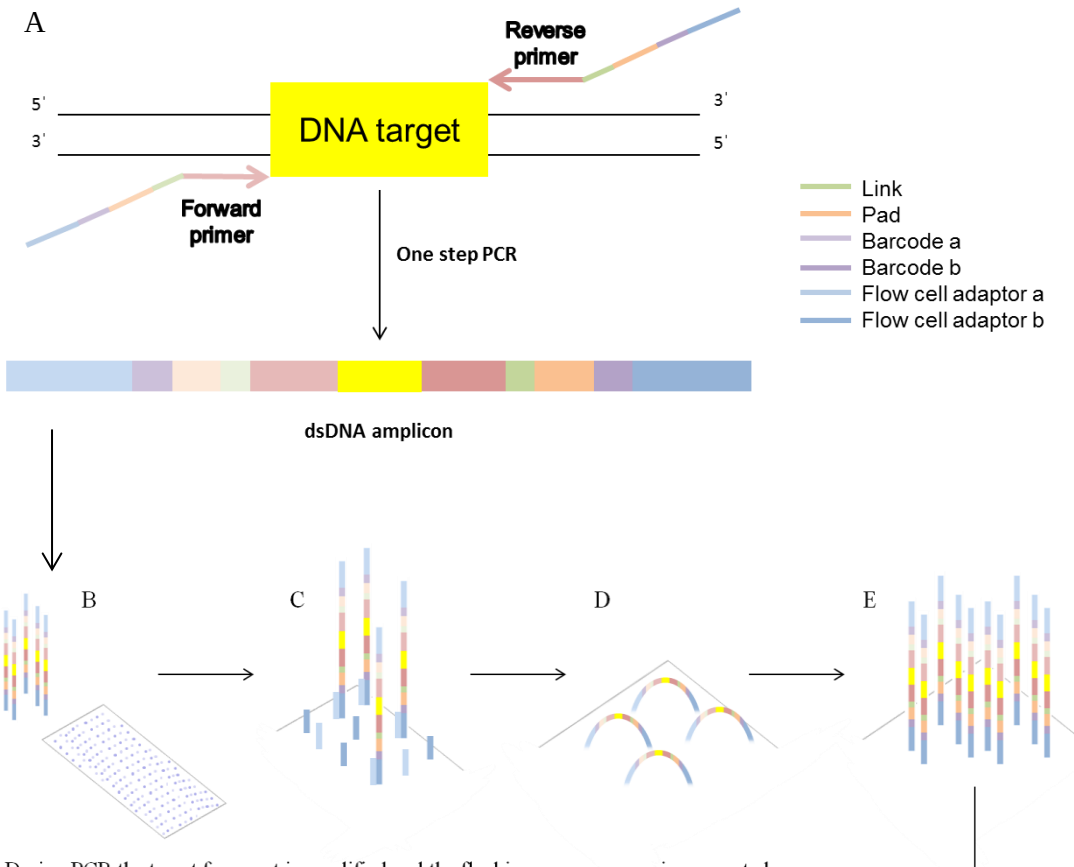


Figure 5.10 Design of modified Edson (Edson, 2004) primer sets, adapted for direct sequencing using the Illumina MiSeq MPS platform. Each modified primer has a flow cell adapter attachment enabling the amplified DNA fragment to hybridise to the flow cell, a unique barcode for amplicon identification, a pad sequence to allow the melting temperature of the amplicon to be manipulated, a linker sequence to attach the additional sections to the primer, and the primer to target the specific region of DNA to be amplified.

There are two types of oligo adaptors that are used to hybridise the DNA strands to the flow cell surface of the MiSeq platform (see Figure 5.11), the sequences for which were released by Illumina in 2014 (Illumina, 2014a). One version of the oligo adaptor is attached to the forward primer and the second version is incorporated with the reverse primer. Five nucleotide index sequences were designed using the online tool COMAIWIKI Barcode Generator (Comai, 2012), one for each of the forward and reverse primer sequences that together make up a unique sample identifier. The design of these barcodes took into consideration the following factors: that the detection of individual clusters on the flowcell occurs within the first four sequencing cycles and therefore there needs to be enough diversity between these bases to enable cluster detection (Mitra et al., 2015); that at least one of the two nucleotides for detection by each of the red (A/C) and green (G/T) lasers is

included for each sequencing cycle in order to ensure correct registration, and that there should be an approximately 25% distribution of each nucleotide across the index (Kozich et al., 2013). The following design parameters were entered into the barcode generator: a length of five nucleotides, a minimum of two nucleotides genetic difference between indices and a GC content of 20-80%. The suggested indices sequences with the associated primers are listed in appendix 8.21 and for the index combinations utilised for each sample see Table 5-3.

Two, 10-nt pad sequences were designed to manipulate the melting temperature (Tm) of the sequencing primers (see below), so that they fall within the required annealing temperature of 65-70 °C for the Illumina platform. Two linker nucleotides, which are not complementary to the two bases immediately prior to the primer, are used to link the primer with flanking sequences (Kozich et al., 2013). The target-specific primers are taken from the paper of Edson et al. (Edson, 2004), which are designed to target and amplify degraded HVSI mtDNA in four overlapping fragments, with a maximum amplicon size of 132-169 bp.



A. During PCR the target fragment is amplified and the flanking sequences are incorporated into the PCR product. This results in amplicons that comprise the target DNA fragment and both forward and reverse primers with their respective attached sequences.

B. Single strands of the amplified DNA are washed across the flow cell and it's lawn of oligonucleotides inside the MiSeq.

C. The flow cell attachment sites on each strand hybridise with their complimentary oligo on the flow cell itself.

D. DNA strands form bridges through hybridisation of the free oligonucleotide at the 3' end with its compliment on the flow cell. Bridge amplification using a DNA polymerase enzyme then allows for mutli-replication of the DNA template.

E. Clusters of copies of each single DNA strand are formed across the surface of the flow cell.

F. Reversible dye terminators for each base are washed through the flow cell and a single base, complimentary to the DNA template, is incorporated. After each addition, lasers are used to capture the base specific fluorescent signal of the base, which after many cycles, allows for the identification of the sequence.

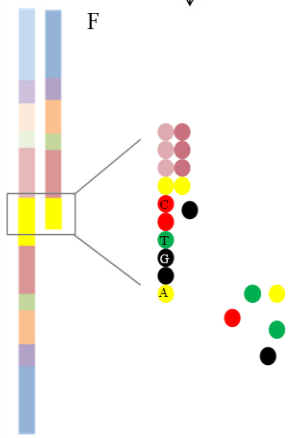


Figure 5.11 Diagram depicting the use of the modified primers for PCR, and the key stages during the sequencing of the resulting amplicons on the Illumina MiSeq platform.

Table 5-3 List of all primer/index combinations to identify each amplicon and its source. Prior to use with samples, all primer/index combinations were tested using 20 µl reactions with 0.5 ng of 007 control DNA as described in Section 2.9.3. PCR product was visualised using gel electrophoresis as described in Section 2.13.1.2.

	mtGenome HVS I region			
Sample	AMPLICON A	AMPLICON B	AMPLICON C	AMPLICON D
005B	AF1/AR1	BF1/BR1	CF1/CR1	DF1/DR1
007B	AF2/AR1	BF2/BR1	CF2/CR1	DF2/DR1
013B	AF3/AR1	BF3/BR1	CF3/CR1	DF3/DR1
017B	AF1/AR2	BF1/BR2	CF1/CR2	DF1/DR2
020B	AF2/AR2	BF2/BR2	CF2/CR2	DF2/DR2
EXT Neg A	AF3/AR2	BF3/BR2	CF3/CR2	DF3/DR2

Table 5-4 Table listing the sequences and Tm of each sequencing read 1 (SR1) and read 2 (SR2) primers for each target fragment (denoted by letters A-D). Read 1 primers consist of the pad-link-forward primer sequence used in the amplification of each of the target sequences whilst Read 2 is the same but utilises the reverse primer rather than the forward primer.

	ID	PAD	LINK	PRIMER	Tm °C
READ 1	SR1A	ATTCATGCAT	TG	CCCAAAGCTAAGATTCTAAT	68.09
	SR1B	ATTCATGCAT	AT	CACCATGAATATTGTACGGT	66.96
	SR1C	ATTCATACAT	AG	CCCCATGCTTACAAGCAAGT	68.83
	SR1D	ATTCATGCAT	AT	CACTAGGATACCAACAAACC	66.14
READ 2	SR2A	ACCTGAGTCG	TG	TACTACAGGTGGTCAAGTAT	66.48
	SR2B	ACTTGAATAT	CC	GGAGTTGCAGTTGATGT	67.16
	SR2C	ACCTGAGTCG	CC	TGGCTTTATGTACTATGTAC	67.36
	SR2D	ACTTGAATAT	AT	GAGGATGGTGGTCAAGGGAC	68.65

Table 5-5 Table listing the sequencing index primer (SIP) for each target fragment (denoted by letters A-D). This primer is the reverse complement of the Read 2 sequencing primer and is used to read the index which was originally incorporated with the forward primer during amplification.

	ID	PRIMER	LINK	PAD	Tm °C
INDEX	SIPA	ATACTTGACCACCTGTAGTA	CA	CGACTCAGGT	66.48
	SIPB	ACATCAACTGCAACTCC	GG	ATATTCAAGT	67.16
	SIPC	GTACATAGTACATAAAGCCA	GG	CGACTCAGGT	67.36
	SIPD	GTCCCTTGACCACCATCCTC	AT	ATATTCAAGT	68.65

Samples 005B, 007B, 013B, 017B, 020B & Batch 1 Negative A were each set-up for amplification of four overlapping fragments of mtGenome HVSI region using the primers, which had been adapted for use on the MiSeq Benchtop sequencer, as listed in

Table 2-19. Reactions were set-up using 5 µl of DNA extract per primer set, in a total reaction volume of 20 µl as outlined in Table 2-12 and Table 2-13, and amplified on a Veriti

thermal cycler using the parameters outlined in Table 2-21 and Table 2-22. AmpFlSTR® Control DNA 007 was used as a PCR positive control for each primer set with a total DNA input of 50 pg.

On completion of the first amplification step, samples were transferred to the NUomics sequencing facility at Northumbria University. Briefly, PCR product was cleaned-up using a Sequal Prep normalisation plate and then DNA fragment size was checked using the Agilent High Sensitivity DNA Kit and 2100 Bioanlayzer (Agilent, Santa Clara, CA, USA). Samples were then pooled into four groups, following which the DNA of each pool was quantified using Kapa qPCR library quantification kit. Samples were then normalised to the lowest concentration of the four pools and then combined to make a single pool prior to diluting to 3.5 pM and denaturing the samples. The pool was then spiked with 5% PhiX to provide greater variation in the sample pool, thus ensuring that the lasers can distinguish between the clusters during the sequencing reads. A MiSeq Reagent Kit V2, 2 x 250 cycle cartridge was utilised for sequencing which was spiked with the required read 1 read 2 and index matching for the mitochondrial primers used, the run then took approximately forty hours to complete (G. Holt, NUomics, 2015 pers. comm. 05.08.2015).

5.3.7 Data analysis

5.3.7.1 Statistics tests

The Kruskal-Wallis test was selected to determine if a statistical difference existed in the DNA yield between extraction methods as it is a non-parametric test, thus allowing for several independent and randomly selected samples to be compared. This test was conducted using R version 3.1.2 (2014-10-31) -- "Pumpkin Helmet".

5.3.7.2 Analysis of data generated by the MiSeq MPS platform

Following sequencing using the MiSeq platform, samples could not be demultiplexed and consequently just the two read data files were available for analysis. Fortunately due to the short fragment length of the mtDNA regions amplified, the 250 sequencing read cycles also included the sequence data of the modified primers, including the unique barcode identifiers. Using a number of different bioinformatics tools it was possible to trim each of the sequences to remove additional bases such as the flowcell attachment site and then manipulate the files so that the forward and reverse sequences were in the same file. It was then possible to demulitplex the samples using the unique barcodes included in the sequence. Once demulitplexed the forward and reverse reads were then separated and all of the barcode, pad and linker sequences were removed to leave just the mtDNA target sequences. Finally the trimmed sequences were aligned against the Hg38 reference genome. A number of

bioinformatics tools were utilised for this analysis the details of which, along with a step-by-step guide to the manipulations carried out, can be found in appendix 8.23.

5.4 Results

5.4.1 Extraction technique comparison

Five bone samples were selected at random from the Brazilian sample set and each was extracted with three different methods, using an equivalent mass of pulverised sample. Extraction method one is a phenol-chloroform based protocol outlined by Kemp et al. (Kemp et al., 2007), method 2 utilises a silica column for DNA purification as described by Yang et al. (Yang et al., 1998), and method 3 relies on the use of a silica suspension as described by Rohland and Hofreiter (Rohland and Hofreiter, 2007). The nDNA and mtDNA in each extract was quantified using the Quantifiler Trio Kit or mtDNA qPCR assay respectively. The quantity of nDNA in sample 13B2 was an order of magnitude greater than the nDNA results for any other sample; a mixed nDNA profile was observed in subsequent DNA profiling of this sample, thus demonstrating that a contamination event had occurred and so this value was excluded from analysis. In order to determine if any one of the three extraction methods statistically resulted in a consistently higher quantity of DNA, the results shown in Table 5-6 were subjected to a Kruskal-Wallis test using R. This test is a non-parametic test, which allows for the comparison of more than two groups for an independent variable (Kruskal and Wallis, 1987). In order to use this test, four assumptions must be met: the dependant variable must be measured at the ordinal or continuous level (mass of DNA in ng/µl was measured in the extracted samples in this data-set); the independent variable must be two or more groups (three extraction methods were tested in this stu); there should be no relationship between the results of each group (each extraction method was tested independently) and each group should be expected to have a normal distribution.

The null hypothesis for this test is that there is no difference in the yield of either nDNA or mtDNA between extraction methods, which was accepted in both cases $P = 0.74$ and $P = 0.29$ (see appendix 8.22 for details).

Table 5-6 Summary of Quantifiler Trio and mtDNA qPCR assay results for the comparison of three different extraction methods; method 1- phenol-chloroform-based, method 2 – silica-column-based and method 3 – silica-suspension-based. Each sample was quantified in triplicate and the average and standard deviation of the results is given. A negative result is denoted by (-)

	Method 1			Method 2			Method 3		
Sample	**Quant mean (ng/µl)**	**Quant SD**	**mtDNA copy number**	**Quant mean (ng/µl)**	**Quant SD**	**mtDNA copy number**	**Quant mean (ng/µl)**	**Quant SD**	**mtDNA copy number**
005B1	0.0232	0.0015	**8.00E+00**	0.0400	0.0126	**1.15E+06**	0.0493	0.0003	**3.12E+05**
007B1	0.0010	0.0006	**3.84E+04**	0.0005	0.0003	**9.55E+04**	0.0011	0.0004	**4.02E+04**
011B1	0.0012	0.0006	**-**	0.0079	0.0034	**-**	0.0066	0.0009	**1.00E+02**
013B1	0.0497	0.0093	**3.20E+05**	0.4657	0.0319	**4.00E+05**	0.0690	0.0139	**1.48E+05**
020B1	0.0014	0.0007	**9.71E+03**	0.0436	0.0030	**1.50E+05**	0.0081	0.0011	**4.19E+04**

5.4.2 Brazilian samples

5.4.2.1 Quantifiler Trio and mtDNA assay results

All submitted samples were quantified in triplicate using both the Quantifiler Trio kit and the in-house mtDNA qPCR assay; for a summary of results see Table 5-7. Amplification of DNA was detected for all samples with an average of 0.1 ng/µl with a range of 0.0008-0.54 ng/µl. Of the eighteen samples that could be categorised by DI value as described in Section 5.3.4, three samples showed no degradation, four samples mild degradation, nine samples were grouped as degraded and two samples as severely degraded. Six samples did not have DI values assigned, one of which, 006B, showed a sign of PCR inhibition as a high IPC Ct value was observed. Sex was assigned to eighteen of the twenty-four samples; fourteen males, one female, four potential mixes and five unassigned. Mitochondrial amplification was observed in eleven of the twenty-four samples with an average of 1.31×10^5 and minimum and maximum values of 1.89×10^3 and 4.40×10^5 copies of mtDNA respectively.

Two-extraction negative controls were processed alongside each batch of bone samples, a negative deer bone control (Neg A) and a reagent blank control (Neg B), in addition to which a no template control (NTC) was included during quantification. Nuclear DNA was amplified in six of the twelve samples (see Table 5-8), the quantification value was < 1 pg/µl in all cases with the exception of Method 2 Neg A, for which a value of < 2 pg/µl was observed. mtDNA was amplified in ten samples, the remaining two being NTC samples., with an average of 8.28×10^3 and minimum and maximum values of 1.31×10^3 and 1.21×10^4 copies of mtDNA respectively.

Table 5-7 Summary of Quantifiler Trio results for Brazilian skeletal samples. In cases where multiple extractions were tested, the highest short fragment quantity was selected. The short autosomal target is 80 bp in length, the long autosomal is 214 bp and the Y target is 75 bp. A negative result is denoted by (-)

Sample	Extraction method	Short autosomal		Long autosomal		Y target				Internal positive control		mtDNA
		Quant mean (ng/µl)	Quant SD	Quant mean (ng/µl)	Quant SD	Quant mean (ng/µl)	Quant SD	DI	Sex	Ct value	Ct mean	Copy number
001B	3	0.0433	0.0048	0.0050	0.0010	0.0346	0.0039	8.6	M	27.25	0.81	-
002B	3	0.0443	0.0033	0.0071	0.0013	0.0348	0.0051	6.2	M	28.01	0.58	-
002T	1	0.0236	0.0091	-	-	0.0299	0.0051		M	26.00	1.90	-
003B	3	0.1054	0.0234	0.0236	0.0044	0.0716	0.0101	4.5	M	28.34	0.16	-
004B	3	0.3959	0.0662	0.0749	0.0121	0.2416#	0.0143	5.3	M/F	28.59	0.19	3.04E+04
005B	3	0.0493	0.0003	0.0195	0.0024	0.0331	0.0023	2.5	M/F	27.69	0.30	3.12E+05
006B	3	0.0015	0.0003	-	-	0.0042	0.0040		-	38.11^	-	-
007B	3	0.0011	0.0004	0.0009	0.0007	-	-	1.2	-	27.84	0.10	4.02E+04
007T	1	0.0008#	0.0007	-*	-	-	-		-	27.00	0.39	1.40E+04
008B	3	0.5359	0.0099	0.546	0.0094	0.4952	0.0127	1.0	M	27.46	0.58	-
009B	1	0.0030	0.0018	0.0005	0.0004	0.0036	0.0025	5.7	M	26.18	1.01	1.74E+04
009T	1	0.0017#	0.0006	-	-	0.0016	0.0005		-	26.13	1.34	1.89E+03
010B	3	0.1175	0.0062	0.0315	0.0003	0.1004	0.0078	3.7	M	28.18	0.06	-
011B	2	0.0079	0.0034	0.0012	0.0010	0.0055	0.0030	6.8	M	27.52	0.51	-
011T	2	0.0037	0.0011	-	-	0.0011	0.0006		M/F	27.33	0.20	-
012B	1	0.0015#	0.0006	-	-	-	-		-	26.60	0.97	1.20E+04
013B	2	0.4657	0.0319	0.1041	0.0070	0.4745	0.0374	4.5	M	26.62	0.62	4.00E+05
014B	3	0.4257	0.0557	0.0475	0.0087	0.2045	0.0747	9.0	M/F	28.19	0.53	-
015B	3	0.0368	0.0048	0.0159	0.0023	-	-	2.3	F	28.13	0.09	-
016B	3	0.0118	0.0026	0.0037	0.0015	0.0084	0.0021	3.2	M	27.69	0.61	1.80E+04

017B	1	0.0351	0.0033	0.0030	0.0025	0.0269	0.0021	11.8	M	26.67	0.75	4.40E+05
018B	3	0.0831	0.0096	0.0086	0.0040	0.0667	0.0026	9.7	M	27.74	0.49	-
019B	3	0.0261	0.0043	0.0015[#]	0.0006	0.0231	0.0042	17.3	M	27.97	0.05	-
020B	2	0.0436	0.0030	0.0342	0.0025	0.0541	0.0081	1.3	M	27.28	0.04	1.50E+05

*only one result recorded [#]two results out of three [^]inhibited sample

Table 5-8 Quantifiler Trio results for all negative control samples Brazil samples. A negative result is denoted by (-)

Sample	Short autosomal		Long autosomal		Y target		mtDNA
	Quant mean (ng/µl)	**Quant SD**	**Quant mean (ng/µl)**	**Quant SD**	**Quant mean (ng/µl)**	**Quant SD**	**Copy number**
Method 1 Neg A	0.0008	0.00001	-	-	-	-	6.77E+03
Method 1 Neg B	-	-	-	-	-	-	8.38E+03
Method 2 Neg A	0.0017	0.00064	-	-	-	-	1.21E+04
Method 2 Neg B	-	-	-	-	-	-	9.11E+03
Method 1/2 NTC	0.0002	0.00003	-	-	-	-	-
Method 3a Neg A	0.0005	0.00032	-	-	-	-	1.29E+04
Method 3a Neg B	-	-	-	-	0.0003	0.0001	8.35E+03
Method 3a NTC	-	-	-	-	-	-	1.31E+03
Method 3b Neg A	0.0006	0.00014	-	-	-	-	8.47E+03
Method 3b Neg B1	-	-	-	-	0.0008	0.0002	8.79E+03
Method 3b NegB2	0.0005	0.00008	0.0003	0.000001	-	-	6.64E+03
Method 3b NTC	-	-	-	-	-	-	-

5.4.2.2 Illumina protocol

The Illumina protocol was optimised for each primer mix using 250 pg 007 control DNA including the primer annealing temperature (56 – 62 ˚C), $MgCl_2$ concentration (4 – 8 mM), primer concentration (0.03 – 0.06 µM) and the use of a PCR additive (0.5X Qiagen Q Solution), and the sensitivity of each primer set was then demonstrated down to approximately 16 pg of DNA as shown in Figure 5.12. The band intensity for Primer mix MIII was relatively less than the other three sets; however, amplified product is present down to the lowest DNA input. Although assay optimisation has reduced the presence of secondary PCR product, faint bands can still be observed for primer mixes MI, MII and MIV as shown Figure 5.12. Having discussed this with the NUomics team, it was decided due to the relatively larger fragment size of this product, it could be bioinformatically removed post-sequencing and thus should not impact downstream analyses.

Following this six of the Brazil sample DNA extracts were selected, representing the range of DNA quantification values obtained using the Quantifiler Trio Kit, for amplification using the optimised protocol. As the electrophoresis gel image shown in Figure 5.13, no amplified product was observed for any of these samples.

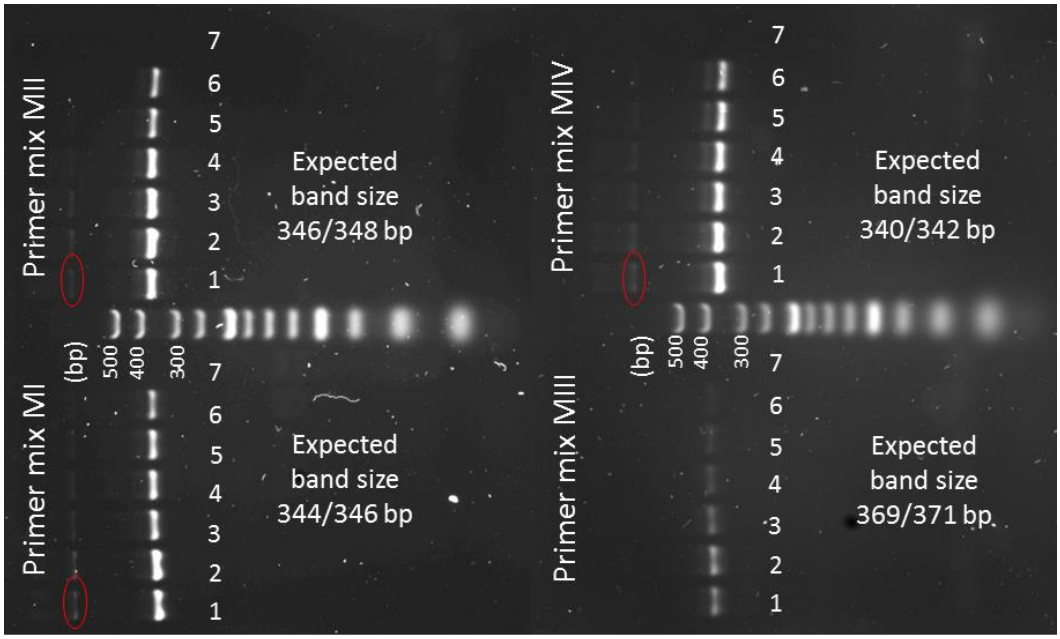


Figure 5.12 Sensitivity tests for Illumina suggested primer sets using a 1 in 2 serial dilution of 007 control DNA with the following approximate DNA inputs; 500 pg (1), 250 pg (2), 125 pg (3), 63 pg (4), 32 pg (5), 16 pg (6) and a negative control (7). DNA Hyperladder™ 25bp was used to to measure approximate DNA fragment size. The two values provided for expected band size represent the duplex tagged primers within each amplicon mix. Amplified DNA was visible down to a DNA input of approximately 16 pg of input DNA for all primer mixes, whilst no band was visible in any of the negative controls. A secondary PCR product can be observed (examples circled in red) for primer mixes MI, MII and MIIV, particularly at the higher DNA inputs.

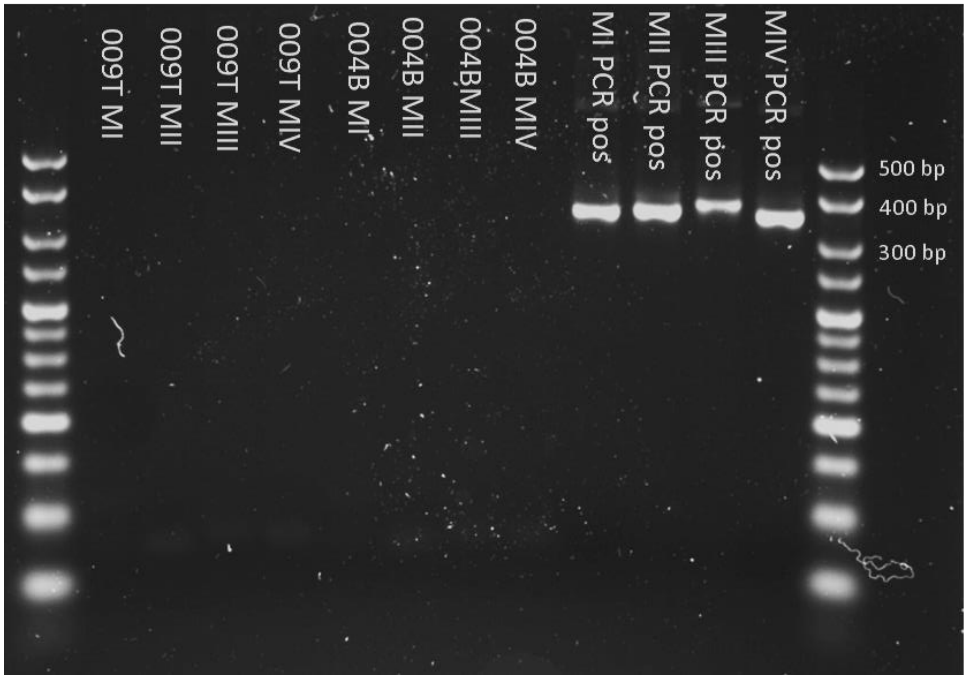


Figure 5.13 Example of results for the amplification of samples 004B, 008B, 009T, 019B, 013B, 014B and the PCR positive controls amplified using Illumina's suggested primer sets. There was no amplified DNA visible for any of the samples tested.

5.4.2.3 Edson primer set

In order to test the theory that the DNA extracted from the Brazil bone samples was too degraded for amplification using the Illumina primer mixes, the same sample set of 009T, 019B, 004B, 014B, 013B and 008B were amplified using Edson primer set E (Edson, 2004) which has a smaller target fragment size of 125 bp. As shown in Figure 5.14, amplified DNA of the expected size is present for samples 009T, 004B and 013B. Faint bands were also visible for sample 014B and the 'A' negative control for extraction 1, the latter indicating the presence of contaminating DNA. These results reflect the qPCR results and demonstrate that the use of a shorter target fragment size is apposite for degraded samples.

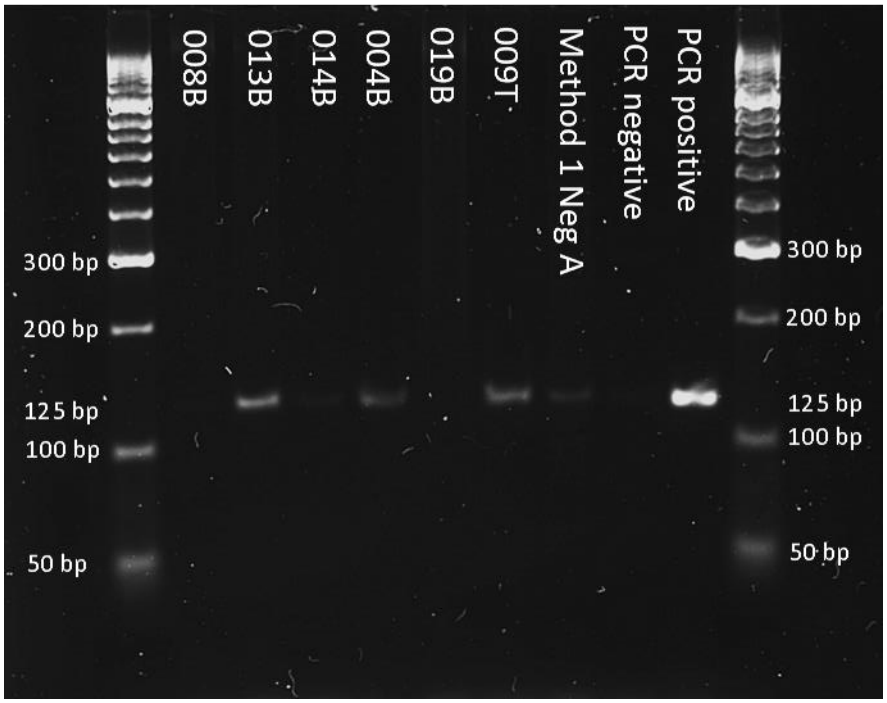


Figure 5.14 Results for the amplification of samples 004B, 008B, 009T, 019B, 013B and 014B, using primer set E from (Edson, 2004). PCR product of the expected 125 bp fragment size is observed for samples 009T, 004B and 013B. Faint bands are also visible for sample 014B and the method 1 extraction negative A control sample.

5.4.2.4 MiSeq results

The HVSI region of samples 005B, 007B, 013B, 017B, 020B and Method 1 Neg A were sequenced using the MiSeq platform. Samples could not be demultiplexed and consequently just the two read data files were available for analysis. These were analysed as described in section 5.3.7.2 and appendix 8.23.

Once demultiplexed, the number of indexed and unmatched reads for each amplicon was plotted using Microsoft Excel 2010 as shown in Figure 5.15. A total of 287,446 single reads were obtained in this MiSeq run, of which 234,283 were indexed and 53,163 were unmatched. The unmatched reads represented 10-25% of the total read count for each amplicon, whilst the average number of single indexed reads per amplicon is 58,571 ± 29,556.

Following the trimming of data to remove the barcodes, pad sequences etc. as previously described in section 5.3.7.2, > 98% of the raw sequence reads were utilised for alignment of amplicons A, C and D with the human mtGenome reference and 0.6% of the raw sequence reads generated for amplicon B as shown in Figure 5.16.

BWA-MEM files were opened using the Integrative Genomics Viewer program (v2.3.57) (Robinson et al., 2011, Thorvaldsdottir et al., 2013) and aligned against the human (hg38) mtGenome as the reference sequence. An example of the Viewer is shown in Figure 5.17. This was a proof of concept study to demonstrate the use of the adapted primer set method for the amplification of degraded DNA using MPS. As such, individual sequences and heteroplasmy were not considered at this point, but all DNA fragments were successfully aligned to the human mtDNA rCRS.

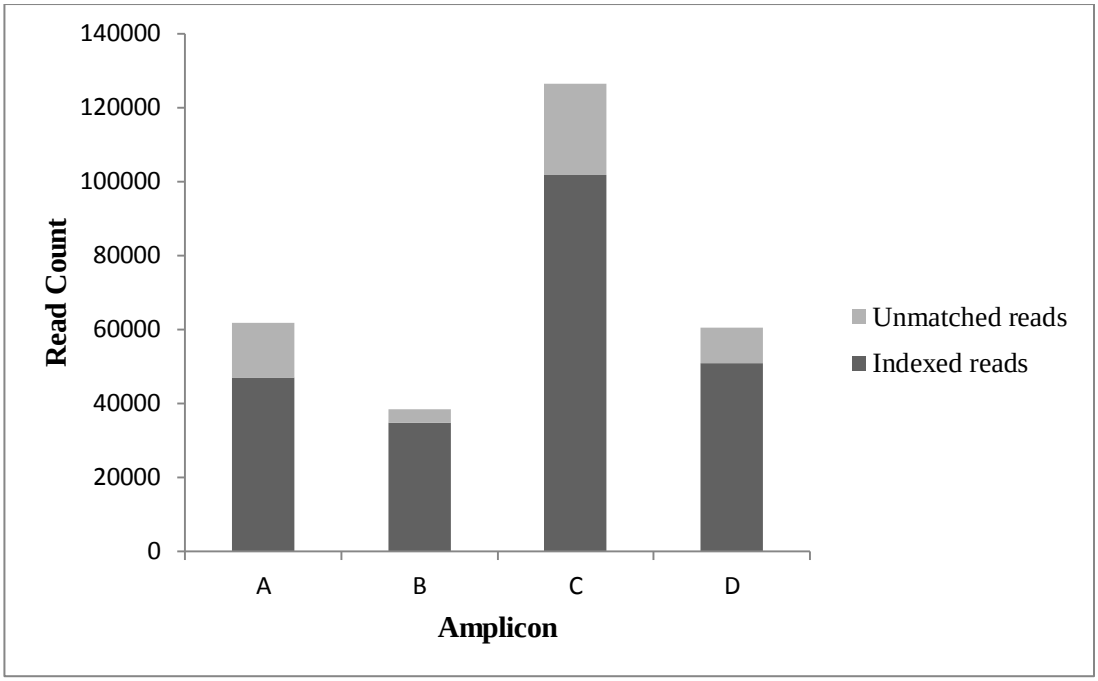


Figure 5.15 Stacked column chart showing the indexed and unmatched read counts obtained for each amplicon post demulitplexing

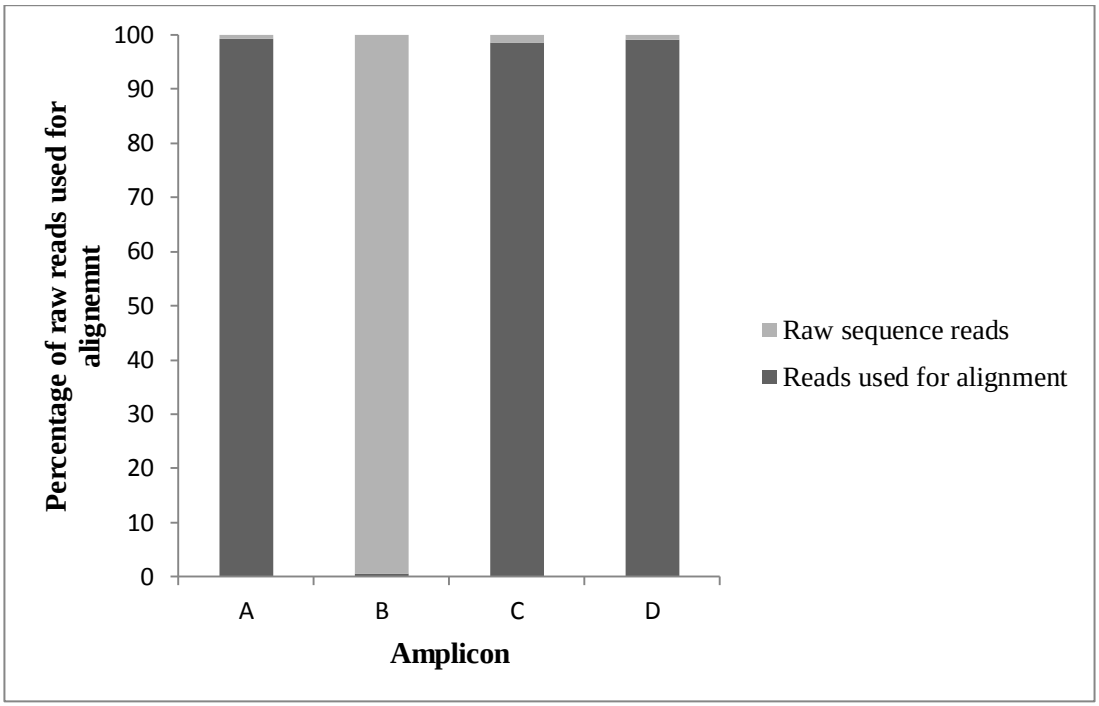


Figure 5.16 Stacked column chart showing the number of reads that were utilised for alignment as a percentage of the total reads generated for each amplicon

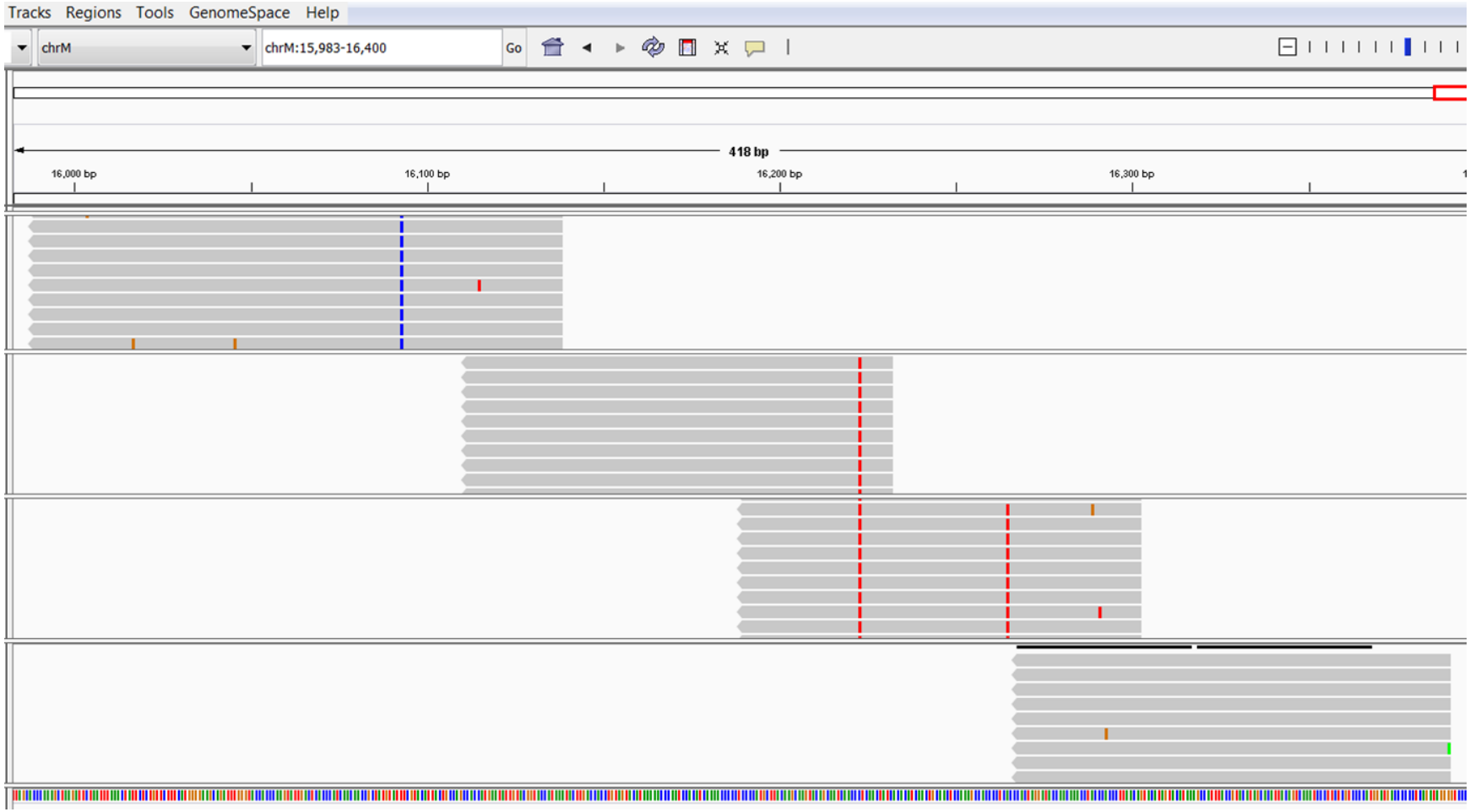


Figure 5.17 Example of Integrative Genomics Viewer showing all for amplicons for sample 020B aligned to the human mtGenome. Differences from the reference sequence are shown in a base-dependent colour while conserved bases are shown in grey.

5.4.3 Vindolanda samples

5.4.3.1 Quantifiler Trio results

The DNA extracts from the Vindolanda samples, including both femur DNA extracts, were quantified in triplicate using the Quantifiler Trio kit, the results for which are summarised in Table 5-9. Nuclear DNA was amplified in all three DNA extracts with 35 pg/µl and 71 pg/µl recorded for the femur sample from extracts 1 and 2 respectively, and a DNA concentration of 56 pg/µl for the tooth extract. Male DNA was amplified in both of the femur DNA extracts and the DI values, calculated as described in Section 5.3.4, were consistent with one another with values of 9.1 and 9.3, indicating the DNA is in a degraded state. The sex of the tooth sample was also recorded as male, however the DI value of 2.5 indicates that the extracted DNA is in a mildly degraded state, and thus of better quality than the femur sample.

Two extraction negative controls were processed alongside each sample, a negative deer bone control (Neg A) and a reagent blank control (Neg B). Nuclear DNA was detected in the Femur 1 DNA extract to a concentration of <0.5 pg/µl, whilst no DNA was detected in the Femur 2 negative controls. DNA was amplified in both of the tooth negative controls at concentrations of 6 pg/µl and 1 pg/µl for Neg A and Neg B respectively.

The IPC values for all samples were within the expected range and so no indication of PCR inhibition was observed.

Table 5-9 Summary of Vindolanda Quantifiler Trio results

	Short autosomal target		Long autosomal target		Y target				IPC	
Sample	**Quant mean (ng/µl)**	**Quant SD**	**Quant mean (ng/µl)**	**Quant SD**	**Quant mean (ng/µl)**	**Quant SD**	**DI**	**Sex**	**Ct value**	**Ct SD**
Femur 1	**0.0349**	**0.0020**	**0.0038**	**0.0019**	**0.0420**	**0.0020**	**9.1**	**Male**	**27.62**	**0.61**
F1NA	0.0003	0.0001	0.0001	0.00002	-	-			27.41	0.81
F1NB	-	-	-	-	-	-			27.66	0.07
Femur 2	**0.0711**	**0.0115**	**0.0076**	**0.0029**	**0.0753**	**0.0096**	**9.3**	**Male**	**27.62**	**0.21**
F2NA	-	-	-	-	-	-			27.15	0.88
F2NB	-	-	-	-	-	-			27.88	0.16
Tooth	**0.0562**	**0.0030**	**0.0226**	**0.0035**	**0.0617**	**0.0048**	**2.5**	**Male**	**28.29**	**0.02**
TNA	0.0062	0.0022	0.0036	0.0002	0.0066	0.0031			28.03	0.30
TNB	0.0013	0.0006	0.0010	0.0002	0.0015	0.0001			28.02	0.13

5.4.3.2 AmpFℓSTR® NGM™ PCR amplification kit

NGM™ DNA profiles shown in Table 5-10 for both the tooth and femur samples were generated through consensus using a minimum peak amplitude of 25 RFU; for example EPGs of both samples see Figure 5.18 and Figure 5.19. Twenty-seven alleles were amplified from the tooth DNA extract and fifteen alleles from the femur extract. In addition to this, XY was present at Amelogenin in both samples, confirming the biological sex as male. In both samples a decrease in peak height was observed as the amplicon size of the amplified allele increased (Figure 5.20). The decline in peak height is more severe in the femur sample which, along with a maximum amplified amplicon size of approximately 250 bp, is reflective of the degraded quality of DNA observed with the Quantifiler Trio kit. An outlier was removed from tooth results at locus D8 (peak height 544 RFU); this allele had an increased peak height relative to the other alleles in both amplifications of the DNA extract and indicates that this locus may be homozygous.

Table 5-10 Vindolanda samples NGM™ consensus DNA profiles. 'F' denotes alleles for which consensus was not attained or no amplification was observed

Locus	Tooth		Femur	
AMEL	X	Y	X	Y
D10S1248	13	15	13	14
vWA	17	18	17	F
D16S539	9	12	11	F
D2S1338	20	25	F	F
D8S1179	14	F	11	16
D21S11	28	F	29	30
D18S51	16	20	F	F
D22S1045	11	16	15	17
D19S433	13	14	F	F
THO1	9.3	F	8	F
FGA	20	24	F	F
D2S441	10	14	10	11.3
D3S1358	15	16	15	18
D1S1656	14.3	17.3	F	F
D12S391	17	19	F	F

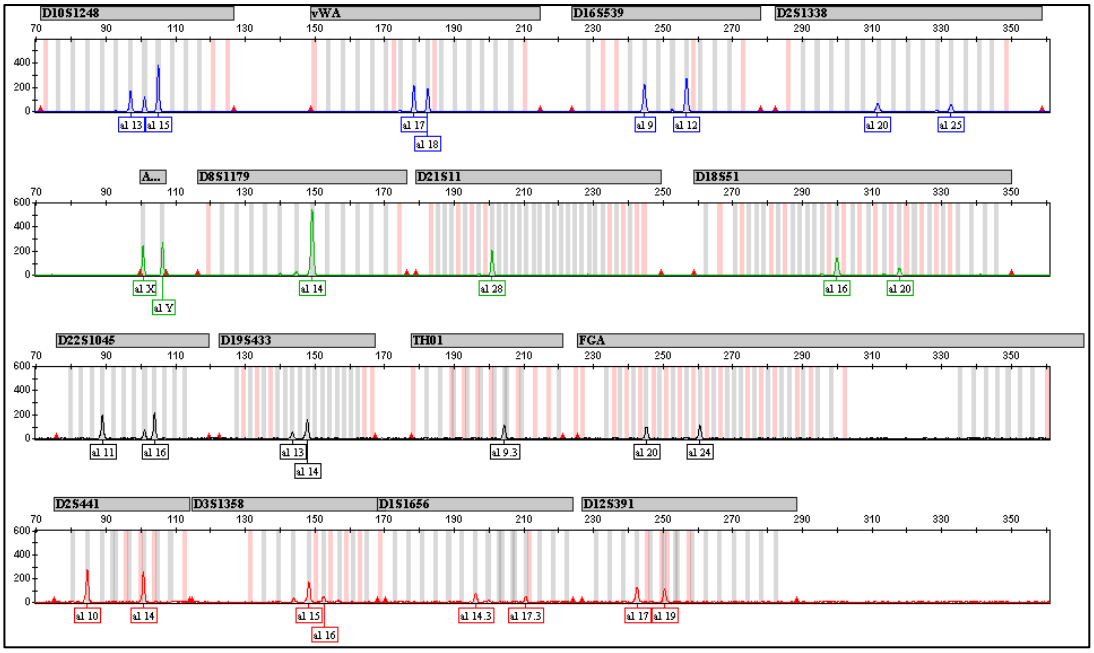


Figure 5.18 EPG for Vindolanda tooth DNA (amplified product a) profile generated using NGM™ amplification kit

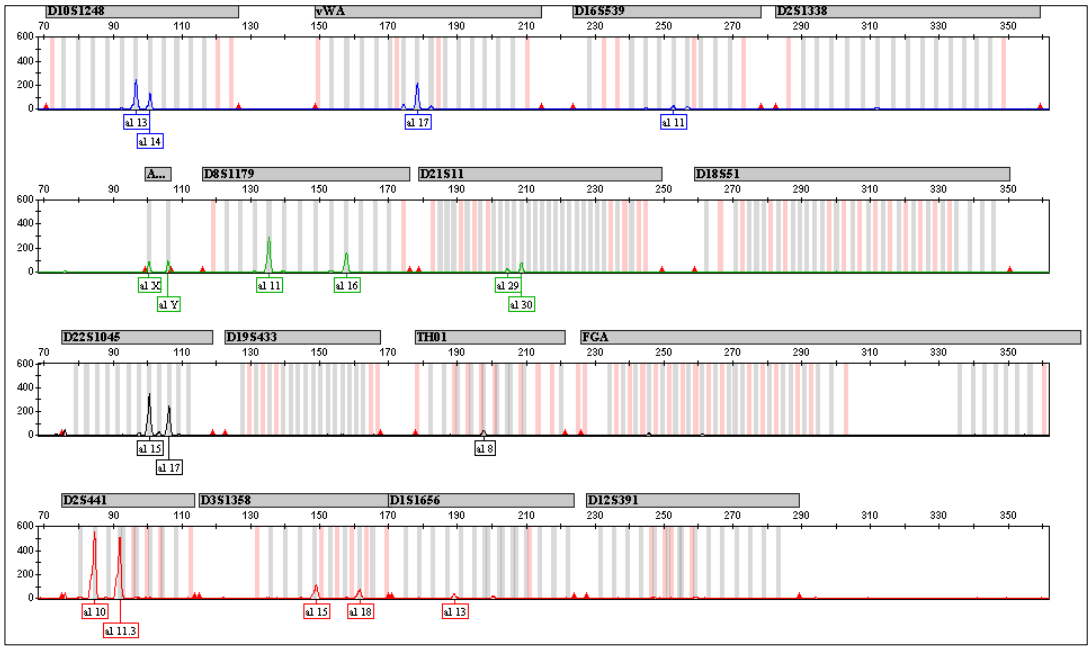


Figure 5.19 EPG for Vindolanda femur (extract 2) DNA profile generated using NGM™ amplification kit

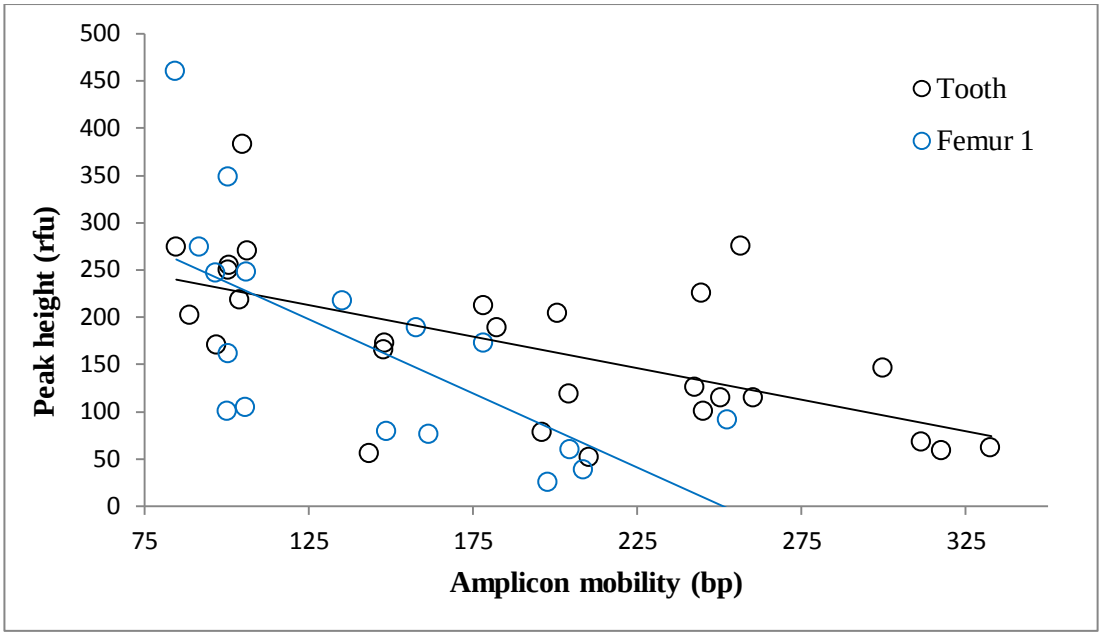


Figure 5.20 Relationship between size of amplified STR alleles and peak height (RFU) for both tooth and femur samples. It can be seen that preferential amplification of the lower molecular weight alleles has occurred with higher rfu values for loci with a molecular weight of <125 bp. A decrease in the peak height is observed as the molecular weight of the amplicons increases.

5.4.3.3 ForenSeq™ DNA Signature Kit

The ForenSeq™ DNA Signature Kit (ForenSeq™) results provided by Illumina Inc. for the amplification of each marker type observed in the DNA profile from the Vindolanda tooth DNA extract are summarised in Table 5-11. A total of 58% of all possible loci were observed to have amplified, of which 72% were <150 bp in fragment size. Of all the markers amplified the highest proportion present (44%) were identity SNPs as shown in Figure 5.21. No DNA amplification was observed in either of the two femur DNA extracts.

Table 5-11 Summary of ForenSeq™ result for Vindolanda tooth

Marker Type	Number of loci observed	Percentage of total possible loci amplified (%)
Autosomal STRs	15	56
X-STRs	3	43
Y-STRs	5	21
Identity SNPs	59	62
Ancestry SNPs	35	56
Phenotypic SNPs	18	82

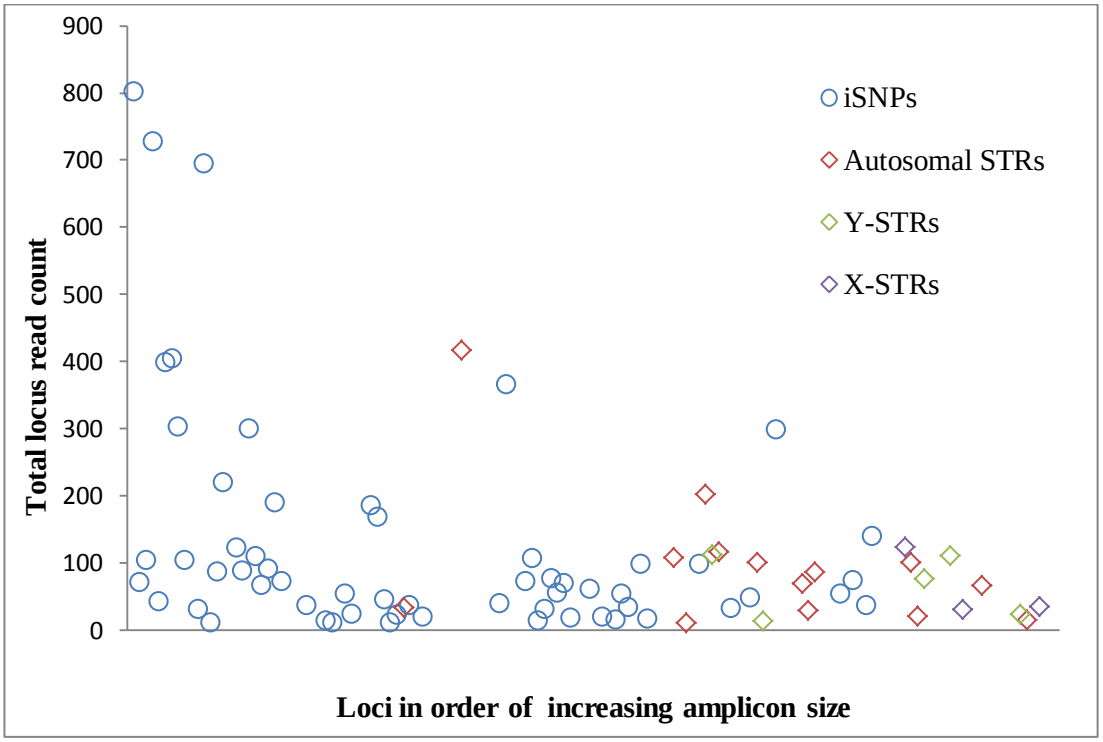


Figure 5.21 Marker types amplified in tooth sample in ascending order of amplicon size. It can be seen that the read count is relatively consistent across all amplicon sizes.

5.4.3.4 Biogeographic results for Vindolanda samples

To obtain an indication of biogeographic ancestry, both the tooth and femur DNA extracts were amplified using the AmpFℓSTR® Yfiler® PCR Amplification Kit. A consensus haplotype for each sample was generated using alleles with a minimum peak height of 25 rfu, which were observed in each of the amplified profiles but not in the consensus negative controls as described in the guidelines set by Gill et al. (Gill et al., 2000b). The resulting consensus haplotype for the femur, shown in

Table 5-12, and the combined YFiler® and ForenSeq™ kits Y-STR haplotypes, Table 5-13, were entered into the Haplogroup Predictor (Athey, 2015) (http://www.hprg.com/hapest5/index.html, accessed on 18/07/2015). This is an online tool that can be used to indicate the haplogroup of the individual based on their Y-STR DNA profile. This tool calculates a 'goodness of fit' score (Athey, 2005) and a probability using a Bayesian approach (Athey, 2006) across twenty-one haplogroups, which >98% of the Western European population fall into (Athey, 2005). Allele frequencies for STRs associated with these haplogroups have been calculated using public databases, and are used as a point of comparison to see which group the unknown sample best fits into.

As only three alleles are available for analysis from the femur sample the probability of the individual belonging to a particular haplogroup is spread across a number of different haplogroups with the highest probability of 38% assigned to group I2a1. The tooth however has been assigned to haplogroup R1b with a probability of 97% based on the seven alleles combined using both YFiler® and ForenSeq™ results (see Table 5.13).

ForenSeq™ Primer Set B contains fifty-six ancestry informative SNP markers, 63% of which were successfully amplified from the tooth DNA extract. Principal component analysis was utilised by Illumina Inc. to evaluate the results as shown in Figure 5.22. This involves reducing the overall complexity of the dataset whilst retaining the most variation between samples that is possible. This way several inter-correlated variables can be assessed using patterns of similarity to extract pertinent information from the data (Wold et al., 1987).

In order to generate accurate PCA plots, the relevance, in terms of the populations sample, and quality of the data is of vital importance. Illumina generate their PCA plots using data from the 1000 genomes project, which ran between 2008-2015 to create a catalogue of human genetic variation and genotype data (The Genomes Project, 2015, Sudmant et al., 2015). Data has been collected from 26 populations from within Africa, Europe, East Asia, South Asia and the Americas, sampling approximately 500 individuals per continental

group. Each continental group contained approximately five subpopulations, the number of samples from those in within Europe are 180 British individuals in England and Scotland (GBR), 192 Finnish in Finland (FIN), 121 individuals from Iberian populations in Spain (IBS), 205 Toscani in Italia (TSI) and 184 UTAH residents (CEPH) with Northern and Western European ancestry (The Genomes Project, 2015). Whilst there are gaps in the data for populations such as Native Australian, North Africa and Southeast Asian (Phillips, 2015), it is the GBR and TSI populations that are of direct relevance to this study.

The biogeographic origin of the individual from which the tooth sample was taken was estimated to be European, with the distance to the nearest centroid (the weighted mean of the data set) given as 2.40. The closest reference samples were from the TSI (Toscani in Italia) grouping, indicating that the individual is more closely related to samples in this region than other reference samples within the centroid.

Table 5-12 Consensus profiles generated for both the tooth and femur DNA extracts using the AmpFℓSTR® Yfiler® PCR Amplification Kit. Loci included in the minimal haplotype for ancestry analysis are denoted by an asterisk.

Consensus haplotype	DYS389I*	DYS393*	DYS391*	DYS635	DYS437
Tooth	13	-	11	23	15
Femur	-	13	10	-	15

Table 5-13 Y-STR haplotypes generated from the tooth DNA extract using either the AmpFℓSTR® Yfiler® PCR Amplification Kit or the ForenSeq™ DNA Signature Kit. Loci included in the minimal haplotype for ancestry analysis are denoted by an asterisk.

Amplification Kit used	DYS389 I*	DYS391 *	DYS63 5	DYS43 7	DYS43 8	DYS57 0	DYS57 6
Yfiler	13	11	23	15			
ForenSeq	13	11			12	17	19

Table 5-14 Results of the Haplogroup Predictor online tool (Athey, 2015) for both the femur sample (left) and tooth (right) Y-STR haplotypes. The femur is predicted to belong to haplogroup I2a1 (38%) followed by I2a (11.9%) and R1b (10.5%), whilst the results for the tooth are more conclusive with a probability of 99.7% that the indivdiual is in haplogroup R1b.

	Femur sample		Tooth sample	
Haplogroup	**Fitness score**	**Probability (%)**	**Fitness score**	**Probability (%)**
E1b1a	50	1.9	20	0.0
E1b1b	27	0.6	7	0.0
G2a	29	0.7	5	0.0
G2c	21	0.5	0	0.0
H	6	0.0	1	0.0
I1	38	2.0	3	0.0
I2a (xI2a1)	87	11.9	23	0.0
I2a1	100	38.0	3	0.0
I2b (xI2b1)	68	10.1	1	0.0
I2b1	37	0.8	11	0.0
J1	11	0.1	2	0.0
J2a1b	41	2.1	8	0.0
J2a1h	11	0.0	1	0.0
J2a1 x J2a1-bh	40	1.9	3	0.0
J2b	21	0.3	1	0.0
L	39	1.0	19	0.0
N	9	0.0	10	0.0
Q	69	8.4	28	0.2
R1a	26	0.4	18	0.0
R1b	73	10.5	57	97.0
T	67	8.6	4	0.0

Figure 5.22 PCA plot of biogeographical results for tooth sample as provided by Illumina using the ForenSeq™ Signature Prep Kit

5.4.3.5 Authentication of results

Two fragments of the mtGenome were targeted; nucleotide positions 220-389 and 403-569 which cover substitutions 185A and 228A, and 462T, 482C and 489C respectively. Both the forward and reverse strands for each of the amplified fragments in both femur extract 1 and femur extract 2 were sequenced, resulting in a minimum of X2 and a maximum of X4 coverage for each substitution, and the consensus haplotype was recorded as shown in Table 5-15. The two sets of results agree as all five substitutions recorded by York University were found to be present in both of the femur DNA extracts, as well as an additional substitution of 263G as shown in Figure 5.23.

Table 5-15 Substitutions in the control region of the mtGenome as recorded by both the previous work by York University and as part of this study. Positions 73G, 295T and 16076T were not tested (n/t) as part of this study.

	mtGenome Positions								
York University	73G	185A	228A	-	295T	462T	482C	489C	16076T
Current study	n/t	185A	228A	263G	n/t	462T	482C	489C	n/t

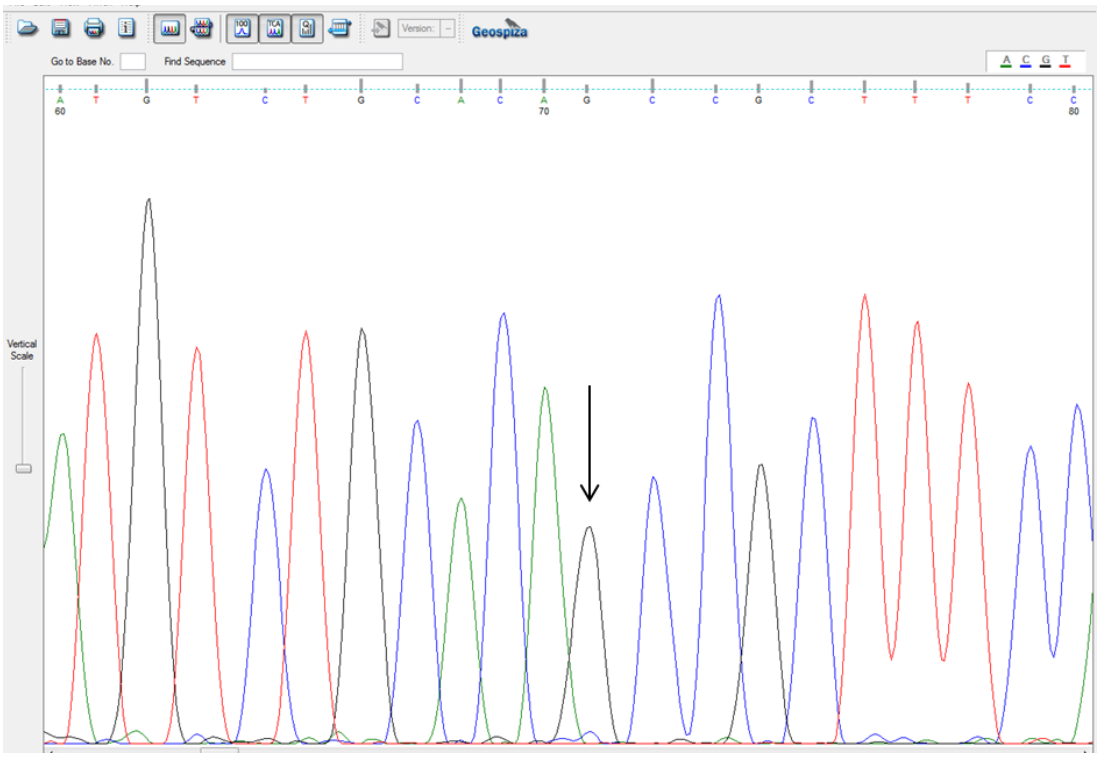


Figure 5.23 Image of the electropherogram for the mtDNA sequence amplified from Femur extract 1 (forward strand) as an example to show the presence of substitution 263G, indicated by arrrow, as compared with the rCRS.

5.5 Discussion

5.5.1 Brazilian sample set

Due to the expected degraded state of the DNA in the Brazilian sample set it was determined that the use of a more specialised extraction protocol would help maximise the DNA yield from the samples and so methods utilised for ancient DNA work, which is also subject to degradation, were reviewed. Three well documented methods were selected for comparison; a phenol-chloroform-based extraction, a silica-column-based extraction and a silica-suspension-based extraction. A selection of five bone samples were randomly chosen from the Brazilian set of remains and extracted using each of the three methods to determine if any one protocol consistently yielded a higher quantity of mtDNA or nDNA. Using the Kruskal-Wallis test; the null hypothesis that there was no difference between DNA extraction methods was accepted for both mtDNA ($p=0.3$) and nDNA ($p=0.7$). It is difficult to assess the different methods in terms of DNA recovery given the low average DNA quantification value of 20 pg/µl (sd 0.02) for all samples across the three methods. It was decided that method 3 would be used for further extractions due to a very marginal increase in quantification values in three of the five tested samples, along with ease of handling during the extraction protocol.

Over half of the samples for which the DI value could be calculated were categorised as degraded or severely degraded which, given that the *post mortem* interval (PMI) on these samples is only eight to fourteen years, demonstrates the detrimental effects of a hostile environment. The environmental conditions in Brazil are such that the rate of putrefaction and skeletisation will be rapid, creating a limited timeframe for the binding and mineralisation of both DNA and mtDNA within the bone infrastructure (Campos et al., 2012, Jans et al., 2004). As an exponential increase in microbial activity with increasing temperature has been demonstrated, it is also likely that microbial activity and bioerosion will have played a significant role in the degradation of DNA within this sample set (Campos et al., 2012, Child, 1995, Jans et al., 2004).

All four of the tooth samples tested in this sample set had no detected amplification of the LA marker using the Quantifiler Trio kit and therefore do not have DI values assigned. As previously discussed, it has been demonstrated that DNA in teeth samples is usually preserved better than in femur bone and so it is possible that the LA target was not amplified in these samples, along with 012B, due to the low DNA concentration rather than an accurate representation of the quality of the contained DNA. Having said this, however, all four of the tooth samples were prepared by removing the inner material without destroying the tooth casing and whilst this approach has been successfully demonstrated, the

combination of a low mass of tooth powder, together with the expected DNA degradation suggests that this approach may not be effective under these circumstances.

The use of quantification values as low as those observed in many of the Brazilian sample to assign sex presents a difficulty for interpretation, for example, if 10 pg of total DNA and 6 pg of male DNA was recorded, this could be interpreted as a mixture or attributed to stochastic variation, but it is difficult to determine at what values the results can be accepted as reliable. It was decided to err on the side of caution for this study and so the sex was not assigned to any sample containing <2 pg/µl total DNA concentration, which accounted for five samples. Sex was only assigned as single source if an overlap was observed between the SA quantification value mean ± SD and the Y quantification value mean ± SD. Of the remaining nineteen samples, fourteen were determined to be male, one was assigned as female (zero quantification value for the Y marker) and four were tentatively assigned as M/F mixtures. The information submitted with the Brazilian sample set included the sex for each individual, sixteen males and four females, based on anthropological examination for comparison with the DNA results. Sample 015 was confirmed as female, with a zero quantification result for the Y marker, whilst amplification of the Y target in sample 019 resulted in a male designation thus contradicting the paperwork. It should be noted however, that whilst the total and Y quantification values were 26 pg/ µl and 23 pg/ µl respectively, the DI value for this sample was 17.3 indicating extreme degradation and no amplification was observed using the NGM amplification kit. Although no Y target amplification was observed for the remaining two samples that had been designated female, 007 and 012, both had total quantification values of <2 pg/µl (both had a zero quantification value for the Y marker) and so no sex was assigned.

It had been originally planned to use the Illumina suggested protocol for D-loop amplification (Illumina, 2013) to prepare the DNA library for sequencing of the mtGenome control region on the MiSeq platform. However, despite the demonstrated sensitivity of the assay down to an approximate DNA input of 16 pg, no amplified product was observed for any of the tested sample DNA extracts. The DNA in this sample set has been shown to be moderately-severely degraded as reflected in the low quantification values for the long autosomal marker, a 214 bp target, amplified during qPCR, therefore the 236-256 bp target range of the Illumina proposed primer sets was not attainable for these samples. The presence of mtDNA albeit in a degraded state, was demonstrated in four of these six samples using the Edson primer set, which were designed for degraded DNA samples (Edson, 2004) having a target of just 125 bp.

Following the success of the Edson primer set, an ‘in-house’ assay was developed through adaptation of the 16S protocol ustilsed by Kozich et al. (Kozich et al., 2013), whereby primers were extended to include all the necessary attachments for MPS on the Illumina MiSeq platform. Five samples and one extraction negative control were selected for analysis using the in-house assay. Following sequencing, Illumina’s integrated software ‘Basespace’ should demulitplex samples based on the index reads, but due to software issues the six samples in this study could not be automatically demultiplexed. Fortunately, the sequencing method and amplicon design enabled the entire target sequence and the reverse index to be sequenced during read 1, and the target sequence plus the forward index to be sequenced in read 2, which allowed for samples to be manually demultiplexed as described in appendix 8.23.

The total number of reads generated during this run was 287,446, 82% of which were an exact match for one of the utilised index sequences. The remaining 18% are unmatched reads, resulting from the stringent filtering process applied during demulitplexing, which did not allow for any mismatches to be accepted within the index sequence. Amplicon C had twice the number of associated reads relative to the other amplicons, which could be the effects of preferential amplification and sequencing as this had the smallest target amplicon size of 132 bp. Whilst over 34,000 indexed reads were obtained for Amplicon B, just 0.6% of these could be aligned against the rCRS. The drop in read number occurred during the combining of the paired end reads that was carried out as described in step four of appendix 8.23 as part of the DNA fragment identification process. This most likely due to either an inconsistency in the primer set-up during preparation for loading to the MiSeq flowcell, or a homopolymorphic region between positions 16185-16194 (CCCCCTCCCC) in the template sequence, which has caused strand bias during amplification, and therefore the subsequent loss of reads as only those combined in pairs were advanced through the process. Despite the requirement for further optimisation in order to better balance the coverage across the four amplicons, all paired reads were successfully aligned against the human rCRS. This method has many advantages over the commonly used two-step PCR method; the potential for further exacerbating amplification bias within the sample is reduced, DNA fragments are indexed in the first PCR reducing the risk of sample confusion, a lower number of required manipulations minimises the potential for sample loss or contamination and there are also associated time and cost efficiency gains. This method will prove particularly useful for degraded DNA, as a high number of primers can be multiplexed to amplify the required target region in a number of short fragments, thereby increasing the chance of success.

5.5.2 Vindolanda sample set

The human skull, from which the tooth submitted for DNA analysis was extracted, was excavated from a ditch dated to the Severan period, circa 208-212 CE based on associated coin evidence (Birley, 2009). The individual had suffered severe *peri mortem* injury, with evidence of at least three blows to the head with a sharp weapon as well as decapitation and mounted on a pole, possibly as a trophy (Birley, 2009) or warning to others. There are two conflicting theories on the origins of this individual; firstly that that he was a native Briton, possibly one of the revolting Maeatae killed at the hands of the imperial expedition who had been deployed to eradicate them, or secondly that he was a disgraced member of the Roman army (A. Birley 2015 pers. comm. 16th, April).

The femur bone was recovered as part of a partial skeleton, excavated from one of the earliest fort ditches dated circa 110 AD. The skeletal remains were found strewn along a section of the ditch suggesting that the body may have been in flowing water at some point *post mortem* (A. Birley/T. Buck 2015 pers. comm. 16th, April). It was highly unusual for the Romans during this period to bury their dead so the discovery of remains from this era is very rare. Aside from this, the presence of general debris in the area suggested that this individual had been left without a burial of any kind, once again raising questions about whether this individual was a native who died at the hands of the army or perhaps a disgraced Roman (A. Birley 2015 pers. comm. 16th April). In lieu of the skull or pelvis, anthropologist Dr Trudi Buck had tentatively assigned female sex to the remains based on the size of the femur bone, although this was complicated by the open ends of the bone, as whilst the length was on the larger size for a female, it was on the short size for a Roman soldier, who would be expected to have a height over 6 ft (T. Buck 2015 pers. comm. 16th, April).

Previous analysis by York University had already established the presence of mtDNA within the femur sample through the amplification and sequence of the control region. The resulting haplotype was then used to predict the individual's mitochondrial haplogroup as J1c1, which is not very common in Europe, with an observed frequency of approximately 9% for the whole of haplogroup J (http://www.mitomap.org/MITOMAP/GBFreqInfo, last accessed 28/05/2016). Although relatively uncommon, this group is widely dispersed across Europe (Gonzalez Fortes, 2014) and therefore not highly informative in this case.

In order to address the questions associated with both of these samples, particularly in terms of sex confirmation for the femur sample, and specific biogeographic ancestry data along with the potential for additional phenotypic information for both samples, the decision was taken to pursue the amplification of DNA markers.

The immediate environment in which the sample is situated *post mortem* does have a demonstrable impact on the persistence of DNA within the sample (Gilbert et al., 2003, Lamers et al., 2009, Bollongino et al., 2008, Burger et al., 1999). Clarifying the extent to which different variables influence DNA persistence is problematic due to the small sample sets available from each environment, contamination issues and the lack of taphonomic information available in most cases. Nuclear DNA was however detected in both of the femur and tooth DNA extracts with values of 35 pg/µl, 71 pg/µl and 56 pg/µl respectively. The persistence of DNA in these two specimens is likely due to both the anaerobic environment in which both samples were kept and the temperate nature of the UK weather. An anaerobic environment, such as that described in the some strata of the Vindolanda site, will reduce the rate of DNA degradation though chemical modifications due to oxidation and hydrolysis, particularly depurination which can lead to further fragmentation of the DNA (Briggs et al., 2007, Ginolhac et al., 2012) as well as limiting bioerosion due to microbial activity (Campos et al., 2012, Child, 1995), thus better preserving the integrity of the DNA relative to an aerobic environment. This is in stark contrast to the severe degradation observed in the Brazilian sample set, thus demonstrating the effects of the environment are more influential on the rate of degradation than PMI.

Considering that the tooth sample is eighteen hundred years in age, the DNA within this sample was remarkably well preserved, showing signs of only mild degradation, whereas the nineteen hundred year old femur sample yielded DNA that only just fell short of the severely degraded DNA category (Vernarecci et al., 2015). Aside from the approximate one hundred year difference in specimen age, it is well documented that tooth samples contain a higher quality of DNA than bone samples, due to the enamel coating of the tooth sample and encasement of the root section within the jaw bone, both of which facilitate the preservation of DNA through limiting access to it by the surrounding microbiome, oxygen and water.

Interestingly, the Quantifiler® Trio result indicated that the femur sample, which had a female sex assigned anthropologically, actually originated from a male individual. This was supported with the amplification of Amelogenin using the AmpFℓSTR® NGM™ PCR Kit (NGM™) in each of the femur DNA extracts and also through amplification of Y-STRs using the AmpFℓSTR® Yfiler® PCR Kit (YFiler®). Two negative controls were processed for each of the femur extractions; of these four, amplified DNA was only detected in one control with a Quantifiler® value of <0.5 pg/µl, thus indicating that the DNA amplified in each of the extracts was endogenous to the sample and not from a foreign source. On further discussion with Dr Buck, the possibility was raised that due to the shorter stature of this individual, he may in fact have been a slave brought in to serve the garrison officers. The biogeographic ancestry of this individual may shed further light on this proposal, however

unfortunately no DNA amplification was observed using the ForenSeq™ DNA Signature Kit (ForenSeq™) for this sample. Given the demonstrated amplification of DNA using Quantifiler® Trio, in addition to consensus amplification using both the NGM™ and YFiler® amplification kits, the most likely explanation for this is there is a PCR inhibitor present in the DNA extract which did not affect the Thermo Scientific kits, but which prevented amplification using the ForenSeq™ chemistry. This may be due to a difference in the polymerase used during the PCR or that the complexity of a multiplex this size is unable to tolerate inhibitors as well as the smaller PCR-CE assays (N. Oldroyd, Illumina Inc. 2015 pers. comm. 24th June).

PCR inhibitors associated with skeletal remains include calcium and collagen from the remains themselves, and humic acid absorbed from the surrounding soil; it is possible that one or more of these may have been co-extracted with the sample DNA. A paper by Opel et al. (Opel et al., 2010) used qPCR to demonstrate that each of these molecules utilises differing mechanisms of inhibition; calcium competes with the magnesium ions thereby reducing the efficacy of the PCR polymerase, collagen binds with the DNA template hindering the progress of the polymerase as it extends the primers, whilst humic acid acts to bind with sequence-specific areas of the DNA strand, thus restricting access to the template by the polymerase. It is not possible to determine the extent to which these inhibitors may be affecting PCR using the ForenSeq™ kit, as the degraded nature of the DNA within the femur sample is already subject to stochastic effects resulting from amplification.

It is worth noting at this point that although DNA was extracted from both the tooth and femur extracts using the same protocol, DNA from the tooth extract was successfully amplified using ForenSeq™. This may be down to the fact that the better quality of DNA in the tooth sample made it more amenable to amplification in the presence of the same inhibitors that prevented the degraded femur DNA from undergoing PCR. As the two samples were excavated from differing locations and strata at the Vindolanda site, however, they may consequently have had different types or quantities of inhibitors co-extracted with the samples. Alternatively, as less collagen is present in teeth than in bone, along with the possibility that the enamel layer of the tooth sample may have offered protection against the imbibing of humic acid from the surrounding environment, this may have resulted in less of these inhibitors being present compared with the femur extract. A greater mass of the powdered femur sample (1.5 g) was also used for the extraction process than the tooth sample (1.1 g) which, together with one or more of the aforementioned possibilities, may also account or the difference in amplification success between the two samples.

Optimisation of the PCR assay can facilitate DNA amplification in the presence of inhibitors; however, in this case the ForenSeq™ the reaction components are already finely balanced in order to accommodate the large number of loci in the multiplex and so it would be preferable to remove the inhibitors from the DNA extract itself. Although a second attempt to amplify the femur sample using the ForenSeq™ kit will be carried out, it will not be completed within the time frame of this study.

As expected, the results of the Quantifiler® Trio and NGM™/YFiler® kits confirmed the presence of male DNA in the tooth sample. As previously discussed, the biogeographical ancestry of this sample is of particular interest and so both Y-STRs and aSNPs were employed to draw conclusions about the individual in question.

The Y-STR haplotype determined by combining both the YFiler® and ForenSeq™ Y-STR results, was entered into both the YHRD (Willuweit and Roewer, 2015) (release 50) and the online Haplogroup Predictor tool (Athey, 2015). The ancestry analysis performed on the YHRD website is based on a minimal haplotype, made up of eight loci, of which only two were amplified in each of the tooth and femur samples and so no meaningful information could be obtained from these results. The online Haplogroup predictor tool takes into account all of the available loci which enabled the assignment of haplogroup R1b to the tooth sample with a 99.7% probability (Athey, 2015). One of two major sub-clades of the haplogroup R1, R1b is found predominately in Western Europe (Myres et al., 2011), which supports the theory that the individual was either native to the UK or had travelled from Europe. The results for the femur sample are inconclusive given that only three consensus alleles were amplified and available for analysis.

Illumina's ForenSeq™ Universal Analysis Software utilises principal component analysis (PCA) to determine the biogeographical ancestry based on the generated aSNP profile. Amplification of the tooth DNA extract resulted in 63% of the aSNPs being successfully amplified with the individual being placed as European with the distance to the nearest centroid given as 2.40. The closest reference samples were from the TSI (Toscani in Italia) grouping, indicating that the individual is more closely related to samples in this region than other reference samples within the centroid. The statistical calculations utilised for these projections are not yet published for further analysis (L. Devesse 2105 pers. comm. 23rd October).

Wherever possible, it is good practice to carry out DNA analyses on ancient samples in two independent laboratories in order to support the authentication of any DNA results obtained. Due to the precious nature of these aged samples, the quantity of sample available for testing was limited to one tooth from the skull and a small section of the femur, and so it was not

possible to carry out independent tests in this manner. However, as previously mentioned, DNA was extracted from this same femur sample by York University for mtDNA sequencing of the control region and, although no attempt was made to amplify DNA on this occasion, this represented results from an independent laboratory. Two sections of the mtGenome were consequently amplified to enable comparison with five out of the eight differences from the rCRS that were recorded by York University (Gonzalez Fortes, 2014); all five of these substitutions were confirmed in each of the femur extracts. An additional substitution of 263G was identified in both femur extracts which was not included in the report. This was queried with Dr Gloria Gonzalez Fortes who carried out the original mtDNA sequencing work at York University, who reported that there was a gap in her consensus sequence between positions 247 and 267 and consequently position 263 was not reported (G. Gonzalez 2015 pers. comm. 24th February).

It is expected that the increased mtGenome copy number per cell along with its inherent resistance to degradation relative to nDNA, would result in detection of mtDNA in all samples for which DNA was detected. However in this study the presence of DNA was detected in all twenty-four of the submitted Brazilian samples, whilst the mtDNA target was only amplified in eleven of these. As discussed for the ForenSeq kit above, the most probable explanation for this is that the in-house mtDNA qPCR assay is less tolerant to the presence of inhibitors than the Quantifiler® Trio chemistry. Signs of inhibition were detected in only one of the samples using the Quantifiler® Trio internal positive control which had a Ct value or 38.11, well above the calculated flag threshold of 30.71 for this run. If this in indeed the case then given that the internal positive control values did not indicate any additional samples as showing signs of inhibition, then it can be concluded that the Quantifiler® Trio results accurately represent the quantity of DNA within the samples, but the mtDNA copy numbers may be significantly under-represented in this study.

Sample 013B was extracted using three different methods; the Quantifiler® Trio value for the silica-column-based method was an order of magnitude higher than for the other two methods, whereas the variation within other samples was very limited. Subsequent DNA profiling using both the NGM and ForenSeq™ kits demonstrated the presence of a DNA mixture that could be attributed to the operator. Although the quantity of DNA was elevated, there was no indication of a male/female mix in this sample as the quantification value for both the Y and short autosomal targets was 0.47 ng/µl. It is possible that the sample was contaminated post-quantification, however with a male/female mix the ratio of X:Y DNA is 3:1 and therefore an increased representation of the X markers would be expected. Conversely both of the consensus NGM profiles and the ForenSeq™ data, the Y marker had

the higher RFU (Hb = 0.66 and 0.42) or read count (Hb = 0.36). This serves to highlight the stochastic effects that can be observed when working with low level of DNA template.

As discussed in Section 5.1, one of the advantages of MPS is the sequence data obtained for each STR as it allows for the detection of isolalleles. An example of this was seen in the results of the ForenSeq™ amplification on the Vindolanda tooth DNA extract. The low read counts obtained for this sample means that the designation of locus D13S317 allele 11, would have been recorded as a possible heterozygote with allelic dropout. The sequence data however indicates the presence of two distinct sequences, with a read count of thirty-nine for the expected 'TATC' repeat and forty-two copies of 'AATC', a SNP variant in the flaking region immediate to the 3′ end of the repeating unit.

5.6 Conclusions

Of the three DNA extraction methods selected, there was not one that consistently yielded a higher quantity or quality of DNA relative to the other two tested.

The removal of dentin/cementum from tooth samples in order to preserve the casing is unlikely to yield amplifiable DNA in sub-optimal samples, due to the low mass of material produced using this method.

Nuclear DNA was detected in all of twenty of the Brazilian samples and both the Vindolanda femur and tooth samples. The detrimental effects of a hot and humid environment on buried skeletal remains is clear as the samples from Brazil were observed to be of low quantity and poor quality despite a *post mortem* interval of just 8 -14 years.

Endogenous DNA is believed to have been extracted and amplified from both the Vindolanda tooth and femur samples, demonstrating that the anaerobic conditions at the Vindolanda site favour DNA preservation. When compared with the Brazilian sample set, this also shows that given good preservation conditions, the age of the sample, in this case 1800 – 1900 years, is of less consequence in terms of the quantity and quality of extracted DNA, than temperature, humidity and an aerobic environment.

Although Quantifiler Trio is sensitive to <1 pg/µl DNA, assigning sex at this level should be treated with caution. The results from this study suggest that the Quantifiler Trio assay chemistry is highly tolerant of PCR inhibitors, and so may not be reflective of likelihood of obtaining a PCR result if using assays, either developed in-house or commercially available, which rely on other chemistries that may be less tolerant to the presence of PCR inhibitors.

The Illumina suggested protocol for amplification of mtGenome D-loop region was shown to be unsuitable for amplification of moderately-severely degraded DNA, presumably owing to the relatively large target fragment size of >230 bp.

Adapting primers for a one-step amplification reaction in which all components necessary for direct flowcell attachment were added to the target fragment, proved a sensitive and time/cost effective method of targeted sequencing on the MiSeq MPS platform.

The multi-marker ForenSeq™ Signature Preparation Kit enables a more comprehensive evaluation of DNA than any other assay currently available on the forensic market. The inclusion of identity SNPs alongside the STR markers ensures a greater chance of amplification of degraded forensic material. The capability to amplify the ancestry and phenotypic SNPs are particularly useful in its application to bioarchaeological investigations that would benefit from a forensic approach.

6. Conclusions

The overall aim of this thesis was to explore the use of enhanced experimental strategies for the analysis of sub-optimal biological material that is encountered during forensic investigations. In order to fulfil this aim two main areas of research were selected, namely ltDNA analysis, using single-cell DNA analysis as an example of an underexploited source of DNA in forensic casework, and severely degraded DNA samples, in this instance obtained through analysis of bone/teeth samples, which are aged or have been subjected to hostile environmental conditions.

The first area of research focusing on single-cell DNA analysis was further separated into two sections; the initial part aimed to investigate the analysis and interpretation of DNA derived from individual single-cells in order to determine their efficacy for use in forensic casework, whilst the second section then considered a novel application of this work for use in forensic casework.

The difficulties associated with single-cell DNA analysis has resulted in the under-exploitation of this sample type within a forensic context, however, given the development in our knowledge on the processing and interpretation of ltDNA samples, alongside the increased sensitivity of the commercially available assays, a reinvestigation of this potentially valuable source of DNA is warranted. In order to address the aim of the first part of this study it was first necessary to develop a method by which individual single-cells could be detected and transferred from a surface to a PCR tube without sample loss. This was vital to ensure that any generated DNA data could be confidently attributed to a single-cell as the developed protocol included a visual check of the sample under the microscope immediately prior to transfer into the PCR tube. J-LAR adhesive tape and a 1 µl volume of Sticky Stuff Remover were found to be compatible with downstream PCR, whilst the use of the ForensicGEM® Tissue extraction reagents facilitated direct amplification of DNA using the AmpFℓSTR® SGM Plus® PCR Amplification Kit. Following development, this protocol for the detection, recovery and amplification was utilised to generate DNA profiles for fifty individual single-cells, enabling an assessment of the effects of artefacts associated with DNA profiling at this low quantity of starting template. Overall, the results for this data set demonstrate that the frequency and intensity of artefacts resulting from the stochastic effects of PCR are comparable with those observed using DNA extracted from any ltDNA samples and can, therefore, be interpreted using the same parameters. Therefore, although a consensus DNA profile cannot be generated from a single-cell, the move towards computer-based statistical likelihood data analysis and evaluation for all ltDNA samples should facilitate the increased use of single-cells as a source of DNA in forensic casework. An overall drop-in rate of <1% demonstrates that provided good DNA anti-contamination practice is used, the impact of this unavoidable artefact is minimal. Finally an initial trial of

several recently developed, and purportedly more sensitive, amplification assays were tested using individual single-cells as a source of DNA. Of these kits Promega's ESX-16 amplification assay, used in conjunction with ForensicGEM extraction reagents, appears to be particularly effective for amplification of DNA from individual single-cells.

As an application of the aforementioned single-cell DNA analysis work, the second section of this area of research tested the hypothesis that when textile fibres are transferred from a garment to a second surface, they are also acting as a vector for the wearer's own DNA, via cells that have adhered to the fibre surfaces. This would offer forensic scientists a method by which transferred fibres could be targeted on fibre tapings, in order to obtain a DNA profile. Rather than the use of DNA evidence at the expense or destruction of the fibre evidence, as so often happens in forensic casework. The intention of this work was that the two evidence types should be examined with causing detrimental effects to the other so that both could be utilised alongside one another in forensic casework.

It was quickly established that the use of a DAPI-based fluorescent stain, in place of the previously used haematoxylin and eosin, was an effective method for the visualisation of cells on the surface of textile fibres without staining the fibre itself, thereby showing no detrimental effect on the MSP spectral profile. Furthermore it was demonstrated that the DAPI stain did not cause PCR inhibition and so was compatible with downstream DNA analyses. Prior to testing of the hypothesis, it was deemed necessary to first establish the frequency with which nucleated cells are likely to be present on the outside front surface of an individual's garment, in order to inform expectations on the availability of cells for transfer to a second surface. It was demonstrated through the staining of entire fibre tapes that the garments sampled for this study had on average over one thousand nucleated cells present and as such represent a source of DNA for possible transfer to a recipient surface. Of particular interest, was the presence of clusters of nucleated cells in addition to the individual cells, which were often observed, wrapped around the fibre surface therefore representing a high chance of transfer. A total of ten fibre/cell transfer experiments were then conducted in order to test the previously stated hypothesis, during which time a number of developments to the protocol established in the previous chapter were made in order to account for the extended length of time required for the examination, staining, searching and recovery of fibres relative to the initial single-cell experiments. Whilst it has been demonstrated that nucleated cells can be visualised *in situ* on fibre tapes and recovered for DNA analysis without adversely affecting subsequent fibre analysis, the results of this study suggest that if fibres are in fact acting as a vector for the transfer of DNA then they do so rarely and as such, specifically targeting transferred fibres for an associated DNA profile is unlikely to yield a result.

The second area of research is focused on DNA contained within human skeletal remains that has severely degraded as a result of temporal or environmental factors, and therefore represents a significant challenge for forensic DNA analysis. The aim of this part of the thesis is to investigate the efficacy of novel and emerging MPS techniques in the targeted sequencing of degraded DNA samples, using existing laboratory strategies as a point of comparison. In order to address the aims of this section, two distinct sample sets were obtained; twenty samples (8-14 years PMI) excavated from a site in Brazil, which had been subjected to extremely hostile environmental conditions, and two samples from the Vindolanda Roman archaeological dig site, which are over 1800 years in age.

The starting point for this work was a comparison of commonly used methods used for the extraction of DNA from skeletal material. Of the three DNA extraction methods selected, there was not one that consistently yielded a higher quantity or quality of DNA relative to the other two tested.

Quantifiler® Trio, a recently introduced real-time PCR quantification assay, was utilised to quantify the DNA in all samples. Nuclear DNA was detected in all of twenty of the Brazilian samples although the detrimental effects of a hot and humid environment on buried skeletal remains is clear, as the samples from Brazil were observed to be of low quantity and poor quality despite a *post mortem* interval of just 8 -14 years. DNA was also detected in both of the Vindolanda samples. This DNA is believed to be endogenous to the samples, demonstrating that the anaerobic conditions at the Vindolanda site favour DNA preservation. When compared with the Brazilian sample set, this also shows that given good preservation conditions, the age of the sample, in this case 1800 – 1900 years, is of less consequence in terms of the quantity and quality of extracted DNA, than temperature, humidity and an aerobic environment. Quantifiler® Trio purports to be sensitive to <1 pg/µl DNA, assigning sex at this level should be treated with caution. The results from this study suggest that the Quantifiler® Trio assay chemistry is highly tolerant of PCR inhibitors, and so may not be reflective of likelihood of obtaining a PCR result if using assays, either developed in-house or commercially available, which rely on other chemistries that may be less tolerant to the presence of PCR inhibitors.

Following DNA extraction and quantification, a variety of strategies were tested for the amplification and targeted sequencing of the degraded samples. Initially the Illumina suggested protocol for amplification of mtGenome D-loop region was to be used for the amplification of the Brazilian samples, however this method was shown to be unsuitable for amplification of moderately-severely degraded DNA, owing to the relatively large target fragment size of >230 bp. It was decided to adapt a number of primers that amplified a

shorter fragment length, thus facilitating the amplification of the degraded samples, for direct flowcell attachment by adding all the necessary components to the target fragment in a one-step PCR. This method, which has not been used in a forensic context at the time of writing, was successful and proved to be a sensitive and time/cost effective method of targeted sequencing on the MiSeq MPS platform. The Vindolanda samples were amplified using AmpFℓSTR® NGM™ and Yfiler® PCR Amplification Kits, which are existing PCR-CE methods as well as the newly released ForenSeq™ Signature Preparation Kit. This multi-marker assay enabled a more comprehensive evaluation of DNA than any other assay currently available on the forensic market. The inclusion of identity SNP alongside the STR markers ensures a greater chance of amplification of degraded forensic material. The capability to amplify the ancestry and phenotypic SNPs is particularly useful in its application to bioarchaeological investigations that would benefit from a forensic approach.

There are many advantages to support the introduction of MPS into routine forensic casework. The ability to multiplex different forensic markers of interest offers the opportunity to obtain more information from a smaller quantity of DNA, running one rather than multiple reactions, also saving time and potentially cost in the process. Removing the restriction on the number of dye lanes associated with the use of CE, enables an increased number of STR loci to be included for analysis; including loci in use around the world to facilitate global crime fighting efforts in areas such as terrorism, drug trafficking and paedophilia, and also miniSTRs /identity SNPs to aid detection of low template and degraded DNA samples. Phenotypic SNPs for hair, eye and skin tone have the potential use for intelligence purposes when trying to identify an unknown offender; although this is of course somewhat limited in that artificial change can be made to these features. Ancestry information can also be utilised to infer the ethnicity of an individual using ancestry SNPs, Y-STRs and mtDNA. In addition to the number of markers used, the generation of actual sequence information for STRs rather than fragment size only as used with CE, increases the ability to distinguish between two seemingly homozygote alleles as well as identifying the proportion of a peak that may be stutter of the subsequent allele, both of which enables a better resolution of mixed DNA samples.

However, rather than a simple extension of current techniques, as the introduction of DNA 17 in England and Wales in recent years was, MPS is a complete deviation from current methods, with the interpretation process almost unrecognisable. The platforms used for MPS are expensive and unfamiliar to many forensic scientists, requiring a huge investment in terms of equipment and staff training by forensic providers for successful implementation, which may not be desirable or feasible given the current state of the forensic market within the UK. In addition to this, extensive validation processes will be required in order to

properly accredit all the new methods used, which again are expensive in terms of time and cost. Discussions around the need for the additional information that could be obtained from MPS will be essential, particularly with respect to at what point in the investigation phenotypic and ancestry information should be generated, how this information should be stored and who should have access to it? Changes to the National DNA Database will also be required in order to utilise the additional STR and identity SNP data obtained from every sample submitted.

Even with obvious advantages over current techniques, completing all the validation processes and developing a good level of expertise for use within the criminal Justice System, along with the understanding and trust of the public, police and other interested parties, means that the introduction of MPS into routine casework will be a long process.

6.1 Further work

Following the completion of this study there are a number of suggested topics of research, which would further contribute to knowledge and understanding in the areas focused on during this thesis:

- As computer-based statistical programmes are being developed with increasing capabilities, an investigation into the statistical evaluation of single-cell DNA profiles is essential for use in forensic casework.
- As discussed in Section 4.4, it would be interesting to carry out RNA work on clusters of cells detected on the outside surface of garments, in order to determine the body fluid of origin.
- Further detailed investigation into DNA extraction methods for degraded skeletal material would be beneficial for both forensic and bioarchaeological fields.
- Primer sets adapted for the MiSeq platform should be extended to enable amplification of the entire control region of the mtGenome, all of which will require optimisation in order to maximise results.

7. References

1993. Daubert v. Merrell Dow Pharmaceuticals. U.S. 579.

Aird, D., Ross, M. G., Chen, W. S., Danielsson, M., Fennell, T., Russ, C., Jaffe, D. B., Nusbaum, C. & Gnirke, A. 2011. Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. *Genome Biol,* 12**,** R18.

Akulova, V., Vasiliauskiene, D. & Talaliene, D. 2002. Further insights into the persistence of transferred fibres on outdoor clothes. *Sci Justice,* 42**,** 165-71.

Albert, T. J., Molla, M. N., Muzny, D. M., Nazareth, L., Wheeler, D., Song, X., Richmond, T. A., Middle, C. M., Rodesch, M. J., Packard, C. J., Weinstock, G. M. & Gibbs, R. A. 2007. Direct selection of human genomic loci by microarray hybridization. *Nat Methods,* 4**,** 903-5.

Alessandrini, F., Cecati, M., Pesaresi, M., Turchi, C., Carle, F. & Tagliabracci, A. 2003. Fingerprints as evidence for a genetic profile: morphological study on fingerprints and analysis of exogenous and individual factors affecting DNA typing. *J Forensic Sci,* 48**,** 586-92.

Allentoft, M. E., Collins, M., Harker, D., Haile, J., Oskam, C. L., Hale, M. L., Campos, P. F., Samaniego, J. A., Gilbert, M. T., Willerslev, E., Zhang, G., Scofield, R. P., Holdaway, R. N. & Bunce, M. 2012. The half-life of DNA in bone: measuring decay kinetics in 158 dated fossils. *Proc Biol Sci,* 279**,** 4724-33.

Anderson, S., Bankier, A. T., Barrell, B. G., De Bruijn, M. H., Coulson, A. R., Drouin, J., Eperon, I. C., Nierlich, D. P., Roe, B. A., Sanger, F., Schreier, P. H., Smith, A. J., Staden, R. & Young, I. G. 1981. Sequence and organization of the human mitochondrial genome. *Nature,* 290**,** 457-65.

Anderson, T. D., Ross, J. P., Roby, R. K., Lee, D. A. & Holland, M. M. 1999. A validation study for the extraction and analysis of DNA from human nail material and its application to forensic casework. *J Forensic Sci,* 44**,** 1053-6.

Andrews, R. M., Kubacka, I., Chinnery, P. F., Lightowlers, R. N., Turnbull, D. M. & Howell, N. 1999. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat Genet,* 23**,** 147.

Ankel-Simons, F. & Cummins, J. M. 1996. Misconceptions about mitochondria and mammalian fertilization: implications for theories on human evolution. *Proc Natl Acad Sci U S A,* 93**,** 13859-63.

Aponte, R. a. G., K.B. Duewer, D.L. Coble, M.D. Vallone, P.M. 2015. Sequence-based analysis of stutter at STR loci: Characterization and utility. *Forensic Science International: Genetics Supplement Series*.

Athey, T. W. 2005. Haplogroup prediction from Y-STR values using an allele-frequency approach. *J Genet Geneal,* 1**,** 1-7.

Athey, T. W. 2006. Haplogroup prediction from Y-STR values using a Bayesian-allele-frequency approach. *J Genet Geneal,* 2**,** 34-39.

Athey, W. 2015. *Y Haplogroup Prediction from Y-STR Values* [Online]. http://www.hprg.com/hapest5/index.html [Date last accessed 18/07/2015.

Bada, J., Shou, M. Y., Hare, P. E., Hoering, T. C., King, K., Carnegie Institution of Washington. & National Science Foundation (U.S.) 1980. *Biogeochemistry of amino acids : papers presented at a conference at Airlie House, Warrenton, Virginia, October 29 to November 1, 1978,* New York, Wiley.

Balding, D. J. & Nichols, R. A. 1994. DNA profile match probability calculation: how to allow for population stratification, relatedness, database selection and single bands. *Forensic Sci Int,* 64**,** 125-40.

Bayley, H. 2015. Nanopore sequencing: from imagination to reality. *Clin Chem,* 61**,** 25-31.

Bergquist, P. L., Love, D. R., Croft, J. E., Streiff, M. B., Daniel, R. M. & Morgan, W. H. 1987. Genetics and potential biotechnological applications of thermophilic and extremely thermophilic micro-organisms. *Biotechnol Genet Eng Rev,* 5**,** 199-244.

Biermann, T. W. 2007. Blocks of colour IV: The evidential value of blue and red cotton fibres. *Science & Justice,* 47**,** 68-87.

Bintz, B. J., Dixon, G. B. & Wilson, M. R. 2014. Simultaneous detection of human mitochondrial DNA and nuclear-inserted mitochondrial-origin sequences (NumtS) using forensic mtDNA amplification strategies and pyrosequencing technology. *J Forensic Sci,* 59**,** 1064-73.

Biosystems, A. 2015. AmpFlSTR® NGM™ PCR Amplification Kit User Guide. https://tools.lifetechnologies.com/content/sfs/manuals/cms_074250.pdf: Life Technologies.

Birley, R. 2007. *Vindolanda : the home of Britain's finest treasure,* Greenhead, Roman Army Museum Publications.

Birley, R. 2009. *Vindolanda : a Roman frontier fort on Hadrian's wall,* Stroud, Amberley.

Blanchard, J. L. & Schmidt, G. W. 1996. Mitochondrial DNA migration events in yeast and humans: integration by a common end-joining mechanism and alternative perspectives on nucleotide substitution patterns. *Mol Biol Evol,* 13**,** 893.

Blankenberg, D., Gordon, A., Von Kuster, G., Coraor, N., Taylor, J., Nekrutenko, A. & Galaxy, T. 2010. Manipulation of FASTQ data with Galaxy. *Bioinformatics,* 26**,** 1783-5.

Bogenhagen, D. F. 2012. Mitochondrial DNA nucleoid structure. *Biochimica Et Biophysica Acta-Gene Regulatory Mechanisms,* 1819**,** 914-920.

Bollongino, R., Tresset, A. & Vigne, J. D. 2008. Environment and excavation: Pre-lab impacts on ancient DNA analyses. *Comptes Rendus Palevol,* 7**,** 91-98.

Boore, J. L. 1999. Animal mitochondrial genomes. *Nucleic Acids Res,* 27**,** 1767-80.

Bornman, D. M., Hester, M. E., Schuetter, J. M., Kasoji, M. D., Minard-Smith, A., Barden, C. A., Nelson, S. C., Godbold, G. D., Baker, C. H., Yang, B., Walther, J. E., Tornes, I. E., Yan, P. S., Rodriguez, B., Bundschuh, R., Dickens, M. L., Young, B. A. & Faith, S. A. 2012. Short-read, high-throughput sequencing technology for STR genotyping. *Biotech Rapid Dispatches,* 2012**,** 1-6.

Borsting, C., Fordyce, S. L., Olofsson, J., Mogensen, H. S. & Morling, N. 2014. Evaluation of the Ion Torrent HID SNP 169-plex: A SNP typing assay developed for human identification by second generation sequencing. *Forensic Sci Int Genet,* 12**,** 144-54.

Borsting, C. & Morling, N. 2015. Next generation sequencing and its applications in forensic genetics. *Forensic Sci Int Genet*.

Bouwman, A. S. & Brown, T. A. 2005. The limits of biomolecular palaeopathology: ancient DNA cannot be used to study venereal syphilis. *Journal of Archaeological Science,* 32**,** 703-713.

Bowman, R. D. & Davidson, N. 1972. Hydrodynamic shear breakage of DNA. *Biopolymers,* 11**,** 2601-24.

Briggs, A. W., Stenzel, U., Johnson, P. L., Green, R. E., Kelso, J., Prufer, K., Meyer, M., Krause, J., Ronan, M. T., Lachmann, M. & Paabo, S. 2007. Patterns of damage in genomic DNA sequences from a Neandertal. *Proc Natl Acad Sci U S A,* 104**,** 14616-21.

Bright, J. A., Stevenson, K. E., Coble, M. D., Hill, C. R., Curran, J. M. & Buckleton, J. S. 2014. Characterising the STR locus D6S1043 and examination of its effect on stutter rates. *Forensic Sci Int Genet,* 8**,** 20-3.

Brown, T. A., Tkachuk, A. N., Shtengel, G., Kopek, B. G., Bogenhagen, D. F., Hess, H. F. & Clayton, D. A. 2011. Superresolution fluorescence imaging of mitochondrial nucleoids reveals their spatial range, limits, and membrane interaction. *Mol Cell Biol,* 31**,** 4994-5010.

Brown, W. M., George, M. & Wilson, A. C. 1979. Rapid Evolution of Animal Mitochondrial-DNA. *Proceedings of the National Academy of Sciences of the United States of America,* 76**,** 1967-1971.

Brudevold, F., Steadman, L. T. & Smith, F. A. 1960. Inorganic and organic components of tooth structure. *Ann N Y Acad Sci,* 85**,** 110-32.

Budowle, B., Eisenberg, A. J. & Van Daal, A. 2009a. Validity of low copy number typing and applications to forensic science. *Croat Med J,* 50**,** 207-17.

Budowle, B., Onorato, A. J., Callaghan, T. F., Della Manna, A., Gross, A. M., Guerrieri, R. A., Luttman, J. C. & Mcclure, D. L. 2009b. Mixture interpretation: defining the relevant features for guidelines for the assessment of mixed DNA profiles in forensic casework. *J Forensic Sci,* 54**,** 810-21.

Burbano, H. A., Hodges, E., Green, R. E., Briggs, A. W., Krause, J., Meyer, M., Good, J. M., Maricic, T., Johnson, P. L. F., Xuan, Z. Y., Rooks, M., Bhattacharjee, A., Brizuela, L., Albert, F. W., De La Rasilla, M., Fortea, J., Rosas, A., Lachmann, M., Hannon, G. J. & Paabo, S. 2010. Targeted Investigation of the Neandertal Genome by Array-Based Sequence Capture. *Science,* 328**,** 723-725.

Burger, J., Hummel, S., Herrmann, B. & Henke, W. 1999. DNA preservation: A microsatellite-DNA study on ancient skeletal remains. *Electrophoresis,* 20**,** 1722-1728.

Butler, J. M. 2015a. The future of forensic DNA analysis. *Philos Trans R Soc Lond B Biol Sci,* 370.

Butler, J. M. 2015b. U.S. initiatives to strengthen forensic science & international standards in forensic DNA. *Forensic Sci Int Genet,* 18**,** 4-20.

Butler, J. M., Buel, E., Crivellente, F. & Mccord, B. R. 2004. Forensic DNA typing by capillary electrophoresis using the ABI Prism 310 and 3100 genetic analyzers for STR analysis. *Electrophoresis,* 25**,** 1397-412.

Butler, J. M., Mccord, B. R., Jung, J. M., Lee, J. A., Budowle, B. & Allen, R. O. 1995. Application of dual internal standards for precise sizing of polymerase chain reaction products using capillary electrophoresis. *Electrophoresis,* 16**,** 974-80.

Butler, J. M., Shen, Y. & Mccord, B. R. 2003. The development of reduced size STR amplicons as tools for analysis of degraded DNA. *J Forensic Sci,* 48**,** 1054-64.

Butler, J. M. H., C.R. 2012. Biology and Genetics of New Autosomal STR loci Useful for Forensic DNA analysis. *Forensic Sci. Rev.,* 24**,** 15-26.

Calabrese, F. M., Simone, D. & Attimonelli, M. 2012. Primates and mouse NumtS in the UCSC Genome Browser. *BMC Bioinformatics,* 13 Suppl 4**,** S15.

Calloway, C. D., Reynolds, R. L., Herrin, G. L., Jr. & Anderson, W. W. 2000. The frequency of heteroplasmy in the HVII region of mtDNA differs across tissue types and increases with age. *Am J Hum Genet,* 66**,** 1384-97.

Campos, P. F., Craig, O. E., Turner-Walker, G., Peacock, E., Willerslev, E. & Gilbert, M. T. 2012. DNA in ancient bone - where is it located and how should we extract it? *Ann Anat,* 194**,** 7-16.

Cantrell, S., Roux, C., Maynard, P. & Robertson, J. 2001. A textile fibre survey as an aid to the interpretation of fibre evidence in the Sydney region. *Forensic Science International,* 123**,** 48-53.

Capaldi, R. A. 1990. Structure and function of cytochrome c oxidase. *Annu Rev Biochem,* 59**,** 569-96.

Carelli, V. 2015. Keeping in shape the dogma of mitochondrial DNA maternal inheritance. *PLoS Genet,* 11**,** e1005179.

Carracedo, A., Bar, W., Lincoln, P., Mayr, W., Morling, N., Olaisen, B., Schneider, P., Budowle, B., Brinkmann, B., Gill, P., Holland, M., Tully, G. & Wilson, M. 2000. DNA commission of the international society for forensic genetics: guidelines for mitochondrial DNA typing. *Forensic Sci Int,* 110**,** 79-85.

Carracedo, A. L., M.V. . Development of New STRs for Forensic Casework: Criteria for Selection, Sequencing & population data and forensic validation

Population Data and Forensic Validation. Ninth International Symposium on Human Identification, 1998 Orlando, CA. Promega website: Promega, 89-17.

Carter, P. J., Cheng, C. C. & Thorp, H. H. 1996. Oxidation of DNA Hairpins by Oxoruthenium(IV): Effects of Sterics and Secondary Structure. *Inorg Chem,* 35**,** 3348-3354.

Catalano, M. J., Liu, S., Andersen, N., Yang, Z. Y., Johnson, K. M., Price, N. E., Wang, Y. S. & Gates, K. S. 2015. Chemical Structure and Properties of Interstrand Cross-Links Formed by Reaction of Guanine Residues with Abasic Sites in Duplex DNA. *Journal of the American Chemical Society,* 137**,** 3933-3945.

Caulfield, J. L., Wishnok, J. S. & Tannenbaum, S. R. 1998. Nitric oxide-induced deamination of cytosine and guanine in deoxynucleosides and oligonucleotides. *J Biol Chem,* 273**,** 12689-95.

Cavelier, L., Johannisson, A. & Gyllensten, U. 2000. Analysis of mtDNA copy number and composition of single mitochondrial particles using flow cytometry and PCR. *Experimental Cell Research,* 259**,** 79-85.

Chaisson, M. J. & Tesler, G. 2012. Mapping single molecule sequencing reads using basic local alignment with successive refinement (BLASR): application and theory. *Bmc Bioinformatics,* 13.

Child, A. M. 1995. Towards an Understanding of the Microbial Decomposition of Archaeological Bone in the Burial Environment. *Journal of Archaeological Science,* 22**,** 165-174.

Churchill, J. D., Chang, J., Ge, J., Rajagopalan, N., Wootton, S. C., Chang, C. W., Lagace, R., Liao, W., King, J. L. & Budowle, B. 2015. Blind study evaluation illustrates utility of the Ion PGM system for use in human identity DNA typing. *Croat Med J,* 56**,** 218-29.

Clarke, J., Wu, H. C., Jayasinghe, L., Patel, A., Reid, S. & Bayley, H. 2009. Continuous base identification for single-molecule nanopore DNA sequencing. *Nature Nanotechnology,* 4**,** 265-270.

Clayton, T. M., Whitaker, J. P., Sparkes, R. & Gill, P. 1998. Analysis and interpretation of mixed forensic stains using DNA STR profiling. *Forensic Sci Int,* 91**,** 55-70.

Coble, M. D. & Butler, J. M. 2005. Characterization of new miniSTR loci to aid analysis of degraded DNA. *J Forensic Sci,* 50**,** 43-53.

Coble, M. D., Loreille, O. M., Wadhams, M. J., Edson, S. M., Maynard, K., Meyer, C. E., Niederstatter, H., Berger, C., Berger, B., Falsetti, A. B., Gill, P., Parson, W. & Finelli, L. N. 2009. Mystery solved: the identification of the two missing Romanov children using DNA analysis. *PLoS One,* 4**,** e4838.

Collins, F. S., Green, E. D., Guttmacher, A. E., Guyer, M. S. & Institute, U. S. N. H. G. R. 2003. A vision for the future of genomics research. *Nature,* 422**,** 835-47.

Collins, F. S., Lander, E. S., Rogers, J., Waterston, R. H. & Conso, I. H. G. S. 2004. Finishing the euchromatic sequence of the human genome. *Nature,* 431**,** 931-945.

Coloplast. 2013. *Brava Ahesive Remover* [Online]. http://www.coloplast.co.uk/brava-adhesive-remover-en-gb.aspx#section=product-description_3 [Date last accessed 01/03/2013 2013]].

Comai, L. H., T. 2012. *Barcode Generator* [Online]. http://comailab.genomecenter.ucdavis.edu/index.php/Barcode_generator [Date last accessed 16/10/2014.

Cook, R., Webbsalter, M. T. & Marshall, L. 1997. The significance of fibres found in head hair. *Forensic Science International,* 87**,** 155-160.

Cooke, M. S., Evans, M. D., Dizdaroglu, M. & Lunec, J. 2003. Oxidative DNA damage: mechanisms, mutation, and disease. *Faseb Journal,* 17**,** 1195-1214.

Cooper, A. & Poinar, H. N. 2000. Ancient DNA: do it right or not at all. *Science,* 289**,** 1139.

Cowen, S., Debenham, P., Dixon, A., Kutranov, S., Thomson, J. & Way, K. 2011. An investigation of the robustness of the consensus method of interpreting low-template DNA profiles. *Forensic Sci Int Genet,* 5**,** 400-6.

Dabney, J. & Meyer, M. 2012. Length and GC-biases during sequencing library amplification: a comparison of various polymerase-buffer systems with ancient and modern DNA sequencing libraries. *Biotechniques,* 52**,** 87-94.

Dalsgaard, S., Rockenbauer, E., Buchard, A., Mogensen, H. S., Frank-Hansen, R., Borsting, C. & Morling, N. 2014. Non-uniform phenotyping of D12S391 resolved by second generation sequencing. *Forensic Sci Int Genet,* 8**,** 195-9.

Daly, D. J., Murphy, C. & Mcdermott, S. D. 2012. The transfer of touch DNA from hands to glass, fabric and wood. *Forensic Sci Int Genet,* 6**,** 41-6.

Davis, C., Peters, D., Warshauer, D., King, J. & Budowle, B. 2015. Sequencing the hypervariable regions of human mitochondrial DNA using massively parallel sequencing: Enhanced data acquisition for DNA samples encountered in forensic testing. *Leg Med (Tokyo),* 17**,** 123-7.

Davison, P. F. 1959. The Effect of Hydrodynamic Shear on the Deoxyribonucleic Acid from T(2) and T(4) Bacteriophages. *Proc Natl Acad Sci U S A,* 45**,** 1560-8.

Dayama, G., Emery, S. B., Kidd, J. M. & Mills, R. E. 2014. The genomic landscape of polymorphic human nuclear mitochondrial insertions. *Nucleic Acids Res,* 42**,** 12640-9.

De-Solve-It. 2012. *De-Solv-it Cleaning Technology Sticky Stuff Remover Label Information* [Online]. http://www.lakeland.co.uk/content/documents/8976_Label.pdf [Date last accessed 15/03/2016.

De Wael, K. V. D., K. Gason, F. 2015. Discrimination of reactively-dyed cotton fibres with thin layer chromatography and UV microspectrophotometry. *Science & Justice*.

Deng, J., Shoemaker, R., Xie, B., Gore, A., Leproust, E. M., Antosiewicz-Bourget, J., Egli, D., Maherali, N., Park, I. H., Yu, J., Daley, G. Q., Eggan, K., Hochedlinger, K., Thomson, J., Wang, W., Gao, Y. & Zhang, K. 2009. Targeted bisulfite sequencing reveals changes in DNA methylation associated with nuclear reprogramming. *Nat Biotechnol,* 27**,** 353-60.

Di Martino, D., Giuffre, G., Staiti, N., Simone, A., Le Donne, M. & Saravo, L. 2004a. Single sperm cell isolation by laser microdissection. *Forensic Sci Int,* 146 Suppl**,** S151-3.

Di Martino, D., Giuffre, G., Staiti, N., Simone, A., Todaro, P. & Saravo, L. 2004b. Laser microdissection and DNA typing of cells from single hair follicles. *Forensic Sci Int,* 146 Suppl**,** S155-7.

Diep, D., Plongthongkum, N., Gore, A., Fung, H. L., Shoemaker, R. & Zhang, K. 2012. Library-free methylation sequencing with bisulfite padlock probes. *Nat Methods,* 9**,** 270-2.

Divne, A. M., Edlund, H. & Allen, M. 2010. Forensic analysis of autosomal STR markers using Pyrosequencing. *Forensic Sci Int Genet,* 4**,** 122-9.

Downey, L. P., J, Hermanson, S. Steffen, L. Sprecher, C. Mclaren, R.S. Storts, D.R. 2015. *Development of PowerSeq™ Systems for Forensic Identification Using Next Generation Sequencing* [Online]. http://www.promega.com/~/media/files/resources/posters/development%20of%20powerseq%20systems%20for%20forensic%20identification%20using%20next%20generation%20sequencing%20poster.pdf [Date last accessed 22/10/2015.

Drohat, A. C. & Maiti, A. 2014. Mechanisms for enzymatic cleavage of the N-glycosidic bond in DNA. *Organic & Biomolecular Chemistry,* 12**,** 8367-8378.

Duncan, B. K. & Miller, J. H. 1980. Mutagenic deamination of cytosine residues in DNA. *Nature,* 287**,** 560-1.

Dutta, S. & Gates, K. S. 2006. Interstrand cross-links generated by abasic sites in duplex DNA. *Chemical Research in Toxicology,* 19**,** 1689-1689.

Eastburn, D. J., Huang, Y., Pellegrino, M., Sciambi, A., Ptacek, L. J. & Abate, A. R. 2015. Microfluidic droplet enrichment for targeted sequencing. *Nucleic Acids Res*.

Edson, S. M. C., M.D. Parson, T.J. Barritt, S.M. 2004. Naming the Dead — Confronting the Realities of Rapid Identification of Degraded Skeletal Remains. *Forensic Sci Rev* 16**,** 63.

Eduardoff, M., Santos, C., De La Puente, M., Gross, T. E., Fondevila, M., Strobl, C., Sobrino, B., Ballard, D., Schneider, P. M., Carracedo, A., Lareu, M. V., Parson, W. & Phillips, C. 2015. Inter-laboratory evaluation of SNP-based forensic identification by massively parallel sequencing using the Ion PGM. *Forensic Sci Int Genet,* 17**,** 110-21.

Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G., Peluso, P., Rank, D., Baybayan, P., Bettman, B., Bibillo, A., Bjornson, K., Chaudhuri, B., Christians, F., Cicero, R.,

Clark, S., Dalal, R., Dewinter, A., Dixon, J., Foquet, M., Gaertner, A., Hardenbol, P., Heiner, C., Hester, K., Holden, D., Kearns, G., Kong, X. X., Kuse, R., Lacroix, Y., Lin, S., Lundquist, P., Ma, C. C., Marks, P., Maxham, M., Murphy, D., Park, I., Pham, T., Phillips, M., Roy, J., Sebra, R., Shen, G., Sorenson, J., Tomaney, A., Travers, K., Trulson, M., Vieceli, J., Wegener, J., Wu, D., Yang, A., Zaccarin, D., Zhao, P., Zhong, F., Korlach, J. & Turner, S. 2009. Real-Time DNA Sequencing from Single Polymerase Molecules. *Science,* 323**,** 133-138.

Ekblom, R. & Galindo, J. 2011. Applications of next generation sequencing in molecular ecology of non-model organisms. *Heredity (Edinb),* 107**,** 1-15.

Enari, M., Sakahira, H., Yokoyama, H., Okawa, K., Iwamatsu, A. & Nagata, S. 1998. A caspase-activated DNase that degrades DNA during apoptosis, and its inhibitor ICAD. *Nature,* 391**,** 43-50.

Etok, S. E., Valsami-Jones, E., Wess, T. J., Hiller, J. C., Maxwell, C. A., Rogers, K. D., Manning, D. a. C., White, M. L., Lopez-Capel, E., Collins, M. J., Buckley, M., Penkman, K. E. H. & Woodgate, S. L. 2007. Structural and chemical changes of thermally treated bone apatite. *Journal of Materials Science,* 42**,** 9807-9816.

Excoffier, L., Smouse, P. E. & Quattro, J. M. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics,* 131**,** 479-91.

Fazekas, A., Steeves, R. & Newmaster, S. 2010. Improving sequencing quality from PCR products containing long mononucleotide repeats. *Biotechniques,* 48**,** 277-85.

Feig, D. I., Sowers, L. C. & Loeb, L. A. 1994. Reverse chemical mutagenesis: identification of the mutagenic lesions resulting from reactive oxygen species-mediated damage to DNA. *Proc Natl Acad Sci U S A,* 91**,** 6609-13.

Findlay, I., Mathews, P. & Quirke, P. 1998. Multiple genetic diagnoses from single cells using multiplex PCR: Reliability and allele dropout. *Prenatal Diagnosis,* 18**,** 1413-1421.

Findlay, I., Taylor, A., Quirke, P., Frazier, R. & Urquhart, A. 1997. DNA fingerprinting from single cells. *Nature,* 389**,** 555-6.

Fleming, A. M., Alshykhly, O., Zhu, J., Muller, J. G. & Burrows, C. J. 2015. Rates of Chemical Cleavage of DNA and RNA Oligomers Containing Guanine Oxidation Products. *Chem Res Toxicol.*

Fong, W. & Inami, S. H. 1986. Results of a Study to Determine the Probability of Chance Match Occurrences between Fibers Known to Be from Different Sources. *Journal of Forensic Sciences,* 31**,** 65-72.

Foran, D. R. 2006. Relative degradation of nuclear and mitochondrial DNA: an experimental approach. *J Forensic Sci,* 51**,** 766-70.

Fordyce, S. L., Avila-Arcos, M. C., Rockenbauer, E., Borsting, C., Frank-Hansen, R., Petersen, F. T., Willerslev, E., Hansen, A. J., Morling, N. & Gilbert, M. T. 2011. High-throughput sequencing of core STR loci for forensic genetic investigations using the Roche Genome Sequencer FLX platform. *Biotechniques,* 51**,** 127-33.

Fordyce, S. L., Mogensen, H. S., Borsting, C., Lagace, R. E., Chang, C. W., Rajagopalan, N. & Morling, N. 2015. Second-generation sequencing of forensic STRs using the Ion Torrent HID STR 10-plex and the Ion PGM. *Forensic Sci Int Genet,* 14**,** 132-40.

Forster, L., Thomson, J. & Kutranov, S. 2008. Direct comparison of post-28-cycle PCR purification and modified capillary electrophoresis methods with the 34-cycle "low copy number" (LCN) method for analysis of trace forensic DNA samples. *Forensic Sci Int Genet,* 2**,** 318-28.

Freeman, B. A. & Crapo, J. D. 1982. Biology of disease: free radicals and tissue injury. *Lab Invest,* 47**,** 412-26.

Freese, E. & Cashel, M. 1964. Crosslinking of Deoxyribonucleic Acid by Exposure to Low Ph. *Biochim Biophys Acta,* 91**,** 67-77.

Frudakis, T., Venkateswarlu, K., Thomas, M. J., Gaskin, Z., Ginjupalli, S., Gunturi, S., Ponnuswamy, V., Natarajan, S. & Nachimuthu, P. K. 2003. A classifier for the SNP-based inference of ancestry. *J Forensic Sci,* 48**,** 771-82.

Fu, Q. M., Meyer, M., Gao, X., Stenzel, U., Burbano, H. A., Kelso, J. & Paabo, S. 2013. DNA analysis of an early modern human from Tianyuan Cave, China. *Proceedings of the National Academy of Sciences of the United States of America,* 110**,** 2223-2227.

Fungtammasan, A., Ananda, G., Hile, S. E., Su, M. S., Sun, C., Harris, R., Medvedev, P., Eckert, K. & Makova, K. D. 2015. Accurate typing of short tandem repeats from genome-wide sequencing data and its applications. *Genome Res,* 25**,** 736-49.

Gaines, M. L., Wojtkiewicz, P. W., Valentine, J. A. & Brown, C. L. 2002. Reduced volume PCR amplification reactions using the AmpFlSTR Profiler Plus kit. *J Forensic Sci,* 47**,** 1224-37.

Garvin, A. M., Holzgreve, W. & Hahn, S. 1998. Highly accurate analysis of heterozygous loci by single cell PCR. *Nucleic Acids Research,* 26**,** 3468-3472.

Gelardi, C., Rockenbauer, E., Dalsgaard, S., Borsting, C. & Morling, N. 2014. Second generation sequencing of three STRs D3S1358, D12S391 and D21S11 in Danes and a new nomenclature for sequenced STR alleles. *Forensic Science International-Genetics,* 12**,** 38-41.

Gettings, K. A., R.A. Vallone, P.M. Butler, J.M. 2015. STR allele sequence variation: Current knowledge and future issues. *Forensic Science International: Genetics*.

Gibb, A. J., Huell, A. L., Simmons, M. C. & Brown, R. M. 2009. Characterisation of forward stutter in the AmpFlSTR SGM Plus PCR. *Sci Justice,* 49**,** 24-31.

Gilbert, M. T., Hansen, A. J., Willerslev, E., Rudbeck, L., Barnes, I., Lynnerup, N. & Cooper, A. 2003. Characterization of genetic miscoding lesions caused by postmortem damage. *Am J Hum Genet,* 72**,** 48-61.

Giles, R. E., Blanc, H., Cann, H. M. & Wallace, D. C. 1980. Maternal inheritance of human mitochondrial DNA. *Proc Natl Acad Sci U S A,* 77**,** 6715-9.

Gill, P. 2001. An assessment of the utility of single nucleotide polymorphisms (SNPs) for forensic purposes. *Int J Legal Med,* 114**,** 204-10.

Gill, P., Brown, R. M., Fairley, M., Lee, L., Smyth, M., Simpson, N., Irwin, B., Dunlop, J., Greenhalgh, M., Way, K., Westacott, E. J., Ferguson, S. J., Ford, L. V., Clayton, T., Guiness, J. & Technical, U. K. D. N. a. W. G. 2008. National recommendations of the Technical UK DNA working group on mixture interpretation for the NDNAD and for court going purposes. *Forensic Sci Int Genet,* 2**,** 76-82.

Gill, P. & Buckleton, J. 2010. A universal strategy to interpret DNA profiles that does not require a definition of low-copy-number. *Forensic Sci Int Genet,* 4**,** 221-7.

Gill, P., D'aloja, E., Andersen, J., Dupuy, B., Jangblad, M., Johnsson, V., Kloosterman, A. D., Kratzer, A., Lareu, M. V., Meldegaard, M., Phillips, C., Pfitzinger, H., Rand, S., Sabatier, M., Scheithauer, R., Schmitter, H., Schneider, P. & Vide, M. C. 1997a. Report of the European DNA profiling group (EDNAP): an investigation of the complex STR loci D21S11 and HUMFIBRA (FGA). *Forensic Sci Int,* 86**,** 25-33.

Gill, P., Fereday, L., Morling, N. & Schneider, P. M. 2006a. The evolution of DNA databases--recommendations for new European STR loci. *Forensic Sci Int,* 156**,** 242-4.

Gill, P., Fereday, L., Morling, N. & Schneider, P. M. 2006b. New multiplexes for Europe-amendments and clarification of strategic development. *Forensic Sci Int,* 163**,** 155-7.

Gill, P., Foreman, L., Buckleton, J. S., Triggs, C. M. & Allen, H. 2003. A comparison of adjustment methods to test the robustness of an STR DNA database comprised of 24 European populations. *Forensic Sci Int,* 131**,** 184-96.

Gill, P., Ivanov, P. L., Kimpton, C., Piercy, R., Benson, N., Tully, G., Evett, I., Hagelberg, E. & Sullivan, K. 1994. Identification of the remains of the Romanov family by DNA analysis. *Nat Genet,* 6**,** 130-5.

Gill, P., Jeffreys, A. J. & Werrett, D. J. 1985. Forensic application of DNA 'fingerprints'. *Nature,* 318**,** 577-9.

Gill, P., Puch-Solis, R. & Curran, J. 2009. The low-template-DNA (stochastic) threshold--its determination relative to risk analysis for national DNA databases. *Forensic Sci Int Genet,* 3**,** 104-11.

Gill, P., Sparkes, R., Fereday, L. & Werrett, D. J. 2000a. Report of the European Network of Forensic Science Institutes (ENSFI): formulation and testing of principles to evaluate STR multiplexes. *Forensic Science International,* 108**,** 1-29.

Gill, P., Sparkes, R. & Kimpton, C. 1997b. Development of guidelines to designate alleles using an STR multiplex system. *Forensic Sci Int,* 89**,** 185-97.

Gill, P., Whitaker, J., Flaxman, C., Brown, N. & Buckleton, J. 2000b. An investigation of the rigor of interpretation rules for STRs derived from less than 100 pg of DNA. *Forensic Sci Int,* 112**,** 17-40.

Ginolhac, A., Vilstrup, J., Stenderup, J., Rasmussen, M., Stiller, M., Shapiro, B., Zazula, G., Froese, D., Steinmann, K. E., Thompson, J. F., Al-Rasheid, K. a. S., Gilbert, T. M. P., Willerslev, E. & Orlando, L. 2012. Improving the performance of true single molecule sequencing for ancient DNA. *Bmc Genomics,* 13.

Gnirke, A., Melnikov, A., Maguire, J., Rogov, P., Leproust, E. M., Brockman, W., Fennell, T., Giannoukos, G., Fisher, S., Russ, C., Gabriel, S., Jaffe, D. B., Lander, E. S. & Nusbaum, C. 2009. Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nat Biotechnol,* 27**,** 182-9.

Goldenberger, D., Perschil, I., Ritzler, M. & Altwegg, M. 1995. A simple "universal" DNA extraction procedure using SDS and proteinase K is compatible with direct PCR amplification. *PCR Methods Appl,* 4**,** 368-70.

Gonzalez Fortes, G. 2014. Genetic analysis in archaeological remains from Vindolanda Museum (UK). York University.

Götherström, A. C., M.J. Angerbjörn, A. Lidén, K. 2002. Bone preservation and DNA amplification. *Archaeometry,* 44**,** 395-404.

Götz, F. 2015. ALLST*R Database for autosomal short tandem repeats. allstr.de: Qualtitype GmbH, Germany.

Green, R. L., Lagace, R. E., Oldroyd, N. J., Hennessy, L. K. & Mulero, J. J. 2013. Developmental validation of the AmpFlSTR(R) NGM SElect PCR Amplification Kit: A next-generation STR multiplex with the SE33 locus. *Forensic Sci Int Genet,* 7**,** 41-51.

Grieve, M. & Deck, S. 1995. A New Mounting Medium for the Forensic Microscopy of Textile Fibers. *Science & Justice,* 35**,** 109-112.

Grieve, M. C. & Biermann, T. 1997a. The population of coloured textile fibres on outdoor surfaces. *Science & Justice,* 37**,** 231-239.

Grieve, M. C., Biermann, T. & Davignon, M. 2003. The occurrence and individuality of orange and green cotton fibres. *Science & Justice,* 43**,** 5-22.

Grieve, M. C. & Biermann, T. W. 1997b. Wool fibres - Transfer to vinyl and leather vehicle seats and some observations on their secondary transfer. *Science & Justice,* 37**,** 31-38.

Grieve, M. C., Biermann, T. W. & Schaub, K. 2005. The individuality of fibres used to provide forensic evidence - not all blue polyesters are the same. *Science & Justice,* 45**,** 13-28.

Grossmann, K. F., Ward, A. M., Matkovic, M. E., Folias, A. E. & Moses, R. E. 2001. S. cerevisiae has three pathways for DNA interstrand crosslink repair. *Mutat Res,* 487**,** 73-83.

Grunenwald, A., Keyser, C., Sautereau, A. M., Crubezy, E., Ludes, B. & Drouet, C. 2014. Adsorption of DNA on biomimetic apatites: Toward the understanding of the role of bone and tooth mineral on the preservation of ancient DNA. *Applied Surface Science,* 292**,** 867-875.

Guo, F., Shen, H., Tian, H., Jin, P. & Jiang, X. 2014. Development of a 24-locus multiplex system to incorporate the core loci in the Combined DNA Index System (CODIS) and the European Standard Set (ESS). *Forensic Sci Int Genet,* 8**,** 44-54.

Hall, D. E. & Roy, R. 2014. An evaluation of direct PCR amplification. *Croat Med J,* 55**,** 655-61.

Hanes, J. W., Thal, D. M. & Johnson, K. A. 2006. Incorporation and replication of 8-oxo-deoxyguanosine by the human mitochondrial DNA polymerase. *J Biol Chem,* 281**,** 36241-8.

Hansson, O. F., M. Heitmann, I.K. Ramse, M. Bouzga, M. 2009. Trace DNA collection—Performance of minitape and three different swabs. *Forensic Science International: Genetics Supplement Series,* 2**,** 189-190.

Haqqi, T., Zhao, X., Panciu, A. & Yadav, S. P. 2002. Sequencing in the presence of betaine: Improvement in sequencing of the localized repeat sequence regions. *J Biomol Tech,* 13**,** 265-71.

Hart, C., Lipson, D., Ozsolak, F., Raz, T., Steinmann, K., Thompson, J. & Milos, P. M. 2010. Single-molecule sequencing: sequence methods to enable accurate quantitation. *Methods Enzymol,* 472**,** 407-30.

Hebert, P. D., Ratnasingham, S. & Dewaard, J. R. 2003. Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proc Biol Sci,* 270 Suppl 1**,** S96-9.

Hedges, R. E. M. & Millard, A. R. 1995. Bones and Groundwater - Towards the Modeling of Diagenetic Processes. *Journal of Archaeological Science,* 22**,** 155-164.

Heinrich, M., Felske-Zech, H., Brinkmann, B. & Hohoff, C. 2005. Characterisation of variant alleles in the STR systems D2S1338, D3S1358 and D19S433. *Int J Legal Med,* 119**,** 310-3.

Henke, W., Herdel, K., Jung, K., Schnorr, D. & Loening, S. A. 1997. Betaine improves the PCR amplification of GC-rich DNA sequences. *Nucleic Acids Res,* 25**,** 3957-8.

Higgins, D., Rohrlach, A. B., Kaidonis, J., Townsend, G. & Austin, J. J. 2015. Differential nuclear and mitochondrial DNA preservation in post-mortem teeth with implications for forensic and ancient DNA studies. *PLoS One,* 10**,** e0126935.

High, K. M., N. Panter, I. Penkman, K.E.H 2015. Apatite for destruction: investigating bone degradation due to high acidity at Star Carr. *Journal of Archaeological Science,* 59**,** 159-168.

Higuchi, R., Von Beroldingen, C. H., Sensabaugh, G. F. & Erlich, H. A. 1988. DNA typing from single hairs. *Nature,* 332**,** 543-6.

Hill, C. R., Kline, M. C., Coble, M. D. & Butler, J. M. 2008. Characterization of 26 miniSTR loci for improved analysis of degraded DNA samples. *J Forensic Sci,* 53**,** 73-80.

Holt, I. J., Harding, A. E. & Morgan-Hughes, J. A. 1988. Deletions of muscle mitochondrial DNA in patients with mitochondrial myopathies. *Nature,* 331**,** 717-9.

Hoss, M., Jaruga, P., Zastawny, T. H., Dizdaroglu, M. & Paabo, S. 1996. DNA damage and DNA sequence retrieval from ancient tissues. *Nucleic Acids Res,* 24**,** 1304-7.

Houck, M. M. 2003. Inter-comparison of unrelated fiber evidence. *Forensic Science International,* 135**,** 146-149.

Iborra, F. J., Kimura, H. & Cook, P. R. 2004. The functional organization of mitochondrial genomes in human cells. *BMC Biol,* 2**,** 9.

Illumina 2010. Illumina Sequencing Technology. http://www.illumina.com/documents/products/techspotlights/techspotlight_sequencing.pdf.

Illumina 2013. Human mtDNA D-loop Hybervariable Region For the Illumina Sequencing Platform. http://support.illumina.com/content/dam/illumina-support/documents/documentation/chemistry_documentation/samplepreps_legacy/human-mtdna-d-loop-guide-15034858-b.pdf: Illumina Inc.

Illumina 2014a. Illumina customer sequence letter. http://supportres.illumina.com/documents/documentation/chemistry_documentation/experiment-design/illumina-customer-sequence-letter.pdf.

Illumina 2014b. Sequencing power for every scale. http://www.illumina.com/content/dam/illumina-marketing/documents/products/brochures/brochure_sequencing_systems_portfolio.pdf.

Illumina 2015a. ForenSeq DNA Signature Prep Guide. February 2015 ed. http://support.illumina.com/content/dam/illumina-support/documents/documentation/chemistry_documentation/forenseq/forenseq-dna-signature-prep-guide-15049528-d.pdf: Illumina Inc.

Illumina 2015b. ForenSeq Universal Analysis Software http://www.illumina.com/content/dam/illumina-marketing/documents/products/datasheets/forenseq-uas-datasheet-1470-2014-002.pdf.

Illumina 2015c. ForenSeq™ DNA Signature Prep Reference Guide. http://support.illumina.com/content/dam/illumina-support/documents/documentation/chemistry_documentation/forenseq/forenseq-dna-signature-prep-guide-15049528-01.pdf.

Ireland, J. B., J. 2015. Admitting scientific expert evidence in the UK: reliability challenges and the need for revised criteria – proposing an Abridged Daubert. *The Journal of Forensic Practice,* 17**,** 3-12.

Jans, M. M. E., Nielsen-Marsh, C. M., Smith, C. I., Collins, M. J. & Kars, H. 2004. Characterisation of microbial attack on archaeological bone. *Journal of Archaeological Science,* 31**,** 87-95.

Jazin, E., Soodyall, H., Jalonen, P., Lindholm, E., Stoneking, M. & Gyllensten, U. 1998. Mitochondrial mutation rate revisited: hot spots and polymorphism. *Nat Genet,* 18**,** 109-10.

Jeffreys, A. J., Wilson, V. & Thein, S. L. 1985. Hypervariable 'minisatellite' regions in human DNA. *Nature,* 314**,** 67-73.

Jensen, M. A., Fukushima, M. & Davis, R. W. 2010. DMSO and betaine greatly improve amplification of GC-rich constructs in de novo synthesis. *PLoS One,* 5**,** e11024.

Jobling, M. A. & Gill, P. 2004. Encoded evidence: DNA in forensic analysis. *Nat Rev Genet,* 5**,** 739-51.

Jones, B. 2015. Nanopore sequencing for clinical diagnostics. *Nature Reviews Genetics,* 16**,** 1-3.

Jones, J. & Coyle, T. 2011. Synthetic flock fibres: A population and target fibre study. *Science & Justice,* 51**,** 68-71.

Just, R. S., Scheible, M. K., Fast, S. A., Sturk-Andreaggi, K., Rock, A. W., Bush, J. M., Higginbotham, J. L., Peck, M. A., Ring, J. D., Huber, G. E., Xavier, C., Strobl, C., Lyons, E. A., Diegoli, T. M., Bodner, M., Fendt, L., Kralj, P., Nagl, S., Niederwieser, D., Zimmermann, B., Parson, W. & Irwin, J. A. 2015. Full mtGenome reference data: development and characterization of 588 forensic-quality haplotypes representing three U.S. populations. *Forensic Sci Int Genet,* 14**,** 141-55.

Kadenbach, B., Jarausch, J., Hartmann, R. & Merle, P. 1983. Separation of mammalian cytochrome c oxidase into 13 polypeptides by a sodium dodecyl sulfate-gel electrophoretic procedure. *Anal Biochem,* 129**,** 517-21.

Kanvah, S., Joseph, J., Schuster, G. B., Barnett, R. N., Cleveland, C. L. & Landman, U. 2010. Oxidation of DNA: Damage to Nucleobases. *Accounts of Chemical Research,* 43**,** 280-287.

Karran, P. & Lindahl, T. 1980. Hypoxanthine in Deoxyribonucleic-Acid - Generation by Heat-Induced Hydrolysis of Adenine Residues and Release in Free Form by a Deoxyribonucleic-Acid Glycosylase from Calf Thymus. *Biochemistry,* 19**,** 6005-6011.

Kaufman, B. A., Durisic, N., Mativetsky, J. M., Costantino, S., Hancock, M. A., Grutter, P. & Shoubridge, E. A. 2007. The mitochondrial transcription factor TFAM coordinates the assembly of multiple DNA molecules into nucleoid-like structures. *Mol Biol Cell,* 18**,** 3225-36.

Kavlick, M. F., Lawrence, H. S., Merritt, R. T., Fisher, C., Isenberg, A., Robertson, J. M. & Budowle, B. 2011. Quantification of human mitochondrial DNA using synthesized DNA standards. *J Forensic Sci,* 56**,** 1457-63.

Kelly, E. & Griffin, R. M. E. 1998. A target fibre study on seats in public houses. *Science & Justice,* 38**,** 39-44.

Kemp, B. M., Monroe, C. & Smith, D. G. 2007. Extraction and Analysis of DNA from Archaeological Specimens. *Archaeological Chemistry: Analytical Techniques and Archaeological Interpretation,* 968**,** 78-98.

Keutenius, E., O'keeffe, P. & Allen, K. 2013. The recovery of fibres from non-textile items using a static charge. *Sci Justice,* 53**,** 171-7.

Kidd, C. B. M. & Robertson, J. 1982. The Transfer of Textile Fibers during Simulated Contacts. *Journal of the Forensic Science Society,* 22**,** 301-308.

Kimpton, C., Fisher, D., Watson, S., Adams, M., Urquhart, A., Lygo, J. & Gill, P. 1994. Evaluation of an automated DNA profiling system employing multiplex amplification of four tetrameric STR loci. *Int J Legal Med,* 106**,** 302-11.

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

King, J. L., Larue, B. L., Novroski, N. M., Stoljarova, M., Seo, S. B., Zeng, X., Warshauer, D. H., Davis, C. P., Parson, W., Sajantila, A. & Budowle, B. 2014a. High-quality and high-throughput massively parallel sequencing of the human mitochondrial genome using the Illumina MiSeq. *Forensic Sci Int Genet,* 12**,** 128-35.

King, T. E., Fortes, G. G., Balaresque, P., Thomas, M. G., Balding, D., Delser, P. M., Neumann, R., Parson, W., Knapp, M., Walsh, S., Tonasso, L., Holt, J., Kayser, M., Appleby, J., Forster, P., Ekserdjian, D., Hofreiter, M. & Schurer, K. 2014b. Identification of the remains of King Richard III. *Nature Communications,* 5.

Kita, T., Yamaguchi, H., Yokoyama, M., Tanaka, T. & Tanaka, N. 2008. Morphological study of fragmented DNA on touched objects. *Forensic Sci Int Genet,* 3**,** 32-6.

Kivisild, T. 2015. Maternal ancestry and population history from whole mitochondrial genomes. *Investig Genet,* 6**,** 3.

Klintschar, M. & Wiegand, P. 2003. Polymerase slippage in relation to the uniformity of tetrameric repeat stretches. *Forensic Sci Int,* 135**,** 163-6.

Kloosterman, A. D. & Kersbergen, P. 2003. Efficacy and limits of genotyping low copy number (LCN) DNA samples by multiplex PCR of STR loci. *J Soc Biol,* 197**,** 351-9.

Knierim, E., Lucke, B., Schwarz, J. M., Schuelke, M. & Seelow, D. 2011. Systematic comparison of three methods for fragmentation of long-range PCR products for next generation sequencing. *PLoS One,* 6**,** e28240.

Korlach, J., Levene, M., Foquet, M., Turner, S. W., Craighead, H. G. & Webb, W. W. 2003. Single molecule DNA sequence profiling in zero-mode waveguides using gamma-phosphate linked nucleotide analogs. *Biophysical Journal,* 84**,** 141a-141a.

Kozich, J. J., Westcott, S. L., Baxter, N. T., Highlander, S. K. & Schloss, P. D. 2013. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Appl Environ Microbiol,* 79**,** 5112-20.

Krings, M., Stone, A., Schmitz, R. W., Krainitzki, H., Stoneking, M. & Paabo, S. 1997. Neandertal DNA sequences and the origin of modern humans. *Cell,* 90**,** 19-30.

Kruskal, W. H. & Wallis, W. A. 1987. Citation Classic - Use of Ranks in One-Criterion Variance Analysis. *Current Contents/Arts & Humanities***,** 20-20.

Kukat, C., Wurm, C. A., Spahr, H., Falkenberg, M., Larsson, N. G. & Jakobs, S. 2011. Super-resolution microscopy reveals that mammalian mitochondrial nucleoids have a uniform size and frequently contain a single copy of mtDNA. *Proc Natl Acad Sci U S A,* 108**,** 13534-9.

Lamers, R., Hayter, S. & Matheson, C. D. 2009. Postmortem miscoding lesions in sequence analysis of human ancient mitochondrial DNA. *J Mol Evol,* 68**,** 40-55.

Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., Fitzhugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., Levine, R., Mcewan, P., Mckernan, K., Meldrim, J., Mesirov, J. P., Miranda, C., Morris, W., Naylor, J., Raymond, C.,

Rosetti, M., Santos, R., Sheridan, A., Sougnez, C., Stange-Thomann, Y., Stojanovic, N., Subramanian, A., Wyman, D., Rogers, J., Sulston, J., Ainscough, R., Beck, S., Bentley, D., Burton, J., Clee, C., Carter, N., Coulson, A., Deadman, R., Deloukas, P., Dunham, A., Dunham, I., Durbin, R., French, L., Grafham, D., Gregory, S., Hubbard, T., Humphray, S., Hunt, A., Jones, M., Lloyd, C., Mcmurray, A., Matthews, L., Mercer, S., Milne, S., Mullikin, J. C., Mungall, A., Plumb, R., Ross, M., Shownkeen, R., Sims, S., Waterston, R. H., Wilson, R. K., Hillier, L. W., Mcpherson, J. D., Marra, M. A., Mardis, E. R., Fulton, L. A., Chinwalla, A. T., Pepin, K. H., Gish, W. R., Chissoe, S. L., Wendl, M. C., Delehaunty, K. D., Miner, T. L., Delehaunty, A., Kramer, J. B., Cook, L. L., Fulton, R. S., Johnson, D. L., Minx, P. J., Clifton, S. W., Hawkins, T., Branscomb, E., Predki, P., Richardson, P., Wenning, S., Slezak, T., Doggett, N., Cheng, J. F., Olsen, A., Lucas, S., Elkin, C., Uberbacher, E., Frazier, M., et al. 2001. Initial sequencing and analysis of the human genome. *Nature,* 409**,** 860-921.

Lareu, M. V., Pestoni, M. C., Barros, F., Salas, A. & Carracedo, A. 1996. Sequence variation of a hypervariable short tandem repeat at the D12S391 locus. *Gene,* 182**,** 151-3.

Larmuseau, M. H., Van Geystelen, A., Kayser, M., Van Oven, M. & Decorte, R. 2015. Towards a consensus Y-chromosomal phylogeny and Y-SNP set in forensics in the next-generation sequencing era. *Forensic Sci Int Genet,* 15**,** 39-42.

Laver, T. H., J. O'neill, P.A. Moore, K. Farbos, A. Paszkiewicz, K. Studholme, D.J. 2015. Assessing the performance of the Oxford Nanopore Technologies MinION. *Biomolecular Detection and Quantification,* 3**,** 1-8.

Lechardeur, D., Drzymala, L., Sharma, M., Zylka, D., Kinach, R., Pacia, J., Hicks, C., Usmani, N., Rommens, J. M. & Lukacs, G. L. 2000. Determinants of the nuclear localization of the heterodimeric DNA fragmentation factor (ICAD/CAD). *Journal of Cell Biology,* 150**,** 321-334.

Leclair, B., Sgueglia, J. B., Wojtowicz, P. C., Juston, A. C., Fregeau, C. J. & Fourney, R. M. 2003. STR DNA typing: increased sensitivity and efficient sample consumption using reduced PCR reaction volumes. *J Forensic Sci,* 48**,** 1001-13.

Legros, F., Malka, F., Frachon, P., Lombes, A. & Rojo, M. 2004. Organization and dynamics of human mitochondrial DNA. *J Cell Sci,* 117**,** 2653-62.

Levene, M. J., Korlach, J., Turner, S. W., Foquet, M., Craighead, H. G. & Webb, W. W. 2003. Zero-mode waveguides for single-molecule analysis at high concentrations. *Science,* 299**,** 682-686.

Levinson, G. & Gutman, G. A. 1987. Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Mol Biol Evol,* 4**,** 203-21.

Li, C. B., Q. Anquan, J. Xu, X. Hu, L. . The combination of single cell micromanipulation with LV-PCR system and its application in forensic science. *In:* Amorim, A. C.-R., F. Morling, N., ed. Progress in Forensic Genetics 13 — Proceedings of the 23rd International ISFG Congress, 2009. Elsevier, 516-517.

Li, H. 2011a. Improving SNP discovery by base alignment quality. *Bioinformatics,* 27**,** 1157-8.

Li, H. 2011b. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics,* 27**,** 2987-93.

Li, H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. http://arxiv.org/abs/1303.3997: Cornell University.

Li, H. & Durbin, R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics,* 25**,** 1754-60.

Li, H. & Durbin, R. 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics,* 26**,** 589-95.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. & Genome Project Data Processing, S. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics,* 25**,** 2078-9.

Li, H. H., Gyllensten, U. B., Cui, X. F., Saiki, R. K., Erlich, H. A. & Arnheim, N. 1988. Amplification and Analysis of DNA-Sequences in Single Human-Sperm and Diploid-Cells. *Nature,* 335**,** 414-417.

Li, M., Schonberg, A., Schaefer, M., Schroeder, R., Nasidze, I. & Stoneking, M. 2010. Detecting heteroplasmy from high-throughput sequencing of complete human mitochondrial DNA genomes. *Am J Hum Genet,* 87**,** 237-49.

Linacre, A., Pekarek, V., Swaran, Y. C. & Tobe, S. S. 2010. Generation of DNA profiles from fabrics without DNA extraction. *Forensic Science International-Genetics,* 4**,** 137-141.

Lindahl, T. 1993. Instability and decay of the primary structure of DNA. *Nature,* 362**,** 709-15.

Lindahl, T. & Karlstro.O 1973. Heat-Induced Depyrimidination of Deoxyribonucleic Acid in Neutral Solution. *Biochemistry,* 12**,** 5151-5154.

Lindahl, T. & Nyberg, B. 1972. Rate of Depurination of Native Deoxyribonucleic Acid. *Biochemistry,* 11**,** 3610-&.

Lindahl, T. & Nyberg, B. 1974. Heat-induced deamination of cytosine residues in deoxyribonucleic acid. *Biochemistry,* 13**,** 3405-10.

Liu, F., Van Duijn, K., Vingerling, J. R., Hofman, A., Uitterlinden, A. G., Janssens, A. C. & Kayser, M. 2009. Eye color and the prediction of complex phenotypes from genotypes. *Curr Biol,* 19**,** R192-3.

Liu, L., Li, Y., Li, S., Hu, N., He, Y., Pong, R., Lin, D., Lu, L. & Law, M. 2012. Comparison of next-generation sequencing systems. *J Biomed Biotechnol,* 2012**,** 251364.

Lo, Y. S., Hsiao, L. J., Cheng, N., Litvinchuk, A. & Dai, H. 2011. Characterization of the structure and DNA complexity of mung bean mitochondrial nucleoids. *Molecules and Cells,* 31**,** 217-224.

Locard, E. 1920. *L'enquète criminelle et les mèthodes scientifiques,* Paris, Flammarion.

Loeb, L. A. & Preston, B. D. 1986. Mutagenesis by Apurinic Apyrimidinic Sites. *Annual Review of Genetics,* 20**,** 201-230.

Loman, N. J., Misra, R. V., Dallman, T. J., Constantinidou, C., Gharbia, S. E., Wain, J. & Pallen, M. J. 2012. Performance comparison of benchtop high-throughput sequencing platforms. *Nat Biotechnol,* 30**,** 434-9.

Lopez, J. V., Yuhki, N., Masuda, R., Modi, W. & O'brien, S. J. 1994. Numt, a recent transfer and tandem amplification of mitochondrial DNA to the nuclear genome of the domestic cat. *J Mol Evol,* 39**,** 174-90.

Lowrie, C. N. & Jackson, G. 1991. Recovery of Transferred Fibers. *Forensic Science International,* 50**,** 111-119.

Ma, Z., Lee, R. W., Li, B., Kenney, P., Wang, Y., Erikson, J., Goyal, S. & Lao, K. 2013. Isothermal amplification method for next-generation sequencing. *Proc Natl Acad Sci U S A,* 110**,** 14320-3.

Malaver, P. C. & Yunis, J. J. 2003. Different dental tissues as source of DNA for human identification in forensic cases. *Croatian Medical Journal,* 44**,** 306-309.

Mamanova, L., Coffey, A. J., Scott, C. E., Kozarewa, I., Turner, E. H., Kumar, A., Howard, E., Shendure, J. & Turner, D. J. 2010. Target-enrichment strategies for next-generation sequencing. *Nat Methods,* 7**,** 111-8.

Mardis, E. R. 2008. The impact of next-generation sequencing technology on genetics. *Trends in Genetics,* 24**,** 133-141.

Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., Bemben, L. A., Berka, J., Braverman, M. S., Chen, Y. J., Chen, Z., Dewell, S. B., Du, L., Fierro, J. M., Gomes, X. V., Godwin, B. C., He, W., Helgesen, S., Ho, C. H., Irzyk, G. P., Jando, S. C., Alenquer, M. L., Jarvie, T. P., Jirage, K. B., Kim, J. B., Knight, J. R., Lanza, J. R., Leamon, J. H., Lefkowitz, S. M., Lei, M., Li, J., Lohman, K. L., Lu, H., Makhijani, V. B., Mcdade, K. E., Mckenna, M. P., Myers, E. W., Nickerson, E., Nobile, J. R., Plant, R., Puc, B. P., Ronan, M. T., Roth, G. T., Sarkis, G. J., Simons, J. F., Simpson, J. W., Srinivasan, M., Tartaro, K. R., Tomasz, A., Vogt, K. A., Volkmer, G. A.,

Wang, S. H., Wang, Y., Weiner, M. P., Yu, P., Begley, R. F. & Rothberg, J. M. 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature,* 437**,** 376-80.

Marnane, R. N., Elliot, D. A. & Coulson, S. A. 2006. A pilot study to determine the background population of foreign fibre groups on a cotton/polyester T-shirt. *Sci Justice,* 46**,** 215-20.

Marshall, P. L., King, J. L. & Budowle, B. 2015. Utility of amplification enhancers in low copy number DNA analysis. *Int J Legal Med,* 129**,** 43-52.

May, R. T., J. 2009. Optimisation of cellular DNA recovery from tape-lifts. *Forensic Science International: Genetics Supplement Series,* 2**,** 191-192.

Mcwilliam, H., Li, W., Uludag, M., Squizzato, S., Park, Y. M., Buso, N., Cowley, A. P. & Lopez, R. 2013. Analysis Tool Web Services from the EMBL-EBI. *Nucleic Acids Res,* 41**,** W597-600.

Mercier, B., Gaucher, C., Feugeas, O. & Mazurier, C. 1990. Direct PCR from whole blood, without DNA extraction. *Nucleic Acids Res,* 18**,** 5908.

Metzker, M. L. 2010. Sequencing technologies - the next generation. *Nat Rev Genet,* 11**,** 31-46.

Meyer, M., Fu, Q., Aximu-Petri, A., Glocke, I., Nickel, B., Arsuaga, J. L., Martinez, I., Gracia, A., De Castro, J. M., Carbonell, E. & Paabo, S. 2014. A mitochondrial genome sequence of a hominin from Sima de los Huesos. *Nature,* 505**,** 403-6.

Meyer, M., Kircher, M., Gansauge, M. T., Li, H., Racimo, F., Mallick, S., Schraiber, J. G., Jay, F., Prufer, K., De Filippo, C., Sudmant, P. H., Alkan, C., Fu, Q. M., Do, R., Rohland, N., Tandon, A., Siebauer, M., Green, R. E., Bryc, K., Briggs, A. W., Stenzel, U., Dabney, J., Shendure, J., Kitzman, J., Hammer, M. F., Shunkov, M. V., Derevianko, A. P., Patterson, N., Andres, A. M., Eichler, E. E., Slatkin, M., Reich, D., Kelso, J. & Paabo, S. 2012. A High-Coverage Genome Sequence from an Archaic Denisovan Individual. *Science,* 338**,** 222-226.

Mikheyev, A. S. & Tin, M. M. Y. 2014. A first look at the Oxford Nanopore MinION sequencer. *Molecular Ecology Resources,* 14**,** 1097-1102.

Misawa, K. 2016. Short tandem repeats in the human, cow, mouse, chicken, and lizard genomes are concentrated in the terminal regions of chromosomes. *Gene Reports*.

Mitchell, D. L. & Nairn, R. S. 1989. The biology of the (6-4) photoproduct. *Photochem Photobiol,* 49**,** 805-19.

Mitra, A., Skrzypczak, M., Ginalski, K. & Rowicka, M. 2015. Strategies for Achieving High Sequencing Accuracy for Low Diversity Samples and Avoiding Sample Bleeding Using Illumina Platform. *Plos One,* 10.

Miyata, T., Hayashida, H., Kikuno, R., Hasegawa, M., Kobayashi, M. & Koike, K. 1982. Molecular clock of silent substitution: at least six-fold preponderance of silent changes in mitochondrial genes over those in nuclear genes. *J Mol Evol,* 19**,** 28-35.

Mizuta, R., Araki, S., Furukawa, M., Furukawa, Y., Ebara, S., Shiokawa, D., Hayashi, K., Tanuma, S. & Kitamura, D. 2013. DNase gamma Is the Effector Endonuclease for Internucleosomal DNA Fragmentation in Necrosis. *Plos One,* 8.

Morozova, O. & Marra, M. A. 2008. Applications of next-generation sequencing technologies in functional genomics. *Genomics,* 92**,** 255-64.

Mouret, S., Baudouin, C., Charveron, M., Favier, A., Cadet, J. & Douki, T. 2006. Cyclobutane pyrimidine dimers are predominant DNA lesions in whole human skin exposed to UVA radiation. *Proc Natl Acad Sci U S A,* 103**,** 13765-70.

Mrevlishvili, G. M. & Svintradze, D. V. 2005a. Complex between triple helix of collagen and double helix of DNA in aqueous solution. *Int J Biol Macromol,* 35**,** 243-5.

Mrevlishvili, G. M. & Svintradze, D. V. 2005b. DNA as a matrix of collagen fibrils. *Int J Biol Macromol,* 36**,** 324-6.

Mulero, J. J., Chang, C. W., Calandro, L. M., Green, R. L., Li, Y., Johnson, C. L. & Hennessy, L. K. 2006. Development and validation of the AmpFlSTR Yfiler PCR amplification kit: a male specific, single amplification 17 Y-STR multiplex system. *J Forensic Sci,* 51**,** 64-75.

Mulero, J. J., Chang, C. W., Lagace, R. E., Wang, D. Y., Bas, J. L., Mcmahon, T. P. & Hennessy, L. K. 2008. Development and validation of the AmpFlSTR MiniFiler PCR Amplification Kit: a MiniSTR multiplex for the analysis of degraded and/or PCR inhibited DNA. *J Forensic Sci,* 53**,** 838-52.

Mulligan, C. M., Kaufman, S. R. & Quarino, L. 2011. The utility of polyester and cotton as swabbing substrates for the removal of cellular material from surfaces. *J Forensic Sci,* 56**,** 485-90.

Myres, N. M., Rootsi, S., Lin, A. A., Jarve, M., King, R. J., Kutuev, I., Cabrera, V. M., Khusnutdinova, E. K., Pshenichnov, A., Yunusbayev, B., Balanovsky, O., Balanovska, E., Rudan, P., Baldovic, M., Herrera, R. J., Chiaroni, J., Di Cristofaro, J., Villems, R., Kivisild, T. & Underhill, P. A. 2011. A major Y-chromosome haplogroup R1b Holocene era founder effect in Central and Western Europe. *Eur J Hum Genet,* 19**,** 95-101.

Nass, M. M. 1966. The circularity of mitochondrial DNA. *Proc Natl Acad Sci U S A,* 56**,** 1215-22.

Nass, M. M. 1969. Mitochondrial DNA. I. Intramitochondrial distribution and structural relations of single- and double-length circular DNA. *J Mol Biol,* 42**,** 521-8.

National Research Council (U.S.). Committee on Identifying the Needs of the Forensic Science Community., National Research Council (U.S.). Committee on Science Technology and Law. & National Research Council (U.S.). Committee on Applied and Theoretical Statistics. 2009. *Strengthening forensic science in the United States : a path forward,* Washington, D.C., National Academies Press.

Naue, J., Horer, S., Sanger, T., Strobl, C., Hatzer-Grubwieser, P., Parson, W. & Lutz-Bonengel, S. 2015. Evidence for frequent and tissue-specific sequence heteroplasmy in human mitochondrial DNA. *Mitochondrion,* 20**,** 82-94.

Navidi, W., Arnheim, N. & Waterman, M. S. 1992. A multiple-tubes approach for accurate genotyping of very small DNA samples by using PCR: statistical considerations. *Am J Hum Genet,* 50**,** 347-59.

Nielsen-Marsh, C. M. & Hedges, R. E. M. 2000. Patterns of diagenesis in bone I: The effects of site environments. *Journal of Archaeological Science,* 27**,** 1139-1150.

Nilsson, M., Malmgren, H., Samiotaki, M., Kwiatkowski, M., Chowdhary, B. P. & Landegren, U. 1994. Padlock probes: circularizing oligonucleotides for localized DNA detection. *Science,* 265**,** 2085-8.

Oberhammer, F., Wilson, J. W., Dive, C., Morris, I. D., Hickman, J. A., Wakeling, A. E., Walker, P. R. & Sikorska, M. 1993. Apoptotic Death in Epithelial-Cells - Cleavage of DNA to 300 and or 50 Kb Fragments Prior to or in the Absence of Internucleosomal Fragmentation. *Embo Journal,* 12**,** 3679-3684.

Okazaki, M., Yoshida, Y., Yamaguchi, S., Kaneno, M. & Elliott, J. C. 2001. Affinity binding phenomena of DNA onto apatite crystals. *Biomaterials,* 22**,** 2459-64.

Olaisen, B., Bar, W., Brinkmann, B., Budowle, B., Carracedo, A., Gill, P., Lincoln, P., Mayr, W. R. & Rand, S. 1998. DNA recommendations 1997 of the International Society for Forensic Genetics. *Vox Sang,* 74**,** 61-3.

Oostdik, K., French, J., Yet, D., Smalling, B., Nolde, C., Vallone, P. M., Butts, E. L., Hill, C. R., Kline, M. C., Rinta, T., Gerow, A. M., Allen, S. R., Huber, C. K., Teske, J., Krenke, B., Ensenberger, M., Fulmer, P. & Sprecher, C. 2013. Developmental validation of the PowerPlex(R) 18D System, a rapid STR multiplex for analysis of reference samples. *Forensic Sci Int Genet,* 7**,** 129-35.

Opel, K. L., Chung, D. & Mccord, B. R. 2010. A study of PCR inhibition mechanisms using real time PCR. *J Forensic Sci,* 55**,** 25-33.

Orlando, L., Ginolhac, A., Zhang, G., Froese, D., Albrechtsen, A., Stiller, M., Schubert, M., Cappellini, E., Petersen, B., Moltke, I., Johnson, P. L., Fumagalli, M., Vilstrup, J. T., Raghavan, M., Korneliussen, T., Malaspinas, A. S., Vogt, J., Szklarczyk, D., Kelstrup, C. D., Vinther, J., Dolocan, A., Stenderup, J., Velazquez, A. M., Cahill, J., Rasmussen, M., Wang, X., Min, J., Zazula, G. D., Seguin-Orlando, A., Mortensen, C., Magnussen, K., Thompson, J. F., Weinstock, J., Gregersen, K., Roed, K. H.,

Eisenmann, V., Rubin, C. J., Miller, D. C., Antczak, D. F., Bertelsen, M. F., Brunak, S., Al-Rasheid, K. A., Ryder, O., Andersson, L., Mundy, J., Krogh, A., Gilbert, M. T., Kjaer, K., Sicheritz-Ponten, T., Jensen, L. J., Olsen, J. V., Hofreiter, M., Nielsen, R., Shapiro, B., Wang, J. & Willerslev, E. 2013. Recalibrating Equus evolution using the genome sequence of an early Middle Pleistocene horse. *Nature,* 499**,** 74-8.

Osier, M. V., Cheung, K. H., Kidd, J. R., Pakstis, A. J., Miller, P. L. & Kidd, K. K. 2001. ALFRED: an allele frequency database for diverse populations and DNA polymorphisms--an update. *Nucleic Acids Res,* 29**,** 317-9.

Ozsolak, F., Platt, A. R., Jones, D. R., Reifenberger, J. G., Sass, L. E., Mcinerney, P., Thompson, J. F., Bowers, J., Jarosz, M. & Milos, P. M. 2009. Direct RNA sequencing. *Nature,* 461**,** 814-U73.

Palmer, R. & Burch, H. J. 2009. The population, transfer and persistence of fibres on the skin of living subjects. *Science & Justice,* 49**,** 259-264.

Palmer, R., Hutchinson, W. & Fryer, V. 2009. The discrimination of (non-denim) blue cotton. *Science & Justice,* 49**,** 12-18.

Palmer, R. & Oliver, S. 2004. The population of coloured fibres in human head hair. *Science & Justice,* 44**,** 83-88.

Palmer, R. & Polwarth, G. 2011. The persistence of fibres on skin in an outdoor deposition crime scene scenario. *Sci Justice,* 51**,** 187-9.

Park, S. J., Kim, J. Y., Yang, Y. G. & Lee, S. H. 2008. Direct STR amplification from whole blood and blood- or saliva-spotted FTA without DNA purification. *J Forensic Sci,* 53**,** 335-41.

Parson, W. 2003. *ENFSI DNA WG STR Population Database* [Online].

Parson, W., Gusmao, L., Hares, D. R., Irwin, J. A., Mayr, W. R., Morling, N., Pokorak, E., Prinz, M., Salas, A., Schneider, P. M. & Parsons, T. J. 2014. DNA Commission of the International Society for Forensic Genetics: Revised and extended guidelines for mitochondrial DNA typing. *Forensic Science International-Genetics,* 13**,** 134-142.

Parson, W., Huber, G., Moreno, L., Madel, M. B., Brandhagen, M. D., Nagl, S., Xavier, C., Eduardoff, M., Callaghan, T. C. & Irwin, J. A. 2015. Massively parallel sequencing of complete mitochondrial genomes from hair shaft samples. *Forensic Sci Int Genet,* 15**,** 8-15.

Parson, W. R., A. 2015. *EMPOP Website* [Online]. http://empop.org/ [Date last accessed 28/04/2016.

Parsons, T. J. & Coble, M. D. 2001. Increasing the forensic discrimination of mitochondrial DNA testing through analysis of the entire mitochondrial DNA genome. *Croat Med J,* 42**,** 304-9.

Parsons, T. J., Muniec, D. S., Sullivan, K., Woodyatt, N., Alliston-Greiner, R., Wilson, M. R., Berry, D. L., Holland, K. A., Weedn, V. W., Gill, P. & Holland, M. M. 1997. A high observed substitution rate in the human mitochondrial DNA control region. *Nat Genet,* 15**,** 363-8.

Payne, B. A., Wilson, I. J., Yu-Wai-Man, P., Coxhead, J., Deehan, D., Horvath, R., Taylor, R. W., Samuels, D. C., Santibanez-Koref, M. & Chinnery, P. F. 2013. Universal heteroplasmy of human mitochondrial DNA. *Hum Mol Genet,* 22**,** 384-90.

Peak, M. J. & Peak, J. G. 1989. Solar-ultraviolet-induced damage to DNA. *Photodermatol,* 6**,** 1-15.

Pesole, G., Gissi, C., De Chirico, A. & Saccone, C. 1999. Nucleotide substitution rate of mammalian mitochondrial genomes. *J Mol Evol,* 48**,** 427-34.

Petricevic, S., Whitaker, J., Buckleton, J., Vintiner, S., Patel, J., Simon, P., Ferraby, H., Hermiz, W. & Russell, A. 2010. Validation and development of interpretation guidelines for low copy number (LCN) DNA profiling in New Zealand using the AmpFlSTR (R) SGM Plus (TM) multiplex. *Forensic Science International-Genetics,* 4**,** 305-310.

Phillips, C. 2015. Forensic genetic analysis of bio-geographical ancestry. *Forensic Science International: Genetics,* 18**,** 49-65.

Pidaparti, R. M., Svintradze, D. V., Shan, Y. & Yokota, H. 2009. Optimization of hydrogen bonds for combined DNA/collagen complex. *J Theor Biol,* 256**,** 149-56.

Porreca, G. J., Zhang, K., Li, J. B., Xie, B., Austin, D., Vassallo, S. L., Leproust, E. M., Peck, B. J., Emig, C. J., Dahl, F., Gao, Y., Church, G. M. & Shendure, J. 2007. Multiplex amplification of large sets of human exons. *Nat Methods,* 4**,** 931-6.

Port, N. J., Bowyer, V. L., Graham, E. A., Batuwangala, M. S. & Rutty, G. N. 2006. How long does it take a static speaking individual to contaminate the immediate environment? *Forensic Sci Med Pathol,* 2**,** 157-63.

Pounds, C. A. 1975. The recovery of fibres from the surface of clothing for forensic examinations. *J Forensic Sci Soc,* 15**,** 127-32.

Pounds, C. A. & Smalldon, K. W. 1975a. The transfer of fibres between clothing materials during simulated contacts and their persistence during wear. Part I--Fibre transference. *J Forensic Sci Soc,* 15**,** 17-27.

Pounds, C. A. & Smalldon, K. W. 1975b. The transfer of fibres between clothing materials during simulated contacts and their persistence during wear. Part III--a preliminary investigation of the mechanisms involved. *J Forensic Sci Soc,* 15**,** 197-207.

Pounds, C. A. & Smalldon, K. W. 1975c. The transfer of fibres between clothing materials during simulated contacts and their persistence during wear. Part III--Fibre persistence. *J Forensic Sci Soc,* 15**,** 29-37.

Price, N. E., Johnson, K. M., Wang, J., Fekry, M. I., Wang, Y. & Gates, K. S. 2014. Interstrand DNA-DNA cross-link formation between adenine residues and abasic sites in duplex DNA. *J Am Chem Soc,* 136**,** 3483-90.

Proff, C. R., M.A. Schneider, P.M. Low volume PCR (LV-PCR) for STR typing of forensic casework samples. *In:* Amorin, A. C.-R., F. Morling, N. , ed. Progress in Forensic Genetics 11, 2006 21st International ISFG Congress. Elsevier, 645-647.

Purschke, M., Laubach, H. J., Anderson, R. R. & Manstein, D. 2010. Thermal injury causes DNA damage and lethality in unheated surrounding cells: active thermal bystander effect. *J Invest Dermatol,* 130**,** 86-92.

Pushkarev, D., Neff, N. F. & Quake, S. R. 2009. Single-molecule sequencing of an individual human genome. *Nature Biotechnology,* 27**,** 847-U101.

Qiagen 2010. Qiagen Multiplex PCR Handbook. https://www.qiagen.com/gb/resources/resourcedetail?id=a541a49c-cd06-40ca-b1d2-563d0324ad6c&lang=en

Qiagen 2015a. Investigator ESSplex SE QS Kit Technical Information. Qiagen Website.

Qiagen 2015b. QIAamp DNA Mini and Blood Mini Handbook. www.qiagen.com: Qiagen.

Quail, M. A., Smith, M., Coupland, P., Otto, T. D., Harris, S. R., Connor, T. R., Bertoni, A., Swerdlow, H. P. & Gu, Y. 2012. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics,* 13**,** 341.

Quick, J., Ashton, P., Calus, S., Chatt, C., Gossain, S., Hawker, J., Nair, S., Neal, K., Nye, K., Peters, T., De Pinna, E., Robinson, E., Struthers, K., Webber, M., Catto, A., Dallman, T. J., Hawkey, P. & Loman, N. J. 2015. Rapid draft sequencing and real-time nanopore sequencing in a hospital outbreak of Salmonella. *Genome Biol,* 16**,** 114.

Ramos, A., Santos, C., Mateiu, L., Gonzalez Mdel, M., Alvarez, L., Azevedo, L., Amorim, A. & Aluja, M. P. 2013. Frequency and pattern of heteroplasmy in the complete human mitochondrial genome. *PLoS One,* 8**,** e74636.

Rastogi, R. P., Richa, Kumar, A., Tyagi, M. B. & Sinha, R. P. 2010. Molecular mechanisms of ultraviolet radiation-induced DNA damage and repair. *J Nucleic Acids,* 2010**,** 592980.

Ravanat, J. L., Douki, T. & Cadet, J. 2001. Direct and indirect effects of UV radiation on DNA and its components. *Journal of Photochemistry and Photobiology B-Biology,* 63**,** 88-102.

Ray, P. F., Winston, R. M. L. & Handyside, A. H. 1996. Reduced allele dropout in single-cell analysis for preimplantation genetic diagnosis of cystic fibrosis. *Journal of Assisted Reproduction and Genetics,* 13**,** 104-106.

Reich, D., Green, R. E., Kircher, M., Krause, J., Patterson, N., Durand, E. Y., Viola, B., Briggs, A. W., Stenzel, U., Johnson, P. L. F., Maricic, T., Good, J. M., Marques-Bonet, T., Alkan, C., Fu, Q. M., Mallick, S., Li, H., Meyer, M., Eichler, E. E., Stoneking, M., Richards, M., Talamo, S., Shunkov, M. V., Derevianko, A. P., Hublin, J. J., Kelso, J., Slatkin, M. & Paabo, S. 2010. Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature,* 468**,** 1053-1060.

Rho, J. Y., Kuhn-Spearing, L. & Zioupos, P. 1998. Mechanical properties and the hierarchical structure of bone. *Med Eng Phys,* 20**,** 92-102.

Ricchetti, M., Tekaia, F. & Dujon, B. 2004. Continued colonization of the human genome by mitochondrial DNA. *PLoS Biol,* 2**,** E273.

Roberts, S. J., Smith, C. I., Millard, A. & Collins, M. J. 2002. The taphonomy of cooked bone: Characterizing boiling and its physico-chemical effects. *Archaeometry,* 44**,** 485-494.

Robertson, J. & Grieve, M. 1999. *Forensic examination of fibres,* London, Taylor & Francis.

Robin, E. D. & Wong, R. 1988. Mitochondrial DNA molecules and virtual number of mitochondria per cell in mammalian cells. *J Cell Physiol,* 136**,** 507-13.

Robinson, J. T., Thorvaldsdottir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G. & Mesirov, J. P. 2011. Integrative genomics viewer. *Nat Biotechnol,* 29**,** 24-6.

Rockenbauer, E., Hansen, S., Mikkelsen, M., Borsting, C. & Morling, N. 2014. Characterization of mutations and sequence variants in the D21S11 locus by next generation sequencing. *Forensic Sci Int Genet,* 8**,** 68-72.

Rohland, N. & Hofreiter, M. 2007. Ancient DNA extraction from bones and teeth. *Nature Protocols,* 2**,** 1756-1762.

Rohland, N., Siedel, H. & Hofreiter, M. 2010. A rapid column-based ancient DNA extraction method for increased sample throughput. *Mol Ecol Resour,* 10**,** 677-83.

Ross, K. S., Haites, N. E. & Kelly, K. F. 1990. Repeated freezing and thawing of peripheral blood and DNA in suspension: effects on DNA yield and integrity. *J Med Genet,* 27**,** 569-70.

Rossmanith, P., Roder, B., Fruhwirth, K., Vogl, C. & Wagner, M. 2011. Mechanisms of degradation of DNA standards for calibration function during storage. *Appl Microbiol Biotechnol,* 89**,** 407-17.

Rothberg, J. M., Hinz, W., Rearick, T. M., Schultz, J., Mileski, W., Davey, M., Leamon, J. H., Johnson, K., Milgrew, M. J., Edwards, M., Hoon, J., Simons, J. F., Marran, D., Myers, J. W., Davidson, J. F., Branting, A., Nobile, J. R., Puc, B. P., Light, D., Clark, T. A., Huber, M., Branciforte, J. T., Stoner, I. B., Cawley, S. E., Lyons, M., Fu, Y. T., Homer, N., Sedova, M., Miao, X., Reed, B., Sabina, J., Feierstein, E., Schorn, M., Alanjary, M., Dimalanta, E., Dressman, D., Kasinskas, R., Sokolsky, T., Fidanza, J. A., Namsaraev, E., Mckernan, K. J., Williams, A., Roth, G. T. & Bustillo, J. 2011. An integrated semiconductor device enabling non-optical genome sequencing. *Nature,* 475**,** 348-352.

Roux, C., Huttunen, J., Rampling, K. & Robertson, J. 2001. Factors affecting the potential for fibre contamination in purpose-designed forensic search rooms. *Science & Justice,* 41**,** 135-144.

Roux, C., Langdon, S., Waight, D. & Robertson, J. 1999. The transfer and persistence of automotive carpet fibres on shoe soles. *Sci Justice,* 39**,** 239-51.

Roux, C. & Margot, P. 1997. The population of textile fibres on car seats. *Science & Justice,* 37**,** 25-30.

Rutty, G. N., Hopwood, A. & Tucker, V. 2003. The effectiveness of protective clothing in the reduction of potential DNA contamination of the scene of crime. *Int J Legal Med,* 117**,** 170-4.

Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B. & Erlich, H. A. 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science,* 239**,** 487-91.

Saks, M. J. & Koehler, J. J. 2005. The coming paradigm shift in forensic identification science. *Science,* 309**,** 892-5.

Salter, M. T. & Cook, R. 1996. Transfer of fibres to head hair, their persistence and retrieval. *Forensic Science International,* 81**,** 211-221.

Samejima, K. & Earnshaw, W. C. 2005. Trashing the genome: The role of nucleases during apoptosis. *Nature Reviews Molecular Cell Biology,* 6**,** 677-688.

Sanger, F., Nicklen, S. & Coulson, A. R. 1977. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A,* 74**,** 5463-7.

Satoh, M. & Kuroiwa, T. 1991. Organization of multiple nucleoids and DNA molecules in mitochondria of a human cell. *Exp Cell Res,* 196**,** 137-40.

Sawaya, S., Bagshaw, A., Buschiazzo, E., Kumar, P., Chowdhury, S., Black, M. A. & Gemmell, N. 2013. Microsatellite tandem repeats are abundant in human promoters and are associated with regulatory elements. *PLoS One,* 8**,** e54710.

Schadt, E. E., Turner, S. & Kasarskis, A. 2010. A window into third-generation sequencing. *Human Molecular Genetics,* 19**,** R227-R240.

Schirmer, M., Ijaz, U. Z., D'amore, R., Hall, N., Sloan, W. T. & Quince, C. 2015. Insight into biases and sequencing errors for amplicon sequencing with the Illumina MiSeq platform. *Nucleic Acids Res,* 43**,** e37.

Schlotterer, C. & Tautz, D. 1992. Slippage synthesis of simple sequence DNA. *Nucleic Acids Res,* 20**,** 211-5.

Schmitt, M. W., Kennedy, S. R., Salk, J. J., Fox, E. J., Hiatt, J. B. & Loeb, L. A. 2012. Detection of ultra-rare mutations by next-generation sequencing. *Proc Natl Acad Sci U S A,* 109**,** 14508-13.

Schotman, T. G. V. D. W., J. 2015. On the recovery of fibres by tape lifts, tape scanning, and manual isolation. *Science & Justice*.

Schumm, J. W. B., J.W. Hennes, L.F. Micka, K.A. Sprecher, C.J. Lins, A.M. Amiott, E.A. Rabbach, D.R. Taylor, J.A. Tereba, A. 1998. Pentanucleotide Repeats: Highly Polymorphic Genetic Markers Displaying Minimal Stutter Artifact. *Genetic Identity Conference Proceedings. Ninth International Symposium on human identification.* USA.

Schwartz, M. & Vissing, J. 2002. Paternal inheritance of mitochondrial DNA. *N Engl J Med,* 347**,** 576-80.

Schwarz, C., Debruyne, R., Kuch, M., Mcnally, E., Schwarcz, H., Aubrey, A. D., Bada, J. & Poinar, H. 2009. New insights from old bones: DNA preservation and degradation in permafrost preserved mammoth remains. *Nucleic Acids Res,* 37**,** 3215-29.

Seah, L. H., Othman, M. I., Jaya, P. & Jeevan, N. H. 2004. DNA profiling on fabrics: an in-situ method. *Progress in Forensic Genetics 10,* 1261**,** 565-567.

Seo, S. B., King, J. L., Warshauer, D. H., Davis, C. P., Ge, J. & Budowle, B. 2013. Single nucleotide polymorphism typing with massively parallel sequencing for human identification. *Int J Legal Med,* 127**,** 1079-86.

Sethi, V. P., E.A. Green, W.M. Ng, J. Kanthaswamy, S. 2013. Yield of Male Touch DNA from Fabrics in an Assault Model. *Journal of Forensic Research,* T1**,** 2157-7145.

Setlow, R. B. 1966. Cyclobutane-type pyrimidine dimers in polynucleotides. *Science,* 153**,** 379-86.

Shao, W., Khin, S. & Kopp, W. C. 2012. Characterization of effect of repeated freeze and thaw cycles on stability of genomic DNA using pulsed field gel electrophoresis. *Biopreserv Biobank,* 10**,** 4-11.

Shapiro, R. & Klein, R. S. 1966. The deamination of cytidine and cytosine by acidic buffer solutions. Mutagenic implications. *Biochemistry,* 5**,** 2358-62.

Shen, J. C., Rideout, W. M., 3rd & Jones, P. A. 1994. The rate of hydrolytic deamination of 5-methylcytosine in double-stranded DNA. *Nucleic Acids Res,* 22**,** 972-6.

Shigenaga, M. K., Park, J. W., Cundy, K. C., Gimeno, C. J. & Ames, B. N. 1990. In vivo oxidative DNA damage: measurement of 8-hydroxy-2'-deoxyguanosine in DNA and urine by high-performance liquid chromatography with electrochemical detection. *Methods Enzymol,* 186**,** 521-30.

Shokolenko, I. N., Wilson, G. L. & Alexeyev, M. F. 2013. Persistent damage induces mitochondrial DNA degradation. *DNA Repair (Amst),* 12**,** 488-99.

Shriver, M. D., Parra, E. J., Dios, S., Bonilla, C., Norton, H., Jovel, C., Pfaff, C., Jones, C., Massac, A., Cameron, N., Baron, A., Jackson, T., Argyropoulos, G., Jin, L., Hoggart, C. J., Mckeigue, P. M. & Kittles, R. A. 2003. Skin pigmentation, biogeographical ancestry and admixture mapping. *Hum Genet,* 112**,** 387-99.

Smith, C. I., Chamberlain, A. T., Riley, M. S., Stringer, C. & Collins, M. J. 2003. The thermal history of human fossils and the likelihood of successful DNA amplification. *Journal of Human Evolution,* 45**,** 203-217.

Smith, P. J. & Ballantyne, J. 2007. Simplified low-copy-number DNA analysis by post-PCR purification. *J Forensic Sci,* 52**,** 820-9.

Steenken, S. & Jovanovic, S. V. 1997. How easily oxidizable is DNA? One-electron reduction potentials of adenosine and guanosine radicals in aqueous solution. *Journal of the American Chemical Society,* 119**,** 617-618.

Strand, M., Prolla, T. A., Liskay, R. M. & Petes, T. D. 1993. Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair. *Nature,* 365**,** 274-6.

Sudmant, P. H., Rausch, T., Gardner, E. J., Handsaker, R. E., Abyzov, A., Huddleston, J., Zhang, Y., Ye, K., Jun, G., Hsi-Yang Fritz, M., Konkel, M. K., Malhotra, A., Stutz, A. M., Shi, X., Paolo Casale, F., Chen, J., Hormozdiari, F., Dayama, G., Chen, K., Malig, M., Chaisson, M. J. P., Walter, K., Meiers, S., Kashin, S., Garrison, E., Auton, A., Lam, H. Y. K., Jasmine Mu, X., Alkan, C., Antaki, D., Bae, T., Cerveira, E., Chines, P., Chong, Z., Clarke, L., Dal, E., Ding, L., Emery, S., Fan, X., Gujral, M., Kahveci, F., Kidd, J. M., Kong, Y., Lameijer, E.-W., Mccarthy, S., Flicek, P., Gibbs, R. A., Marth, G., Mason, C. E., Menelaou, A., Muzny, D. M., Nelson, B. J., Noor, A., Parrish, N. F., Pendleton, M., Quitadamo, A., Raeder, B., Schadt, E. E., Romanovitch, M., Schlattl, A., Sebra, R., Shabalin, A. A., Untergasser, A., Walker, J. A., Wang, M., Yu, F., Zhang, C., Zhang, J., Zheng-Bradley, X., Zhou, W., Zichner, T., Sebat, J., Batzer, M. A., Mccarroll, S. A., The Genomes Project, C., Mills, R. E., Gerstein, M. B., Bashir, A., Stegle, O., Devine, S. E., Lee, C., Eichler, E. E. & Korbel, J. O. 2015. An integrated map of structural variation in 2,504 human genomes. *Nature,* 526**,** 75-81.

Sullivan, K. M., Mannucci, A., Kimpton, C. P. & Gill, P. 1993. A rapid and quantitative DNA sex test: fluorescence-based PCR analysis of X-Y homologous gene amelogenin. *Biotechniques,* 15**,** 636-8, 640-1.

Sun, H., Liu, S., Zhang, Y. & Whittle, M. R. 2014. Comparison of southern Chinese Han and Brazilian Caucasian mutation rates at autosomal short tandem repeat loci used in human forensic genetics. *Int J Legal Med,* 128**,** 1-9.

Sung, T. 1981. *The Washing Away of Wrongs,* Centre for Chinese Studies Publications.

Syed, F., Grunenwald, H. & Caruccio, N. 2009. Optimized library preparation method for next-generation sequencing. *Nature Methods,* 6**,** I-Ii.

Taberlet, P., Griffin, S., Goossens, B., Questiau, S., Manceau, V., Escaravage, N., Waits, L. P. & Bouvet, J. 1996. Reliable genotyping of samples with very low DNA quantities using PCR. *Nucleic Acids Res,* 24**,** 3189-94.

Technologies, L. 2012. 5500 W Series Genetic Analyzers. http://tools.lifetechnologies.com/content/sfs/brochures/5500-w-series-spec-sheet.pdf.

Technologies, L. 2013. *Overview of SOLiD™ Sequencing Chemistry* [Online]. http://www.appliedbiosystems.com/absite/us/en/home/applications-technologies/solid-next-generation-sequencing/next-generation-systems/solid-sequencing-chemistry.html [Date last accessed 20/06/2015.

Technologies, L. 2015. *Ion PGM™ and Ion Proton™ System Chips* [Online]. https://www.lifetechnologies.com/uk/en/home/life-science/sequencing/next-generation-sequencing/ion-torrent-next-generation-sequencing-workflow/ion-torrent-next-generation-sequencing-run-sequence/ion-pgm-ion-proton-system-chips.html [Date last accessed 20/06/2015.

Tewhey, R., Warner, J. B., Nakano, M., Libby, B., Medkova, M., David, P. H., Kotsopoulos, S. K., Samuels, M. L., Hutchison, J. B., Larson, J. W., Topol, E. J., Weiner, M. P., Harismendy, O., Olson, J., Link, D. R. & Frazer, K. A. 2009. Microdroplet-based PCR enrichment for large-scale targeted sequencing. *Nat Biotechnol*, 27**,** 1025-31.

The Genomes Project, C. 2015. A global reference for human genetic variation. *Nature*, 526**,** 68-74.

Thorvaldsdottir, H., Robinson, J. T. & Mesirov, J. P. 2013. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform*, 14**,** 178-92.

Tilstone, W. J. S., K.A. Clark, L.A. 2006. *Forensic Science: An Encyclopedia of History, Methods and Techniques*, ABC-CLIO.

Tomas, C., Mogensen, H. S., Friis, S. L., Hallenberg, C., Stene, M. C. & Morling, N. 2014. Concordance study and population frequencies for 16 autosomal STRs analyzed with PowerPlex(R) ESI 17 and AmpFlSTR(R) NGM SElect in Somalis, Danes and Greenlanders. *Forensic Sci Int Genet*, 11**,** e18-21.

Tucker, V. C., Hopwood, A. J., Sprecher, C. J., Mclaren, R. S., Rabbach, D. R., Ensenberger, M. G., Thompson, J. M. & Storts, D. R. 2011. Developmental validation of the PowerPlex((R)) ESI 16 and PowerPlex((R)) ESI 17 Systems: STR multiplexes for the new European standard. *Forensic Sci Int Genet*, 5**,** 436-48.

Tucker, V. C., Hopwood, A. J., Sprecher, C. J., Mclaren, R. S., Rabbach, D. R., Ensenberger, M. G., Thompson, J. M. & Storts, D. R. 2012a. Developmental validation of the PowerPlex(R) ESX 16 and PowerPlex(R) ESX 17 Systems. *Forensic Sci Int Genet*, 6**,** 124-31.

Tucker, V. C., Kirkham, A. J. & Hopwood, A. J. 2012b. Forensic validation of the PowerPlex(R) ESI 16 STR Multiplex and comparison of performance with AmpFlSTR(R) SGM Plus(R). *Int J Legal Med*, 126**,** 345-56.

Turner, E. H., Lee, C., Ng, S. B., Nickerson, D. A. & Shendure, J. 2009. Massively parallel exon capture and library-free resequencing across 16 genomes. *Nat Methods*, 6**,** 315-6.

Union, T. C. O. T. E. 2008. ACTS ADOPTED UNDER TITLE VI OF THE EU TREATY Decision 2008/615/JHA. *Official Journal of the European Union*, 210**,** 89-100.

Urquhart, A., Kimpton, C. P., Downes, T. J. & Gill, P. 1994. Variation in short tandem repeat sequences--a survey of twelve microsatellite loci for use as forensic identification markers. *Int J Legal Med*, 107**,** 13-20.

Vallone, P. M. & Butler, J. M. 2004. AutoDimer: a screening tool for primer-dimer and hairpin structures. *Biotechniques*, 37**,** 226-31.

Van Bruggen, E. F., Borst, P., Ruttenberg, G. J., Gruber, M. & Kroon, A. M. 1966. Circular mitochondrial DNA. *Biochim Biophys Acta*, 119**,** 437-9.

Van Dijk, E. L., Auger, H., Jaszczyszyn, Y. & Thermes, C. 2014a. Ten years of next-generation sequencing technology. *Trends Genet*, 30**,** 418-26.

Van Dijk, E. L., Jaszczyszyn, Y. & Thermes, C. 2014b. Library preparation methods for next-generation sequencing: tone down the bias. *Exp Cell Res*, 322**,** 12-20.

Van Neste, C., Van Nieuwerburgh, F., Van Hoofstat, D. & Deforce, D. 2012. Forensic STR analysis using massive parallel sequencing. *Forensic Sci Int Genet*, 6**,** 810-8.

Van Oorschot, R. A. & Jones, M. K. 1997. DNA fingerprints from fingerprints. *Nature*, 387**,** 767.

Van Oven, M. 2015. *Phylo Tree mt* [Online]. http://www.phylotree.org/ [Date last accessed 24/05/2015.

Van Oven, M. & Kayser, M. 2009. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum Mutat*, 30**,** E386-94.

Verdon, T. J., Mitchell, R. J. & Van Oorschot, R. A. 2014. Evaluation of tapelifting as a collection method for touch DNA. *Forensic Sci Int Genet,* 8**,** 179-86.

Vernarecci, S., Ottaviani, E., Agostino, A., Mei, E., Calandro, L. & Montagna, P. 2015. Quantifiler(R) Trio Kit and forensic samples management: a matter of degradation. *Forensic Sci Int Genet,* 16**,** 77-85.

Voigt, O. E., V. Worheide, G. 2012. First evaluation of mitochondrial DNA as a marker for phylogeographic studies of Calcarea: a case study from Leucetta chagosensis. *Hydrobiologia* 687**,** 101-106.

Wallace, D. C., Ye, J. H., Neckelmann, S. N., Singh, G., Webster, K. A. & Greenberg, B. D. 1987. Sequence analysis of cDNAs for the human and bovine ATP synthase beta subunit: mitochondrial DNA genes sustain seventeen times more mutations. *Curr Genet,* 12**,** 81-90.

Walsh, P. S., Fildes, N. J. & Reynolds, R. 1996. Sequence analysis and characterization of stutter products at the tetranucleotide repeat locus vWA. *Nucleic Acids Res,* 24**,** 2807-12.

Walsh, S., Liu, F., Wollstein, A., Kovatsi, L., Ralf, A., Kosiniak-Kamysz, A., Branicki, W. & Kayser, M. 2013. The HIrisPlex system for simultaneous prediction of hair and eye colour from DNA. *Forensic Sci Int Genet,* 7**,** 98-115.

Wang, D. Y., Chang, C. W., Lagace, R. E., Calandro, L. M. & Hennessy, L. K. 2012. Developmental validation of the AmpFlSTR(R) Identifiler(R) Plus PCR Amplification Kit: an established multiplex assay with improved performance. *J Forensic Sci,* 57**,** 453-65.

Wang, D. Y., Chang, C. W., Lagace, R. E., Oldroyd, N. J. & Hennessy, L. K. 2011. Development and validation of the AmpFlSTR(R) Identifiler(R) Direct PCR Amplification Kit: a multiplex assay for the direct amplification of single-source samples. *J Forensic Sci,* 56**,** 835-45.

Watson, J. D. & Crick, F. H. 1953. The structure of DNA. *Cold Spring Harb Symp Quant Biol,* 18**,** 123-31.

Watt, R., Roux, C. & Robertson, J. 2005. The population of coloured textile fibres in domestic washing machines. *Science & Justice,* 45**,** 75-83.

Weber-Lehmann, J., Schilling, E., Gradl, G., Richter, D. C., Wiehler, J. & Rolf, B. 2014. Finding the needle in the haystack: Differentiating "identical" twins in paternity testing and forensics by ultra-deep next generation sequencing. *Forensic Science International-Genetics,* 9**,** 42-46.

Weber, J. L. & May, P. E. 1989. Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *Am J Hum Genet,* 44**,** 388-96.

Welch, L. A., Gill, P., Phillips, C., Ansell, R., Morling, N., Parson, W., Palo, J. U. & Bastisch, I. 2012. European Network of Forensic Science Institutes (ENFSI): Evaluation of new commercial STR multiplexes that include the European Standard Set (ESS) of markers. *Forensic Sci Int Genet,* 6**,** 819-26.

Werrett, D. J. 1997. The National DNA Database. *Forensic Sci Int,* 88**,** 33-42.

Westen, A. A., Nagel, J. H., Benschop, C. C., Weiler, N. E., De Jong, B. J. & Sijen, T. 2009. Higher capillary electrophoresis injection settings as an efficient approach to increase the sensitivity of STR typing. *J Forensic Sci,* 54**,** 591-8.

Wheeler, D. A., Srinivasan, M., Egholm, M., Shen, Y., Chen, L., Mcguire, A., He, W., Chen, Y. J., Makhijani, V., Roth, G. T., Gomes, X., Tartaro, K., Niazi, F., Turcotte, C. L., Irzyk, G. P., Lupski, J. R., Chinault, C., Song, X. Z., Liu, Y., Yuan, Y., Nazareth, L., Qin, X., Muzny, D. M., Margulies, M., Weinstock, G. M., Gibbs, R. A. & Rothberg, J. M. 2008. The complete genome of an individual by massively parallel DNA sequencing. *Nature,* 452**,** 872-6.

Whitaker, J. P., Cotton, E. A. & Gill, P. 2001. A comparison of the characteristics of profiles produced with the AMPFlSTR SGM Plus multiplex system for both standard and low copy number (LCN) STR DNA analysis. *Forensic Sci Int,* 123**,** 215-23.

Widlak, P., Li, P., Wang, X. D. & Garrard, W. T. 2000. Cleavage preferences of the apoptotic endonuclease DFF40 (caspase-activated DNase or nuclease) on naked DNA and chromatin substrates. *Journal of Biological Chemistry,* 275**,** 8226-8232.

Wiggins, K. & Drummond, P. 2007. Identifying a suitable mounting medium for use in forensic fibre examination. *Sci Justice,* 47**,** 2-8.

Wiggins, K., Drummond, P. & Champod, T. H. 2004. A study in relation to the random distribution of four fibre types on clothing - (incorporating a review of previous target fibre studies). *Science & Justice.*

Wiggins, K. G. 2001. Forensic textile fiber examination across the USA and Europe. *J Forensic Sci,* 46**,** 1303-8.

Willuweit, S. & Roewer, L. 2015. The new Y Chromosome Haplotype Reference Database. *Forensic Sci Int Genet,* 15**,** 43-8.

Wink, D. A., Kasprzak, K. S., Maragos, C. M., Elespuru, R. K., Misra, M., Dunams, T. M., Cebula, T. A., Koch, W. H., Andrews, A. W., Allen, J. S. & Et Al. 1991. DNA deaminating ability and genotoxicity of nitric oxide and its progenitors. *Science,* 254**,** 1001-3.

Woo, E. J., Kim, Y. G., Kim, M. S., Han, W. D., Shin, S., Robinson, H., Park, S. Y. & Oh, B. H. 2004. Structural mechanism for inactivation and activation of CAD/DFF40 in the apoptotic pathway. *Molecular Cell,* 14**,** 531-539.

Wyllie, A. H., Kerr, J. F. & Currie, A. R. 1980. Cell death: the significance of apoptosis. *Int Rev Cytol,* 68**,** 251-306.

Yang, D. Y., Eng, B., Waye, J. S., Dudar, J. C. & Saunders, S. R. 1998. Technical note: improved DNA extraction from ancient bones using silica-based spin columns. *Am J Phys Anthropol,* 105**,** 539-43.

Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S. & Madden, T. L. 2012. Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics,* 13**,** 134.

Yoo, H. B. L., H.M Yang, I. Kim, S.K. Park, S.R 2011. Flow cytometric investigation on degradation of ma-cro-DNA by common laboratory manipulations. *Journal of Biophysical Chemistry,* 2**,** 102-110.

Yoon, J. K., Ahn, J., Kim, H. S., Han, S. M., Jang, H., Lee, M. G., Lee, J. H. & Bang, D. 2015. microDuMIP: target-enrichment technique for microarray-based duplex molecular inversion probes. *Nucleic Acids Res,* 43**,** e28.

Zeng, X., King, J. L., Stoljarova, M., Warshauer, D. H., Larue, B. L., Sajantila, A., Patel, J., Storts, D. R. & Budowle, B. 2015. High sensitivity multiplex short tandem repeat loci analyses with massively parallel sequencing. *Forensic Sci Int Genet,* 16**,** 38-47.

Zeng, X. K., J.L. Hermanson, S. Patel, J. Storts, D.R. Budowle, B. 2015. Evaluation of the PowerSeq Auto System by massively parallel sequencing. http://www.promega.com/~/media/files/resources/conference%20proceedings/ishi%2025/poster%20abstracts/97%20-%20xiangpei%20zeng.pdf.

Zhang, J. H., Lee, H., Lou, D. W., Bovin, G. P. & Xu, M. 2000. Lack of obvious 50 kilobase pair DNA fragments in DNA fragmentation factor 45-deficient thymocytes upon activation of apoptosis. *Biochemical and Biophysical Research Communications,* 274**,** 225-229.

Zhang, X., Rosenstein, B. S., Wang, Y., Lebwohl, M. & Wei, H. 1997. Identification of possible reactive oxygen species involved in ultraviolet radiation-induced oxidative DNA damage. *Free Radic Biol Med,* 23**,** 980-5.

Zischler, H., Geisert, H., Von Haeseler, A. & Paabo, S. 1995a. A nuclear 'fossil' of the mitochondrial D-loop and the origin of modern humans. *Nature,* 378**,** 489-92.

Zischler, H., Hoss, M., Handt, O., Von Haeseler, A., Van Der Kuyl, A. C. & Goudsmit, J. 1995b. Detecting dinosaur DNA. *Science,* 268**,** 1192-3; author reply 1194.

8. Appendices

8.1 Identifying mtDNA haplotype

In a simple point mutation, the nucleotide position and the nucleotide base that differ from the rCRS are recorded, i.e. at position 263 the rCRS nucleotide is adenine (A), if the queried sequence has guanine (G) at that same position then the designation for that position is written 263G. Insertions are shown through the use of a decimal place after the preceding nucleotide positon, i.e. a thymine (T) insertion after nucleotide position 267 is depicted as 267.1T. If there is a two-base insertion of TT, then the second insertion is shown as 267.2T and so on. In the case of homopolymeric stretches such as the common C stretch at positions 311-315, then the highest number possible is used as the preceding nucleotide position, i.e. 315.1C [58]. A nucleotide position followed by the letters 'DEL' is used to represent a deleted nucleotide, i.e. 389DEL [57] and an 'N' is utilised to denote an ambiguous base call where all four bases are observed [57]. In the case of heteroplasmy, mixed base sites are represented by a standardised letter as set out by the International Union of Biochemistry (IUB). For example, if a mixture of adenine and guanine is observed at position 678, then the notation 678R is used [59]. Upper case letters should be used as standard, with lower cases letters used to denote heteroplasmic events that include an insertion or deletion [57]. RS nucleotide is adenine (A), if the queried sequence has guanine (G) at that same position then the designation for that position is written 263G. Insertions are shown through the use of a decimal place after the preceding nucleotide positon, i.e. a thymine (T) insertion after nucleotide position 267 is depicted as 267.1T. If there is a two base insertion of TT, then the second insertion is shown as 267.2T and so on. In the case of homopolymeric stretches such as the common C stretch at positions 311-315, then the highest number possible is used as the preceding nucleotide position, i.e. 315.1C (Carracedo et al., 2000). A nucleotide position followed by the letters 'DEL' is used to represent a deleted nucleotide, i.e. 389DEL (Parson et al., 2014) and an 'N' is utilised to denote an ambiguous base call where all four bases are observed (Parson et al., 2014).

8.2 Allele frequencies for ltDNA assay loci (Section 3.3.6)

PENTA E		D12S391		D10S1248	
Allele	**Amplicon size**	**Allele**	**Amplicon size**	**Allele**	**Amplicon size**
5	73	15	145	8	79
6	78	16	149	9	83
7	83	17	153	10	87
8	88	17.3	156	11	91
9	93	18	157	12	95
9.1	94	18.3	160	13	99
9.4	97	19	161	14	103
10	98	19.3	164	15	107
10.2	100	20	165	16	111
11	103	21	169	17	115
11.4	107	22	173	18	119
12	108	23	177	19	123
12.1	109	24	181		
12.2	110	25	185		
12.3	111	26	189		
13	113	27	193		
13.2	115	27.2	195		
13.4	117				
14	118				
14.4	122				
15	123				
15.2	125				
15.4	127				
16	128				
16.4	132				
17	133				
17.4	137				
18	138				
18.4	142				
19	143				
19.4	147				
20	148				
20.2	150				
20.3	151				
21	153				
22	158				
23	163				
23.4	167				
24	168				

8.3 Inhibition test data

Table 8-1 Average peak area (RFU) data for alleles amplified during inhibition tests for J-LAR Tape, Sellotape, Sticky stuff remover (SSR) and Adhesive Off (AO). Peak areas for control DNA also recorded. (-) denotes no observed amplification (Figure 3.4 and Figure 3.5)

	Sample	D3		VWA		D16		D2		AMEL		D8	
	CONTROL	24382	20866	47579	35598	27972	26307	5895	15119	12534	19140	22930	31310
Tape	J-LAR TAPE	14257	9447	62327	32929	30605	25436	15963	10303	21849	37404	41702	17143
	SELLOTAPE	8081	11458	10294	22332	8164	13258	3708	2201	16028	16318	8794	12157
Adhesive remover	SSR 1 µl	17726	13917	37075	26336	15222	10008	5843	11378	21623	20871	17945	19389
	SSR 2 µl	-	-	-	-	-	-	-	-	-	-	-	-
	SSR 3 µl	-	-	-	-	-	-	-	-	-	-	-	-
	AO 1 µl	-	-	-	-	-	-	-	-	-	-	-	-
	AO 2 µl	-	-	-	-	-	-	-	-	-	-	-	-
	AO 3 µl	-	-	-	-	-	-	-	-	-	-	-	-

	Sample	D21		D18		D19		THO1		FGA	
	CONTROL	17440	14867	11773	10134	20276	17665	17413	25802	12138	8154
Tape	J-LAR TAPE	32635	31399	10587	13944	23948	23012	27078	20944	21285	14171
	SELLOTAPE	8241	7776	5884	3340	8104	5884	3721	12221	6487	4688
Adhesive remover	SSR 1 µl	11530	14038	7728	9264	18429	16426	13442	17716	8516	6192
	SSR 2 µl	-	-	-	-	-	-	-	-	-	-
	SSR 3 µl	-	-	-	-	-	-	-	-	-	-
	AO 1 µl	-	-	-	-	-	-	-	-	-	-
	AO 2 µl	-	-	-	-	-	-	-	-	-	-
	AO 3 µl	-	-	-	-	-	-	-	-	-	-

8.4 Amplification reaction volume trial data (Figure 3.6)

Table 8-2 Alleles observed using five different reaction volumes for the amplification of DNA from a single-cell (in triplicate)

	AMEL		D3		VWA	D16		D2		D8		D21		D18		D19		THO1		FGA	
Single_cell_5µL_1	-	-	-	-	17	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Single_cell_5µL_2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Single_cell_5µL_3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Single_cell_10µL_1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Single_cell_10µL_2	X	Y	-	17	-	-	-	-	-	12	-	-	32	12	-	13	15	-	-	-	21
Single_cell_10µL_3	X	-	16	17	17	-	-	16	20	-	15	-	-	-	-	-	-	8	-	-	-
Single_cell_15µL_1	X	Y	16	17	17	-	12	-	20	-	15	30	32	12	18	13	15	8	9	19	21
Single_cell_15µL_2	-	-	-	17	17	-	-	-	-	12	-	30	-	-	-	-	-	-	-	-	-
Single_cell_15µL_3	X	Y	-	-	-	8	12	-	-	-	-	-	-	-	18	-	-	8	-	-	-
Single_cell_20µL_1	X	-	-	-	17	-	-	-	-	12	-	-	-	-	-	-	-	-	-	-	-
Single_cell_20µL_2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Single_cell_20µL_3	X	Y	16	17	17	-	12	16	-	12	-	30	32	-	-	13	15	8	9	-	21
Single_cell_25µL_1	X	-	16	-	17	-	-	16	20	12	15	30	32	12	18	13	15	8	9	-	-
Single_cell_25µL_2	X	Y	-	-	17	-	-	-	-	-	-	-	-	12	-	-	15	-	9	-	21
Single_cell_25µL_3	-	-	-	17	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

8.5 Single-cell overall results summary data (Figure 3.7)

Table 8-3 Summary data for the number of alleles amplified and profile type obtained for single-cell tests using a standard CE injection time

Standard CE injection time				Standard CE injection time				Standard CE injection time			
Profile type	Number of alleles present	Total number of alleles possible	Alleles amplified (%)	Profile type	Number of alleles present	Total number of alleles possible	Alleles amplified (%)	Profile type	Number of alleles present	Total number of alleles possible	Alleles amplified (%)
NP	0	19	0	P	5	19	26	P	11	19	58
NP	0	19	0	P	13	19	68	P	11	19	58
NP	0	19	0	NP	0	19	0	P	10	19	53
P	1	19	5	P	2	19	11	P	14	19	74
P	2	19	11	P	8	19	42	P	0	19	0
P	1	19	5	P	3	19	16	NP	0	19	0
P	13	19	68	P	9	19	47	P	9	19	47
NP	0	19	0	P	4	19	21	NP	0	19	0
P	1	19	5	NP	0	19	0	P	13	19	68
P	5	19	26	P	1	19	5	P	4	19	21
P	4	19	21	P	10	19	53	P	3	19	16
NP	0	19	0	P	2	19	11	P	13	19	68
P	9	19	47	P	10	19	53	P	16	19	84
P	5	19	26	P	11	19	58	P	5	19	26
P	7	19	37	P	13	19	68	P	17	19	89
NP	0	19	0	P	2	19	11				
NP	0	19	0	P	3	19	16				

8.6 Summary data for alleles amplified at each locus (Figure 3.8)

Table 8-4 Summary data for the number of alleles amplified at each of the heterozygous loci

	AMEL	D19	D3	D8	THO1	D21	FGA	D16	D18	D2
Total number of observed alleles at each locus	42	41	41	35	28	21	23	24	21	11
Total number of alleles possible	98	98	98	98	98	98	98	98	98	98
Percentage of alleles amplified for each locus (standard injection) (%)	43	42	42	36	29	21	23	24	21	11
Percentage of alleles amplified for each locus (increased injection) (%)	45	44	42	39	31	24	26	27	23	11

8.7 Summary data for relationship between fragment size and allele peak height (Figure 3.9)

Table 8-5 Peak height (Ht in RFU) and fragment size (Fs in bp) for each amplified peak at all locus for each cell, using a standard CE injection time

	D3				D16				D2				AMEL				D8			
Id	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**
5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	79	152.48
6	35	126.18	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
7	200	126.60	349	130.64	-	-	142	256.73	-	-	132	312.33	666	102.79	447	108.37	82	137.73	43	152.77
9	443	126.83	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
10	721	126.78	-	-	-	-	-	-	-	-	-	-	250	102.95	-	-	146	139.50	419	153.36
11	169	126.67	-	-	-	-	-	-	-	-	-	-	-	-	130	108.48	-	-	-	-
13	395	126.72	118	130.87	-	-	145	257.02	-	-	108	312.52	656	102.97	415	108.49	-	-	236	152.58
14	248	126.69	-	-	-	-	-	-	-	-	-	-	244	102.95	230	108.46	152	139.30	-	-
15	467	126.52	-	-	-	-	-	-	-	-	-	-	579	102.80	187	108.40	-	-	-	-

	D21				D18				D19				THO1				FGA			
Id	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**
5	-	-	-	-	-	-	-	-	-	-	-	-	266	174.33	-	-	-	-	-	-
6	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
7	299	208.16	371	216.02	-	-	207	308.85	261	116.24	384	124.17	-	-	-	-	270	220.07	-	-
9	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
10	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	168	220.49	-	-
11	-	-	90	216.09	-	-	-	-	27	124.15	-	-	-	-	-	-	-	-	-	-
13	-	-	-	-	-	-	-	-	272	116.32	179	124.27	-	-	288	182.41	-	-	-	-
14	-	-	-	-	-	-	66	309.21	200	116.34	-	-	500	178.48	-	-	-	-	-	-
15	181	208.12	-	-	-	-	-	-	164	116.30	302	124.17	-	-	364	182.36	-	-	294	228.34

	D3				D16				D2				AMEL				D8			
Id	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**
19	72	126.00	-	-	-	-	-	-	-	-	-	-	150	102.66	230	108.30	-	-	-	-
20	54	126.47	37	130.52	391	240.4	111	256.56	188	294.56	-	-	270	102.73	367	108.26	-	-	255	152.87
22	-	-	-	-	-	-	-	-	-	-	-	-	247	102.6	-	-	-	-	122	152.70
23	135	126.44	106	130.53	-	-	81	256.41	-	-	-	-	159	102.59	169	108.16	124	139.00	-	-
24	-	-	-	-	-	-	-	-	-	-	-	-	-	-	253	108.21	-	-	-	-
25	417	126.48	155	130.50	-	-	-	-	-	-	-	-	284	102.6	312	108.12	144	139	87	152.82
26	680	126.49	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	132	152.73
28	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
29	264	126.48	-	-	-	-	220	256.63	-	-	102	312.03	267	102.61	279	108.25	-	-	-	-
30	-	-	-	-	-	-	-	-	-	-	-	-	279	102.69	205	108.32	125	139.1	-	-

	D21				D18				D19				THO1				FGA			
Id	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**
19	-	-	-	-	-	-	-	-	335	116.32	-	-	88	178.18	323	182.22	-	-	-	-
20	276	208.01	-	-	-	-	106	308.64	194	116.25	110	124.20	-	-	307	182.22	100	219.99	-	-
22	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
23	-	-	-	-	89	282.51	-	-	151	116.22	-	-	-	-	109	182.20	-	-	277	227.90
24	-	-	-	-	-	-	-	-	-	-	159	124.15	-	-	-	-	-	-	124	227.92
25	-	-	-	-	188	282.45	-	-	213	116.21	351	124.19	-	-	317	182.48	-	-	127	227.96
26	-	-	-	-	-	-	-	-	708	116.32	-	-	-	-	-	-	-	-	-	-
28	-	-	-	-	-	-	-	-	43	116.21	-	-	-	-	-	-	-	-	-	-
29	240	208.05	427	216.00	-	-	106	308.28	-	-	382	124.26	346	178.36	262	182.27	-	-	-	-
30	-	-	193	215.97	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

	D3				D16				D2				AMEL				D8			
Id	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**
31	-	-	-	-	84	240.32	249	256.5	-	-	96	312.13	-	-	388	108.27	151	139.00	141	152.72
32	201	126.55	180	130.67	48	240.49	187	256.60	-	-	-	-	-	-	-	-	100	139.10	-	-
33	445	126.54	526	130.60	-	-	174	256.69	182	294.56	-	-	310	102.71	353	108.30	-	-	111	152.75
34	-	-	-	-	-	-	-	-	-	-	-	-	303	102.61	-	-	-	-	105	152.42
35	355	126.42	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	151	152.73
36	376	126.68	676	130.72	-	-	125	256.78	-	-	140	312.53	600	102.78	656	108.43	-	-	216	152.42
37	498	126.65	385	130.77	-	-	-	-	256	294.73	76	312.34	-	-	258	108.40	-	-	170	153.02
38	409	126.58	609	130.69	-	-	-	-	-	-	-	-	-	-	-	-	126	139.20	101	153.00
39	519	126.56	735	130.64	499	240.64	464	256.77	-	-	-	-	-	-	-	-	414	139.20	-	-
40	-	-	-	-	-	-	-	-	-	-	-	-	233	102.85	-	-	-	-	-	-

	D21				D18				D19				THO1				FGA			
Id	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**
31	-	-	125	215.98	-	-	-	-	377	124.22	-	-	-	-	309	182.33	101	219.81	-	-
32	-	-	-	-	-	-	115	308.41	227	116.32	-	-	374	178.26	-	-	134	219.94	157	227.99
33	-	-	216	216.01	261	282.62	105	308.67	-	-	305	124.22	229	178.31	613	182.35	-	-	350	228.00
34	-	-	-	-	-	-	-	-	-	-	170	124.08	-	-	-	-	-	-	-	-
35	-	-	-	-	-	-	-	-	-	-	242	124.20	-	-	-	-	-	-	-	-
36	314	208.13	-	-	-	-	-	-	414	116.33	538	124.28	-	-	267	182.37	147	220.10	-	-
37	-	-	-	-	175	282.83	51	308.71	395	116.27	353	124.26	-	-	-	-	-	-	229	228.26
38	-	-	-	-	-	-	79	309.04	447	116.15	260	124.20	297	178.35	399	182.34	-	-	-	-
39	265	208.16	-	-	-	-	117	308.87	92	116.26	276	124.27	447	178.38	250	182.44	88	220.12	292	228.29
40	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

	D3				D16				D2				AMEL				D8			
Id	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**
41	-	-	-	-	-	-	-	-	-	-	-	-	398	102.83	-	-	-	-	-	-
42	-	-	151	130.73	142	240.65	117	256.77	-	-	90	312.46	-	-	-	-	70	139.20	53	153.10
44	-	-	-	-	108	240.60	368	256.76	53	294.65	-	-	287	102.80	1248	108.40	110	139.16	-	-
45	548	126.56	315	130.61	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
46	-	-	-	-	-	-	-	-	-	-	-	-	343	102.55	-	-	-	-	-	-
47	266	126.44	-	-	265	240.47	477	256.57	-	-	-	-	521	102.57	808	108.19	344	139.00	53	152.48
48	464	123.51	259	130.64	294	240.36	129	256.42	-	-	-	-	511	102.64	489	108.25	112	138.54	224	152.77
49	278	126.40	-	-	-	-	-	-	-	-	-	-	-	-	-	-	147	138.91	-	-
50	252	126.56	593	130.61	1260	240.48	1109	256.61	-	-	-	-	1072	102.66	981	108.29	588	139.00	296	152.87

	D21				D18				D19				THO1				FGA			
Id	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**
41	-	-	-	-	-	-	-	-	-	-	-	-	149	178.32	-	-	-	-	-	-
42	83	208.04	35	215.88	34	282.72	-	-	-	-	-	-	-	-	-	-	93	220.17	-	-
44	262	207.98	269	215.87	198	282.64	-	-	598	116.29	156	124.08	463	178.29	356	182.37	-	-	87	227.89
45	256	207.99	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	102	227.95
46	-	-	179	215.93	-	-	-	-	315	116.22	-	-	-	-	-	-	-	-	-	-
47	260	208.08	190	215.97	165	282.54	-	-	403	116.19	339	124.17	305	178.24	228	182.27	-	-	-	-
48	285	207.96	-	-	340	282.55	78	308.47	532	116.22	295	124.16	632	178.23	294	182.33	386	219.87	281	227.95
49	-	-	-	-	-	-	143	308.52	-	-	223	124.15	-	-	-	-	217	219.87	-	-
50	1166	208.01	798	216.01	136	282.62	268	308.56	715	116.31	208	124.29	2325	178.24	373	182.18	680	219.90	481	228.01

Table 8-6 Peak height (Ht in RFU) and fragment size (Fs in bp) for each amplified peak at all locus for each cell, using a increased CE injection time

	D3				D16				D2				AMEL				D8			
Id	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**
5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	271	152.24
6	114	126.25	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
7	574	125.36	1286	129.53	-	-	437	256.6	-	-	366	311.98	937	102.48	913	107.99	373	138.18	157	152.22
8	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	41	140.21	-	-
9	740	125.53	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
10	1152	125.53	-	-	-	-	-	-	-	-	-	-	403	102.74	-	-	374	138.17	911	152.20
11	198	125.48	-	-	-	-	-	-	-	-	-	-	-	-	361	108.06	-	-	-	-
13	1102	125.49	419	129.6	-	-	477	256.51	-	-	350	311.97	1102	102.48	1287	108.06	-	-	1038	152.00
14	606	125.49	-	-	-	-	-	-	-	-	-	-	494	102.48	462	108.06	430	138.18	-	-
15	1228	125.51	-	-	-	-	-	-	-	-	-	-	1211	102.37	376	107.98	-	-	-	-

	D21				D18				D19				THO1				FGA			
Id	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**
5	-	-	-	-	-	-	-	-	-	-	-	-	850	174.09	-	-	-	-	-	-
6	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
7	902	208.01	1117	215.92	-	-	626	308.82	687	115.03	1043	123.06	-	-	-	-	802	219.72	-	-
8	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
9	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
10	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	527	220.11	-	-
11	-	-	286	215.96	-	-	-	-	92	124.10	-	-	-	-	-	-	-	-	-	-
13	-	-	-	-	-	-	-	-	804	116.20	443	124.02	-	-	1030	182.28	84	219.79	-	-
14	-	-	-	-	-	-	233	308.44	606	116.10	-	-	1827	178.23	-	-	-	-	-	-
15	565	207.95	-	-	-	-	-	-	661	116.19	1154	124.13	-	-	1273	182.13	-	-	1045	227.67

	D3				D16				D2				AMEL				D8			
Id	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**
19	133	125.99	-	-	-	-	-	-	-	-	-	-	328	102.32	474	107.88	-	-	-	-
20	73	125.80	134	129.85	1367	240.17	406	256.37	660	294.25	-	-	601	102.38	821	107.90	-	-	975	152.17
22	-	-	-	-	-	-	-	-	-	-	-	-	498	102.35	-	-	-	-	418	152.17
23	444	125.51	418	129.65	-	-	347	256.44	-	-	-	-	428	102.42	459	108.02	606	138.22	-	-
24	-	-	-	-	-	-	-	-	-	-	-	-	-	-	804	108.01	-	-	-	-
25	1237	125.59	494	129.72	-	-	-	-	-	-	-	-	569	102.37	498	107.99	552	138.24	335	152.18
26	1637	125.60	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	478	152.28
28	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
29	701	125.63	-	-	-	-	782	256.32	-	-	333	311.71	1072	102.29	1087	107.86	-	-	-	-
30	-	-	-	-	-	-	-	-	-	-	-	-	1045	102.27	776	107.89	470	138.27	-	-

	D21				D18				D19				THO1				FGA			
	Ht	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**
19	-	-	-	-	-	-	-	-	1343	116.22	-	-	322	178.14	1134	182.16	-	-	-	-
20	933	207.90	-	-	-	-	375	308.13	813	116.18	446	124.13	-	-	1118	182.04	367	219.57	50	227.57
22	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
23	-	-	-	-	406	282.54	-	-	603	116.11	-	-	-	-	519	182.25	-	-	1228	227.79
24	-	-	-	-	-	-	-	-	-	-	507	124.09	-	-	-	-	-	-	437	227.74
25	-	-	-	-	627	282.44	36	308.25	532	116.11	925	124.15	-	-	1071	182.11	37	219.56	398	227.68
26	-	-	-	-	-	-	-	-	1663	116.16	-	-	-	-	-	-	-	-	-	-
28	-	-	-	-	-	-	-	-	115	116.07	-	-	-	-	-	-	-	-	-	-
29	798	207.86	1412	215.79	-	-	379	308.09	-	-	923	123.93	1162	178.02	847	181.98	-	-	-	-
30	-	-	613	215.81	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

	D3				D16				D2				AMEL				D8			
Id	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**
31	-	-	-	-	288	240.26	878	256.32	-	-	290	311.72	-	-	1237	107.93	649	138.28	560	152.18
32	559	125.68	631	129.80	134	240.27	592	256	-	-	-	-	-	-	-	-	382	138.29	-	-
33	1245	125.70	1620	129.78	-	-	591	256.38	555	294.21	-	-	1199	102.24	1402	107.86	-	-	471	152.21
34	-	-	-	-	-	-	-	-	-	-	-	-	1373	102.23	-	-	-	-	524	152.23
35	903	125.64	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	556	152.26
36	1048	125.67	2042	129.83	-	-	411	256.36	-	-	407	311.68	2200	102.26	2497	107.9	163	138.28	930	152.16
37	850	125.66	1252	129.80	-	-	-	-	777	194.20	225	311.83	-	-	821	107.89			742	152.30
38	1164	125.68	1863	129.77	-	-	-	-	-	-	-	-	-	-	-	-	553	138.29	458	152.19
39	931	125.60	1477	129.71	1763	240.08	1640	256.31	-	-	-	-	-	-	-	-	950	138.29	-	-
40	-	-	-	-	-	-	-	-	-	-	-	-	895	102.31	-	-	-	-	-	-

	D21				D18				D19				THO1				FGA			
Id	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**
31	-	-	392	215.83	-	-	-	-	904	123.89	-	-	-	-	1000	182.05	339	219.36	-	-
32	-	-	-	-	-	-	375	308.19	491	115.70	-	-	1178	177.93	-	-	399	219.34	447	227.49
33	-	-	678	215.74	895	282.39	379	308	-	-	672	123.64	753	177.96	2000	181.92	-	-	1079	227.38
34	-	-	-	-	-	-	-	-	-	-	476	123.61	-	-	-	-	-	-	-	-
35	-	-	-	-	-	-	-	-	-	-	476	123.62	-	-	-	-	-	-	-	-
36	927	207.83	-	-	-	-	-	-	882	115.65	1175	123.60	-	-	790	182.06	428	219.35	-	-
37	-	-	-	-	615	282.46	189	308	874	115.57	843	123.69	-	-	-	-	-	-	784	227.51
38	-	-	-	-	-	-	297	308.33	1054	115.41	617	123.51	1015	177.98	1307	181.96	-	-	-	-
39	830	207.87	-	-	-	-	397	308.33	244	115.52	732	123.5	1553	178.03	841	182.02	305	219.36	954	227.5
40	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

	D3				D16				D2				AMEL				D8			
Id	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**
41	-	-	-	-	-	-	-	-	-	-	-	-	1737	102.29	-	-	-	-	-	-
42	-	-	485	129.80	522	240.35	453	256.54	45	294.25	319	311.87	-	-	-	-	297	138.37	234	152.35
44	-	-	-	-	325	240.16	1068	256.38	483	294.24	-	-	919	102.26	3980	107.88	229	138.22	-	-
45	1681	125.68	1053	129.78	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
46	-	-	-	-	-	-	-	-	-	-	-	-	1163	102.32	-	-	-	-	-	-
47	639	125.65	-	-	782	240.39	1414	256.51	-	-	-	-	1845	102.41	2885	108.01	1266	138.27	163	152.36
48	1196	125.78	708	129.92	887	240.33	400	256.51	-	-	-	-	1850	102.40	1779	108.04	400	138.37	770	152.28
49	755	125.71	-	-	-	-	-	-	-	-	-	-	-	-	-	-	553	138.38	-	-
50	538	125.78	1471	129.83	3317	240.41	2879	256.58	-	-	-	-	3368	102.37	3205	107.03	1972	138.36	973	152.30

	D21				D18				D19				THO1				FGA			
Id	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**	**Ht**	**Fs**
41	-	-	-	-	-	-	-	-	-	-	-	-	517	177.93	-	-	-	-	-	-
42	284	207.94	123	215.87	130	282.51	71	308.31	-	-	59	123.49	-	-	-	-	323	219.46	-	-
44	788	207.87	803	215.82	676	282.47	-	-	1916	115.41	504	123.45	1391	177.94	1070	181.98	-	-	266	227.54
45	923	207.89	60	215.86	-	-	-	-	-	-	-	-	-	-	-	-	-	-	368	227.54
46	-	-	497	215.91	-	-	-	-	677	115.44	-	-	-	-	-	-	-	-	-	-
47	754	207.95	550	215.97	508	282.58	-	-	844	115.45	737	123.48	919	178.00	667	181.99	-	-	-	-
48	865	207.89	-	-	1114	282.59	238	308.51	1176	115.21	659	123.55	1886	178.02	878	182.05	1121	219.45	862	227.6
49	-	-	-	-	-	-	510	308.56	-	-	502	123.44	-	-	-	-	648	219.50	-	-
50	2857	208.02	2012	215.94	374	282.61	742	308.46	1427	115.48	432	123.51	6010	178.02	975	181.97	1646	219.53	1180	227.72

8.8 Summary of heterozygous balance data (Figure 3.10 and Figure 3.11)

Table 8-7 Heterozygous balance values for loci with both alleles present using a standard CE injection time. Hb = LMW area/ HMW area

Id	D3	D16	D2	AMEL	D8	D21	D18	D19	THO1	FGA
7	0.60	-	-	0.68	0.58	0.78	-	0.64	-	-
10	-	-	-	-	0.33	-	-	-	-	-
13	-	-	-	0.64	-	-	-	0.67	-	-
14	-	-	-	0.98	-	-	-	-	-	-
15	-	-	-	0.34	-	-	-	0.50	-	-
19	-	-	-	0.59	-	-	-	-	0.27	-
20	0.45	0.29	-	0.71	-	-	-	0.60	-	-
23	0.65	-	-	0.94	-	-	-	-	-	-
24	-	-	-	-	-	-	-	-	-	-
25	0.32	-	-	0.87	0.65	-	-	0.58	-	-
29	-	-	-	0.97	-	0.54	-	-	0.74	-
30	-	-	-	0.74	-	-	-	-	-	-
31	-	0.34	-	-	0.81	-	-	-	-	-
32	0.80	0.30	-	-	-	-	-	-	-	0.83
33	0.87	-	-	0.86	-	-	0.42	-	0.37	-
36	0.59	-	-	0.93	-	-	-	0.75	-	-
37	0.85	-	0.29	-	-	-	0.32	0.96	-	-
38	0.73	-	-	-	0.76	-	-	0.57	0.75	-
39	0.72	0.93	-	-	-	-	-	0.32	0.55	0.29
42	-	0.82	-	-	0.73	0.37	-	-	-	-
44	-	0.29	-	0.23	-	0.96	-	0.26	0.77	-
45	0.54	-	-	-	-	-	-	-	-	-
47	-	0.58	-	0.65	0.14	0.75	-	0.85	0.73	-
48	0.51	0.46	-	0.95	0.56	-	0.23	0.55	0.47	0.76
50	0.46	0.90	-	0.92	0.49	0.69	0.49	0.29	0.16	0.71

Table 8-8 Heterozygous balance values for loci with both alleles present using an increased CE injection time. Hb = LMW area/ HMW area

Id	D3	D16	D2	AMEL	D8	D21	D18	D19	THO1	FGA
7	0.41	-	-	0.97	0.52	0.79	-	0.66	-	-
10	-	-	-	-	0.37	-	-	-	-	-
13	-	-	-	0.85	-	-	-	0.56	-	-
14	-	-	-	0.92	-	-	-	-	-	-
15	-	-	-	0.31	-	-	-	0.56	-	-
19	-	-	-	0.72	-	-	-	-	0.28	-
20	0.54	0.30	-	0.77	-	-	-	0.55	-	0.15
23	0.94	-	-	0.94	-	-	-	-	-	-
24	-	-	-	-	-	-	-	-	-	-
25	0.40	-	-	0.85	0.63	-	-	0.57	-	12.06
29	-	-	-	0.97	-	0.57	-	-	0.74	-
30	-	-	-	0.74	-	-	-	-	-	-
31	-	0.32	-	-	0.93	-	-	-	-	-
32	0.90	0.22	-	-	-	-	-	-	-	0.87
33	0.77	-	-	0.84	-	-	0.44	-	0.37	-
36	0.52	-	-	0.88	0.14	-	-	0.77	-	-
37	0.68	-	0.30	-	-	-	0.30	0.93	-	-
38	0.62	-	-	-	0.94	-	-	0.56	0.76	-
39	0.63	0.94	-	-	-	-	-	0.34	0.54	0.31
40	-	0.86	0.12	-	0.91	0.42	0.54	-	-	-
44	-	0.30	-	0.23	-	0.96	-	0.26	0.77	-
45	0.62	-	-	-	-	0.06	-	-	-	-
47	-	0.55	-	0.65	0.13	0.73	-	0.86	0.73	-
48	0.57	0.46	-	0.96	0.47	-	0.23	0.54	0.47	0.77
50	0.36	0.88	-	0.96	0.50	0.71	0.49	0.30	0.16	0.72

8.9 Summary of alleles observed at each locus (Figure 3.12 and Figure 3.15)

Table 8-9 Summary of alleles amplified at each locus using a standard CE injection time (-) denotes no observed amplification

Id	D3		VWA	D16		D2		**AMEL**		**D8**		**D21**		**D18**		**D19**		**THO1**		**FGA**	
4	-	-	17	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
5	-	-	-	-	-	-	-	-	-	-	15	-	-	-	-	-	-	-	-	-	-
6	16	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
7	16	17	17	-	12	-	20	X	Y	12	15	30	32	-	18	13	15	-	-	19	-
9	16	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
10	16	-	17	-	-	-	-	X	-	12	15	-	-	-	-	-	-	-	-	19	-
11	16	-	17	-	-	-	-	-	Y	-	-	-	32	-	-	-	15	-	-	-	-
13	16	17	17	-	12	-	20	X	Y	-	15	-	-	-	-	13	15	-	9	-	-
14	16	-	-	-	-	-	-	X	Y	12	-	-	-	-	18	13	-	8	-	-	-
15	16	-	17	-	-	-	-	X	Y	-	-	30	-	-	-	13	15	-	9	-	21
19	16	-	17	-	-	-	-	X	Y	-	-	-	-	-	-	13	-	8	9	-	-
20	16	17	17	8	12	16	-	X	Y	-	15	30	-	-	18	13	15	-	9	19	-
22	-	-	17	-	-	-	-	X	-	-	15	-	-	-	-	-	-	-	-	-	-
23	16	17	-	-	12	-	-	X	Y	12	-	-	-	12	-	13	-	-	9	-	21
24	-	-	17	-	-	-	-	-	Y	-	-	-	-	-	-	-	15	-	-	-	21
25	16	17	-	-	-	-	-	X	Y	12	15	-	-	12	-	13	15	-	9	-	21
26	16	-	17	-	-	-	-	-	-	-	15	-	-	-	-	13	-	-	-	-	-
28	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	13	-	-	-	-	-
29	16	-	17	-	12		20	X	Y	-	-	30	32	-	18	-	15	8	9	-	-
30	-	-	-	-	-	-	-	X	Y	12	-	-	32	-	-	-	-	-	-	-	-

Id	D3		VWA	D16		D2		AMEL		D8		D21		D18		D19		THO1		FGA	
31	-	-	17	8	12	-	20	-	Y	12	15	-	32	-	-	-	15	-	9	19	-
32	16	17	17	8	12	-	-	-	-	12	-	-	-	-	18	13	-	8	-	19	21
33	16	17	17	-	12	16	-	X	Y	-	15	-	32	12	18	-	15	8	9	-	21
34	-	-	-	-	-	-	-	X	-	-	15	-	-	-	-	-	15	-	-	-	-
35	16	-	-	-	-	-	-	-	-	-	15	-	-	-	-	-	15	-	-	-	-
36	16	17	17	-	12	-	20	X	Y	-	15	30		-	-	13	15	-	9	19	-
37	16	17	17	-	-	16	20	-	Y	-	15	-	-	12	18	13	15	-	-	-	21
38	16	17	17	-	-	-	-	-	-	12	15	-	-	-	18	13	15	8	9	-	-
39	16	17	17	8	12	-	-	-	-	12	-	30	-	-	18	13	15	8	9	19	21
41	-	-	-	-	-	-	-	X	-	-	-	-	-	-	-	-	-	8	-	-	-
42	-	17	17	8	12	-	20	-	-	12	15	30	32	12	-	-	-	-	-	19	-
44	-	-	17	8	12	16	-	X	Y	12	-	30	32	12	-	13	15	8	9	-	21
45	16	17	-	-	-	-	-	-	-	-	-	30	-	-	-	-	-	-	-	-	21
46	-	-	17	-	-	-	-	X	-	-	-	-	32	-	-	13	-	-	-	-	-
47	16	-	17	8	12	-	-	X	Y	12	15	30	32	12	-	13	15	8	9	-	-
48	16	17	17	8	12	-	-	X	Y	12	15	-	-	12	18	13	15	8	9	19	21
49	16	-	-	-	-	-	-	-	-	12	-	-	-	-	18	-	15	-	-	19	-
50	16	17	17	8	12	-	-	X	Y	12	15	-	32	12	18	13	15	8	9	19	21

Table 8-10 Summary of alleles amplified at each locus using an increased CE injection time (-) denotes no observed amplification

Id	D3		VWA	D16		D2		AMEL		D8		D21		D18		D19		THO1		FGA	
4	-	-	17	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
5	-	-	-	-	-	-	-	-	-	-	15	-	-	-	-	-	-	-	-	-	-
6	16	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
7	16	17	17	-	12	-	20	X	Y	12	15	30	32	-	18	13	15	-	-	19	-
8	-	-	-	-	-	-	-	-	-	12	-	-	-	-	-	-	-	-	-	-	-
9	16	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
10	16	-	17	-	-	-	-	X	-	12	15	-	-	-	-	-	-	-	-	19	-
11	16	-	17	-	-	-	-	-	Y	-	-	-	32	-	-	-	15	-	-	-	-
13	16	17	17	-	12	-	20	X	Y	-	15	-	-	-	-	13	15	-	9	19	-
14	16	-	-	-	-	-	-	X	Y	12	-	-	-	-	18	13	-	8	-	-	-
15	16	-	17	-	-	-	-	X	Y	-	-	30	-	-	-	13	15	-	9	-	21
19	16	-	17	-	-	-	-	X	Y	-	-	-	-	-	-	13	-	8	9	-	-
20	16	17	17	8	12	16	-	X	Y	-	15	30	-	-	18	13	15	-	9	19	21
22	-	-	17	-	-	-	-	X	-	-	15	-	-	-	-	-	-	-	-	-	-
23	16	17	-	-	12	-	-	X	Y	12	-	-	-	12	-	13	-	-	9	-	21
24	-	-	17	-	-	-	-	-	Y	-	-	-	-	-	-	-	15	-	-	-	21
25	16	17	-	-	-	-	-	X	Y	12	15	-	-	12	18	13	15	-	9	19	21
26	16	-	17	-	-	-	-	-	-	-	15	-	-	-	-	13	-	-	-	-	-
28	-	-	-	--	-	-	-	-	-	-	-	-	-	-	-	13	-	-	-	-	-
29	16	-	17	-	12	-	2	X	Y	-	-	30	32	-	18	-	15	8	9	-	-
30	-	-	-	-	-	-	-	X	Y	12	-	-	32	-	-	-	-	-	-	-	-

Id	D3		VWA	D16		D2		AMEL		D8		D21		D18		D19		THO1		FGA	
31	-	-	17	8	12	-	20	-	Y	12	15	-	32	-	-	-	15	-	9	19	-
32	16	17	17	8	12	-	-	-	-	12	-	-	-	-	18	13	-	8	-	19	21
33	16	17	17	-	12	16	-	X	Y	-	15	-	32	12	18	-	15	8	9	-	21
34	-	-	-	-	-	-	-	X	-	-	15	-	-	-	-	-	15	-	-	-	-
35	16	-	-	-	-	-	-	-	-	-	15	-	-	-	-	-	15	-	-	-	-
36	16	17	17	-	12	-	20	X	Y	12	15	30	-	-	-	13	15	-	9	19	-
37	16	17	17	-	-	16	20	-	Y	-	15	-	-	12	18	13	15	-	-	-	21
38	16	17	17	-	-	-	-	-	-	12	15	-	-	-	18	13	15	8	9	-	-
39	16	17	17	8	12	-	-	-	-	12	-	30	-	-	18	13	15	8	9	19	21
40	-	-	-	-	-	-	-	X	-	-	-	-	-	-	-	-	-	-	-	-	-
41	-	-	-	-	-	-	-	X	-	-	-	-	-	-	-	-	-	8	-	-	-
42	-	17	17	8	12	16	20	-	-	12	15	30	32	12	18	-	15	-	-	19	-
44	-	-	17	8	12	16	-	X	Y	12	-	30	32	12	-	13	15	8	9	-	21
45	16	17	-	-	-	-	-	-	-	-	-	30	32	-	-	-	-	-	-	-	21
46	-	-	17	-	-	-	-	X	-	-	-	-	32	-	-	13	-	-	-	-	-
47	16	-	17	8	12	-	-	X	Y	12	15	30	32	12	-	13	15	8	9	-	-
48	16	17	17	8	12	-	-	X	Y	12	15	-	-	12	18	13	15	8	9	19	21
49	16	-	-	-	-	-	-	-	-	12	-	-	-	-	18	-	15	-	-	19	-
50	16	17	17	8	12	-	-	X	Y	12	15	-	32	12	18	13	15	8	9	19	21

8.10 Summary of stutter (%) observed at each locus (Figure 3.16 and Figure 3.17 and Figure 3.18)

Table 8-11 Percentage of stutter observed for each locus per sample using a standard CE injection time

Id	D3		VWA	D16		D8		D21		D18		D19		THO1		FGA	
7	-	15	-	-	-	-	-	12	8	-	13	23	11	-	-	-	-
10	-	6	8	-	-	-	-	-	-	-	-	-	-	-	-	-	-
13	-	5	8	-	-	-	-	-	-	-	-	18	-	-	-	-	-
15	-	7	12	-	-	-	-	11	-	-	-	-	11	-	-	-	9
20	-	-	6	-	-	-	-	-	-	-	-	-	-	-	-	-	-
25	-	9	-	-	-	-	-	-	-	-	-	-	12	-	-	-	-
26	-	13	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
29	-	13	10	-	12	-	-	-	16	-	-	-	-	-	-	-	-
31	-	-	7	-	-	-	-	-	-	-	-	29	-	-	-	-	-
32	-	-	9	-	-	-	-	-	-	-	-	-	-	-	-	-	-
33	-	6	7	-	-	-	-	-	11	-	-	-	-	-	-	-	8
35	-	-	-	-	-	-	-	-	-	-	-	-	15	-	-	-	-
36	-	12	8	-	-	-	-	16	-	-	-	-	9	-	-	-	-
37	-	19	-	-	-	-	-	-	-	-	-	-	10	-	-	-	10
38	-	4	14	-	-	-	-	-	-	-	-	-	11	-	-	-	-
39	-	-	8	-	6	-	-	-	-	-	33	-	-	-	-	-	-
42	-	-	7	-		-	-	-	-	-	-	-	-	-	-	-	-
44	-	-	9	-	7	-	-	-	15	-	-	8	20	-	-	-	-
45	-	3	-	-	-	-	-	-	-	-	-	-	-	-	-	--	-
46	-	-	7	-	-	-	-	-	-	-	-	-	-	-	-	-	-
47	-	-	7	-	9	8	-	-	16	-	-	-	11	-	-	-	-
48	-	6	10	-	-	-	-	7	-	-	-	-	9	-	-	7	8
50	-	-	7	3	6	9	-	8	14	-	13	-	-	3	-	5	7

Table 8-12 Percentage of stutter observed for each locus per sample using an increased CE injection time

Id	D3		VWA	D16		D8		D21		D18		D19		THO1		FGA	
7	-	21	-	-	-	-	-	10	9	-	13	19	10	-	-	-	-
10	-	2	7	-	-	-	-	-	-	-	-	-	-	-	-	-	-
13	-	29	12	-	-	-	-	-	-	-	-	2	-	-	-	-	-
15	-	20	14	-	-	-	-	9	-	-	-	-	13	-	-	-	8
20	-	-	7	-	-	-	-	-	-	-	-	-	-	-	-	-	-
25	-	6	-	-	-	-	-	-	-	-	-	-	10	-	-	-	-
26	-	13	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
29	-	13	11	-	11	-	-	-	14	-	-	-	-	-	-	-	-
31	-	-	6	-	-	-	-	-	-	-	-	28	-	-	-	-	-
32	-	-	9	-	-	-	-	-	-	-	-	-	-	-	-	-	-
33	-	4	6	-	-	-	-	-	10	-	-	-	-	-	-	-	8
35	-	-	-	-	-	-	-	-	-	-	-	-	15	-	-	-	-
36	-	10	7	-	-	-	-	16	-	-	-	-	10	-	-	-	-
37	-	24	-	-	-	-	-	-	-	-	-	-	11	-	-	-	10
38	-	3	13	-	-	-	-	-	-	-	-	-	10	-	-	-	-
39	-	-	8	-	7	-	-	-	-	-	30	-	-	-	-	-	-
42	-	-	7	-	-	-	-	-	-	-	-	-	-	-	-	-	-
44	-	-	9	-	6	-	-	-	14	-	-	8	20	-	-	-	-
45	-	3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
46	-	-	7	-	-	-	-	-	-	-	-	-	-	-	-	-	-
47	-	-	8	-	9	7	-	-	14	-	-	-	11	-	-	-	-
48	-	3	8	-	-	-	-	6	-	-	-	-	9	-	-	5	7
50	-	-	6	2	6	8	-	7	13	-	12	-	-	3	-	5	6

8.11 Summary of observed amplified alleles for assay trial (Figure 3.19)

Table 8-13 DNA profiles from single-cells (SC) amplified using the Promega ESI 16 assay

ESI	AMEL		D3		D19		D2S1338		D22		D16		D18		D1		D10		D2S441	THO1		vWA	D21		D12		D8		FGA	
ESI_SC_b	-	-	16	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ESI_SC_c	X	Y	16	17	-	-	-	-	-	-	-	12	-	-	11	-	-	-	-	8	9	17	30	32	-	-	-	-	-	-
ESI_SC_d	-	Y	16	17	-	15	-	-	-	-	-	-	12	18	-	-	-	-	11	8	-	-	-	-	-	23	-	-	-	-
ESI_SC_e	-	-	-	-	13	-	-	-	-	16	-	-	-	-	-	-	-	-	-	-	-	17	-	-	-	-	-	-	-	-
ESI_SC_f	-	-	-	-	-	-	-	-	13	-	-	-	-	-	-	-	-	-	-	-	-	17	-	-	-	-	-	-	-	-
ESI_SC_g	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ESI_SC_h	-	Y	-	-	-	-	-	-	-	-	8	12	12	18	-	-	-	-	-	-	9	17	-	-	-	-	-	-	19	21
ESI_SC_i	X	Y	16	-	13	15	16	-	-	-	8	12	12	18	11	-	-	16	11	8	9	17	-	-	16	23	-	-	19	21
ESI_SC_j	-	Y	16	17	13	-	-	-	-	-	8	12	-	-	11	15	-	-	-	8	9	17	-	32	-	-	-	-	19	21

Table 8-14 DNA profiles from single-cells (SC) amplified using the Promega ESX 16 assay

ESX	AMEL		D3		THO1		D21		D18		D10		D1		D2S1338		D16		D22		vWA	D8S1179		FGA		D2S441	D12		D19	
ESX_SC_a	X	Y	16	17	8	9	30	-	12	18	13	16	11	15	16	20	8	12	13	16	17	12	15	19	21	11	16	23	-	15
ESX_SC_b	X	Y	16	17	8	9	30	32	12	18	13	16	11	15	16	20	8	12	13	16	17	12	15	19	21	11	-	23	13	15
ESX_SC_c	X	Y	16	17	8	9	-	-	12	-	-	16	11	-	16	-	-	-	13	16	17	-	-	19	21	11	-	23	13	15
ESX_SC_e	X	Y	16	17	8	9	30	32	-	18	13	16	-	15	16	-	8	12	13	16	17	12	-	19	21	11	16	23	13	15
ESX_SC_f	-	Y	16	17	8	9	30	32	-	18	13	16	-	-	16	20	8	12	-	16	17	-	15	-	-	11	-	23	-	-
ESX_SC_g	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ESX_SC_h	-	-	-	17	-	-	30	32	12	18	-	-	11	15	-	-	-	-	13	16	17	-	-	-	21	11	-	23	-	-
ESX_SC_i	X	Y	16	17	8	9	30	32	-	18	13	16	11	15	16	20	8	12	13	16	17	12	15	19	-	11	16	23	13	15
ESX_SC_j	X	Y	16	-	8	-	30	32	12	18	13	16	11	-	-	20	8	12	13	-	17	-	15	19	-	11	16	-	13	15

Table 8-15 DNA profiles from single-cells (SC) amplified using the Applied Biosystems SGM Plus assay

SGM PLUS	D3		VWA		D16		D2		AMEL	D8		D21		D18		D19		THO1		FGA	
SGM_SC_a	15	16	18	19	-	-	-	23	X	10	11	28	30	14	20	14	15	-	9	22	24
SGM_SC_b	15	16	-	19	11	14	-	23	X	-	11	-	-	-	20	14	15	-	9	-	-
SGM_SC_c	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
SGM_SC_d	-	16	18	-	-	14	-	23	X	10	11	-	30	14	-	-	15	-	9	-	24
SGM_SC_e	15	16	-	-	11	14	-	-	X	-	11	-	-	14	-	14	-	7	9	22	24
SGM_SC_f	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
SGM_SC_g	-	-	-	19	-	-	-	-	X	-	-	-	30	-	-	14	-	7	9	22	-
SGM_SC_h	-	-	18	-	-	-	-	-	X	-	-	-	-	-	-	-	15	-	-	-	-
SGM_SC_i	-	-	18	-	-	-	-	23	X	-	11	-	-	-	-	14	15	-	-	-	-
SGM_SC_j	15	16	-	-	11	-	19	23	-	-	11	28	-	-	20	-	15	7	9	22	-

Table 8-16 DNA profiles from single-cells (SC) amplified using the Applied Biosystems NGM assay

NGM	D10	VWA		D16		D2		AMEL	D8		D21		D18		D22		D19		THO1		FGA		D2S441		D3		D1S1656		D12S391	
NGM_SC_a	14	18	19	-	14	-	-	X	10	11	28	-	-	-	15	17	14	-	7	9	-	24	12	14	15	-	11	17	18	-
NGM_SC_b	14	-	19	-	14	19	-	X	10	11	-	-	14	20	15	17	14	15	7	9	22.2	-	12	14	15	-	11	17	18	-
NGM_SC_c	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
NGM_SC_d	-	-	-	-	-	-	-	X	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
NGM_SC_e	14	18	-	11	-	19	23	X	10	-	28	30	-	-	-	17	-	15	7	9	-	-	12	14	15	-	-	17	18	19
NGM_SC_f	-	18	19	-	-	-	23	X	10	11	28	-	-	-	15	-	14	-	-	9	22.2	-	12	14	15	16	-	-	-	19
NGM_SC_g	14	-	-	-	14	-	-	X	10	11	-	30	14	-	-	-	-	-	7	-	-	24	12	14	-	16	-	-	-	-
NGM_SC_h	-	18	-	-	14	-	-	-	10	11	-	30	-	-	15	-	-	-	-	-	-	-	-	14	15	-	11	-	18	19
NGM_SC_i	-	-	-	-	-	-	-	-	-	-	-	-	-	-	15	17	-	-	-	-	-	-	12	14	-	-	-	-	-	-
NGM_SC_j	14	18	19	-	-	-	23	X	10	11	-	-	-	-	15	17	-	-	-	9	-	24	12	14	-	-	-	-	-	-

Table 8-17 DNA profiles from single-cells (SC) amplified using the 'in-house' ltDNA assay. Alleles defined using size as no ladder available

ltDNA assay	Penta E		D12S391		D10S1428		Amelogenin	
ltDNA_SC_1	-	113.06	162.62	182.61	99.11	-	-	113.24
ltDNA_SC_2	-	-	-	-	99.06	-	-	-
ltDNA_SC_3	103.90	-	-	-	99.19	-	107.73	-
ltDNA_SC_4	-	113.30	-	-	99.20	-	-	-
ltDNA_SC_5	103.68	-	162.41	-	99.03	-	-	-
ltDNA_SC_6	-	-	-	-		-	-	-
ltDNA_SC_7	103.83	-	162.46	-	99.16	-	107.64	-
ltDNA_SC_8	-	-	162.47	182.39	99.12	-	107.68	113.11
ltDNA_SC_9	103.67	113.04	-	-	99.02	-	-	112.98
ltDNA_SC_10	-	-	-	-	99.09	102.96	-	-

8.12 Statistics carried out in Chapter 3

Fisher's exact test for count data was carried out using R version 3.1.2 (2014-10-31) -- "Pumpkin Helmet". Copyright (C) 2014 The R Foundation for Statistical Computing.

Locus dropout

Null hypothesis: extending the CE injection time does not significantly reduce locus dropout

	Locus amplified	Locus dropped out
Standard CE injection time	229	200
Extended CE injection time	231	198

```
> Locus.dropout <- read.csv("E:/Locus dropout.csv")
>   View(Locus.dropout)
> challenge.df = matrix(c(229,200,231,198), nrow = 2)
> fisher.test(challenge.df)
data:  challenge.df
```

p-value = 0.9454

```
alternative hypothesis: true odds ratio is not equal to 1

95 percent confidence interval:

 0.7436232 1.2952325

sample estimates:

odds ratio

  0.981445
```

Allele dropout

Null hypothesis: extending the CE injection time does not significantly reduce allele drop out

	Alleles amplified	Alleles dropped out
Standard CE injection time	311	529
Extended CE injection time	322	518

```
> Allele.dropout <- read.csv("E:/Allele dropout.csv")

>   View(Allele.dropout)

> challenge.df = matrix(c(311,529,322,518), nrow = 2)

> fisher.test(challenge.df)

data:  challenge.df
```

p-value = 0.6146

```
alternative hypothesis: true odds ratio is not equal to 1

95 percent confidence interval:

 0.7725123 1.1578405

sample estimates:

odds ratio

 0.9458113
```

Heterozygous balance

Null hypothesis: extending the injection time does not significantly reduce heterozygous imbalance of peaks

	Balance	Imbalance
Standard CE injection time	42	41
Extended CE injection time	47	47

```
> Hb.Imbalance <- read.csv("E:/Hb Imbalance.csv")

>   View(Hb.Imbalance)

> challenge.df = matrix(c(42,41,47,47), nrow = 2)

> fisher.test(challenge.df)

data:  challenge.df
```

p-value = 1

```
alternative hypothesis: true odds ratio is not equal to 1

95 percent confidence interval:

 0.5441173 1.9287896

sample estimates:

odds ratio

  1.024259
```

Presence of stutter peaks

Null hypothesis: extending the CE injection time does not significantly reduce the number of stutter peaks

	Stutter	No stutter
Standard CE injection time	74	198
Extended CE injection time	110	281

```
> Stutter.peaks <- read.csv("E:/Stutter peaks.csv")

>   View(Stutter.peaks)

> challenge.df = matrix(c(74,198,110,281), nrow = 2)

> fisher.test(challenge.df)

data:  challenge.df
```

p-value = 0.8601

```
alternative hypothesis: true odds ratio is not equal to 1

95 percent confidence interval:

 0.6640974 1.3684288

sample estimates:

odds ratio

 0.9547822
```

Presence of stutter peaks over under the 15% threshold

Null hypothesis: extending the CE injection time does not significantly reduce the peak height to below the 15% threshold

	Under 15% threshold	Over 15% threshold
Standard CE injection time	240	32
Extended CE injection time	259	22

> Stutter.peaks.over.threshold <- read.csv("E:/Stutter peaks over threshold.csv", header=FALSE)

> View(Stutter.peaks.over.threshold)

> challenge.df = matrix(c(240,32,259,22), nrow = 2)

> fisher.test(challenge.df)

data: challenge.df

p-value = 0.1514

alternative hypothesis: true odds ratio is not equal to 1

95 percent confidence interval:

0.3425962 1.1677407

sample estimates:

odds ratio

0.6375896

8.13 Participant information and consent forms

Participant Information sheet

Project title: Visualisation, recovery & DNA profiling of human cells deposited onto a surface through contact with skin

Investigator: Victoria Barlow Email victoria.barlow@northumbria.ac.uk

Principle supervisor: Dr Eleanor Graham Email eleanor.graham@northumbria.ac.uk

In order to decide whether or not you would like to participate in this study, please read all the information provided very carefully and raise any questions that you may have with the investigator.

Project Introduction & Objectives

In forensic science we often use the evidence left at the scene or on a victim/suspect to help identify the individual(s) involved. This evidence is often so small as to be invisible to the naked eye and consequently research has to be done to find ways that allow the scientists to 'see' this evidence and interpret it for use in court. One type of such evidence is fibres which are often transferred between the clothes of the victim, suspect and also to/from the scene (i.e. carpets) and can help investigators link these items together. What fibres currently don't help us work out is who was wearing the clothes or has been in contact with the fibres. Fibres are generally collected by placing adhesive tape over an area and pulling it off which should then have any fibres that have been transferred captured on it, known as fibre taping.

What this project aims to do is look at cells under a microscope that have been transferred to the surface of the fibres and to see if a DNA profile can be obtained from them. Factors that may affect the transfer of cells to fibres will also be investigated, for example whether the chemistry of the fibre or the amount of time skin is in contact with the fibres. By investigating these factors and obtaining a DNA profile from these cells will hopefully mean that investigators can use the evidence in court to help in the conviction of criminals.

We use DNA profiling in forensic science to help identify who a particular biological sample came from. In order to do this we target a number of small regions of DNA that are known to vary between individuals called short tandem repeats (STR's). Each repeat is made up of the same short DNA code, and it is the number of repeats that are present at each target,

which allows us to build up your DNA profile. As DNA is inherited from both mother and father, there are two versions of each target.

For example, DNA target 1 – 6, 9

ATAT ATAT ATAT ATAT ATAT ATAT Number of repeats = 6

ATAT ATAT ATAT ATAT ATAT ATAT ATAT ATAT ATAT Number of repeats = 9

The areas of DNA targeted in this study cannot provide us with any additional information such as hair/eye colour, height etc., and will only be used to check that we get the expected profile result during our tests, so that we can rule out any contamination of the samples.

What you will be required to do

You are invited to take part in this study by volunteering to provide samples of cells transferred from the surface of your skin using the method outlined below. This method will not cause any pain or discomfort.

- You will be asked to sign a consent form before participating and fill out a very brief questionnaire detailing very briefly your activities during the time the garment was worn, i.e. 3 hours spent in library, lunch with friends etc.
- You will be given and asked to wear a cotton long sleeved top as part of your normal daily clothing which will allow your cells to transfer naturally to the outside of the fabric. You will then be asked to meet in Ellison building lab A316 and hug a second volunteer for two minutes (one minute with arms around the shoulders and one minute with arms around the side). Once the two minutes is completed you will be asked to deposit the top in a bag and hand back to the investigator. Fibre tapes from the top will then be searched for cells adhered to the surface of the transferred fibres, which will then be recovered for DNA profiling. This part of the study will also require you to provide a buccal swab which is swabbed from the inside of your cheek, which will then be used to compare against the results from the cells taken from the fibres to check that they match. Please note that no personal information other then gender will be determined with your DNA profile. This will take no longer then 10 minutes to complete.

At the time of sampling you will be given a unique reference number and this is the number which will be linked to any data produced to ensure that you remain anonymous at all times. Your personal information will be kept in a secure cabinet which only people directly involved with this project will have access to. Any data that is published in a scientific journal or presented at a conference will use the reference number only. All tissue samples will be kept in a secure freezer and disposed of in accordance with University guidelines once this project has been completed, which is expected to be in October 2015.

Who can participate?

Please only volunteer in this study if you are over the age of 18 years and do not have an allergy to cotton or have any skin conditions which could cause any discomfort during the sampling methods as described above.

What if you change your mind?

If you have any concerns regarding the methods used during these tests or use of your results then you please feel free to raise them with the investigator who will do their best to address your concerns. Should you wish to withdraw from the study at any point however, you are able to do so without providing a reason and without prejudice.

If you have read al the information carefully and would like to participate in this study please contact Victoria Barlow using the email address at the top of the page.

This study has been approved by the Northumbria University Chemical and Forensic Science Department Ethics Committee.

General consent form

School of Life Sciences

INFORMED CONSENT FORM

Project Title: _____

Principal Investigator: _____

Participant Number: _____

	please tick where applicable
I have read and understood the Participant Information Sheet.	☐
I have had an opportunity to ask questions and discuss this study and I have received satisfactory answers.	☐
I understand I am free to withdraw from the study at any time, without having to give a reason for withdrawing, and without prejudice.	☐
I agree to take part in this study.	☐
I would like to receive feedback on the overall results of the study at the email address given below. Email address……	☐

Signature of participant..... Date.....……..

(NAME IN BLOCK LETTERS).....……

Signature of researcher..... Date.....……..

(NAME IN BLOCK LETTERS).....……

Human tissue consent form

School of Life Sciences

Consent Form for Tissue which is being removed and stored

Principal Investigator: _____

Participant Number: _____

I agree that the following tissue or other bodily material may be taken and used for the study:

Tissue/Bodily material	Purpose	Removal Method
Skin cells	To view under microscope attached to fibres and to recover a DNA profile	-Swabbing a non-intimate area of the skin, i.e. hands -Fabric rubbed against non-intimate area of skin, i.e. hands -T-shirt worn for period of time for cells to transfer naturally
Saliva	To obtain a reference DNA profile for interpreting data	-Swab of inside of cheek

I understand that if the material is required for use in any other way than that explained to me, then my consent to this will be specifically sought. I understand that I will not receive specific feedback from any assessment conducted on my samples.

I understand that the University may store this tissue in a Licensed Tissue Bank only for the duration of the study, it will then be destroyed.

Method of disposal:

Clinical Waste ☐

Other ☐

If other please specify.....

Signature of participant..... Date.....……..

Signature of researcher..... Date.....……..

8.14 Summary of data for DNA profiles obtained during two-way cell transfer test (Figure 4.10)

Table 8-18 Table summarising DNA profiles obtained from cells recovered from the surface of textile fibres during two-way transfer test

	Vol 1	Vol 2	Vol 3	Vol 4	Vol 5	Vol 6	**Total**
No. of cell samples recovered	8	17	6	8	4	2	**45**
No. of wearer profiles	1	7	2	3	3	0	**16**
No. of non-wearer profiles	0	0	0	1	0	0	**1**
No. of failed samples	7	10	4	4	1	2	**28**
No. of full profiles obtained	0	1	0	0	0	0	**1**
Partials with 4+ alleles	1	5	2	1	1	0	**10**
Partials with less than 4 alleles	0	1	0	1	2	0	**4**

8.15 Summary of data for the effect of increased SSR exposure (Figure 4.14)

Table 8-19 Peak height (RFU) for amplified alleles following increasing exposure (hrs) to sticky stuff remover (SSR)

Exposure (hrs)	D3		VWA		D16		D2	D8		D21		D18	D19		THO1		FGA	
5	458	386	688	386	398	222	215	456	245	532	249	244	379	517	335	353	173	173
4	318	379	1284	701	371	230	165	402	471	592	340	99	167	489	441	444	165	89
3	749	692	1317	838	69	216	127	380	331	245	215	85	439	656	502	354	140	83
2	1091	991	1587	201	972	897	258	395	231	1161	344	559	461	373	542	650	457	407
1	1300	958	1880	1298	901	371	407	1337	1088	566	717	597	701	1185	1325	1521	484	430
0	1469	1286	2952	2038	749	724	1512	1112	836	1201	519	685	1270	1327	1231	1313	564	336

8.16 Summary of data showing the effectiveness of FAST™ minitapes with MBG water (Figure 4.15)

Table 8-20 Alleles amplified at locus from a single-cell (in triplicate) using different tape and adhesive removers

Sample id	D3		VWA		D16		D2	AMEL	D8		D21		D18	D19		TH01		FGA	
FAST_TAPE_SSR_a	-	16	14	-	-	12	-	X	11	13	-	-	-	13.2	15	-	-	20	-
FAST_TAPE_SSR_b	15	-	-	16	-	12	-	-	-	-	-	30	-	-	15	-	9.3	-	-
FAST_TAPE_SSR_c	-	-	-	-	-	-	-	-	-	13	-	-	-	-	-	-	-	-	-
FAST_TAPE_SDW_a	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
FAST_TAPE_SDW_b	-	-	-	16	-	-	-	-	-	-	-	-	-	-	15	-	-	-	-
FAST_TAPE_SDW_c	15	16	-	-	9	-	-	X	-	-	27	-	-	13.2	15	-	-	-	-
JLAR_TAPE_SSR_a	15	16	14	16	9	12	-	X	11	13	-	30	12	13.2	15	7	-	-	-
JLAR_TAPE_SSR_b	15	-	-	-	-	-	-	X	-	-	-	-	-	-	-	-	-	-	-
JLAR_TAPE_SSR_c	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	20	-
JLAR_TAPE_SDW_a	15	-	-	-	-	-	-	X	11	-	-	-	-	-	-	-	-	-	23
JLAR_TAPE_SDW_b	-	-	-	-	-	-	-	-	11	-	-	-	-	-	-	-	-	-	-
JLAR_TAPE_SDW_c	-	-	14	16	9	12	-	X	11	-	27	30	12	-	15	7	742	20	23

8.17 Summary of data demonstrating the persistence of single-cells in fibre tapes (Figure 4.19)

Table 8-21 Peak heights for alleles amplified from single-cells following a 24 month period in storage on fibre tapes

Sample id	Amel		D19		D3		D8		VWA	TH01		D21		FGA		D16		D18	D2	
24_months_a			1973	2394	9371	11153												1095		
24_months_b	4959		2605	2604	9364	7387	2028				3960				911	4158				
24_months_c	410	3617		3400				2622								3600				
24_months_d			1500			4704			4585		3967									
24_months_e		5670	2853	3778	2438	1486	1226	1292			2088									
24_months_f		6247		3276	1383	3606		1344			3898		3288		2975			2929		
24_months_h	9861	14400	11866	4018	21281	11080	1877	2396	14188	4625	8160	1380	6844	2166		2642	926	2122	982	1145
24_months_i	3645	5695	2284	5998	7219	4392	1876	807	5724	4260	6669		1492	1253	2068	1849	1295			790
24_months_j	8911	5979	4731	9629	9551	5860	6648		11668	3564		2853		2309		3815	868			1914

8.18 Ethical approval for Brazil samples

Sao Paulo

En

FEDERAL UNIVERSITY OF SÃO PAULO – UNIFESP / HOSPITAL SÃO PAULO

CEP RECORD*

DETAILS OF THE RESEARCH PROJECT

Title of Research: HISTOPATHOLOGICAL CHARACTERIZATION OF EFFECTIVE METHODS FOR HUMAN IDENTIFICATION BY DNA ANALYSIS OF HUMAN BONES

Researcher: RAFAEL DIAS ASTOLPHI
...
Proposing Institution: PAULISTA MEDICAL SCHOOL
....

Comments and Considerations concerning the Research

LETTER OF APPROVAL PRESENTED / HC CEP RIBEIRÃO PRETO MEDICAL SCHOOL-USP
RE: INSTITUTION AS CO-PARTICIPANT

Considerations concerning the Terms of obligatory documentation
OBLIGATORY DOCUMENTS ADEQUATELY PRESENTED

Conclusions or Pending Problems
AMENDMENT APPROVED

Status of Record
APPROVED

SAO PAULO, 19th July 2013

José Osmar Medina Pestana
(COORDINATOR)

(11)5539-7162 E-mail: cepunifesp@unifesp.br

***CEP: Comitê de Ética em Pesquisa / Research Ethics Committee**

Northumbria University ethics approval

From: Daniele Castagnolo
Sent: 09 December 2014 14:58
To: Martin Evison
Subject: RE: Ethics proposal - DNA analysis of RP skeletal material

Hello Martin,

The project has been approved without any revision and it can start

Best wishes

daniele

From: Martin Evison
Sent: 05 December 2014 16:47
To: Daniele Castagnolo
Cc: victoria.barlow; Marco Aurelio Guimarães (maguima250770@gmail.com); raffaela francisco; 'Edna Sadayo Miazato Iwamura'; Ed Schwalbe; Darren Smith; Eleanor Graham
Subject: RE: Ethics proposal - DNA analysis of RP skeletal material

Dear Daniele,

Please find attached a request for Ethical Approval for proposed collaborative research on skeletal material from Ribeirão Preto, SP, Brazil.

It is intended to cover Vicki Barlow's related work as part of her PhD project and Dr Edna Iwamura's proposed CNPq "Science Without Borders" visiting scholarship next year, as well as any other likely collaborative work on the material.

The ethics proposal includes summary translations supporting documents from Portuguese (as an Italian speaker you may well be able to understand the originals!).

I also attach the research proposals for Vicki Barlow and Edna Iwamuras projects.

Please let me know if you have any questions.

With best wishes,

Martin.

Professor Martin Evison
Northumbria University Centre for Forensic Science
+44 191 243 7631
www.nucfs.ac.uk

Northumbria University Centre for FORENSICSCIENCE northumbria UNIVERSITY NEWCASTLE

8.19 Permission for sample destruction - Vindolanda

VINDOLANDA
CHARITABLE TRUST

Roman Vindolanda Fort & Museum
The Roman Army Museum

Head Office:
Chesterholm Museum
Bardon Mill, Hexham
Northumberland
NE47 7JN

T: 01434 344 277
F: 01434 344 060
E: info@vindolanda.com
www.vindolanda.com

ROMAN VINDOLANDA FORT & MUSEUM

ROMAN ARMY MUSEUM

Loan of Vindolanda material for research

I, THE UNDERSINGNED, have received the following Vindolanda material for research purposes after which the said material will be returned to the Vindolanda Trust by the specified date unless extended by the Trust.

SF8658 – Tooth from skull ✓
~~Fragments of human long bones and skull~~
~~Jaw bone – V12-89B~~
~~4 long bones – V12-35N~~
~~V12-27N~~
V12-35N – 29180

SIGNED _____

Date 16 OCT 2014

Contact details:
Eleanor AM Graham, PhD
Senior Lecturer In Forensic Science
Northumbria University
Faculty of Health & Life Sciences
A320 Ellison Building
Newcastle upon Tyre
NE1 8ST, UK

Tel: 0191 243 7651
Email: eleanor.graham@northumbria.ac.uk

Request authorised for the Vindolanda Trust by;

Destructive sampling authorised: (Yes) No ARB. (initials by VT staff)

Andrew Birley – Director of Excavations _____

Return date of material by or before 16/01/2015.

Director: Patricia Birley MBE JP. [Registered Charity No 500210]

REGISTERED MUSEUM | LOTTERY FUNDED

8.20 Illumina ForenSeq™ Signature Prep Kit Workflow

Information taken directly from the ForenSeq™ DNA Signature Prep Reference Guide (Illumina, 2015c).

1. **Amplify and tag targets**

This process amplifies and tags the gDNA using a ForenSeq oligonucleotide primer mix with regions specific to DNA sequences upstream and downstream of STRs and SNPs.

2. **Enrich targets**

This process amplifies the DNA and adds Index 1 (i7) adapters, Index 2 (i5) adapters, and sequences required for cluster amplification. The index adapters tag DNA with a unique combination of index sequences, which allow data from each tagged library to be separated during later analysis.

3. **Purify libraries**

This process uses SPB (Sample Purification Beads) to purify the amplified libraries from the other reaction components.

4. **Normalize libraries**

This process prepares DNA libraries for cluster generation to make sure that libraries of varying yields are equally represented within the sequencing run. This process assures that samples with varying input amounts or sample types achieve consistent cluster density to optimize the resolution of individual samples when pooled together. By normalizing the concentration of the libraries, while preserving the content of each library, post-PCR quantification and individual PCR product normalization are not necessary.

5. **Pool libraries**

This process combines equal volumes of normalized library to create a pool of libraries that are sequenced together on the same flow cell.

6. **Denature and dilute libraries**

This process dilutes the libraries in HT1 (hybridization buffer), adds HSC (Human Sequencing Control), and heat denatures the libraries in preparation for sequencing.

8.21 Primer design for adapted Edson primer sequences

The colour coding of the difference parts of the sequence correspond with the diagrams shown in Figure 5.10 and Figure 5.11

Primer ID	FLOW ATTACHMENT SITE	BARCODE	PAD	LINKER	PRIMER
AF1	AAT GAT ACG GCG ACC ACC GAG ATC TAC AC	CGCTG	ATTCATGCAT	TG	CCC AAA GCT AAG ATT CTA AT
AF2	AAT GAT ACG GCG ACC ACC GAG ATC TAC AC	AGGCG	ATTCATGCAT	TG	CCC AAA GCT AAG ATT CTA AT
AF3	AAT GAT ACG GCG ACC ACC GAG ATC TAC AC	GCGGT	ATTCATGCAT	TG	CCC AAA GCT AAG ATT CTA AT
AR1	CAA GCA GAA GAC GGC ATA CGA GAT	AATAC	ACCTGAGTCG	TG	TAC TAC AGG TGG TCA AGT AT
AR2	CAA GCA GAA GAC GGC ATA CGA GAT	TTTTC	ACCTGAGTCG	TG	TAC TAC AGG TGG TCA AGT AT
BF1	AAT GAT ACG GCG ACC ACC GAG ATC TAC AC	GTGAC	ATTCATGCAT	AT	CAC CAT GAA TAT TGT ACG GT
BF2	AAT GAT ACG GCG ACC ACC GAG ATC TAC AC	CCCTA	ATTCATGCAT	AT	CAC CAT GAA TAT TGT ACG GT
BF3	AAT GAT ACG GCG ACC ACC GAG ATC TAC AC	AGGTC	ATTCATGCAT	AT	CAC CAT GAA TAT TGT ACG GT
BR1	CAA GCA GAA GAC GGC ATA CGA GAT	CACTT	ACTTGAATAT	CC	GGA GTT GCA GTT GAT GT
BR2	CAA GCA GAA GAC GGC ATA CGA GAT	ATGCT	ACTTGAATAT	CC	GGA GTT GCA GTT GAT GT
CF1	AAT GAT ACG GCG ACC ACC GAG ATC TAC AC	CCTGT	ATTCATACAT	AG	CCC CAT GCT TAC AAG CAA GT
CF2	AAT GAT ACG GCG ACC ACC GAG ATC TAC AC	CGATC	ATTCATACAT	AG	CCC CAT GCT TAC AAG CAA GT
CF3	AAT GAT ACG GCG ACC ACC GAG ATC TAC AC	GTCCA	ATTCATACAT	AG	CCC CAT GCT TAC AAG CAA GT
CR1	CAA GCA GAA GAC GGC ATA CGA GAT	GGTAA	ACCTGAGTCG	CC	TGG CTT TAT GTA CTA TGT AC
CR2	CAA GCA GAA GAC GGC ATA CGA GAT	GATTC	ACCTGAGTCG	CC	TGG CTT TAT GTA CTA TGT AC
DF1	AAT GAT ACG GCG ACC ACC GAG ATC TAC AC	CTGGT	ATTCATGCAT	AT	CAC TAG GAT ACC AAC AAA CC
DF2	AAT GAT ACG GCG ACC ACC GAG ATC TAC AC	GCATC	ATTCATGCAT	AT	CAC TAG GAT ACC AAC AAA CC
DF3	AAT GAT ACG GCG ACC ACC GAG ATC TAC AC	GGAAC	ATTCATGCAT	AT	CAC TAG GAT ACC AAC AAA CC
DR1	CAA GCA GAA GAC GGC ATA CGA GAT	ATTCC	ACTTGAATAT	AT	GAG GAT GGT GGT CAA GGG AC
DR2	CAA GCA GAA GAC GGC ATA CGA GAT	AAGGA	ACTTGAATAT	AT	GAG GAT GGT GGT CAA GGG AC

8.22 Kruskal Wallis test for DNA extraction methods

Kruskal Wallis test conducted using R version 3.1.2 (2014-10-31) -- "Pumpkin Helmet"

Copyright (C) 2014 The R Foundation for Statistical Computing

Platform: i386-w64-mingw32/i386 (32-bit)

```
> BrazilSA <- read.csv("C:/Users/vicki/Desktop/BrazilSA.csv", header=FALSE)
>   View(BrazilSA)
> r = c(t(as.matrix(BrazilSA)))
> f = c("V1", "V2", "V3")
> k = 3
> n = 5
> tm = gl(k, 1, n*k, factor(f))
> kruskal.test(r ~tm)


Kruskal-Wallis rank sum test
data:  r by tm
Kruskal-Wallis chi-squared = 0.5914, df = 2, p-value = 0.744
> BrazilmtDNA <- read.csv("C:/Users/vicki/Desktop/BrazilmtDNA.csv", header=FALSE)
>   View(BrazilmtDNA)
> r = c(t(as.matrix(BrazilmtDNA)))
> f = c("V1", "V2", "V3")
> k = 3
> n = 5
> tm = gl(k, 1, n*k, factor(f))
> kruskal.test(r ~tm)
        Kruskal-Wallis rank sum test
data:  r by tm
Kruskal-Wallis chi-squared = 2.4394, df = 2, p-value = 0.2953
```

8.23 Steps utilised to analyse raw MiSeq data

Following sequencing using the MiSeq platform, samples (005B, 007B, 013B, 017B, 020B and Method 1 Neg A) could not be demultiplexed and consequently just the two read data files were available for analysis. The following steps were carried out to analyse the data:

1. Both read files were uploaded to Galaxy (www.usegalaxy.org) using the Upload File tool (Blankenberg et al., 2010)

Sample files converted to Galaxy compatible fastqsanger files using FASTQ Groomer (Blankenberg et al., 2010) 2098148 sanger reads were converted for each read file (1.1 GB). An example of the data is shown in Figure 8.1.

TTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATC
AACAACCGCTATGTATTTCGTACATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAATACTT
GACCACCTGTAGTACACGACTCAGGTGTATTATCTCGTATGCCGTCTTCTGCTTGTCAAATTACACTC
TATTACCACATCCTTCCTCTTCCTTCCCTCCTATTTTTAACACCCC

TTATGGTACCGTACAATATTCATGGTGGCTGGCAGTAATGTACGAAATACATAGCGGTTGTTGATG
GGTGAGTCAATACTTGGGTGGTACCCAAATCTGCTTCCCCATGAAAGAACAGAGAATAGTTTAAAT
TAGAATCTTAGCTTTGGGCAATGCATGAATCAGCGGTGTAGATCTCGGTGGTCGCCGTATCATTAAT
ACAAATAAACAAAAGAGTTTGGGTGTTCGTTTCGCGGTATGCATAGGTCGAT

Figure 8.1 The fastqsanger format R1 data contains the target DNA sequence (highlighted blue) and the reverse complement of the reverse primer/ linker/ pad/ index/ and flow cell adaptor. The fastqsanger format R2 data contains the target DNA sequence (highlighted blue) and the reverse complement of the forward primer/ linker/ pad/ index and flow cell adaptor. Additional bases (not highlighted) were present at the end of each read after the flow cell adaptor sequence.

2. Illumina adapters were removed from the end of every read in each file using Clip (based on the FASTX-Toolkit by Assaf Gordon) as shown in Figure 8.2.

TATGGTACCGTACAATATTCATGGTGGCTGGCAGTAATGTACGAAATACATAGCGGTTGTTGATGG
GTGAGTCAATACTTGGGTGGTACCCAAATCTGCTTCCCCATGAAAGAACAGAGAATAGTTTAAATT
AGAATCTTAGCTTTGGGCAATGCATGAATCAGCG

Figure 8.2 Sequence with Illumina adapters and additional bases removed

3. In order for the two reads to be joined, the index for the reverse read (R2) needed to be positioned at the 5' end of the sequence for identification at a later stage, which was accomplished using the Reverse-Complement tool (FASTX-Toolkit)

4. The paired end reads in both R1 and R2 files were then combined using FASTQ joiner (Blankenberg et al., 2010) in order to place the forward and reverse indexes in the sample sequence for sample identification and demulitplexing, shown in Figure 8.3.

CGCTGATTCATGCATTGCCCAAAGCTAAGATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGAAG
CAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTACATTACTGC
CAGCCACCATGAATATTGTACGGTACCATAATTAAACTATTCTCTGTTCTTTCATGGGGAAGCAGAT
TTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTACATTACTGCCAGCC
ACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAGTACACGACTCAGGTGTATT

Figure 8.3 The reverse complement of the R2 file reads were joined with the R1 files so that both indexes were present for sample identification and demulitplexing

5. A text file containing the reverse indexes (in reverse complement) was uploaded to Galaxy using Upload File
6. The Barcode Splitter tool (FASTX-Toolkit) was selected to split the joined file by the reverse index at the 3′ sequence end and the number of allowed mismatches was set to zero
7. This produced two files, i.e. AR1 and AR2, now in FASTQ txt format which were downloaded and extracted
8. The AR1 file was uploaded using Upload File and then the format converted for later use with the barcode splitter tool using FASTQ to FASTA (Blankenberg et al., 2010)
9. A text file for each of the forward indexes, AF1, AF2 and AF3, was uploaded to Galaxy using Upload File
10. The Barcode Splitter tool was then utilised to split the converted AR1 file by the three forward indexes at the 5′ sequence end and the number of allowed mismatches was set to zero
11. Steps 9-11 were then repeated for the AR2 file resulting in a total of six FASTQ txt 8) files, one for each sample, which were all downloaded and renamed; AF1/AR1, AF2/AR1, AF3/AR1, AF1/AR2, AF2/AR2, AF3/AR2.
12. The AF1/AR1 FASTQ txt file was uploaded using Upload File and the format converted using FASTQ to FASTA
13. Combine FASTA and QUAL (Blankenberg et al., 2010) was used to convert the AF1/AR1 file once again into the FASTQ sanger format required for the following steps
14. FASTQ splitter (Blankenberg et al., 2010) was utilised to separate the still-joined paired end reads back into the forward (LHS) and reverse (RHS) reads for alignment

15. The Reverse-Complement of the RHS read file was obtained to position the barcode once again at the 3′sequence end (previously changed in step 4)
16. FASTQ Trimmer (Blankenberg et al., 2010) was used to remove seventeen bases (barcode + pad + link) from the 3′ end of each read in both the LHS and RHS files
17. The reads in each file were then aligned to the Human reference sequence (hg38) using BWA-MEM (Li, 2013, Li and Durbin, 2009, Li and Durbin, 2010) which produces a BAM format file
18. Steps 13-18 was repeated for each of the six sample files generated in step 12
19. Steps 6-18 were repeated for each of the remaining three amplicons
20. Flagstat (Li, 2011b, Li, 2011a, Li et al., 2009) was used to tabulate the descriptive statistics for the BAM dataset for each individual sample file