Ph.D.

Saving Bambi: Whole genome sequencing of *Dama dama* (fallow deer) and an investigation into their profound lack of genetic diversity.

Rebecca Barnard (BSc. Hons, MSc)

A thesis submitted in partial fulfilment for the requirements for the degree of Doctor of Philosophy, at the University of Lancashire

December 2025

University of Lancashire

Declaration

Type of Award: PhD
School: School of Natural Sciences

1. **Concurrent registration for two or more academic awards**

I declare that while registered as a candidate for the research degree, I have not been a registered candidate or enrolled student for another award of the University or other academic or professional institution

2. **Material submitted for another award**

I declare that no material contained in the thesis has been used in any other submission for an academic award and is solely my own work

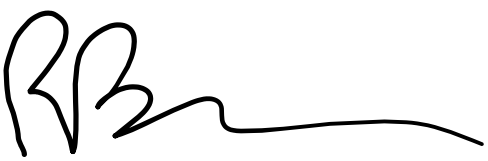
3. **Collaboration**

Where a candidate's research programme is part of a collaborative project, the thesis must indicate in addition clearly the candidate's individual contribution and the extent of the collaboration. Please state below:

4. **Use of a Proof-reader**

No proof-reading service was used in the compilation of this thesis.

'I confirm that this report is all my own work and that all references and quotations from both primary and secondary sources have been fully identified and properly acknowledged in footnotes and bibliography.'



Signed..... **Date**09/09/2025.....

Print name: Rebecca Barnard

Acknowledgements

This PhD project was a collaboration and I could not have completed it without the guidance, help and support from some very important people. Firstly, I would like to thank my Supervisor Dr Judith Smith, who picked me up half way through my PhD after my initial supervisor had left. Judith helped guide me to build the final project, without her knowledge and ideas my PhD would not have made it to where it is today.

Secondly, I thank Dr Sibte Hadi who gave me the inspiration to do the PhD and has continually provided support throughout the project. I also would like to thank The University of Central Lancashire for all their funding support throughout the project but especially for providing the funds for the whole genome sequencing, without which this project would not have been possible.

I would like to acknowledge all the park rangers and deer stalkers who collected samples for this project during the cull season. They all went out of their way to collect the samples, document the information related to each sample and photograph each specimen, for which I am very grateful. Special thanks to Tony Hatton from Royal Parks Richmond. Tony collected the sample which went on to be DNA extracted for the whole genome sequencing. Due to his dedication to the sampling for this project, I was able to collect the sample from him in London within an hour of the cull, and have it extracted within 48 hrs, ensuring freshness of the sample leading to the best yield of high quality DNA for sequencing.

I would like to thank my Parents, Susan and Peter Banard for their unwavering faith in me and their continued support throughout my university life, without them I wouldn't have been able to go to university and follow my dreams.

Last but not least I thank my partner Will, for his constant love, dedication and support. Without him I couldn't have done what I have achieved at UCLan. Will has been my biggest cheer leader and my constant throughout. Not only has he kept me sane with his endless words of encouragement, but he has provided many long days and nights of tech support and IT help. This was particularly with the bioinformatic command line programs. Without Wills IT skills I couldn't have completed the analysis to as high standard, for which I will be ever grateful.

Abstract

On average, 55,000 deer are illegally killed in the UK each year. Deer which are killed illegally are often done so in a cruel and inhumane way, causing unnecessary distress and suffering to the animal. Furthermore, the illegal killing of deer negatively impacts herd numbers and the natural environment of the deer. Deer are legally killed each year by licenced game keepers in order to keep a control of the ever-growing size of deer herds in the UK. Illegally killing deer disrupts this process and thus affects the livelihood of these workers. Illegally killed deer are often gall roached on site and the sought-after meat and antlers are taken, whilst leaving the carcass to rot where slaughtered. More often than not, the perpetrators get away with their crimes, free to commit again.

During the initial phase of work, five samples were collected and extracted via an optimised method to ensure DNA quantity and quality for whole genome sequencing. The entire genome of the fallow deer was sequenced via Pac Bio Sequencing, using a tissue sample from an individual from Richmond Park, London. This enabled the investigation of the fallow deer genome for STRs within nuclear DNA, in order to design new primers. This overcame issues found with cross-species amplification between fallow deer and primers developed for other species of deer in the UK during a previous project. The Genome is 3,108,385,535 bp in length with the sequence assembled into 35 Chromosomes including an assembly of the X and Y allosomes. The final assembly also had a BUSCO completeness score of 99%. The annotated genome was found to consist of 34,891 genes, 22,157 of which are protein coding genes. The final BUSCO v4.1.4 completeness score for the annotation was 99%. Predicted genes have an average transcript length of 37,945bp and on average each gene has 10 exons.

The second study involved mining the genome for STRs; in order to best assess the entire UK population, 364 samples were collected in total across 15 geographical locations. Out of the 364 samples, the best 190 were used for downstream analysis. Using the genome sequenced in study 1, a total of 21,580 tetra and penta nucleotide STRs were located within the genome and 264 primers were developed. Out of the 132 designed primers for tetranucleotide markers, 90 were tested and 9 were found to be polymorphic. Polymorphism was detected via fragment analysis on the ABI3500 followed by Sanger Sequencing. The lack of polymorphic loci discovered can be attributed to the low genetic diversity of the fallow deer species. It is well documented that *Dama dama* have low genetic diversity caused by a genetic bottle neck during the Mesolithic (8000 BC – 2700 BC) and Neolithic (10,200 BC - 2000 BC) periods. This lack of genetic diversity which is still observed today implicates the likelihood of a forensically significant STR multiplex for fallow deer individual identification.

Oxford Nanopore Sequencing via the Native Barcoding kit was used to develop a novel method for rapidly screening samples for polymorphism. The results for the NGS rapid polymorphism screening study showed that this new method is an excellent way of quickly screening for genetic variation at known STR loci. This method is quicker and overall more cost effective because more loci can be screened at once against an increased number of samples (223 in this study). Therefore, loci can be easily included or omitted, compared to traditional methods of gel electrophoresis and CE. In terms of polymorphism in fallow deer, the nanopore results showed that there was majorly no genetic variation, however Loci Fallow70, Fallow89, Fallow118, Fallow124 and Fallow129 showed increased variation compared to Sanger Sequencing results.

The mitochondrial genome for *Dama dama* is 16,322bp in length, consisting of 13 protein coding genes, 2 rRNAs and 22 tRNAs. The notable difference with the mitogenome for *Dama dama* is that it is the smallest compared to all other *Cervidae* mitochondrial genomes analysed in this study. Overall, the phylogenetic and divergence time analysis has provided support for previously published theories but has also given a new perspective on the divergence of the *Dama dama* species and its sub-species *Dama mesopotamica*. Findings suggest that as Megaloceros giganteus is the older of the Dama species, *Dama mesopotamica* has held onto more ancestral genetic sequence than *Dama dama*, and that it is *Dama dama* that diverged first and has mutated away from the ancestral lineage, around 5.35 million years ago.

This project was the first to utilise the full sequence of the D-Loop to assess population diversity and analyse population signature haplotypes. This study also compared the use of Sanger sequencing to NGS via Oxford Nanopore for the application of SNP and InDel detection. Interestingly, only 6 sites of variation were identified as a result of the sanger sequencing. However, this was latterly increased to 47 sites of variation as a result of the Nanopore sequencing. This increase in variation detected is likely due to the increased sequencing depth achieved with Oxford Nanopore Sequencing. Variation calling via nanopore sequencing was achieved at a 300x sequencing depth compared to 1x with Sanger Sequencing. The level of variation observed in the NGS sequences compared to Sanger equates to a staggering 783% increase.

This project is the first to fully sequence the genome of fallow deer and provides the groundwork for further research involving this species. The result of this project benefits the welfare of the fallow deer species across the world, contributing to the global effort to combat wildlife crime as well as the conservation of the species.

Table of Contents

List of Tables

List of Figures

1 Introduction

1.1 Wildlife Crime: A Global Issue

Wildlife crime is an umbrella term for activities involving a fully protected species (CMS, 2019). For example hunting a protected species for sport, unauthorised trapping and netting of protected bird species, consumption or illegal trade of protected species (this includes protected plants and trees), hunting outside of legal seasons (e.g. Deer culling season), poisoning or snaring bird species, killing of dolphins for bait or consumption, killing of turtles and the poaching of turtle eggs (CMS, 2019). These are just a small selection of the illegal acts which are associated to wildlife crime. These acts are committed against national and international laws and regulations which have been set out to protect and sustain natural resources (CMS, 2019). However, as stated by the United Nations Office on Drugs and Crime (UNODC), offering an international definition of Wildlife crime is difficult because all countries protect its own fauna and flora differently. For other types of organised crime, including drug, human and firearm trafficking there are international bodies that define them for global understanding. Whereas there is no universally excepted legislation for wildlife crime, thus there is no globally accepted definition.

Up until the 1980's there was little concern for wildlife crime (Neme and Leakey, 2014). There was a lack of science surrounding wildlife forensics, scientists were reluctant to provide and argue evidence at court. As more exotic animals have been discovered over the years, due to our better understanding of the need to protect biodiversity, there has been an increase in wildlife crime, including poaching, and trafficking. With this, there has also been an increase in the efforts to combat it, including the development of forensic wildlife DNA strategies. The Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES) came into force in 1975 and started to record the amount of illegal trade in protected wild fauna and flora (Figure 1.1). CITES is an international agreement bound between governments with the aim to protect species affected by international trade to ensure they do not become threatened (CITES (The Convention on International Trade in Endangered Species of Wild Fauna and Flora), 2019). Currently CITES offers protection and monitoring of over 37,000 different species of plants and animals, encompassing all types of trade involving these.

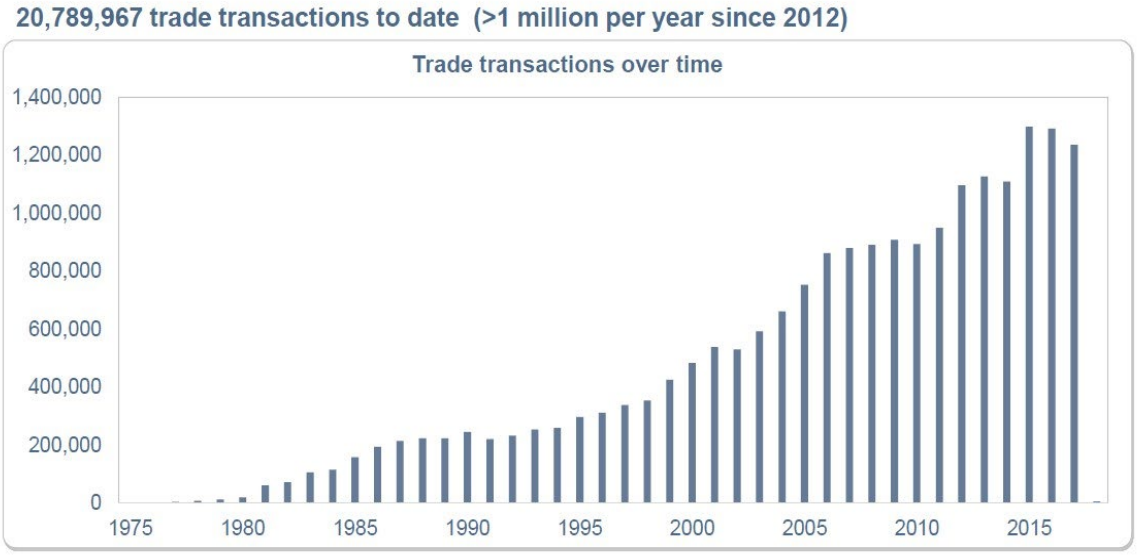


Figure 1.1: *The number of illegal trade transactions of protected species per year since 1975, globally (CITES (The Convention on International Trade in Endangered Species of Wild Fauna and Flora), 2019).*

Figure 1.1 shows the numbers of transactions involving the species CITES is concerned with. Since 2012 the number of illegal trade transactions has surpassed 1 million a year. This is an unprecedented amount of trade involving protected species being smuggled across borders and sold on the black market. The main cause for the increase in Wildlife crime is due to it still being legal in some countries and lacking enforcement in others. Many countries and governments require education on the impact of wildlife crime and more must be done by CITES and accompanying conventions to resolve this (UNODC (United Nations on Drugs and Crime), 2016).

According to the wildlife crime report from UNODC, up until 2021, the number of seizures as a result of illegal trade stood at almost 13 million specimens Worldwide (UNODC (United Nations on Drugs and Crime), 2024). This equates to 2,400 tons and a global worth of over US$20 billion a year, second to illegal drugs trade (Linacre and Tobe, 2011; Smart *et al.*, 2021; UNODC (United Nations on Drugs and Crime), 2024). According to the latest UNDOC report, published in 2024, The most traded animal is Rhino, followed very closely by Pangolin (only 1% difference) (Figure 1.2), between 2005 and 2018 it was recorded that 1,277 seized shipments of ivory, totalled 227.9 metric tons. This equates to ivory poached from an estimated 157,000 elephants, averaging 17,000 elephants a year (UNODC (United Nations on Drugs and Crime), 2020). According to the latest figures, the annual illicit income from elephant ivory equates to more than $400 million and for Rhino horn it is estimated to be $280 million (UNODC (United

Nations on Drugs and Crime), 2024). However, the most illegally traded plant species is Cedar and other Sapindales followed by Rosewood (Figure 1.2) (UNODC (United Nations on Drugs and Crime), 2024). According to a past report, the tropical furniture production worth over $65 billion a year (UNODC (United Nations on Drugs and Crime), 2016). Figure 1.2 shows the latest figures for the percentage share of species type for flora and flora, among seizures between 2015 and 2021.

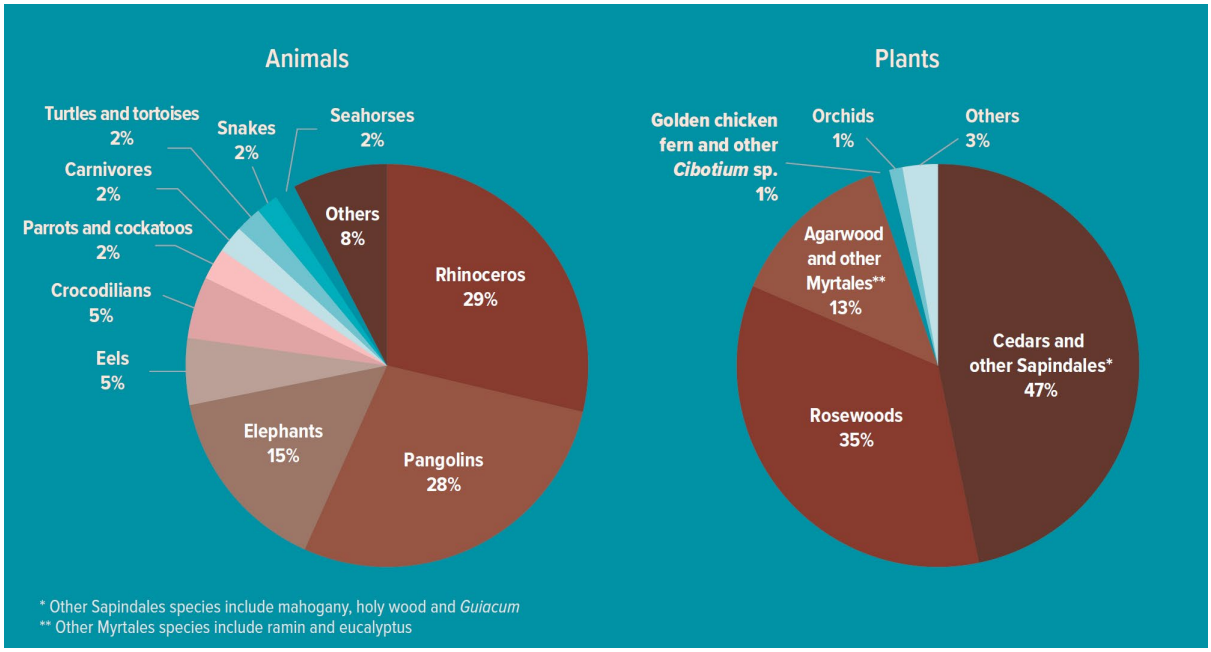


Figure 1.2: *Pie chart to show share of type of flora and fauna among total seizures between 2015 and 2021 (UNODC (United Nations on Drugs and Crime), 2024).*

Illegal wildlife shipments are predominantly seized, as reported by country, in the UK, Europe, Australia, New Zealand, India, China, and USA, see Figure 1.3. This is likely due to the heightened security and border controls at these countries (UNODC (United Nations on Drugs and Crime), 2016; UNODC (United Nations on Drugs and Crime), 2024). However, illegal wildlife shipments are largely seized, by countries identified by source, in Russia, Canada, Australia, India, China, Europe, South Africa and USA (Figure 1.4) with the greatest increase shown across Africa compared to figure 3. This is likely due to the high level of horn, ivory, and fur poaching in these countries (UNODC (United Nations on Drugs and Crime), 2016; Smart *et al.*, 2021). However, wildlife crime primarily originates from Asia, due to the largest illegally traded wildlife being elephant and rosewood which are native to these countries. Nonetheless, it is echoed across the world, see Figure 1.3 and 1.4 for maps to show countries effected by wildlife crime, the darker the colour the higher the level of wildlife crime.

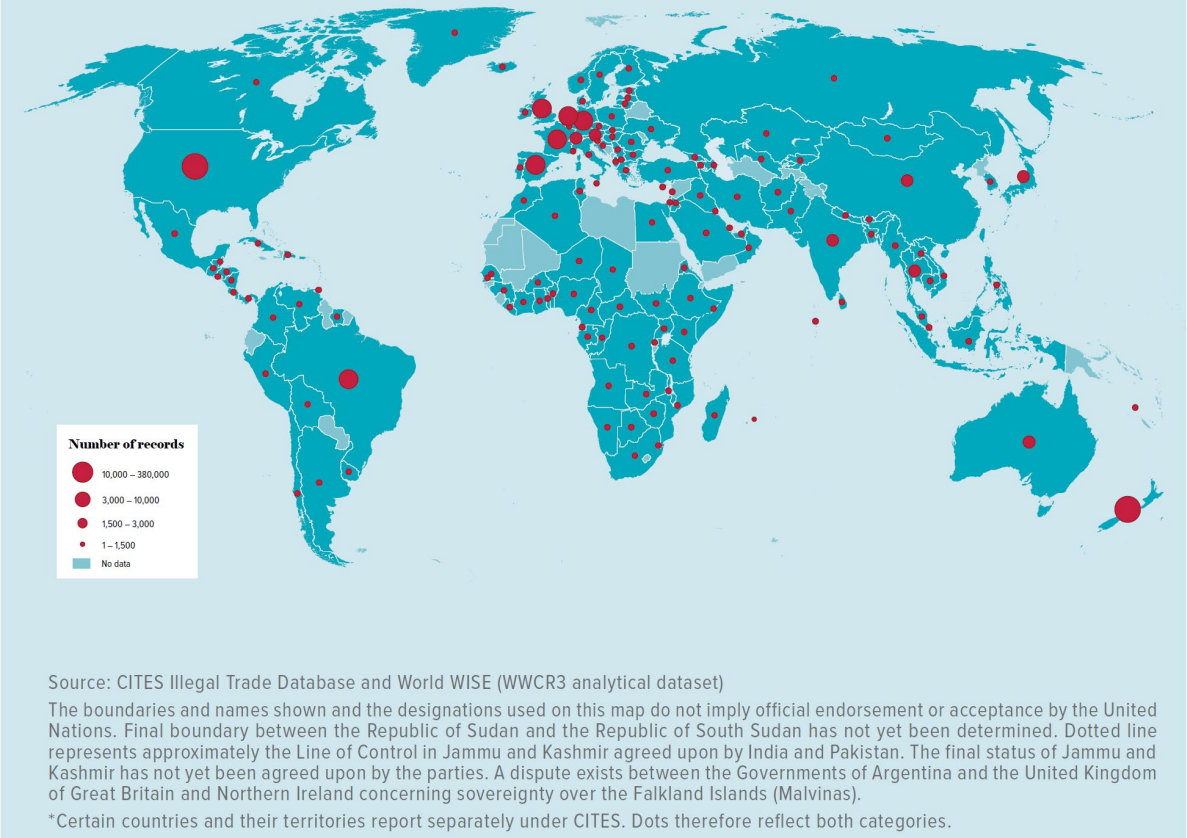


Figure 1.3: *Total number of seizures reported by country between 2015 and 2021, using data from CITES Illegal Trade Database and World WISE (WWCR3 analytical dataset) (UNODC (United Nations on Drugs and Crime), 2024).*

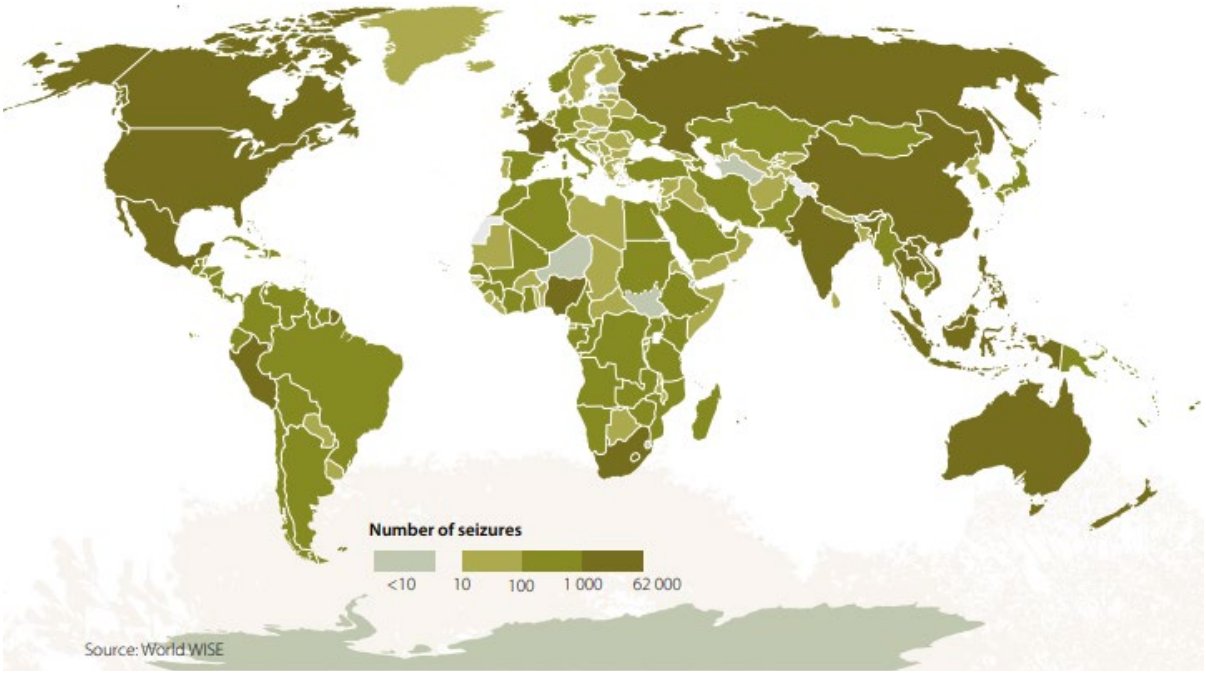


Figure 1.4: *Total number of seizures by countries identified as source between 2004 and 2015, using data from the World WISE database (UNODC (United Nations on Drugs and Crime), 2016).*

1.2 Wildlife Crime in the UK

The UK is not known for ivory, horn, or big cat poaching but it is widely used as a transit country for smugglers to pass through illegal shipments onto its destination country (Illes, 2016). The UK is also a destination country, with much illegal trade being dealt here for sale and consumption. Between 2007 and 2014, 2,853 seizures were made at UK borders, most of which occurred at Heathrow airport, see Figure 1.5 (Illes, 2016). One of the most significant seizures being that of 13 San Salvador Rock Iguanas in 2014 (Illes, 2016).

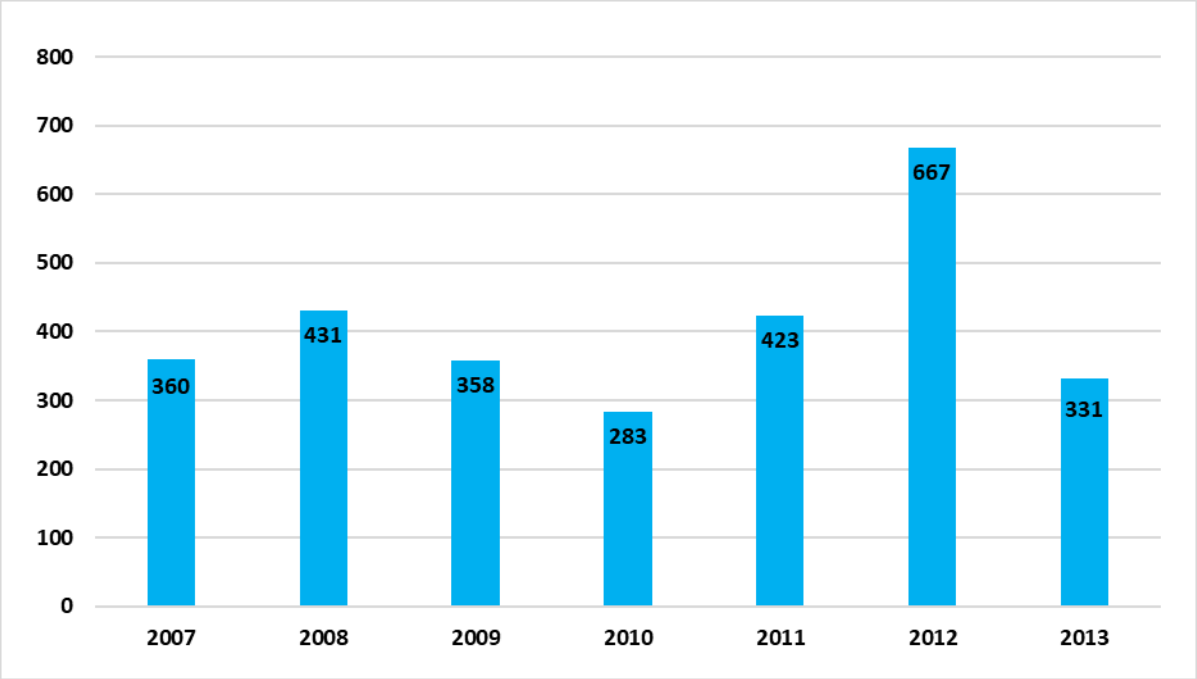


Figure 1.5: *Graph to show number of seizures in the UK between 2007 and 2012, Data from CITIES UK Biennial Reports 2007/2008, 2009/2010 and 2011/2013 (Illes, 2016).*

Figure 1.5 shows the distribution of the 2,853 seizures over 7 years (2007 up to the start of 2014). Out of the 2,853 seizures, 10% were associated to tiger, elephant, and rhino illegal products (Illes, 2016). In 2005 the UK enforced the CITIES convention which increased the maximum prison sentence from 2 years to 5 years for wildlife crimes and the level of fine is now unlimited (Illes, 2016). In 2013, the UK had 17 successful criminal cases and were slightly more varied in previous years, Table 1.1 shows examples of prosecutions in the UK between the years 2007 and 2020 and the sanctions corresponding to crime. Following 2013 until the most recent data published for 2023, there was a further 8,205 seizures made by the UK border force (TRAFFIC, 2021; Wildlife and Countryside Link, 2022; Wildlife and Countryside Link, 2023; Wildlife and Countryside Link, 2024). However, getting these criminal cases to successful prosecution is further on the

decline, during the years 2013 and 2023, only 84 successful prosecutions were obtained, as result of the 8,205 seizures (Wildlife and Countryside Link, 2023; Wildlife and Countryside Link, 2024). Other Wildlife crimes in the UK are also linked to the persecution of badgers, raptors, and bats, as well as poaching of hare, salmon, and deer (Illes, 2016). Deer are one of the most poached animals in the UK with over 50,000 illegally killed every year. In 2023, the London Met Police announced the closure of their wildlife crime task force, this is the opposite to the response that is required. Awareness of wildlife crime in the UK must continue to be raised. Government budgets need to be increased to allow police forces to grow teams dedicated in dealing with wildlife crime (TRAFFIC, 2021; Wildlife and Countryside Link, 2022; Wildlife and Countryside Link, 2023; Wildlife and Countryside Link, 2024).

Table 1.1: *Examples of convictions made in the UK, related to wildlife crime, between the years 2007 and 2016 Data taken from CITES Biennial Reports 2007/2008, 2009/2010, 2011/2012, 2013/2014, 2015/2016 (Illes, 2016).*

Crime	Sentence	Year
Selling elephant ivory, hippo ivory and sperm whale teeth.	2 years in prison and £1,480 fine	2007/2008
Attempted smuggling of rhino horns from the UK to China.	12 months in prison	2009/2010
Illegal importation of bird eggs from the US and Australia and the illegal sale of the eggs.	9 months in prison and £3,607 fine	2009/2010
Attempt to sell two ring tailed lemurs on the internet.	£500 fine	2011/2012
3 attempts to evade ivory export restrictions by using forged documents.	£555 fine	2013/2014
Attempted smuggling of rhino horns from the UK to China.	12 months in prison	2013/2014
The illegal sale of barn owls.	£600 fine	2013/2014
The illegal sale of elephant ivory, hippopotamus ivory, and sperm whale teeth.	£1480 fine*	2013/2014
Attempted sale of 4 hippo tusks; 3 rhino horns, and 2 elephant tusks.	14 months in prison	2015/2016
Illegal trade of Ivory items	$2311 fine and unpaid work	2019/2020

**2 years to pay the fine, if not will face imprisonment.*

1.2.1 Crime Against Deer

Deer in the UK are a defining part of the British natural and cultural heritage, with over two million deer ((SPCA, 2020) residing in the UK (Parkes and Thornley, 2008). Venison is a rich source of food and has been hunted since the stone age (Parkes and Thornley, 2008). Since the 11th century hunting as a sport in the UK has been conducted involving several animals, including deer. Hunting parties were full of pomp and circumstance, to go hunting was a symbol of one's wealth and status. Following this, deer were seen as a symbol of wealth, and during the 12th century large estates began to emerge on which deer were kept privately (Parkes and Thornley, 2008). Deer kept on large estates held an ornamental value, often, breeds of deer were imported from outside of the UK for a more exotic effect, increasing the owner's wealth and status (Parkes and Thornley, 2008). Six breeds of deer currently reside naturally in the UK, of which, deer native to the UK include, red deer (*Cervus elaphus*) and roe deer (*Capreolus capreolus*), however, fallow deer (*Dama dama*), sika deer (*Cervus nippon*), Chinese-water deer (*Hydropotes inermis*), and muntjac deer (*Muntiacus reevesi*) were once imported (Parkes and Thornley, 2008). By the 16th century, wolves had become extinct therefore, deer had no natural predators causing an increase in their numbers. With the increase of numbers came an increase of illegal activity.

A conflict exits between the management of deer and the animal's potential as a sporting activity (Parkes and Thornley, 2008). The law, therefore, plays an important part in keeping the balance and protecting the welfare of the deer. To help with this, in the UK there are several deer societies and bodies which support the law and protect the welfare of deer, these include; The British Deer Society (B.D.S), The Northern Ireland Deer Society, The Deer Initiative (DI), The Deer Commission for Scotland (D.C.S), The British Association for Shooting and Conservation (B.A.S.C), The Countryside Alliance, The U.K. Association for Professional Deer Managers (U.K.A.P.D..M.), The British Farmers Association, The National Game Dealers' Association (N.G.D.A) and The St. Hubert Club of Great Britain, all of which are concerned with the protection, management and conservation of deer in the UK and welcome a project such as this.

In the UK, the three main species of deer affected by poaching are, red deer, roe deer and fallow deer, see section 1.2.1.1. This is due to their size and beauty, making them desirable to poachers for meat, hide and antlers (Szabolcsi *et al.*, 2014). Furthermore, illegal supposedly aphrodisiac drugs are developed from the velvet which coats the antlers (Milner *et al.*, 2006). Red deer and roe deer have already got STR multiplexes available for individual identification within these

species (Deer plex (Szabolcsi *et al.*, 2014) and STRoe (Morf *et al.*, 2021), respectively). Therefore, there is a gap in the research here for fallow deer, with no such kit available for their individual identification.

In the UK there is clear legislation surrounding the culling of deer, which includes when it is legal to do so and what is the appropriate approach. Legislation varies slightly between Northern Ireland, Scotland and England and Wales, but they all set out to achieve the same thing, protection for the welfare of deer. There are open and closed seasons when deer in the UK can be legally culled, see Table 1.2. When deer are killed outside of these permitted seasons, during closed seasons, it is a criminal offence and perpetrators will be prosecuted. Open and closed seasons are used to control the culling of deer in the UK to ensure a stable management of population size (The British Association for Shooting and Conservation (BASC), 2013).

Table 1.2: *Statutory open Seasons for Deer (dates inclusive) (The British Association for Shooting and Conservation (BASC), 2013).*

England/Wales	Scotland	Northern Ireland
Red Deer		
Stags		
Aug 1 – April 30	July 1 – Oct 20	Aug 1 – April 30
Hinds		
Nov 1 – Mar 31	Oct 21 – Feb 15	Nov 1 – Mar 31
Sika Deer		
Stags		
Aug 1 – April 30	July 1 – Oct 20	Aug 1 – April 30
Hinds		
Nov 1 – Mar 31	Oct 21 – Feb 15	Nov 1 – Mar 31
Roe Deer		
Bucks		
April 1 – Oct 31	April 1 – Oct 20	--
Does		
Nov 1 – Mar 31	Oct 21 – Mar 31	--
Fallow Deer		
Bucks		
Aug 1 – April 30	Aug 1 – April 30	Aug 1 – April 30
Does		
Nov 1 – Mar 31	Oct 21 – Feb 15	Nov 1 – Mar 31

The main legislation to cover England and Wales in the protection of deer is the Deer Act, 1991 which is the revised version of the original deer act published in 1963. The main amendment to the law made it illegal to kill, poach and sell venison attained illegally (Deer act, 1991). There are some caveats to the legislation however, deer can be killed outside of the open seasons if it is to protect against crops and property, spread of disease, or if a deer is injured and it is to reduce prolonged suffering (Joint Nature Conservation Committee, 2017). However, this must be carried out by licenced, trained personnel and clear evidence for reasoning must be given (Parkes and Thornley, 2008; Joint Nature Conservation Committee, 2017). Further to this, the legislation also states that the authorised killing of dear is prohibited in the last hour before sunrise and the first hour after sunset as well as several methods and weapons used to kill dear including dogs, traps, bows and arrows, poisons, drugs and spears are also forbidden (Parkes and Thornley, 2008). Firearms may only be used by licenced hunters. The only circumstance when prohibited weapons may be used, and the person not prosecuted, is when an injured or suffering deer needs to be killed in the most humane way possible, but as mentioned, evidence of reasoning must be given (Parkes and Thornley, 2008; Joint Nature Conservation Committee, 2017).

Of all information gathered by the National Wildlife Crime Unit (NWCU) between the years 2011 – 2013, 44% was linked to poaching, particularly deer (Illes, 2016). Approximately 50,000 deer are killed illegally in the UK each year (Wyatt, 2016). Deer poachers vary, some do it for the thrill of the chase others just for extra money, for some it is their living (Parkes and Thornley, 2008). In less remote places poachers take cover in their vehicle, shoot from their vehicle, and then put the slaughtered deer in their vehicle and escape. However, in rural cases poachers will camp out in the hills (Parkes and Thornley, 2008). Once the deer is killed, they gralloch it on site rather than taking the whole carcass away with them. Gangs of poachers are common, they get to know the ground and in one-night gain six to seven deer (Parkes and Thornley, 2008). The gangs sell the venison on the black market to hotel chains and other traders. If poachers are caught, they simply say they were saving a wounded or suffering deer from further pain (Parkes and Thornley, 2008). Poachers have one aim, maximum profit from little effort resulting in no care for animal welfare and immense cruelty to deer.

To confirm illegal poaching has taken place important questions must be addressed such as, what species is it? What was the geographical origin of the deer? And what is its individual identification? DNA offers good potential in answering these questions to support a prosecution.

1.2.1.1 Fallow Deer

Fallow deer are one of the most beautiful species of deer with prominent antlers and distinctive white markings on their bodies, see Figure 1.6 (The British Deer Society (BDS), 2016). Figure 1.6 shows the common colouring of fallow deer, however, there are three other variations; known as Menil (paler colour with white spots visible all year round), melanistic (black or chocolate coloured coat), and white (white or sand colour coat, due to colour not albinism) (The British Deer Society (BDS), 2016; Chapman and Chapman, 1997). Their tails are the longest compared to other species of deer in the UK, often with a black stripe and an inverted horseshoe on their rump (The British Deer Society (BDS), 2016). Fallow deer males, known as bucks, have magnificent antlers and are the only species in the UK to have palmate antlers which are full sized, once the deer is three to four years old, at almost a meter in length (The British Deer Society (BDS), 2016).

Figure 1.6: *A fallow deer Buck (The British Deer Society (BDS), 2016).*

Fallow Deer are the most widely distributed species of deer in the UK populating, England, Wales, Northern Ireland, and parts of Scotland. Genetic analysis date this species of deer in the UK back to roman times (The British Deer Society (BDS), 2016; Baker *et al.*, 2017). However, Roman fallow deer went extinct with the collapse of the Roman Empire. Fallow deer were later re-introduced, from the eastern Mediterranean, during the 11th century (Baker *et al.*, 2017). However, fallow deer originally came from Turkey and Asia with original ancestors being

found in Iraq and Palestine (Randi and Apollonio, 1988; Arslangündoğdu *et al.*, 2010). The distribution of fallow deer as we know it today (Figure 1.7) is attributed to the decline of deer parks after the 15th century as they became less fashionable. Many of the parks fell into poor condition and deer were able to escape and run wild which created the heredity for the wild deer in the UK today (Chapman and Chapman, 1997; The British Deer Society (BDS), 2016; Baker *et al.*, 2017).

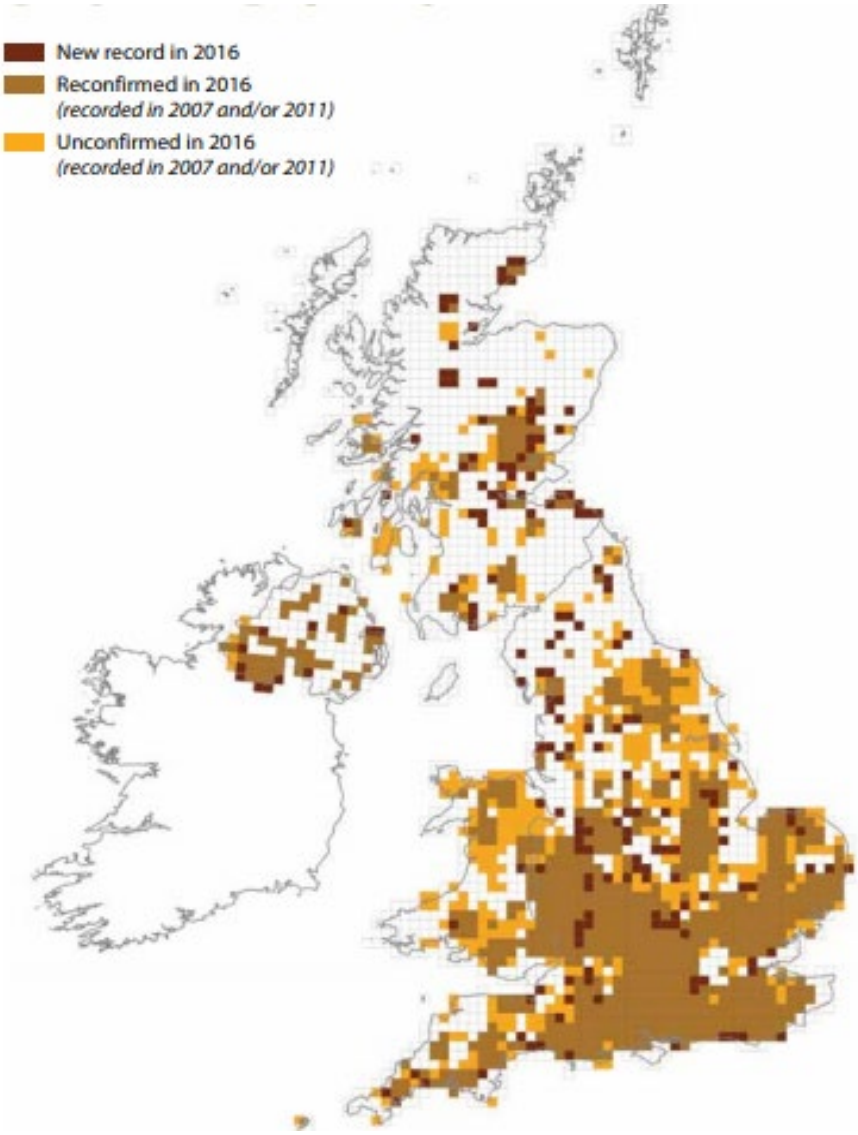


Figure 1.7: *Map to show fallow deer distribution in the UK (The British Deer Society (BDS), 2016).*

European fallow deer (*Dama dama dama*) are one of the main species of deer in the UK (Gill and Morgan, 2009). A sub-species of fallow deer, also considered a separate species by some, is the Persian fallow deer (*Dama mesopotamica*) and is listed as endangered on the IUCN red list (Werner *et al.*, 2015). Due to captivity in Mesolithic (8000 BC – 2700 BC) or Neolithic (10,200 BC - 2000 BC) periods,

fallow deer have the lowest genetic variation, compared to red and roe deer, due to genetic bottlenecking (Baker *et al.*, 2017). It is also believed that, around c. 1000BC fallow deer almost became extinct (Arslangündoğdu *et al.*, 2010). Nearly all species were extinct from northern regions of Europe towards the end of the Pleistocene glaciation period, the last one which Europe experienced was around 10,000 years ago, the only individuals to survive were in southern Europe in refugia (Arslangündoğdu *et al.*, 2010). The climate warmed and species migrated north to colonise however, during the Neolithic period fallow deer from wild populations in the Middle East were introduced by humans into Europe but in captivity (Chapman and Chapmen, 1980). This introduction of captivity of fallow deer led to a genetic bottleneck in this species, reducing their genetic diversity compared to other species in the UK (Ludwig *et al.*, 2011).

This lack of genetic variability can still be seen in fallow deer today. Early studies determined the genetic variability of fallow deer using allozymes (Pemberton and Smith, 1985; Hartl, Schleger and Slowak, 1986; Randi and Apollonio, 1988). Since then there have been further studies which have confirmed low genetic variation using microsatellites in fallow deer, however, mostly dinucleotide STRs were used, see section 1.4, (Baker *et al.*, 2017), and mtDNA sequencing (Ludwig *et al.*, 2011). Baker *et al.* (2017) found that the fallow deer in the UK had the greatest deviation from Hardy-Weinberg Equilibrium (HWE), compared to all other locations sampled, see Table 1.3. They concluded this low genetic diversity to be due to several populations of fallow deer grouped together as one, also known as the Wahlund effect (Wahlund, 1928). This especially low genetic diversity found in English fallow deer increases the value of widening the sampling area to outside of the UK for this project to achieve a representative sample to develop primers that work in all populations of fallow deer across the world.

It is feared that the low level of genetic variation may affect fallow deer in the future in terms of evolution with increased homozygosity and suffering from effects of inbreeding depression (Baker *et al.*, 2017). Furthermore, during a previous project (Barnard *et al.*, n.d.) lack of conservation between other deer species and fallow deer was identified due to difficulties in cross species amplification. The study attempted to use primers developed for red deer and roe deer to amplify DNA in fallow deer. A total of 117 primers were selected from the literature and tested on fallow deer DNA. Out of this, only five primers were suitable, most did not amplify at all. This was concluded to be due to lack of genetic conservation between red and roe deer with fallow deer, this was likely due to greater genetic distance between these species and is corroborated by

Baker *et al.* (2017). Table 1.3 shows the diversity statistics for the microsatellites tested in Baker *et al.* (2017). The low genetic variation and lack of genetic conservation across deer species identified by previous research is based on very few loci/individuals and the use of dinucleotide loci. It was apparent that new tetra-nucleotide STR markers were needed to be developed from the newly sequenced fallow deer genome.

Table 1.3: *Microsatellite diversity statistics for fallow deer samples at each location (Baker et al., 2017).*

Country	Location	N Samples	A*	AR	F_{IS}	H_o	H_e	P-Values**
Spain								
Central	North Madrid and Toledo	15	1.7	1.51	0.21	0.16	0.20	0.75
S West	Cadiz Huelva and Jaen	25	2.3	1.75	0.13	0.20	0.23	0.83
North	Aiguamolls and Rialp	15	1.6	1.47	-0.21	0.21	0.18	0.91
	Asturias	15	1.8	1.60	0.15	0.15	0.17	0.22
Portugal								
E Central	Castelo Branco	13	1.7	1.61	0.15	0.19	0.23	0.89
Italy								
North	Grosetto	15	2.3	2.21	0.04	0.42	0.43	0.25
	Siena	14	2.6	2.42	0.08	0.45	0.48	0.63
	San Rossore	14	2.2	2.05	0.11	0.37	0.41	0.19
England								
S Central	Essex	15	2	1.88	-0.03	0.34	0.33	0.29
East	Oxfordshire (P)***	9	2.2	2.13	0.28	0.25	0.33	0.04
	Cambridge and Bedfordshire	15	2	1.90	-0.04	0.35	0.34	0.08
North	Shropshire (P) and Cheshire (P)	9	2.2	1.76	-0.12	0.29	0.26	0.18
	Lincolnshire (P)	15	2.1	1.9	0.01	0.30	0.31	0.18
	Norfolk (P)	24	2.2	2	0.08	0.35	0.38	0.07

Ireland								
N Central	Roscommon, Galway, and Clare	13	2.1	1.97	0.19	0.25	0.31	0.64
East	Wicklow	13	1.8	1.62	0.29	0.16	0.22	0.08
South	Waterford, Kilkenny, and Tipperary	10	2	1.94	0.29	0.26	0.36	0.45
Sweden								
South	Kristianstad, Maltesholm	15	2.1	1.97	-0.03	0.35	0.34	0.36
Hungary								
East	Gyula	7	2.2	2.10	-0.14	0.34	0.30	0.93
West	Labod	7	2.1	2.07	-0.06	0.39	0.37	0.52
Canada								
South	British Colombia, Sidney Island	24	1.6	1.46	-0.09	0.15	0.14	0.83
Turkey								
	Antalya, Düzlerçami (P)	24	1.7	1.53	-0.22	0.22	0.18	<0.01
Rhodes		24	2.7	2.12	0.23	0.26	0.34	0.01
Bulgaria								
South	Eastern Rhodopes	14	1.7	1.62	0.10	0.19	0.21	0.95
Totals and averages		**364**	**2.04**	**1.86**	**0.06**	**0.28**	**0.29**	

**Abbreviations: A, number of alleles; AR, allelic richness; F_{IS}, inbreeding coefficient; H_e, expected heterozygosity; H_o, observed heterozygosity.*

***P-values are indicated for multilocus Hardy–Weinberg equilibrium tested against the null hypothesis.*

****Samples originating from animals in enclosed parks are denoted parenthetically by 'P' after location names.*

1.3 Wildlife Genetics: Background and Role in Combating Wildlife Crime

Wildlife crime is of growing concern and effects most countries across the globe. There are many organisations which are targeted at combating wildlife crime, including: The Lusaka Agreement Task Force, The INTERPOL (International Criminal Police Organisation) Wildlife Enforcement team, The NCUW (National Wildlife Crime Unit) and The Association of Southeast Asian Nations (ASEAN) Wildlife Enforcement Network (Scanlon and Farroway, 2016). These organisations investigate the entire supply chain of wildlife crime, from poachers to buyers. As with INTERPOL, they conduct numerous operations in order to disrupt the supply chain of illegal wildlife in order to dismantle criminal networks, apprehend illegal shipments and bring the perpetrators to justice (International Criminal Police Organisation (INTERPOL), 2018). INTERPOL have had many successful international operations including: Project Predator, Project Wisdom, Operation Thunderstorm, Operation Thunderbird, Operation Worthy II, and Operation Paws. These operations have resulted in over 5,000 seizures of illegal wildlife shipments, and almost 4,000 arrests with an estimated 80% conviction rate, which for wildlife crime is high (International Criminal Police Organisation (INTERPOL), 2018).

In 2010 the first collaboration of governmental organisations was formed for the sole purpose of combating wildlife crime; this is known as The International Consortium on Combating Wildlife Crime (ICCWC). The ICCWC is a joint consortium involving INTERPOL, The United Nations Office on Drugs and Crime (UNODC), The Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES) Secretariat, The World Customs Organization (WCO) and The World Bank, see Figure 1.8 for a map to show the ICCWC global network (Scanlon and Farroway, 2016; The International Consortium on Combating Wildlife Crime (ICCWC), 2016). The global network spans across the world with headquarters in the USA and Europe. Having this strong global presence is to act as a deterrent for criminals seeking to participate in such crimes as well as to strengthen the criminal justice system by ensuring that perpetrators of wildlife crimes will face serious sanctions for their actions (The International Consortium on Combating Wildlife Crime (ICCWC), 2016).

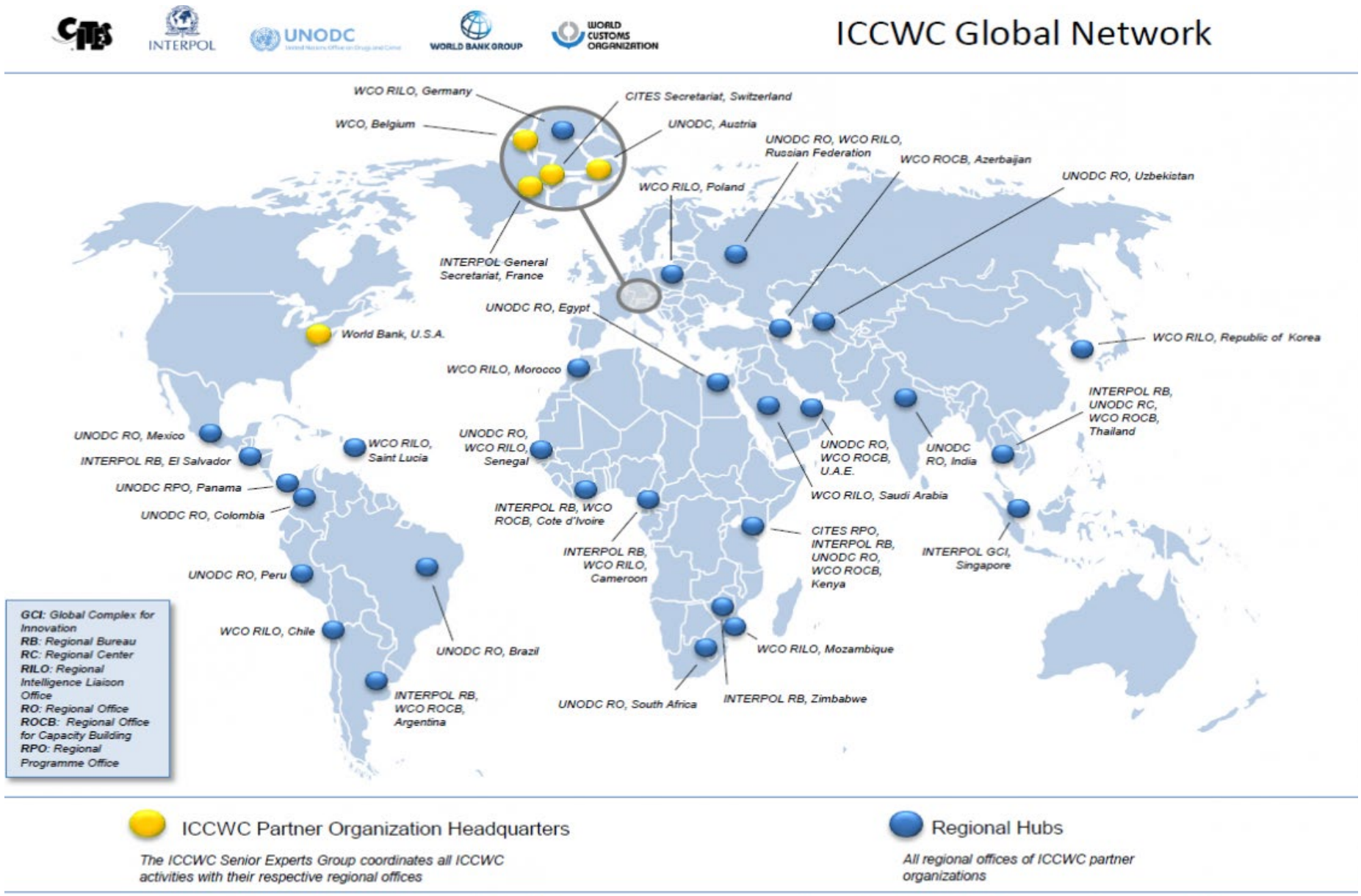


Figure 1.8: *Map of the ICCWC global network (The International Consortium on Combating Wildlife Crime (ICCWC), 2016).*

As well as seizure data and the type of conviction expected from different crimes, as discussed in sections 1.1 and 1.2, criminal justice data is as important which is derived from arrests, prosecutions, and convictions. Criminal justice data complements seizure data in order to fully understand the levels of wildlife crime and whether the efforts to combat wildlife crime are in fact working (UNODC (United Nations on Drugs and Crime), 2016). As mentioned, for some of INTERPOL's projects and operations they achieve an 80% conviction rate. However, in reality it is often much lower than this. Exact global conviction rates are not yet available, efforts are being made by World WISE to improve the way this data is collated and recorded so it can be included in the future UNODC World Wildlife Crime Reports (UNODC (United Nations on Drugs and Crime), 2016; UNODC (United Nations on Drugs and Crime), 2020).

In the UK the numbers of prosecutions and subsequent convictions has been recorded by The House of Commons Library (Finlay, Sutherland and Sturge, 2019). Between the years 2013 to 2018 there were a recorded 654 prosecutions, yet these only resulted in 359 convictions, this conviction rate is only ~50%. Table 1.4 shows the number of prosecutions and the associated convictions for the years 2013 to 2018, organised by type of offence. As the table shows, the most common crimes are those committed under the Hunting Act 2004, The deer Act 1991, and The Protection of Badger Act 1996. The most common being, hunting a wild animal with a dog, which, as mentioned in section 1.2.1, is a common yet prohibited way of poaching deer.

Table 1.4: *Prosecutions and convictions associated to wildlife crime offences for England and Wales for the years 2013 to 2018, taken from The House of Commons Library (Finlay, Sutherland and Sturge, 2019) and updated to include the latest figures for 2018. Data Source: Ministry of Justice Experimental statistics: Prosecutions and Convictions data tool.*

	2013	2014	2015	2016	2017	2018	Total
Prosecutions	**180**	**89**	**135**	**110**	**71**	**69**	**654**
Offences under Wildlife and Countryside Act 1981	11	5	3	8	5	4	**32**
Offences under Deer Act 1991	9	5	13	25	6	4	**55**
Offences under Protection of Badgers Act 1992	50	18	26	16	6	13	**116**
Offences under Wild Mammals (Protection) Act 1996	0	0	1	4	2	1	**7**
Offences under Hunting Act 2004: of which	**110**	**64**	**92**	**57**	**52**	**47**	**375**

- Hunting a wild mammal with a dog	87	60	84	49	49	45	**329**
- Knowingly permitting land to be entered or used in the course of hunting a wild mammal with dogs	0	1	0	0	0	0	1
- Knowingly permitting a dog to be used in the course of hunting a wild mammal	0	0	0	2	2	2	4
- Participating in a hare coursing event	12	0	8	6	1	0	27
- Attending a hare coursing event	11	3	0	0	0	0	14
Convictions	**99**	**49**	**71**	**66**	**38**	**36**	**359**
Offences under Wildlife and Countryside Act 1981[(b)]	9	3	0	0	3	2	**15**
Offences under Deer Act 1991	8	0	11	18	5	3	**42**
Offences under Protection of Badgers Act 1992	26	11	14	13	6	8	**70**
Offences under Wild Mammals (Protection) Act 1996	0	0	1	4	2	1	**7**
Offences under Hunting Act 2004(c): of which	**56**	**35**	**45**	**31**	**22**	**22**	**189**
- Hunting a wild mammal with a dog	44	35	41	29	21	19	170
- Knowingly permitting land to be entered or used in the course of hunting a wild mammal with dogs	0	0	0	0	0	0	0
- Knowingly permitting a dog to be used in the course of hunting a wild mammal	0	0	0	2	1	2	3
- Participating in a hare coursing event	4	0	4	0	0	0	8
- Attending a hare coursing event	8	0	0	0	0	0	8

The figures presented relate to defendants for whom these offences were the principal offences for which they were dealt with. When a defendant has been found guilty of two or more offences it is the offence for which the heaviest penalty is imposed. Where the same disposal is imposed for two or more offences, the offence selected is the offence for which the statutory maximum penalty is the most severe.

Globally, much is being done to combat wildlife crime, yet with only around half of all wildlife crime offences actually making it to a successful conviction more must be done. This poor conviction rate, which is reflected world-wide, is due to several factors for example, there is no structured system for recording wildlife crime, this impedes the analysis of trends and recognising where the hotspots are and thus the areas lacking in judicial and scientific infrastructure (Wildlife and Countryside Link, 2018). Furthermore, if trends cannot be analysed, areas which seek most attention cannot be appropriately prioritised, and resources cannot be allocated effectively. Lack of successful conviction is also linked to the police lacking the ability and willingness to address wildlife crime effectively due to restrictive budgets and resources (Wildlife and Countryside Link, 2018). Evidence supporting prosecutions against wildlife crimes is still fairly new, therefore there is much uncertainty surrounding its admissibility, especially as DNA evidence is not available for all species, lack of evidence will cause many prosecutions to fall through (Wildlife and Countryside Link, 2018). The Lack of comprehensive sentencing guidelines and the ever more common use of the internet to facilitate and commit wildlife crime, just add to the mounting problems surrounding the fight against wildlife crime and needs addressing in order to combat it (Wildlife and Countryside Link, 2018).

Recommendations in order to improve the situation in the UK are set out on an annual basis by The Wildlife and Countryside Link Organisation, this is the largest coalition for protection of the environment and wildlife, comprising of 58 organisations (Wildlife and Countryside Link, 2018; Wildlife and Countryside Link, 2019). Amongst other recommendations, they recently identified the need for regular funding to go to the NWCU in order for them to continue their investigations into wildlife crime and for the resources and effective assistance needed to do this, and for there to be a push in raising awareness of wildlife crime (Wildlife and Countryside Link, 2019). The general public often know nothing of the wildlife crime situation in the UK, let alone the whole world, thus there must be a drive to send information out and to inform people on how to report wildlife crime, if they suspect it (Wildlife and Countryside Link, 2019). In order to remedy the concerns surrounding the admissibility of evidence in court, further forensic scientific research is required (Iyengar, 2014) (see section 1.3.2 for species identification, 1.3.3 for geographical origin, and 1.3.4 for individual identification) such as the project concerned here (See sections 1.3.4 and 1.4), as well as further specified forensic wildlife casework laboratories (see section 1.3.1) (Ogden, 2010; Linacre *et al.*, 2011).

1.3.1 Forensic Wildlife Laboratories

As part of reassuring the admissibility of evidence in court, it must be first assured that the evidence has been generated in a secure, accredited, and accomplished laboratory.

There are four main options for carrying out forensic wildlife work, which are outlined in Figure 1.9. In order for a facility to undertake forensic wildlife genetic case work there are several aspects which must be considered. Quality assurance must be in place, in order for results from the laboratory to be accepted by a court of law quality assurance must be of a very high standard which must be proven (Ogden, 2010). Quality assurance ensures the prevention of errors by checking analysis, rectifying mistakes, and feeding back any improvements to the process prior to the release of the results in the case notes for court. Accreditation to the International Standardisation Organisation (ISO) 17025 will prove that the laboratory works to a high standard.

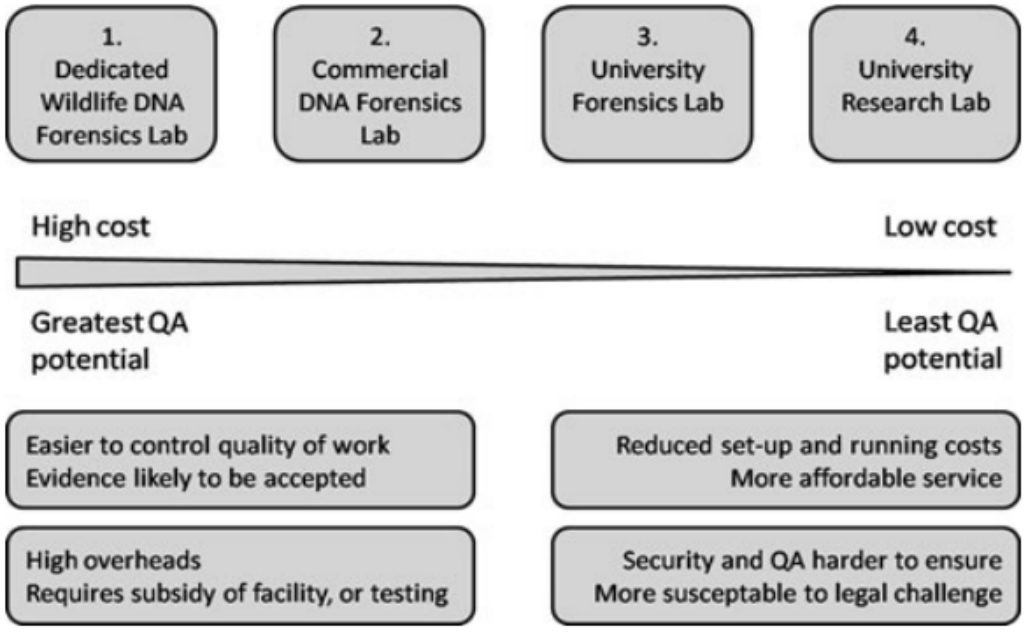


Figure 1.9: Outline of the 4 main approaches for undertaking forensic wildlife DNA analysis (Ogden, 2010).

ISO 17025 also sets out guidance on appropriate laboratory layout for all forensic case work DNA laboratories.. This is to ensure a standard workflow can be upheld which limits contamination and unnecessary errors (International Organisation for Standardisation, 2017). The laboratory is set out in a way that a sample can flow through the laboratory without being contaminated by staff, this is maintained by the positive and negative air pressures for the separated pre-PCR and post-PCR laboratories, respectively.

Funding is a second important factor which must be considered by a facility wishing to work on forensic wildlife genetics. University laboratories are often able to offer the service for free or a subsidised fee due to core funding that the organisation received, however with these facilities QA is often compromised (Ogden, 2010). The other funding option is that the laboratory is fully commercialised and will therefore increase the cost to the consumer, but the laboratory will work on a full cost recovery basis. However, quality assurance is much higher, therefore results are more likely to be accepted in court (Ogden, 2010). Lastly, it is key to consider staffing. It is often said that it is easier to train a wildlife geneticist to do forensic case work than to teach a human forensic geneticist to do wildlife genetics (Ogden, 2010). This must be considered when developing new facilities in countries which do not have an existing one because staff with the appropriate skills and interests may be scarce. However, the best choice would be an experienced wildlife geneticist with forensic interests and history of good quality assurance (Ogden, 2010).

Terry Grosz (A desk officer for the United States Law Enforcement's endangered species Fish and Wildlife Service (FWS)) understood the importance of wildlife crime and the protection of endangered species and wanted to reduce the impact of crimes against wildlife as much as possible, so he fought for a dedicated wildlife laboratory in the United States of America (Neme and Leakey, 2014). He won his plea and the first forensic laboratory of its type, dedicated solely to wildlife forensics was established in the 1980s (Neme and Leakey, 2014). Nowadays there are several laboratories across the world which are either solely dedicated to wildlife forensics, such as the RhODIS laboratories in South Africa who are dedicated to fighting rhino poaching (Erhodis.org, 2010) and the National Wildlife Forensic laboratory Malaysia who deal with a broad range of casework types for forensic wildlife genetics (Wildlife.gov.my, 2015); or laboratories which have specific departments within human forensic laboratories which undertake wildlife forensic casework; such as Cellmark laboratories in the UK (Cellmarkforensics.co.uk, 2016). A full list of laboratories which undertake work in forensic wildlife conservation are recorded on the website for The Convention of International Trade in Endangered species of Wild Fauna and Flora (CITES), these facilities are authorised by CITES to complete this type of work (CITES (The Convention on International Trade in Endangered Species of Wild Fauna and Flora), 2019).

As the field of forensic wildlife genetics expands more dedicated laboratories will be required and these factors will need to be considered. Wildlife genetics is far

more complicated than human forensics with a far wider target species range and the requirement of new testing kits and methodologies to be developed for these species (Ogden, 2010). Priority given to wildlife forensics is often far lower than human forensics, therefore there is great issue in resourcing the laboratories as its maintenance is seen as expensive (Ogden, 2010). Conventionally, most wildlife forensic genetics has been carried out by academics with expert knowledge in the subject area, however, this reduced integrity of the results and evidence produced as often staff are not forensically trained and the laboratories are only research grade (Ogden, 2010). As recommended by LINK, section 1.3, with the spreading of knowledge on forensic wildlife crime, the importance of funding, and further resources to combat it, it is hopeful that forensic wildlife laboratories will become common practise in the world of forensic science.

Forensic wildlife genetics laboratories are often tasked with three main areas of analysis:, species identification, geographical origin and individual identification of crime scene and case-type samples (Ogden and Linacre, 2015). These topics arise in various scenarios such as determining if items have been obtained illegally i.e. poaching, this may raise all three topics as the further the evidence can be traced back the stronger the evidence is against the suspect.

1.3.2 Guidelines for the use of Non-Human DNA for forensic purposes.

Non-human DNA in forensics can be either the illegal trade, possession or illegally obtained species which is submitted as evidence in a case which may either be against a person, property, or cases of animal cruelty, including instances where the offender is the animal (Linacre *et al.*, 2011). Sample collection and transportation, laboratory analysis, statistical analysis, reference material and how the results are presented as evidence all come under scrutiny by lawyers in court trying to find doubt in the evidence (Ogden, Dawnay and McEwing, 2009). New methodologies using this kind of evidence must be validated for forensic use. For a new multiplex kit to be accepted by the field of forensic science and for it to be accredited for use in a court of law, the multiplex kit must be trusted to be reliable when called upon in forensic casework at court. If the intention is for use in forensic casework, such as in this study, it must be validated in accordance with the Scientific Working Group for DNA Analysis Methods (SWGDAM).

SWGDAM set out guidelines which DNA analysis methodology must meet to be validated for use in forensic science (SWGDAM, 2012). SWGDAM separated the validation into developmental and internal. Developmental validation must demonstrate precision, accuracy and reproducibility of procedures prior to their use on forensic case work (SWGDAM, 2003). The same standard of result must be achievable regardless of the analyst or laboratory. Internal validation must be completed by in-house testing of the laboratory to demonstrate the limitations and the reliability of the developed procedure (SWGDAM, 2003). Included in validation, the study must also be subjected to peer review to test the technologies and methodologies before they can be implemented on forensic case work (SWGDAM, 2012). The peer review shall further demonstrate the precision and accuracy of the kit developed, as mentioned in section 1.4, providing evidence that the results achieved first time are reproducible. Once the procedure has been validated, both internally and developmentally and passes the peer review, the procedure is validated for use on forensic case work and therefore, can withstand any scrutiny it may face in court providing strong and reliable evidence (Ogden, Dawnay and McEwing, 2009). SWGDAM have published further guidelines to this, which cover most aspects of DNA analysis for forensic purposes including individual identification analysis and can be found on their website (www.swgdam.org/publications, 2020).

Furthermore, there are 13 recommendations for the forensic use of non-human DNA, set out by the International Society of Forensics Genetics (ISFG) (Linacre *et al.*, 2011). The 13 recommendations cover all aspects of development of new

methodologies applicable to forensic science (Linacre *et al.*, 2011), however the following are specific to individual identification:

> *"**Recommendation #6:** Primers used to amplify polymorphic DNA should be tested to ensure specificity and reproducibility and should be published in the public domain.*
> ***Recommendation #7:** If repeat-based polymorphic loci are used for individualization, tetrameric short tandem repeat systems should be used preferentially.*
> ***Recommendation #8:** Sequenced allelic ladders are essential for the accurate designation of alleles and should be used in all STR typing. The number of repeats should be the basis of reporting of results rather than using only the size based on the number of base pairs of any samples tested.*
> ***Recommendation #9:** In relationship testing, the mutation probabilities of the STR alleles should be estimated if encountered, or at least the probability of a mutational event occurring should be considered when there is genetic inconsistency at a single or few loci while all other loci show genetic consistency.*
> ***Recommendation #10:** Relevant population and forensic genetic parameters including allele frequencies should be estimated.*
> ***Recommendation #11:** A kinship factor should be determined and applied in any calculation. The type of kinship factor applied should be stated clearly and justification should be made for the factor incorporated." (Linacre et al., 2011, pg. 503)*

Furthermore, the ISFG recommendations cover reporting of results (recommendation 12) and if a laboratory is planning to do continued work with non-human DNA it is recommended (recommendation 13) that they seek accreditation to the ISO17025 standard (Morling *et al.*, 2002; Linacre *et al.*, 2011), as discussed in section 1.3.1. ISO17025 sets out the requirements of a laboratory's competence and calibration for forensic purposes, if met the laboratory will achieve accreditation from the United Kingdom Accreditation Service (UKAS). It is recognised that the area of forensic science dealing with non-human DNA is limited compared to human forensics, however this is rapidly changing. The ISFG recommendations enable laboratories to infiltrate good practice measures and help them recognise the same level of analysis should be employed as is with human forensic samples (Linacre *et al.*, 2011). The rate of the growth of the wildlife forensic field is not set to slow down, with more laboratories meeting international standards the better chance it has in achieving successful prosecutions in the court of law.

1.3.3 Species Identification

Often species identification is done via observation of morphological traits, however for most forensic purposes this will not be possible because the full animal is not present (samples are often blood swabs, tissues, bone, horn and tusk items) (Linacre and Tobe, 2011; Ghemrawi *et al.*, 2021). Species identification is important in its own right and has may applications within wildlife forensics, for example: determining if an item is made from a protected substance like ivory, Halal testing, and food forensics. Ensuring that a sample is of forensic importance saves police and law enforcement time, money and resources investigating something which is not actually illegal, for example pewter or bone can be made to look very similar to ivory and would take genetic testing to determine what species it is or if it is in fact animal derived at all (Linacre and Tobe, 2011; Ghemrawi *et al.*, 2021). Halal testing is also crucial for the Muslim community, to ensure products are not being sold to them which falsely contain pig substances as this would be against their religion (Murugaiah *et al.*, 2009). Therefore, species ID is key to tell vital differences.

Food forensics is highly important across the globe and concerns the lack of attention paid to the labelling of food products. This is not just an isolated issue, over the years there have been high profile cases portrayed in the media. For example, there have been 2 cases of donkey/horse meat found in food products (Primrose, Woolfe, and Rollinson, 2010; O'Mahony, 2013) which had been sold in the major retailers in the UK. Every country has, or should have, a governing body for food safety which checks what species are within food products to ensure accurate labelling of products containing meat or fish (Chin Chin *et al.*, 2016). The incorrect labelling of seafood products, for example, has a major impact on accurate representation of stock numbers of sustainable fisheries which negatively impacts on conservation efforts to increase numbers of endangered species (Chin Chin *et al.*, 2016). Incorrectly labelling food items is also a major risk to health as consumers may be eating items that they are allergic to and therefore will cause the consumer to lose confidence in the supplier (Chin Chin *et al.*, 2016). Not correctly labelling seafood products in accordance with the guidelines, as mentioned, is hazardous to health and therefore, is a criminal offence and will result in prosecution.

A genetic marker for species identification must have low variation within a species but must be significantly variant between species to allow for species differentiation (Goodwin, Linacre and Hadi, 2011; Smart *et al.*, 2021). The

variation between species must be significant enough to allow differentiation of closely related species (Goodwin, Linacre and Hadi, 2011; Ghemrawi *et al.*, 2021; Smart *et al.*, 2021). This is vital in order to distinguish between closely related species and species that have separate sub-species within them where one specie is not protected but another is. For example, fallow deer are not endangered whereas the sub-species of fallow deer, *Dama mesopotamica* or Persian fallow deer, are critically endangered, therefore they are a highly protected species. For the purpose of forensic samples to establish whether there is protected species present, mtDNA sequencing is used for species identification. Mitochondrial DNA does not undergo recombination; therefore, all maternal relatives share the same sequence unless a mutation arises (Linacre and Tobe, 2011). Furthermore, the core function of mtDNA genes means they are highly conserved across the animal kingdom yet the long evolutionary history of the mtDNA genome means sufficient variation has arisen between species, making it ideal for species identification. Additionally, mtDNA is protected by a protein coat surrounding the mitochondria, therefore, there is less risk of degradation increasing the chance of extracting quality DNA from forensic samples (Linacre and Tobe, 2011).

Species identification in forensic science is becoming more common, however, there is not one standard gene used for this analysis. Mitochondrial D-loop sequences contained within the Control Region (CR), NADH dehydrogenase subunit 1 (ND1), 12s and 16s rRNA, Cytochrome C Oxidase Subunit I (COI) and Cytochrome b (Cyt b) have all been used for species identification for forensic purposes (Table 1.5) (Linacre and Tobe, 2011; Ghemrawi *et al.*, 2021). The 12s and 16s rRNA, D-loop sequences and ND1 mtDNA genes have a lower discriminatory power between species but higher intra-species variation (Clifford *et al.*, 2004; Zhang *et al.*, 2006; Ghemrawi *et al.*, 2021), therefore they are less useful compared to the Cyt b and COI genes which are commonly observed in the literature (Table 1.5). The best marker for species identification may vary between taxa depending on how much sequence is needed for an accurate ID as well as the availability of good reference data housed in databases such as the Barcode of Life Data System (BOLD: https://id.boldsystems.org/) and the National Center for Biotechnology Information (NCBI: https://blast.ncbi.nlm.nih.gov/Blast.cgi).

Table 1.5: *Examples of species identification for forensic purposes using mtDNA genes between 2005 to 2020.*

Species	MtDNA Marker	Sequence Length	Type of Sample	Reference
Chinese Sika Deer	CR	1079 bp	Skin, blood	Wu *et al.* (2005)
Great White Shark	Cyt *b*	511 bp	Tissue from dried fins	Shivji *et al.* (2005)
Stone Cray Fish	COI	650bp	Muscle	(Schubart and Huber, 2006)
Roe Deer	Cyt *b*	900 bp	Meat, hair	An *et al.* (2007)
Tiger	Cyt b	147bp-151bp	Traditional east Asian medicine	Linacre and Tobe (2008)
Asiatic Black Bear	Cyt *b*	175 bp	Bile crystals	Peppin *et al.* (2008)
Guanaco	Cyt *b*	774 bp	Meat	Marín *et al.* (2009)
Whale species	Cyt *b*	400 bp	Meat (sashimi)	Baker *et al.* (2010)
Commercial fish products in Italy	Cyt *b* COI	300 bp 600 bp	Muscle	Filonzi *et al.* (2010)
Crocodile	COI	645 bp	Crocodile skin handbag	Eaton *et al.* (2010)
Asian Elephant	CR	137 bp	Ivory idol	Gupta, Thangaraj, *et al.* (2011)
Parrots and Cockatoos	12S Cyt *b*	230 bp 500 bp	Embryonic tissue	Coghlan *et al.* (2011)
Reedbuck	COI	650 bp	Meat, carcass	Dalton and Kotze (2011)
Lowland Tapir	Cyt *b*	1070 bp	Meat	Sanches *et al.* (2011)
Indo-Chinese Spitting Cobra	Cyt *b*	472 bp	Hair, tanned leather	Gaur *et al.* (2012)
Cypriot Mouflon	Cyt *b*	1140 bp	Blood stains from suspect's	Barbanera *et al.* (2012)

			boot & jeans, carcasses	
Scarlet Macaw	COI	648 bp	Feathers	Abe, Hayano and Inoue-Murayama (2012)
Sturgeons and Paddlefish	Cyt *b* CR	850 bp 725 bp	Caviar	Doukakis *et al.* (2012)
Elephant	Cyt *b* CR	357 bp 377–630 bp	Ivory seals	Lee *et al.* (2013)
Southeast Asian Monitor Lizards	ND1	1181 bp	Muscle, liver, skin, scale clippings	Welton *et al.* (2013)
Indian Wild Pigs and Domestic Pigs	Cyt *b*	421 bp	Solid tissue	Gupta *et al.* (2013)
Squid (*Cephalopoda Teuthida*)	COI	~660 bp	Mantle Tissue	Katugin *et al.* (2015)
Lobster (*Panulirus homarus*)	COI	650 bp	Tissue	Al-Senaidi *et al.* (2015)
Several species of Crustacea	COI	584 bp–712 bp	Tissue	Mantelatto *et al.* (2016)
Fish	COI	>600 bp	Tissue	Bingpeng *et al.* (2018)
Spiney Lobster	COI	230 bp	Tissue	Govender *et al.* (2019)
Dolphins and other Cetacean	Cyt b	148 bp	Muscle tissues	Lopez-Oceja *et al.* (2019)
Flies	COI	651 bp	Tissue	Fuentes-López *et al.* (2020)
Ray	COI	599 bp	Tissue	Rodrigues Filho *et al.* (2020)

**Abbreviations: CR, Control region; Cyt b, Cytochrome b; COI, Cytochrome C Oxidase Subunit I; ND1, NADH dehydrogenase subunit 1*

1.3.4 Geographical Origin

In order to determine popular trade routes of goods sold on the black market it is essential to determine the geographical origin of the material. This will highlight areas of increased illegal activity; therefore, it will indicate where an increase in security is required (Wasser *et al.*, 2007). Table 1.6 shows examples of previous work that has been completed in order to determine species geographical origin from case-type samples.

Table 1.6: *Examples of geographical location analysis using mtDNA and STRs between 2007 and 2016.*

Species	DNA Marker	Type of Sample	References
African Elephant	STRs	Tusk	Wasser *et al.* (2007)
Seahorses	Cyt b and CR	Tail and Tissue	Sanders et al. (2008)
Italian Fish	Cyt b and COI	Muscle	Filonzi et al. (2010)
Chimpanzees	STRs	Blood	Ghobrial et al. (2010)
Moose	STRs	Solid Tissue	Ball et al. (2011)
Elephant	Cyt b and CR	Ivory Seals	Lee et al. (2013)
SE Asian Monitor Lizard	ND1	Muscle, Liver, Skin and Scale Clippings	Welton et al. (2013)
Leopard	STRs	Solid Tissue	Mondol et al. (2014)
Rhino	Cyt b and 12s rRNA	Horn	Kumar Jha et al (2016)

**Abbreviations: STRs, Short Tandem Repeats; CR, Control region; Cyt b, Cytochrome b; COI, Cytochrome C Oxidase Subunit I; ND1, NADH dehydrogenase subunit 1*

Geographical location assessment can be done using mitochondrial DNA (mtDNA) sequencing or STR analysis, as shown by Table 1.6. mtDNA is the common analysis used, however, STRs are used also (Ghobrial *et al.*, 2010; Sanders *et al.*, 2008). Regions of the Cytochrome b (Cyt b) gene and Cytochrome C Oxidase Subunit I (COI) gene on the mtDNA genome can be sequenced to assess the geographical origin. The control region (CR) of the D-loop is utilised if

the full genetic structure of the sample is intact, due to its size (<1000bp) by assessing ancestry using haplotype analysis (Terencio *et al.*, 2012; Iyengar, 2014).

A haplotype is a genetic lineage which usually gives a family level identification and a haplogroup corresponds to collection of haplotypes that have a distinct geographical/population level distribution. Defining an individual's haplotype also known as mitotype (mitochondrial haplotype) enables the haplogroup to which it belongs to be identified and thus, geographical origin to be established (Grahn *et al.*, 2015). However, this requires a lot of prior work, the mitotypes for different populations of a species must be known in order to assign the forensic sample to one of them.

Grahn *et al.* (2015) has successfully analysed the mitotypes of domestic cats from across the world for forensic applications. They found that 83% of all 1395 samples of domestic cats analysed belonged to 12 major mitotypes. Ottolini *et al.* (2017) specified this for UK cat populations. Using cat hair found on suspects clothing Ottolini *et al.* (2017) showed it to be possible to link a suspect to a crime scene or victim via matching mitotypes. Frequency of the mitotype was assessed and the strength of evidence determined. The evidence was found to be of high evidential value because the mitotype of the cat was rare (B-UK1, a subtype of the common mitotype B), therefore the random match probability (RMP) was low, meaning it was more likely that secondary transfer occurred from the victims cat to suspect than from another random, unrelated cat (Ottolini *et al.*, 2017). However, not all mitotypes are as rare, therefore this type of evidence requires a more discriminatory test for it to be used more frequently in forensic investigations (Ottolini *et al.*, 2017). By sequencing additional parts of the mtDNA genome, i.e. the COI region and the Cyt b region, the discriminatory power of the evidence can be increased, this is something that is currently being investigated (Ottolini *et al.*, 2017). Forensic haplotyping can be employed with wildlife crime cases and is investigated in fallow deer as part of this project.

However, forensic samples are often degraded therefore, complete CR sequencing is often not possible (Iyengar, 2014). In this instance short tandem repeat (STR) genotyping can be employed for a more specific geographical assignment. STRs are a more reliable and specific method of geographical location assignment. This is done via assigning individual multi-locus genotypes to a particular population (Wasser *et al.*, 2004; Wasser *et al.*, 2007; Wasser *et al.*, 2008). The studies by Wasser *et al.* explore the use of 16 microsatellite loci to infer geographical origin of the sample. The Wasser *et al* (2007) study showed that using STR's, geographical origin could be located within 500km, making it possible to differentiate between forest and savannah elephant ivory which not only has a forensic significance, but also has a substantial conservation management

significance. The study used allele frequencies and the likelihood of an individual's genotype to assign it to a population (Wasser *et al.*, 2004; Wasser *et al.*, 2007; Wasser *et al.*, 2008). Therefore, with a high degree of confidence, it is possible for an individual to be excluded from all but one population, therefore the individual tested will belong to that population (Iyengar, 2014). Furthermore, Wasser *et al.* show the importance of a large reference database with the increase in accuracy of results based of reference data and those without. The downside is that prior work is needed, as with mtDNA geographical assessment, to create an allele frequency database which is time consuming and costly, both of which is not usually in abundance during a forensic case.

Geographical origin can be used to map trade routes of samples from the location they were seized back to where it originated. Figure 1.10 is a map of common trade routes, which import and export from Europe, most of which show the origin as Asia (Banos-Ruiz, 2017 When shipments pass though main ports, documents are falsified in order to mask the origin country, which is why DNA analysis is key for assessing world trade routes and determining where increased security is needed in order to catch the perpetrators before the shipment leaves the country of origin (UNODC (United Nations Office on Drugs and Crime), 2017).

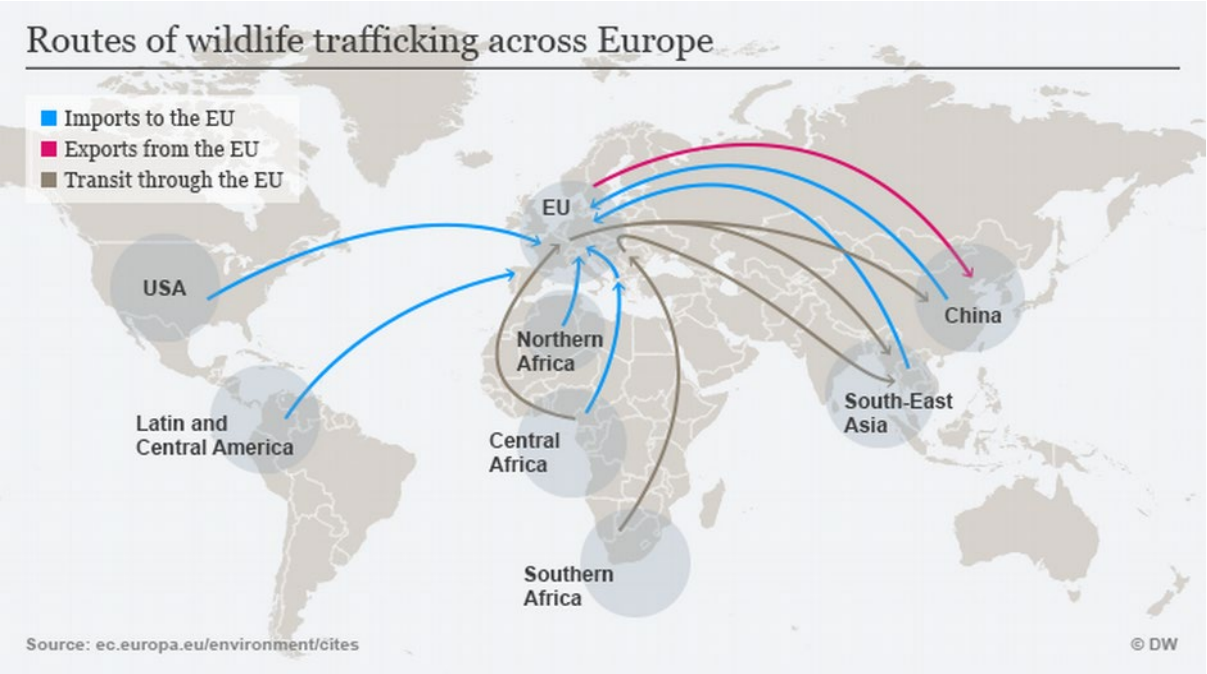


Figure 1.10: *Trade routes of illegally obtained wildlife goods across the world importing and exporting from the UK (Banos-Ruiz, 2017).*

1.3.5 Individual identification

The identification of an individual is the crucial question asked by human forensics in order to identify the perpetrator of a crime. This is also important in wildlife forensics. However, in wildlife forensics, it is often also necessary to trace illegally obtained items, such as horn and ivory, back to its origin which too can be answered by individual identification (Smart *et al.*, 2021). DNA can be retrieved from untreated horn and ivory, blood from weapons, ivory, horn and bone statues. The DNA would be extracted and then STR profiled using the same methodology as for human DNA analysis (Iyengar, 2014). However, new STR multiplex kits need to be developed for every species of animal that needs to be profiled for forensic purposes (Harper *et al.* 2013). There are several STR multiplex kits already available for a broad range of species, although, the majority of kits are not commercially available, the only kits currently commercially available are for cats and dogs). See Table 1.7 for publications which have successfully identified individuals of particular species using STR profiling.

If a database is available of individual DNA profiles from animals in national parks and reserves then material such as ivory, which is seized from a shipment, can be traced back to the place it came from by matching the individual it belongs to (Harper *et al.* 2013). This will confirm that the evidence has been illegally poached or blood on weapons is from an animal that has not been lawfully killed. This does require extensive DNA testing of all animals in nature reserves and parks and their profile uploaded to a database. However, only species at risk of poaching would need to be profiled, such as rhinoceros (Harper *et al.* 2013). The benefits far out way the time, expense and effort which is involved in this analysis. STR profiling is highly discriminatory, therefore, acts as robust evidence in court against the accused if they are guilty (Peng *et al.*, 2015). However, more often than not, police request a sample to be analysed which is from a species which has not been worked with prior to the case, therefore, the selection of primers, fragment analysis and sequence analysis would all need to be conducted first before the sample can be STR profiled which, again, is time consuming (Johnson, Wilson-Wilde and Linacre, 2014). The development and validation of STR multiplexes for key species would be an efficient way of speeding up wildlife forensics, such as the rhino STR kit developed by the team behind RhODIS® (Rhino DNA Index System).

Table 1.7: Examples of publications which have used STRs to achieve individual identification for forensic purposes.

Species	Type of Sample	Number of STR Loci Used	Reference
Wild Boar	Blood stains, carcass	12	Lorenzini (2005)
NE Brown Bear	Hair, blood stain	12	Eiken *et al.* (2009)
Wolf	Teeth	12	Caniglia *et al.* (2010)
Tiger	Claw and decomposed skin	7	Gupta, *et al.* (2011)
Sardinian Mouflon	Blood stains from scene of crime, carcass	16	Lorenzini *et al.* (2011)
White Tailed Black Cockatoos	Feathers	20	White *et al.* (2012)
Rhino	Horn	22	Harper *et al.* (2013)
Wolves and Dogs	Saliva samples from carcasses	12	Caniglia *et al.* (2013)
Tortoise	Buccal swabs	14	Mucci, Mengoni and Randi (2014)
Red Deer	Tissue or Blood	10	Szabolcsi *et al.* (2014)
Parrot	Blood tissue	106 (~15 per species)	Jan and Fumagalli (2016)
Hen Harrier	Tissue (eggshell, feathers, buccal swab)	8	Van Hoppe *et al.* (2016)
Python	Tissue	24	Ciavaglia and Linacre (2018)
Black Bear	Tissue or Blood	11	Meredith, Adkins and Rodzen (2020)

RhODIS, in 2013, developed an STR multiplex kit for rhino, primarily in South Africa, but they are beginning to broaden their scope to the rest of Africa and India due to its success (Harper *et al.* 2013; Harper *et al.* 2018). They created a database of rhino from national parks and reserves in South Africa (Harper *et al.*

2013). DNA sampling was done using a field recovery kit developed by RhODIS, this enabled vets and nature reserve staff to collect DNA from the animals any time they were close enough to do so (either the rhino was undergoing veterinary treatment or was killed by poachers and a carcass had been discovered). Table 1.8 shows the success that that RhODIS has had so far with their work. They have processed 5,800 Rhino poaching cases which included the production of 120 case reports for police personnel to use to achieve a prosecution (Harper *et al.* 2018). So far, 9 successful prosecutions have been achieved and more are in progress (Harper *et al.* 2018). The success which has been shown in Table 1.8 expresses the major benefits of this work and research by RhODIS for combating wildlife crime, therefore, there is a demand for this tool for more species also effected by criminal activity. So far 9 successful prosecutions have been achieved and more are in progress (Harper *et al.* 2018). The success which has been shown in Table 1.8 expresses the major benefits of this work and research by RhODIS for combating wildlife crime, therefore, there is a demand for this tool for more species also effected by criminal activity.

Table 1.8: *The success of the RhODIS multiplex kit along with the corresponding case outcomes (Harper et. al, 2018).*

Match result	Poaching site	Species / subspecies	Match probabilities	Status of case
2 horns matched carcass 1 and1 horn matched carcass 2	KNP, SA	White rhinoceros (C.s. simum)	4.20×10^{-9} 2.03×10^{-10}	2012/08/23: 29 years and 3 months
Horn matched carcass	Hoedspruit, SA	White rhinoceros (C.s. simum)	3.80×10^{-8}	2013/03/28: 15 years each
2 horns matched carcass 1 and 1 horn matched carcass 2	Waterberg, SA	White rhinoceros (C.s. simum)	1.96×10^{-8} 1.35×10^{-8}	2012/11/14: 10 years
2 horns matched carcass 1 and 1 horn matched carcass 2	KNP, SA	Black rhinoceros (D.b. minor)	4.18×10^{-12} 1.03×10^{-12}	2013/08/15: 14 years
The profile from clothing matched carcass	Limpopo, SA	White rhinoceros (C.s. simum)	1.19×10^{-8}	2015/02/24: 8 years
3 horns matched 3 carcasses	ORTIA, HiP, SA	White rhinoceros (C.s. simum)	8.79×10^{-8} 1.45×10^{-9a} 8.08×10^{-8}	2016/11/01: R800 000 fi ne or 6 years
Horn matched blood on carpet	OPC, Kenya	Black rhinoceros (D.b. michaeli)	8.98×10^{-22}	2017/05/12: 11 years
14 horns with 2 horns matched to a carcass	ENP, Namibia	Black rhinoceros (D.b. bicornis)	4.74×10^{-13b}	2016/10/30: 14 years
6 horns with 2 horns matched to a carcass	KNP, SA	White rhinoceros (C.s. simum)	4.55×10^{-9}	2014/01/16: 15 months

1.4 Short Tandem Repeat (STR) Multiplexing

STR multiplexing is the term given to the amplification of several regions of DNA at once via Polymerase Chain Reaction (PCR) (Goodwin, Linacre and Hadi, 2011). A reaction mixture contains several primers, which have been optimised to work together, to amplify more than one STR locus. Multiple loci are needed in order to get an accurate individual identification, with a high power of discrimination (Goodwin, Linacre and Hadi, 2011).

STR's have a repeat motif of 2bp – 6bp (Merkel and Gemmell, 2008). The alleles in human DNA profiling kits are usually less than 350bp's (Goodwin, Linacre and Hadi, 2011). The entire human genome has been sequenced and around 20 STR loci are utilised for forensic use (Goodwin, Linacre and Hadi, 2011). A 4-5bp repeat motif is desirable which is classified as one of the following simple repeats; simple repeats with non-consensus repeats, a compound repeat or complex repeats (Goodwin, Linacre and Hadi, 2011). Tetra- and penta-nucleotide repeats are preferred over di- and tri-nucleotide repeats as they are increasingly more discriminatory as there is less chance of two individuals sharing the same region, and there is less risk of stutter (15% stutter with tetra-nucleotides but 30% with di- and tri-nucleotides) with larger repeat units making it easier to interpret sample mixtures (Butler *et al.*, 2014; Butler, 2005; Gill *et al.*, 1994). Tetra-nucleotide repeats are preferred as penta-nucleotides are rare, however, di-nucleotide repeats are commonly used in non-human analysis as they are more regularly found (Berger *et al.*, 2014; Linacre *et al.*, 2011). For a highly discriminatory multiplex, approximately 15 to 20 loci should be amplified using the kit, the more loci analysed and profiled the more discriminatory the STR-multiplex becomes.

To create a multiplex there are several features which are required of STRs which make them desirable for multiplexing (Goodwin, Linacre and Hadi, 2011). These include alleles which must be discrete and different to one another. STR loci must be highly discriminatory, polymorphic and robust, they must not have genetic linkage to STR loci also under analysis, artefact formation such as spikes and blobs, pull up peaks, stutter, non-template addition, sample overloading, elevated baselines, allelic dropout and locus dropout must be minimal and STR loci must also be able to amplify when run in a multiplex (Goodwin, Linacre and Hadi, 2011). Primer dimers can be an issue, to avoid primer dimers excessively large regions of complementarity should be avoided that causes the primers to bind to each other rather than the DNA region of interest (Butler, 2005).

Each newly developed multiplex kit will require optimisation to some degree for the reaction components and the PCR parameters to ensure the kit is working at its best (Butler, 2005). Multiplex optimization is far more complex than a singleplex reaction as there are several primers each may have slightly varied annealing temperatures, but all must work at the same temperature (Butler, 2005). Several stages of optimization will be needed to ensure a well-balanced multiplex PCR with each gaining similar yields (similar peak heights on an electropherogram – RFU) is achieved by having the correct primer concentrations (van Hoppe *et al.*, 2016). After the multiplex is optimised and is well balanced it must be tested for precision and accuracy via reproducibility and repeatability testing by a separate scientist on 2 different occasions to see if the same results are gained, the same results should be achieved regardless of the operator or laboratory (van Hoppe *et al.*, 2016). Furthermore, creating a new multiplex for a non-human species must be validated for forensic use prior to casework, see section 1.3.2.

However, even though knowledge on human STRs is advanced there is yet to be few polymorphic STR loci found in deer or fallow deer in particular as their genome was yet to be fully sequenced. In these instances, cross species amplification can be adopted. Using primers previously developed for the identification of individuals in closely related species can save time and money. However, there are a few issues which must be addressed with cross-species amplification in closely related species. It is not always 100% successful. There are conserved primer binding regions which makes cross-species amplification possible, however, there may be mismatches in these regions causing an increase of null results such as studies on red, roe, fallow and mule deer (Poetsch *et al.*, 2001; Jobin, Patterson and Zhang, 2008).

Primer binding regions are conserved within a species genome and may be common across closely related species. This allows primers to work across several species, creating the potential for a multiplex involving more than one species as a result (Poetsch *et al.*, 2001). Some loci are more conserved than others and extensive primer testing is required to determine conserved regions (Küpper *et al.*, 2008). Conserved loci are termed type 1 and type 2 (O'Brien *et al.*, 1993). Type 1 loci are gene coding evolutionary conserved loci and type 2 are anonymous DNA markers such as microsatellites (Slate *et al.*, 1998). Type 2 have a greater level of polymorphism than type 1 within a species (Slate *et al.*, 1998). It has been found that conserved STR loci will amplify across a larger taxonomic range and the use of conserved primer binding sites improves success of cross-species

amplification (Moore *et al.*, 1991; FitzSimmons, Moritz and Moore, 1995; Gortari *et al.*, 1997). Improving this further by using highly conserved regions of STR flanking sequence during primer design, increasing the number of species the locus will amplify (Küpper *et al.*, 2008). If no band is visible on an agarose electrophoresis gel it can be deducted that the loci is not conserved in the tested species (Slate *et al.*, 1998).

As previous research found (Barnard *et al.*, n.d.), due to the low genetic diversity of fallow deer their genome is highly conserved, therefore, using primers developed for red and roe deer did not work. To overcome this, this project used Next generation Sequencing (NGS) to sequence the whole genome of fallow deer in order to design primers that were known to work. Whole genome sequencing (WGS) was required to locate trinucleotide, tetranucleotide and pentanucleotide STR's to develop accurate, highly discriminatory primers. However, this is expensive, and many projects lack funding they are unable to perform this, which is why there are fewer studies involving the individual identification aspect of forensic wildlife genetics compared to the growing number of species identification research projects (Vanek *et al.*, 2019). See section 1.5 for further discussion on NGS and WGS.

An allele frequency database is crucial for the calculation of profile frequency and likelihood ratio's because these calculations are based of the individuality of the profile within the population, in other words, the likelihood that the case profile will match that of another unrelated individual from the same population (Goodwin, Linacre and Hadi, 2011). Profile calculations enable the strength of evidence of a match between two profiles to be deduced to assess the significance of the match which is essential for court reporting and therefore, how powerful it is in gaining a successful prosecution (Buckleton, Bright and Taylor, 2016). The underpinning of all profile calculations are fundamental to forensic genetics is the Hardy-Weinberg Law (HW Law) (Hardy, 1908; Stern, 1943; Goodwin, Linacre and Hadi, 2011). The HW Law facilitates the statistical representation of the relationship between allele and genotype frequency within an ideal population. An ideal population is said to be in Hardy-Weinberg equilibrium (HWE) which then allows the genotype frequencies to be predicted from known allele frequencies, hence the requirement for an allele frequency database (Goodwin, Linacre and Hadi, 2011; Stern, 1943; Hardy, 1908). Each allele has a particular frequency within a population denoting how common or how rare the particular allele is enabling the frequency of the final profile to be calculated. Profile frequency is calculated by multiplying all the genotype proportions together (p^2x2pq), known as the product rule, which have been calculated using Hardy-

Weinberg equations and an appropriate population allele frequency database (Goodwin, Linacre and Hadi, 2011; Stern, 1943; Hardy, 1908). The profile frequency is also known as the Random Match Probability (RMP). The rarer the alleles within the profile are, the rarer the profile will become and therefore, the RMP will decrease. Rare alleles, which are not included in the allele frequency database, can be given the frequency of 0.01 as a minimum allele frequency or by using the minimum allele count 5/2N (N= number of individuals in the database) known as the allele ceiling principle (Goodwin, Linacre and Hadi, 2011; National Research Council, 1996). The Likelihood ratio is calculated by dividing the RMP by 1.

Other calculations would include FST (ϴ) and FIS. Where ϴ= FST when a sub-population is in Hardy-Weinberg equilibrium because FST is a correction to lower the likelihood of two individuals sharing the same DNA due to shared ancestry within the same sub-population (Weir, 2012). FIS is the inbreeding coefficient and accounts for inbreeding therefore, at homozygous loci, RMP is increased because within animal populations inbreeding is common, therefore, the chance of finding a homozygous locus increases (Ayres and Overall, 1999). To correct for the presence of subpopulations, important for the fallow deer population (see section 1.2.1.1) a theta (ϴ) value is applied to compensate for inbreeding, as mentioned, because if a sub-population is present there will be a higher degree of relatedness between these individuals compared to the entire population. For homogenous populations, a theta value of 0.01 is commonly used, however, for more isolated populations, for example fallow deer, a theta value of 0.03 may be used in the equations shown in Figure 1.11 (National Research Council, 1996; Balding and Nichols, 1994):

For homozygotes:

$$\text{Profile frequency} = \frac{[2\theta + (1-\theta)p_i][3\theta + (1-\theta)p_i]}{(1+\theta)(1+2\theta)}$$

For heterozygotes:

$$\text{Profile frequency} = \frac{[2\theta + (1-\theta)p_i][\theta + (1-\theta)p_j]}{(1+\theta)(1+2\theta)}$$

Figure 1.11: *Correction for subpopulation presence equations using the Balding Nichols method (Balding and Nichols, 1994).*

Furthermore, profile frequency corrections can be implemented to correct for sampling bias, known as the Balding correction (Balding, 1995). This is an issue sometimes unavoidable in forensic wildlife studies, when it comes to available samples.

1.5 Next Generation Sequencing (NGS)

The original widely used method of sequencing called Sanger Sequencing was developed in the 1970's (Sanger, Nicklen and Coulson, 1977; van Dijk *et al.*, 2014). Since its introduction, it has successfully been implemented to whole genome sequence (WGS) the human genome as part of the Human Genome Project, which was finally completed in 2004 (Venter *et al.*, 2001; Morey *et al.*, 2013). Since the completion of the Human Genome Project there have been enormous advances in sequencing, Sanger Sequencing has a low throughput and is very expensive to run, now a genome can be sequenced for less than $1000 using some technologies, comparing favourably to the original cost of $3 billion (Service, 2006; Haynes *et al.*, 2019). Furthermore, The Human Genome Project took 13 years to complete, nowadays the human genome can be sequenced in less than 48hrs using Next Generation Sequencing (NGS) technologies (Haynes *et al.*, 2019). WGS is the sequencing of an entire genome of any species. This can be done via Sanger sequencing (first generation sequencing), as mentioned, by NGS or newer Third Generation Sequencing technologies which are continually breaking boundaries.

Next Generation Sequencing (NGS) is also referred to as Massively Parallel Sequencing (MPS) or Second-Generation Sequencing by some authors (Morey *et al.*, 2013). A historic day for NGS in the field of forensic science was the 31st August 2011 at the 24th ISFG conference (de Knijff, 2019). The conference included three presentations targeted at dealing with the application of MPS in forensic science (de Knijff, 2019). NGS is described as the sequencing of up to hundreds of millions of amplified sequences at the same time thus, increasing throughput and reducing the need for fragment cloning used in Sanger Sequencing by adopting loop array sequencing enabling the parallel analysis of several samples (Yang, Xie and Yan, 2014). NGS produces relatively short reads and is inadequate in producing high-quality genome assemblies as structural variation remains unresolved (Chaisson, Wilson and Eichler, 2015). Despite this NGS has revolutionised the ability to quickly find genetic variation to aid developmental studies into forensic biology, genetic diagnostics, biotechnology, microbiology, and evolution (Morey *et al.*, 2013; Yang, Xie and Yan, 2014). NGS has a much higher throughput compared to Sanger Sequencing and thus can complete tasks much faster, an ideal attribute for forensic applications, see Figure 1.12 for the main differences between NGS and Sanger Sequencing when analysing STRs. As shown by Figure 1.12, NGS is able to identify much deeper sequence variation, rather than just determining the locus homozygous, as shown with traditional Sanger Sequencing on the right, reads 1 -6 show to have additional SNP variation at base

8, read 1 and 10 has a further SNP variant at the 3' end of the sequence and read 12 has a SNP variant at the 5' end (de Knijff, 2019). Therefore, a total of five additional sequence variants were detected via NGS, as shown by box 2 in Figure 1.12, this provides further discriminatory power to DNA evidence (de Knijff, 2019), including the potential to determine differences between monozygotic twins.

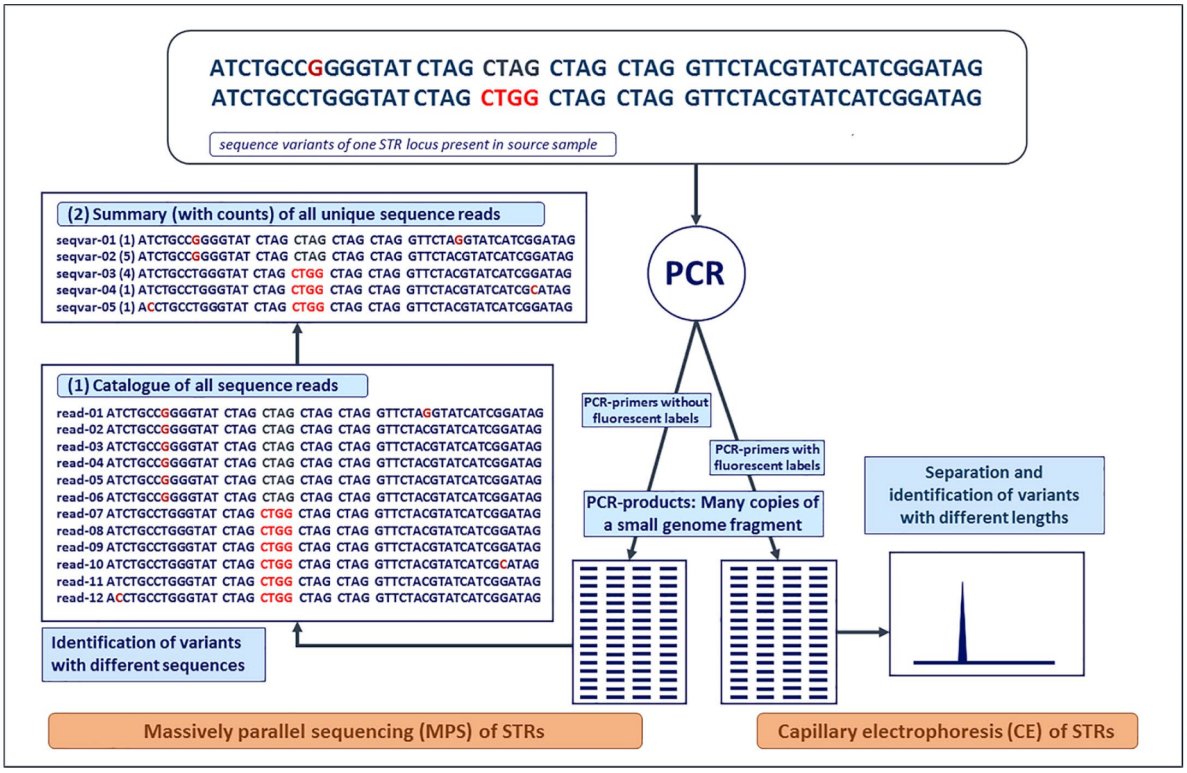


Figure 1.12: *the main differences between Next Generation Sequencing (NGS) and Sanger sequencing when analysing STRs (de Knijff, 2019).*

NGS methodologies are continually advancing as new technologies are frequently developed further increasing speed of sequencing and reducing associated costs (Morey *et al.*, 2013). Morey *et al.* (2013) reviews, in depth, the difference between Targeted enrichment strategies, which include enrichment by PCR, hybridisation capture and circularisation, and clonal amplification methodologies which include, bridge-PCR and emulsion PCR to perform NGS. The methodology depends on the purpose of the sequencing, it is discussed in the review which method is best used in each instance (Morey *et al.*, 2013). Many sequencing platforms now exist for NGS from many leading companies including Illumina, ThermoFisher Scientific, Pacific Biosciences and Oxford Nanopore, see Table 1.9 for an overview of the top sequencers currently available from these companies (Haynes *et al.*, 2019). NGS platforms are still expensive to buy, with prices ranging from around £2,000 into the hundreds of thousands, therefore it is not always viable for small

research laboratories to buy these platforms. Laboratories which have a turnover of, on average, 18,000 genomes a year, at an additional cost of £1,000 a genome may be able to justify the purchase (Haynes *et al.*, 2019).

Table 1.9: *Overview of the currently available leading NGS and TGS sequencing platforms.*

Provider	Sequencer	Cost	Max Read Length (bp)	Max sequence yield per run (Gb)
Illumina	iSeq 100 System	£17,774	2 × 150	1.2
	MiSeq System	£84,409	2 × 300	15
	NextSeq 550 System	£228,951	2 × 150	120
	NextSeq 1000 and 2000 Systems	£174,835 £278,904	2 × 300	540
ThermoFisher Scientific	Ion GeneStudio S5 System	£60,710	600	15
	Ion Torrent Genexus System	£240,375	400	20
Pacific Biosciences	PacBio Sequel IIe	£398,150	20,000	30
	Vega System	£136,020	19,630	60
	Revio system	£626,650	20,000	480
Oxford Nanopore	MinION	£1,607	20,000	48
	GridION X5	£56,299	4,000,000	240
	PromethION	£361,523	4,000,000	13,300

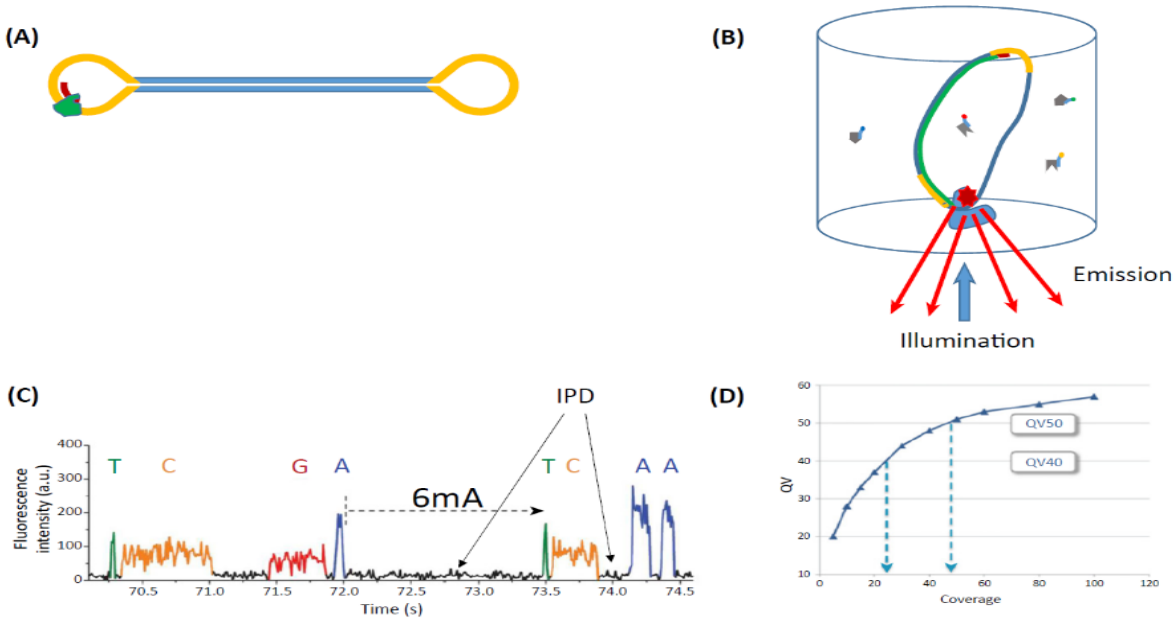
Previously, the main disadvantage of NGS is the short-read lengths which limits the technology's ability to produce high-quality genome assemblies. Third Generation Sequencing (TGS) over comes this and is starting to supersede NGS as more technologies and methodologies are being developed (Levene *et al.*, 2003; Morey *et al.*, 2013). TGS has the advantage of being able to sequence single molecules with longer read lengths, compared to NGS (Levene *et al.*, 2003). Furthermore, prior to sequencing there is no need for PCR amplification which avoids introduction of PCR artefacts and less risk of contamination due to the limited amount of sample manipulation (Morey *et al.*, 2013). Moreover, TGS and NGS use adopt sequencing-by-synthesis chemistries however, the methods of detection differ. Rather than detecting DNA bases based upon chemical incorporation, as with NGS, TGS technologies require an unmodified DNA strand

to detect via the physical recognition of bases (Pettersson, Lundeberg and Ahmadian, 2009). After the incorporation of a base, traditional methods of sequencing require 'wash and scan' steps which is time and reagent dependant, however, TGS does not need this thus is quicker and cheaper, therefore, TGS offers high quality, long read sequencing at reduced costs (Pettersson, Lundeberg and Ahmadian, 2009; (Schadt, Turner and Kasarskis, 2010; Morey *et al.*, 2013). New challenges are raised with TGS due to the difference in data generated as a result of sequencing, TGS has the ability to detect kinetic activity due to influences of secondary structures on sequencing speed and in the polymerase, therefore a new set of challenges is created for bioinformatics.

Several companies are currently developing and improving TGS technologies, however, Morey *et al.* (2013) and, later, van Dijk *et al.* (2018) has reviewed and provided comparisons of the technologies currently available. PacBio was the first company to release a TGS technology based upon 'Single-molecule real-time' SMRT sequencing (Eid *et al.*, 2009), following this in 2014 Oxford Nanopore Technologies (ONT) announced nanopore sequencing (Jain *et al.*, 2015), both of which produce long-reads (van Dijk *et al.*, 2018), see Table 1.9. See Figure 1.13 and Figure 1.14, respectively, for schematics of both these main TGS technologies. Less traditional methods of sequencing have also been developed as TGS technologies which include, the use of electron microscopy for the direct imaging of DNA. This technology been developed by Halcyon Molecular and ZS Genetics and allows the chemical detection and direct imaging of atoms in order to identify nucleotides (Lundquist *et al.*, 2008; Krivanek *et al.*, 2010). However, this methodology involves a PCR step which, as mentioned, is a significant disadvantage. Furthermore, a final TGS technology is being developed by IBM involving a nanostructured sequencing device in order to undertake Transistor-Mediated DNA Sequencing (Morey *et al.*, 2013). This technology electronically detects individual bases of a single molecule of DNA (Krems *et al.*, 2009). This technology is said to have extremely high throughput, several million bases per second, and the label-free optics-free characteristic makes it highly inexpensive compared to other technologies already on the market (Luan *et al.*, 2010; Morey *et al.*, 2013). However, due to a failed partnership with Roche and too many technical risks, IBM have yet been able to get this technology ready for commercial sale, However IBM continue to research the collaboration between DNA sequencing and nanotechnology (Herper, 2013). TGS is a very new

technology in comparison to Sanger Sequencing and NGS, it is continually evolving and the literature is constantly updating.

Figure 1.13: *Schematic representation of 'Single-molecule real-time' SMRT sequencing by PacBio (van Dijk et al., 2018).*

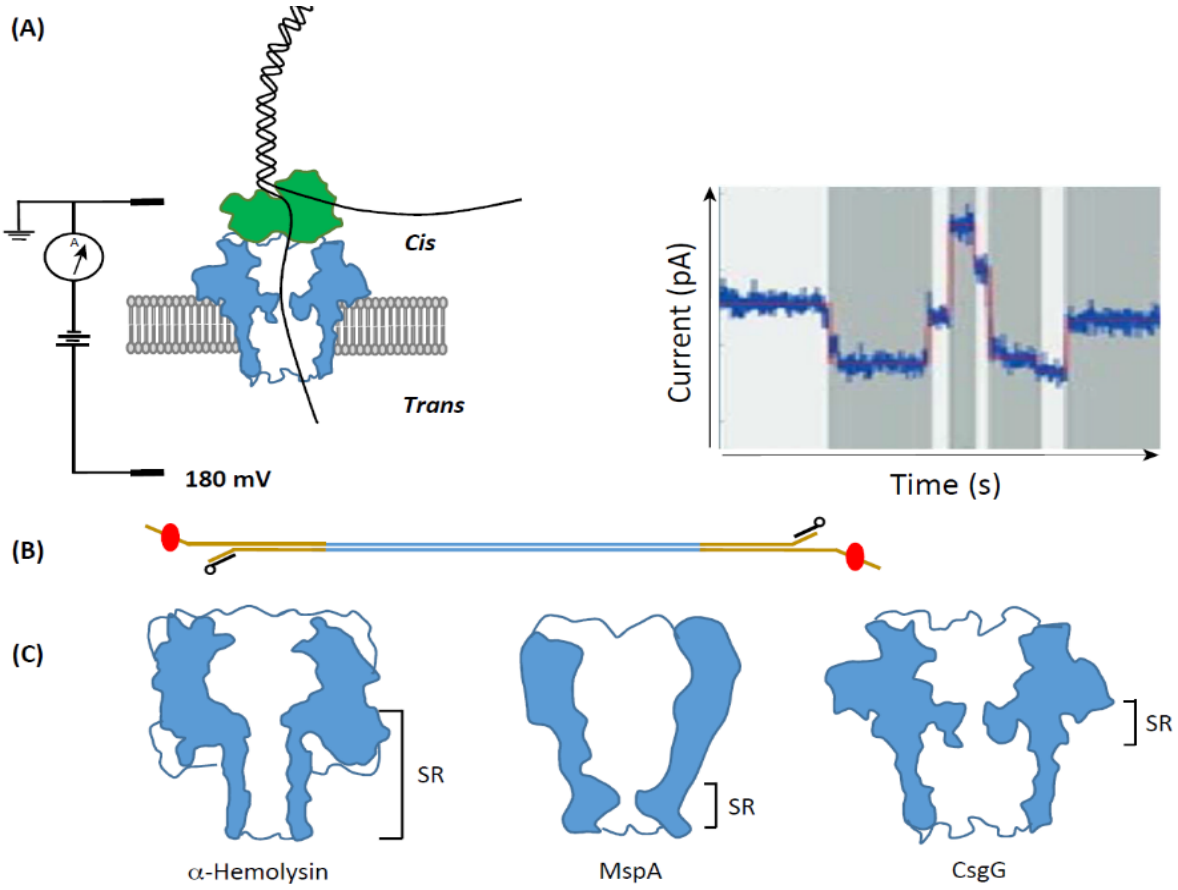


"(A) Library preparation comprises the ligation of hairpin adapters (yellow) to double-stranded DNA molecules (blue), thereby creating circular molecules called 'SMRTbells'. Next, a primer (red) and a polymerase (green) are annealed to the adapter.

(B) Schematic representation of a zero-mode waveguide (ZMW), a nanoscale observation chamber. The polymerase–primer–SMRTbell complex binds to the bottom of the ZMW through biotin–streptavidin chemistry. Note, however, that not all ZMWs will contain a DNA molecule because the library is loaded by diffusion. The polymerase incorporates fluorescently labeled nucleotides emitting a fluorescent signal on illumination from below. These signals are recorded by a camera in real time in a process called a 'movie'.

(C) In a movie, not only the fluorescence colour is registered, but also the time between nucleotide incorporations, called the interpulse duration (IPD) (black). The presence of an epigenetic modification, such as 6-methyladenosine (6 mA), results in a delayed IPD. Adapted with permission from Pacific Biosciences.

(D) Multiple 'passes' of the circular library can be combined into a circular consensus sequence (CCS) that increases in accuracy as the number of passes increases. Accuracy is expressed as the quality value (QV). Note that at ~25 passes, the accuracy reaches 99.999% (QV40), which is similar to the accuracy of Illumina sequencing. At ~50 passes, accuracy can even reach 99.9999% (QV50). Adapted with permission from Pacific Biosciences. The data indicated in the figure are based on a bacterial genome run on the Sequel system with 2.1 chemistry and 5.1 Sequel software." - (van Dijk et al., 2018, pg.670)



"(A) In an Oxford Nanopore Technologies (ONT) flow cell, two chambers (cis and trans) filled with ionic solutions are separated by a membrane containing a CsgG nanopore (blue; R9 chemistry). A nucleic acid (black) is electrophoretically driven through the pore in a controlled manner owing to the presence of a 'motor' protein (green). Note that the nucleic acid is unwound on translocation and only one strand passes through the pore. As the DNA or RNA translocates through the pore, current shifts are recorded in real time and are characteristic for particular k-mer sequences. The current shifts are graphically represented in a 'squiggle plot'.

(B) Typical nanopore library. Double-stranded DNA fragments (blue) often undergo an optional DNA repair step, as single-stranded nicks will lead to premature termination of nanopore sequencing. Then, the extremities are processed to create suitable substrates for ligation of adapters (brown). The adapters have 50 protruding ends to which a 'motor' protein is bound (red); this extremity will enter the pore first and thus sequencing occurs in the 50-to-30 direction. To the other strand of the adapter, an oligonucleotide with a cholesterol moiety (black) is hybridized, which will tether the library molecules to the membrane and increase the efficiency of nanopore sequencing.

(C) Different types of nanopores. The a-hemolysin pore, the MspA pore, and the CsgG pore, which is currently being used by ONT (R9 chemistry). The narrow 'sensing regions' (SRs) of the different pores are indicated; note that the MspA and CsgG pores have shorter SRs than a-hemolysin. As a result, a smaller number of nucleotides contribute to the signal, leading to more accurate base determination." - (van Dijk et al., 2018, pg.674)

Figure 1.14: *Schematic representation of nanopore sequencing by Oxford Nanopore Technologies (ONT) (van Dijk et al., 2018).*

Recently, two main NGS platforms have been utilised for forensic DNA analysis, these are the MiSeq FGx Forensic Genomics Systems by Illumina and the Ion S5 by ThermoFisher Scientific (Butler and Willis, 2020). NGS has been implemented in forensics for species identification, GMO testing, individual identification, biogeographical ancestry, mitochondrial genome analysis, Y chromosome analysis, microbiological analysis, Epigenetics useful to distinguish between monozygotic twins and phenotyping (Yang, Xie and Yan, 2014; Haynes *et al.*, 2019; Butler and Willis, 2020). Using NGS to identify species increases the degree of accuracy and the ability to distinguish between highly related species and sub-species. NGS has the ability to screen multiple genomic regions at the same time which enables the identification of all components of a food product, whether it be plant, animal, microalgae or fungal (Haynes *et al.*, 2019). This identifies if food products are incorrectly labeled which poses a risk to health or contains food stuffs which are illegal, such as registered endangered species (Chin Chin *et al.*, 2016). The ability to screen multiple genomic regions allows for the simultaneous sequencing of STRs and SNPs along with their associated stutter products which increases the level of genetic data generated in a single test vastly increasing the discriminatory power of a multiplex, this is applicable to both human and non-human DNA analysis (Butler and Willis, 2020).

However, there are challenges associated with the use of NGS on forensic case work, including the lack of nomenclature and reporting standards associated to lack of QMS and validation for forensic reporting using this technology (Butler and Willis, 2020). Secondly, there is a lack of compatibility with existing DNA databases, the data generated by NGS and TGS is far more in depth and complex compared to standard CE therefore, the national DNA database does not currently have the infrastructure to house it (de Knijff, 2019; Butler and Willis, 2020). Furthermore, unlike with CE, there is no standard nomenclature for STR alleles analysed using NGS, this further adds to the incompatibility with current DNA databases (de Knijff, 2019). Lastly, the lack of population data available for NGS means that population calculations cannot be satisfactorily undertaken as well as a lack of legislation surrounding the use of NGS is hampering the use of this technology in forensic reporting (de Knijff, 2019; Butler and Willis, 2020). The more forensic studies which utilise NGS for the above applications, the available population data will naturally increase and the legislation surrounding NGS in forensics will be updated.

In wildlife conservation genetics, reference genomes are highly sought after and are considered the gold standard base from which a wide range of population,

disease related, and forensic studies can be conducted. We are currently in the middle of a sixth mass extinction event (Ceballos *et al.*, 2015), it is estimated that during the next few decades over one million animal and plant species are at risk of extinction demonstrating that conservation efforts are now more essential than ever (Brandies *et al.*, 2019). Only a few of the endangered species listed on the IUCN as threatened have utilised the powerful technology available to us today, as mentioned. It is estimated that only 1% of the 13505 species listed as threatened have a fully sequenced genome available (Brandies *et al.*, 2019). The lack of reference genomes available is often due to the lack of understanding surrounding the potential of a reference genome, even if expertise wasn't the primary cause, lack of funds may still hinder the ability to generate reference genome which is not often available to small research laboratories and conservation teams (Fuentes-Pardo and Ruzzante, 2017; Brandies *et al.*, 2019). On NCBI there are 1842 animal genomes currently available, however only 6% of these are associated to threatened and endangered species listed on the IUCN Red List. However, the Wellcome Sanger Institute Tree of Life programme aims to change this by sequencing approximately 70,000 species genomes from Britain and Ireland (Wellcome Sanger Institute, 2019). Developing a reference genome which is capable of determining aspects of a species evolutionary past and biology as well as identifying variants, long repetitive regions, indels and mononucleotide regions (which can create gaps if sequencing resolution is low) is a costly task and often resides to institutions where considerable funding available such as the Wellcome Sanger Institute Tree of Life programme. An average eukaryotic genome 2.5Ggb in size is estimated to cost around $30,000 (Lewin *et al.*, 2018), however, if a high-resolution reference genome is required, it could cost even more than this for a reference genome at 100x coverage.

The value of a reference genome has been shown to outweigh the extreme cost of achieving it. In the case of the Tasmanian devil, the publication of the genome has shown to be crucial in its conservation (Miller *et al.*, 2011; Murchison *et al.*, 2012), and has led to its survival (Brandies *et al.*, 2019). This reference genome has been used by scientists all over the world and is regarded as the gold standard for reference genomes used in conservation genetics. As well as primer design this reference genome has helped to address a plethora of questions, see Table 1.10 for examples of the questions which have been addressed and how they were addressed (Brandies *et al.*, 2019). As shown by Table 1.10, the reference genome for the Tasmanian devil has been highly beneficial. These studies could not have been conducted without it and is likely that the species would now be extinct if it weren't for the publication of the reference genome. Despite not being directly linked to forensics, 22 microsatellite markers have been identified in a quicker and

more cost-effective way as a result of the reference genome (Gooley *et al.*, 2017). The Tasmanian devil is highly conserved and lacks genetic variation (Jones *et al.*, 2003), in this way this species can be compared to fallow deer, which this project is involved in. Therefore, a reference genome for fallow deer will be equally beneficial, not just for forensic purposes but for wider genetic studies (Miller, 2013).

Plans are in motion for the International Organisation for Standardisation (ISO) to coordinate the standardisation of NGS across a wide area of genomic disciplines (Haynes *et al.*, 2019). Despite this, others have published reference standards for the appropriate use of NGS (Hardwick, Deveson and Mercer, 2017; Mahamdallie *et al.*, 2018). The lack of Quality Assurance (QA), Validation and Quality Management System (QMS) for NGS associated with forensic genetic analysis specifically is effecting the power of NGS in criminal cases and the reliability of the technology cannot be backed up without it (de Knijff, 2019). In order for NGS to be accommodated by forensics as routine practice for DNA analysis de Knijff (2019) has put forward seven recommendations which he believes is essential for the implementation of NGS into regular forensic case work; (1) There should be standard nomenclature for MPS based STR alleles, some minimal STR allele nomenclature has been suggested by the DNA commission of the international society for forensic genetics (Parson *et al.*, 2016). To add to this, STRSeq is an international collaboration which is placing efforts into sequence diversity in common STR markers (Gettings *et al.*, 2017) and a revised guide for NGS of STRs has been developed (Phillips *et al.*, 2018) as well as ways of compacting the data (Young, Faris and Armogida, 2019). These are great advancements in STR nomenclature for NGS, yet the standardisation of these will make them useful for forensic DNA analysis for case work. Currently, the only standardised set of recommendations for the analysis of STRs using NGS has been published by SWGDAM via an amendment to their original guidelines surrounding the analysis of STRs in forensic science (SWGDAM, 2019). (2) It is recommended to set a standard for the minimum number of reads, this will provide an understanding as to what is needed to accurately call an STR allele under a plethora of conditions, forensic DNA samples are extremely varied, conditions are rarely the same (de Knijff, 2019). (3) Documentation of barcoding and sample-pool strategies to provide data on the full-spectrum of non-target or error reads (de Knijff, 2019). (4) Recommendations surrounding the NGS methodology used should be provided (de Knijff, 2019). (5) Recommendations are required for the formats needed to hold all NGS results (de Knijff, 2019). (6)

Standard requirements of bioinformatics tools used to analyse NGS results are needed (de Knijff, 2019). Finally, (7) Older software, originally developed for CE STR evidential value-based matching needed to be updated to include new allele designations via NGS (de Knijff, 2019). Since NGS and TGS technology is extremely good at deciphering mixed source DNA samples and revealing genetic information which cannot be gained using basic CE methods; QMS, QA, validation and standardisation of the NGS technologies are soon to be rectified for these methodologies will have a ground-breaking application to forensic science which cannot be ignored (de Knijff, 2019).

Sequencing technology has come on leaps and bounds since the introduction of Sanger Sequencing in the 1970's. The often-degraded nature of forensic DNA samples along with mixed source samples often pose as a challenge for the traditional sequencing methods adopted by forensic routine practise, NGS technologies offers a solution to this problem (Yang, Xie and Yan, 2014). The ability to distinguish between mixed source is offering great advancements within food forensics as it is now possible to determine all components within a food product (Haynes *et al.*, 2019), and it has been shown possible to find variation between monozygotic twins (Weber-Lehmann *et al.*, 2014; Yang, Xie and Yan, 2014). Furthermore, the generation of reference genomes, has been made far easier and is now more achievable with the introduction of the $1000 genome. However, a high-quality genome will still cost much more than this, however, the impact of reference genomes for conservation often outweighs the cost (Brandies *et al.*, 2019). As forensic NGS kits become more commercially available it is likely costs will continue to reduce. However at this present time, the compatibility of NGS and cost of the technology will impede the ability of NGS to replace conventional STR profiling any time soon (Yang, Xie and Yan, 2014). Moreover, until proper guidelines are in place for the use of NGS technology forensic case work will still continue to use current methodologies for now. NGS has also been crucial in the development of TGS technologies and continues to help the advancement of understanding living systems (Morey *et al.*, 2013; van Dijk *et al.*, 2018). NGS and TGS technologies have shown, what was once impossible has been made possible, advancements are happening fast and the future is hard to predict, yet exciting, however, it is clear that NGS will become an increasingly used routine tool (Morey *et al.*, 2013; van Dijk *et al.*, 2018; Haynes *et al.*, 2019; de Knijff, 2019; Butler and Willis, 2020).

Table 1.10: *Examples of how the Tasmanian Devil reference genome has been utilised in order to aid in the conservation of this species (Brandies et al., 2019).*

Reference Genome Use	Conservation Questions Addressed	Conservation Actions	Conservation Outcomes	Reference
Microsatellite development	Were the founders related?	Resolved relatedness of founders	Tool for selecting individuals for translocations based on genetic complementation	Hogg *et al.* (2015)
Genome-wide SNP analysis	Does the metapopulation have equal founder representation to ensure the maintenance of gene diversity? Is inbreeding accumulating in group housing and Maria island insurance populations?	Resolved parentage in group housing within the metapopulation Reconstructed pedigree of island population Informed translocation recommendations	Improved maintenance of genetic diversity across captive populations Increased genetic diversity of hybrid individuals at wild release sites	Gooley *et al.* (2017) Farquharson *et al.* (2019) Hogg *et al.* (2020)
The characterization of DFTD strains	How many DFTD strains exist?	Appropriate management of wild populations	Assisted in managing the spread of new DFTD strains	Hogg *et al.* (2016)
The characterization of immune genes	Can we develop a vaccine for DFTD? Can we improve Tasmanian devil immune diversity?	Immunization development and deployment Immune gene diversity analysis for informed translocation recommendations	Improved immune responses of devils released to the wild Improved immunogenetic diversity of released Tasmanian devils and their resultant offspring	Pye *et al.* (2018) Grueber *et al.* (2019
Primer design and SNP panel development				
Targeted SNP analysis				
Development of blocking primer for metagenomics diet analysis	What constitutes the complete diet of Tasmanian devils on Maria Island?	Investigating the impact of an introduced carnivore to island wildlife	Mitigation implemented to reduce the impact on highly consumed species	Margres *et al.* (2018) Hohenlohe *et al.* (2019
Alignment of re-sequenced genomes	Are devils evolving host-parasite resistance to DFTD?	Ongoing monitoring to ensure releases do not impact the evolution of potential resistance alleles	Assisted in understanding regions of the genome that are Potentially involved in DFTD resistance	
SNP Analysis, Annotation and GWAS				

Abbreviations: DFTD, Devil Facial Tumour Disease; GWAS, Genome-Wide Association Study.

2 The Aim and Objectives of The Study

The aim of this study is to better understand the genetic diversity of the British fallow deer population, either for forensic use or for a conservation standpoint. The project has been split into four studies; these are as follows:

- **Study 1: The Reference Genome.**
  - **Aim:** To sequence the genome of the fallow deer to reference genome standard.
  - **Objectives:**
    - WGS of the fallow deer genome via PacBio Sequel IIe Sequencing, including one sample from the UK.
    - Full genome assessment, including comparison with the Red deer (*Cervus elaphus*) genome. Analysis of genome in order to locate potential tetranucleotide polymorphic STR loci.

- **Study 2: Exploring the best approach for individual identification.**
  - **Aim:** To compare standard PCR and CE to a new novel method of Oxford Nanopore sequencing for the identification of fallow deer individuals.
  - **Objectives:**
    - Design primers and perform PCR optimisation. Screen primers for polymorphism against 26 individuals, via electrophoresis.
    - Select the best primers for inclusion into a multiplex and optimise the multiplex kit.
    - Screen ~250 samples across selected tetranucleotide loci, via amplicon sequencing, on the Oxford Nanopore GridION.
    - Use Bioinformatics to analyse sequencing and detect STR variation.
    - Determine if polymorphism has been missed previously when using lower sensitivity chemistry (Relying on PCR and Gels alone).

- **Study 3: The complete mitochondrial genome of *Dama dama*, and the assessment of genetic divergence between the species and Dama Mesopotamica.**
  - **Aim:** To produce a mitochondrial genome report, with complete genome annotation and phylogenetic analysis.
  - **Objectives:**
    - Isolation of the mitochondrial genome sequence from the WGS in study 1.
    - Conduct phylogenetic analysis comparisons with other *Cervidae* mtDNA genome data from NCBI. Calculate evolutionary divergence from other *Cervidae* species.

- **Study 4: Haplotype diversity of the British Fallow deer population.**
  - **Aim:** To determine haplotype diversity in the British fallow deer population and compare traditional Sanger sequencing to a novel Oxford nanopore sequencing method
  - **Objectives:**
    - Mitochondrial D-loop sequencing of ~200 British fallow deer on the Oxford Nanopore GridION and to determine the haplotype of each sample.
    - To conduct a population study of British fallow deer population and compare to the European fallow population.
    - Compare traditional Sanger sequencing to Oxford nanopore for the analysis of haplotype variation.

3 Study 1: The Reference Genome

3.1 Introduction

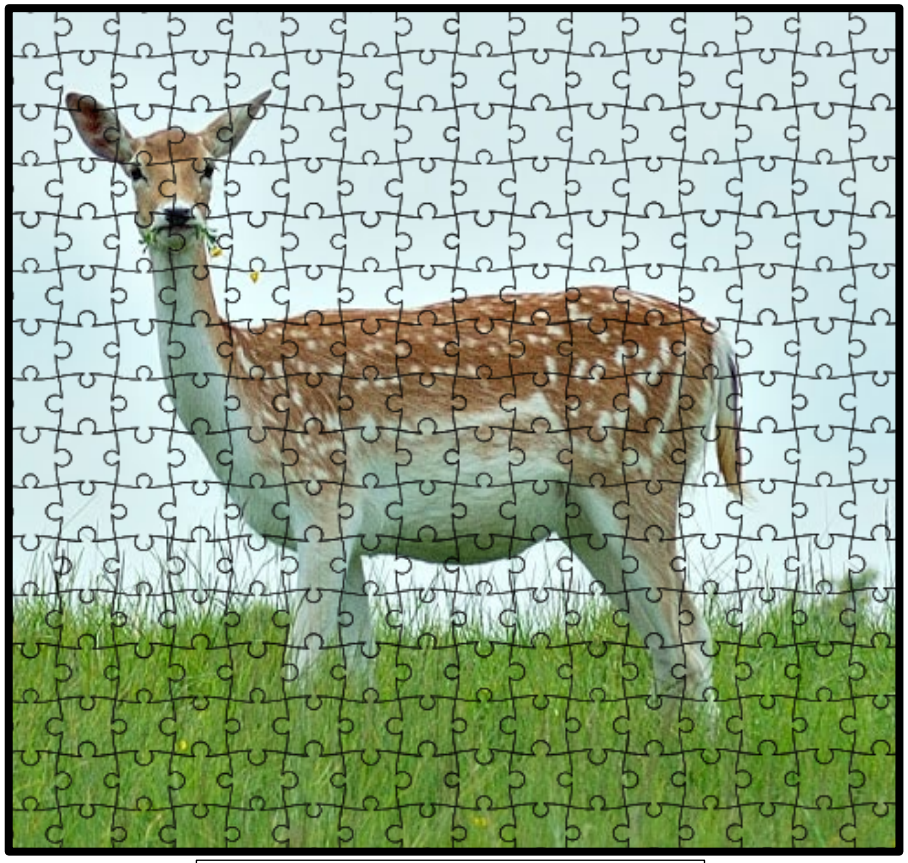
A reference genome is the annotated assemblage of chromosomal sequences, including genes, within an individual from a particular species. This makes the resultant reference genome a standard, to which all other sequencing of the same species, can be aligned to (Brandies *et al.*, 2019). Reference genome sequencing differs from normal WGS as the accuracy is greatly improved, ideally the reference genome should be gap free and completely annotated, with the lowest number of errors as possible (Xu *et al.*, 2016; Brandies *et al.*, 2019). The Human Genome Project set the bench mark for the ideal error rate at 1 error per 10,000 bases which equates to an error rate of 0.01% (Xu *et al.*, 2016). However, current NGS and TGS technologies often have higher error rates, yet they are far quicker at producing results, The Human Genome Project took 13 years to complete the human genome, currently you can achieve a highly accurate reference genome within 2 days (Haynes *et al.*, 2019).

Reference genomes are usually constructed via de novo sequence assembly and was utilised in this study. De novo sequence assembly is used when no reference genome already exists for the species in question (Nanoporetech.com. 2017). The sequences gained from NGS or TGS are assembled as contigs. The quality of the coverage gained from de novo sequence assembly is determined by the number of gaps in the contigs (Illumina, 2015). De novo sequence assembly generates highly accurate reference sequences providing the data to map genomes of novel organisms, enables the clarification of highly similar or repetitive regions and helps to identify structural variants and complex rearrangements, including inversions, translocations and deletions (Illumina, 2015). The more accurate the genome assembly, which is more, correct, complete and contiguous the easier it is to characterise genes and genomic regions, analyse structural variants and analyse evolutionary changes (Brandies *et al.*, 2019).

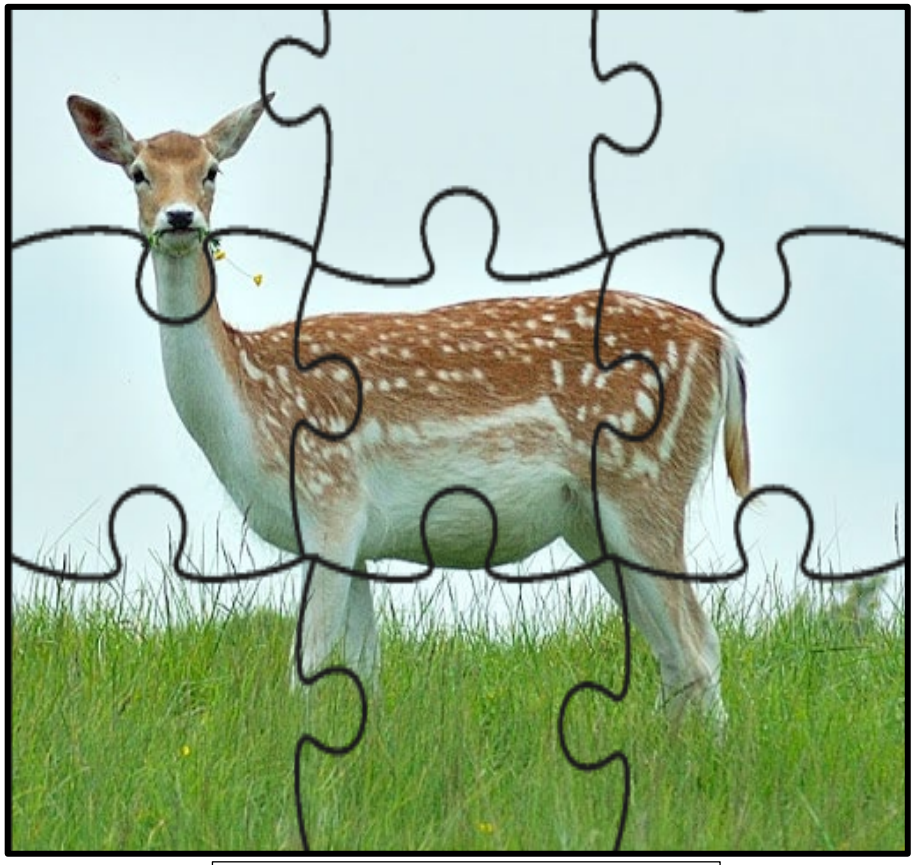
Reference genomes are highly important in conservation genetics, offering insights into a species biology and evolution (Brandies *et al.*, 2019). Many studies have been conducted into fallow deer via the Dama International Fallow Deer Project, however most of these studies have been archaeological based using Isotope analysis on bones to determine age and geographical origin (Miller, 2013). Little has been done with DNA and no one is yet to fully sequence the fallow deer genome. A reference genome for fallow deer makes many research themes possible, including geographical assessment and accurate species determination, the latter

is particularly important in determining ingredients in food products (Haynes *et al.*, 2019). Geographical assessment via genetics is not only forensically applicable but will help in answering many questions linked to the historical origin of fallow deer (Baker *et al.*, 2017). Furthermore, many areas in conservation genetics can be studied utilising the reference genome including: effective population size (Ne), hybridization, population substructure, kinship, evolutionary history, population connectivity, local adaptation, adaptive genetic variation, and inbreeding (Fuentes-Pardo and Ruzzante, 2017). The reference genome can be used to compare the European fallow dear to its close relative the Persian fallow deer, which has often caused debate as to whether it is a true sub-species or a separate species in its own right. Understanding the genetics of different geographical populations of fallow deer will also give a more informed approach to herd management for use in deer parks across the UK and Europe.

This study utilised Pacific Biology (PacBio) technology (Table 1.10 and Figure 1.13 – section 1.5) to achieve a highly accurate reference genome. The PacBio Sequel IIe platform is highly accurate and has a >99% raw read accuracy, this equates to a less than 1% error rate resulting in a highly accurate reference genome (Pacb.com. 2022). PacBio Sequel IIe is long read sequencing which makes genome assembly easier, as the longer the sequence read the more the sequences overlap. This makes it easier to assemble the DNA fragments back into the original order, much like a jigsaw puzzle, the larger the pieces (sequence fragments) the fewer pieces there are to form the final puzzle (the whole genome) (Figure 3.1) (Nanoporetech.com. 2017). With this, a minimum of 30x coverage is expected creating a reference genome exhibiting very few gaps and errors. This study presents a full genome assessment including the presence of STRs, evaluating their abundance, location, and structure. This will provide the groundwork for assessing their potential as a tool for population genetic studies and individual identification.



Short Read Sequencing
~ 50 base read
~ 92,000 “pieces”



Long Read Sequencing
~ 500,000 base read
~ 9 “pieces”

Figure 3.1: *Representation of Short read sequencing vs long read sequencing.*

3.2 Method

In this study, The whole genome of the fallow deer was sequenced via PacBio Next Generation Sequencing (NGS) on the Sequel IIe system. DNA was extracted from five fresh muscle samples following an optimised method for the Gentra® Puregene® Tissue Kit and quantified using Nanodrop, Qubit and Agarose Gel Electrophoresis (AGE). Library preparation for sequencing was conducted using the SMRTbell® prep kit 3.0. De novo genome assembly was undertaken via Flye (Kolmogorov et al., 2019) and HiFiasm (Cheng et al., 2021). Analysis and assessment of the full genome was undertaken using BUSCO (Manni et al., 2021) and QUAST (Gurevich et al., 2013) Software's. For genome annotation, protein sequences were aligned to the assembly of the fallow deer genome using the software package, Exonerate (Slater & Birney, 2005). The gene sets were integrated using software package EVM (Haas *et al.*, 2008).

3.2.1 DNA Extraction

For WGS, high molecular weight good quality DNA was required. This is to ensure the genome is fully intact ready for sequencing. Five samples were chosen from the Royal Parks Richmond (Ldn47 a, Ldn45 b, Ldn47 b, Ldn48 b, Ldn48 c). See Figure 3.2 for Specimen Ldn45, Figure 3.3 for Ldn47, and Figure 3.4 for Ldn48. A negative control was included with every batch of extractions. All five DNA extracts were sent to Edinburgh Genomics for QC and the best sample after QC was taken further to WGS. The samples were chosen based upon its geographical location as well as its quality. Triplicate DNA extractions from muscle were performed using the Gentra® Puregene® Tissue Kit, following the manufacturers protocol: DNA Purification from Muscle Tissue Using the Gentra Puregene Tissue Kit page 39-40 of the Gentra® Puregene® Handbook (QIAGEN, Germany), see appendix 10.2.1, with the following modifications:

Step 1: Pestle and Mortar was used rather than Liquid nitrogen.
Step 2: 1.5ml Cell Lysis Solution and 1.5ml of Tissue lysis solution used rather than just 3ml Cell lysis.
Step 2b: 30µl Puregene Proteinase K and 15ul of DTT used rather than just 15ul Puregene Proteinase K. Incubate for 4 hours.
Step 3: 30µl RNase A Solution used rather than 15ul.
Step 5: 1.5 ml Protein Precipitation Solution used rather than just 1ml with an additional 10min incubation on ice.
Step 6: Increase Centrifugation speed to 7000 x g.
Step 9 and 12: Increase Centrifugation speed to 4000 x g
Step 13: Air dry overnight.

Step 15: Incubate at 65°C for 30 min rather than 1 hour.

Figure 3.2: *Specimen Ldn45 Fallow Deer Buck.*

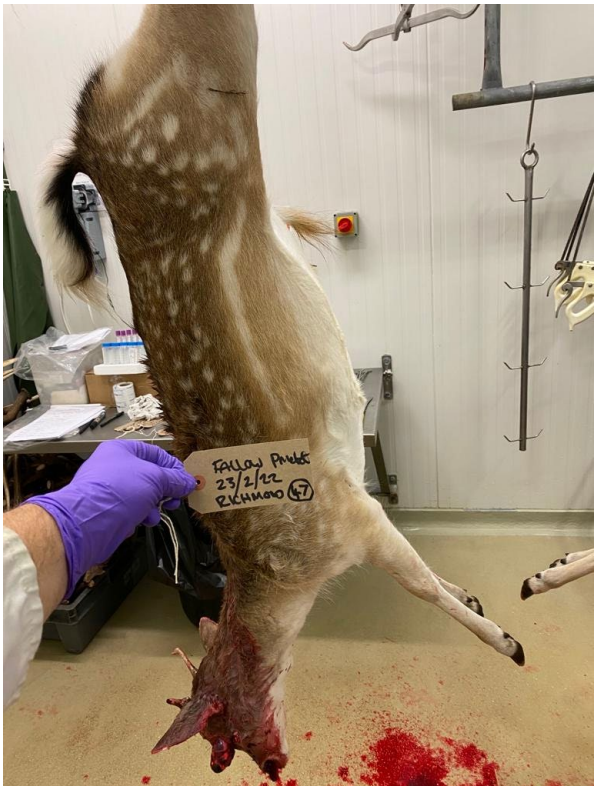


Figure 3.3: *Specimen Ldn47 Fallow Deer Pricket.*

Figure 3.4: *Specimen Ldn48 Fallow Deer Buck.*

3.2.2 DNA Quantification

Extracted DNA was quantified. Three methods of DNA quantification were used to assess sample quality and quantity.

Firstly, DNA was quantified using the NanoDrop 2000/2000c Spectrophotometer (Thermo Fisher Scientific, USA) following the manufactures guidelines, found in section 3, page 1-7 of the NanoDrop 2000/2000c Spectrophotometer user manual version 1.0. Each DNA extract was run three times on the NanoDrop 2000/2000c Spectrophotometer, and an average was made for an accurate quantification result. See results section 3.3.

Secondly, DNA was quantified using the Invitrogen™ Qubit™ 3 Fluorometer using the Qubit™ 1X dsDNA HS Assay Kit (Thermo Fisher Scientific, USA) following the manufactures guidelines, found on page 3-4 of the user guide. Each DNA extract was run three times on the Invitrogen™ Qubit™ 3 Fluorometer, and an average was made for an accurate quantification result. See results section 3.3.

Thirdly, Agarose Gel Electrophoresis (AGE) was used to Quantify DNA on a 1% 1x TAE agarose gel stained with gel red (Biotium) and visualised using a BioRad Molecular Imager® Gel Doc™ XR+ with image lab software. Comparison of band brightness was compared to known concentrations of Control Lambda DNA (Thermo Fisher Scientific, USA). 6 dilutions of control DNA with a final volume of 100µl were set up to 100ng/µl, 50ng/µl, 25ng/µl, 10ng/µl, 5ng/µl and 1ng/µl concentrations. 2µl of tracking dye is added to 4µl of the 6 dilutions, mixed and then 6µl of each can be loaded onto the gel.

3.2.3 Sample Selection Criteria

Samples meeting the following quality criteria were selected for WGS. At least 0.2µg (200ng) of gDNA per sample, and a DNA concentration of 20ng/µl up to 500 ng/µl. The purity of the DNA should be high, the A260/280 reading given by the NanoDrop 2000/2000c Spectrophotometer should be 1.8-2.0 and the A260/230 should be at least 1.8.

3.2.4 Species Confirmation

Five samples which had the best DNA quantification were taken forward for species confirmation (Ldn47 a, Ldn45 b, Ldn47 b, Ldn48 b, Ldn48 c). All five were sent to Edinburgh Genomics for QC and the best sample after QC was taken further to WGS.

3.2.4.1 COI Amplification

PCR was undertaken using the Dawney *et al.* (2007) COI universal primer set (HC02198/LC01490).

HC02198: 5'-TAAACTTCAGGGTGACCAAAAAATCA-3'

LC01490: 5'-GGTCAACAAATCATAAAGATATTGG-3'

A 15µl volume PCR was set up containing 1x Platinum® PCR Multiplex Master Mix (Life Technologies, USA), 3mM of $MgCl_2$ (Fisher Scientific, UK) 0.1uM final concentration of both forward and reverse primer and 6ng DNA extract. A negative control was prepared by substituting the DNA for DNA free water. The following PCR conditions were used: Initial denaturation at 95°C for 15 min; followed by 40 cycles at 94°C for 30 s, 50°C for 30 s, and 72°C for 1 min, followed by a final extension step of 72°C for 15 min. PCR was undertaken on The Applied Biosystems® Veriti® 96-Well Thermal Cycler (Fisher Scientific, UK). Amplified products were checked on a 1.5% 1x TAE gel stained with gel red (Biotium) and visualised using a BioRad Molecular Imager® Gel Doc™ XR+ with image lab software.

3.2.4.2 Purification and Cycle Sequencing

Once amplified and checked on a 1.5% gel, PCR products were purified using The DNA Clean & Concentrator-5 PCR purification kit (Zymo Research, USA) following manufacturers guidelines on page 4 of the DNA Clean & Concentrator-5 handbook, see appendix 10.2.2.

Cycle sequencing was performed using the purified amplified DNA. Cycle sequencing was done using the BigDye™ Terminator v3.1 Cycle Sequencing Kit following the manufacturers protocol with the modification of using half volume reactions, see appendix 10.2.3.

After cycle sequencing the DNA was purified using the in-house protocol for PCR purification. Per sample reaction, 1µl of 3M NaoAc, 1µl of 20µg/µl Glycogen, 1µl of 100mM EDTA and 30µl of cold absolute ethanol was added. This was vortexed briefly and left overnight to incubate. Following the incubation, the reactions were centrifuged at top speed for 30 minutes at 3°C. The supernatant was then removed, leaving the pellet in place and washed with fresh 70% ethanol. This was then centrifuged at top seed for 15 min. This wash step was repeated twice. Next, to air dry the pellet, the tubes were placed with lids open on a PCR machine set

at 50˚C for 10 min. When ready to run on the ABI 3500, the DNA was suspended in 13µl of HiDi formamide.

3.2.4.3 DNA Sequencing Analysis

Once DNA had been purified after cycle sequencing, the samples were prepared and sequences were analysed via capillary electrophoresis on the Applied Biosystems (ABI) 3500. Pop 6 polymer was used on the following instrument protocol: StdSeq50_POP6_Z. The base calling protocol used was the following: BDTv3.1_PA_Protocol-POP6. The injection time was 8 seconds and run time was 5000s. For the PGEM control, run time was upped to 6000. Sequences were visualised using sequence analysis V6 software. BLAST (Blast.ncbi.nlm.nih.gov, 2019) and BOLD (Barcodinglife.org, 2019) search engines were used to identify the species of the samples.

3.2.5 Whole Genome Sequencing (WGS)

Sample Ldn47 a was chosen for WGS after QC conducted by Edinburgh Genomics. Sequencing platform chosen for this project was PacBio Long-Read sequencing, using the SMRTbell barcoded adapter on the Sequel IIe system. This platform gives very high accuracy, at more than 99% (<1% error rate). At least 30x sequencing coverage is expected. 0.2µg (200ng) of gDNA per sample, at least 20ng/µl up to 500 ng/µl in Tris-HCL (pH 8.0) is required for the sequencing.

Library preparation for sequencing was conducted following manufacturer's instructions for whole genome libraries using the SMRTbell® prep kit 3.0. The Blue Pippin pulsed-field size selector was used as an alternative size selection method for the SMRTbell® prep kit 3.0. SMRT Link Sample Setup, to prepare the SMRTbell library for sequencing, it was conducted following the SMRT link user guide.

Work undertaken by: Edinburgh Genomics at the University of Edinburgh.

90G data contains 5G nucleotides. The genome size is 3G (nucleotides) so the depth is 90/3= 30x.

3.2.6 Genome Assembly and Assessment

De novo genome assembly was undertaken via Flye (Kolmogorov *et al.*, 2019) and HiFiasm (Cheng *et al.*, 2021) bioinformatics software's.

Analysis and assessment of the full genome was undertaken using BUSCO (Manni *et al.*, 2021) and QUAST (Gurevich *et al.*, 2013) Software's. See section 4.3 for the results of the genome assessment.

Genome sequencing and assembly was undertaken by Edinburgh Genomics at the University of Edinburgh.

A repeat-masked genome was firstly achieved using a program called RepeatMasker (Smit *et al.*, 2015) using the Repbase library (Bao *et al.*, 2015). Gene Mapping was then undertaken as follows: Using the repeat-masked genome, two approaches were combined to predict genes. The first approach relies on ab initio prediction. Two different software packages were used including AUGUSTUS (Stanke *et al.*, 2004) and Glimmer-HMM (Majoros *et al.*, 2004). The second approach relies on homolog sequence comparison, where the homolog protein sequences were downloaded from two published homolog species including *Cervus elaphus* (https://www.ncbi.nlm.nih.gov/genome/?term=txid9860[orgn]) and *Dama dama* (https://www.ncbi.nlm.nih.gov/protein/?term=Dama+dama). For genome annotation, these protein sequences were aligned to the assembly of the fallow deer genome using the software package, Exonerate (Slater & Birney, 2005). The gene sets predicted by the two approaches were integrated using software package EVM (Haas *et al.*, 2008) to produce weighted consensus gene models. For the annotation of the non-coding genes, tRNAScan-SE program was used to predict tRNA and rRNAmmer was used to predict rRNA (version 1.2).

The assembled original genome sequence was re-mounted at the chromosome level (based on the results of RagTag (Alonge *et al.*, 2021), 72 contigs were successfully mounted), the reference genome used for mounting the chromosomes was the *Cervus elaphus* genome (33 chr+X+MT) (https://www.ncbi.nlm.nih.gov-/genome/?term=txid9860[orgn]) and the *Cervus canadensis* genome (33chr+ X,Y+MT) (https://www. ncbi.nlm.nih.gov/genome-/34916). The MT genome was mounted to the *Dama dama* MT genome already available (https://www.ncbi.nlm.nih.gov/nucleotide/NC_-020700.1).

For comparative genomics, orthoFinder (version 2.2.6) was used to do phylogenetic orthology inference and results were mapped to a phylogenetic tree using FigTree (version 1.4.3)

All mapping and annotation analysis was performed on Linux Cluster by Fan Liu at the Beijing Institute of Genomics in connection with Naif Arab University for Security Sciences.

3.3 Results and Analysis

3.3.1 DNA Quantification

Samples Ldn47 a, Ldn45 b, Ldn47 b, Ldn48 b, Ldn48 c displayed high levels of purity, as shown by Table 3.1 and therefore were chosen for WGS. As all the samples have good concentrations, the purity was the only criterion that was used to select the sample at this stage. Due to the high DNA concentration, it was hard to accurately estimate the exact concentration using gel electrophoresis therefore, it tended to overestimate the concentration of DNA slightly however, there was concordance between the quantification results given by Nanodrop and Gel electrophoresis by comparison. Figure 3.5 of sample Ldn47 a after DNA extraction which was chosen for WGS after QC at Edinburgh Genomics. All samples were taken forward with the assurance that the negative control had passed. See supplementary data for the quantification gel images for all extracted DNA samples.

Table 3.1: *DNA quantification results for samples Ldn47a, Ldn45 b, Ldn47 b, Ldn48 b and Ldn48 c. DNA quantified via Nanodrop and Qubit.*

Sample	Nanodrop			Qubit
	A260/A280	A260/A230	Yield (ng/µl)	Yield (ng/µl)
Ldn47 a	1.84	1.46	908	629.3
Ldn45 b	1.84	1.97	753	678.0
Ldn47 b	1.81	1.64	965	728.0
Ldn48 b	1.80	1.60	453	378.7
Ldn48 c	1.76	1.59	579	454.7

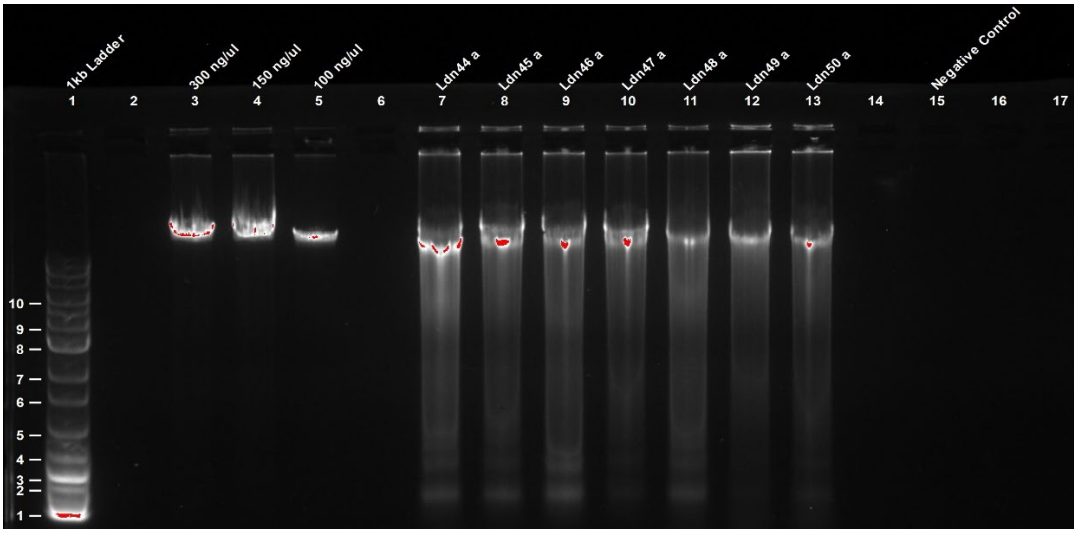


Figure 3.5: *Gel image showing extracted DNA from fallow deer sample Ldn47 (amongst others). From left to right Lane 1: 1kb Ladder, Lane 2: Blank, Lane 3: 300 ng/µl, Lane 4: 150 ng/µl, Lane 5: 100 ng/µl Lane 6: Blank Lane 7: Ldn44a, Lane 8: Ldn45 a, Lane 9: Ldn46 a, Lane 10: Ldn47*

a, Lane 11: Ldn48 a, Lane 12: Ldn49 a, Lane 13: Ldn50 a, Lane 14: Blank and Lane 15: Negative Control. 1% Agarose gel used and 4µl of DNA added.

Visualisation of the DNA extracts shown by the gel image, shows that the DNA is of high molecular weight and of good quality, supporting the quantification results in Table 3.1. This shows that the DNA is more than adequate for achieving good WGS results via NGS. Before the samples were sent for WGS, they were checked to ensure they were definitely of fallow deer origin, this was achieved via species ID sequencing using the COI universal primer set HC02198/LC01490. See section 3.3.2 for the results of the species ID.

3.3.2 Species Identification

3.3.2.1 Polymerase Chain Reaction (PCR) and Purification

Successful COI amplification and purification is shown by clean bright bands at the expected size, 700bp in Figure 3.6. All negative controls were negative.

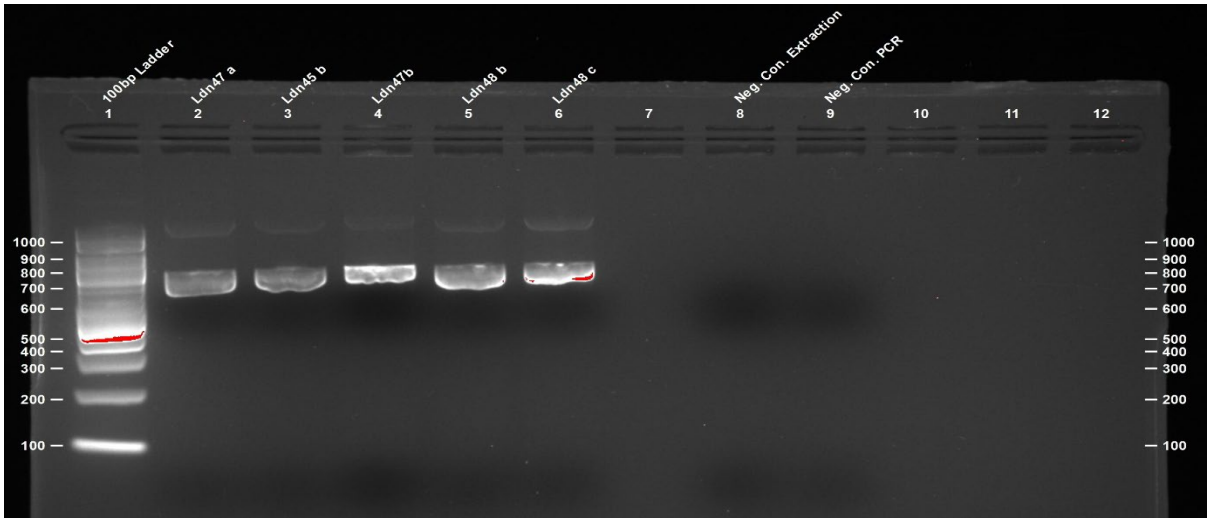


Figure 3.6: *Gel image showing DNA amplification of fallow deer sample Ldn47 a, Ldn45 b, Ldn47 b, Ldn48 b, Ldn48 c using Dawnay et al. (2007) universal primers post PCR purification. From left to right Lane 1: 100bp ladder, Lane 2: Ldn47 a, Lane 3: Ldn45 b, Lane 4: Ldn47 b, Lane 5: Ldn48 b, Lane 6: Ldn48 c Lane 7: Blank, Lane 8: Negative Control: Extraction, Lane 9: Blank and Lane 10: Negative Control: PCR. 1.5% Agarose gel used and 2µl of DNA added.*

3.3.2.2 Sequencing and Species Identification

Species ID, as conducted according to section 3.2.3, was successful. All five samples were identified as fallow deer employing the BLAST species Identification tool on the NCBI website using the GenBank database and the BOLD search engine using the barcoding life database. See Table 3.2 for percentage Identity match to fallow deer. All negative controls returned as negative, and all base lines

of the sample sequences were clean so it can be concluded that these samples are not contaminated. See figures 3.7-3.11 of the sequence electropherograms for samples Ldn47 a, Ldn45 b, Ldn47 b, Ldn48 b, and Ldn48 c. All of these samples, at this stage, would have been ideal for WGS via Pac Bio.

Table 3.2: *Percentage species Identification of DNA samples according to a match with reference fallow deer (Dama dama) COI genomes on BLAST and BOLD.*

Sample Code	Query Length	E-values	Percentage Identity on BLAST	Percentage Identity on BOLD
Ldn47 a	693	0.0	100%	100%
Ldn45 b	650	0.0	100%	100%
Ldn47 b	650	0.0	100%	100%
Ldn48 b	650	0.0	100%	100%
Ldn48 c	500	0.0	100%	100%
Negative Control - Extraction	-	-	Negative	Negative
Negative Control - PCR	-	-	Negative	Negative

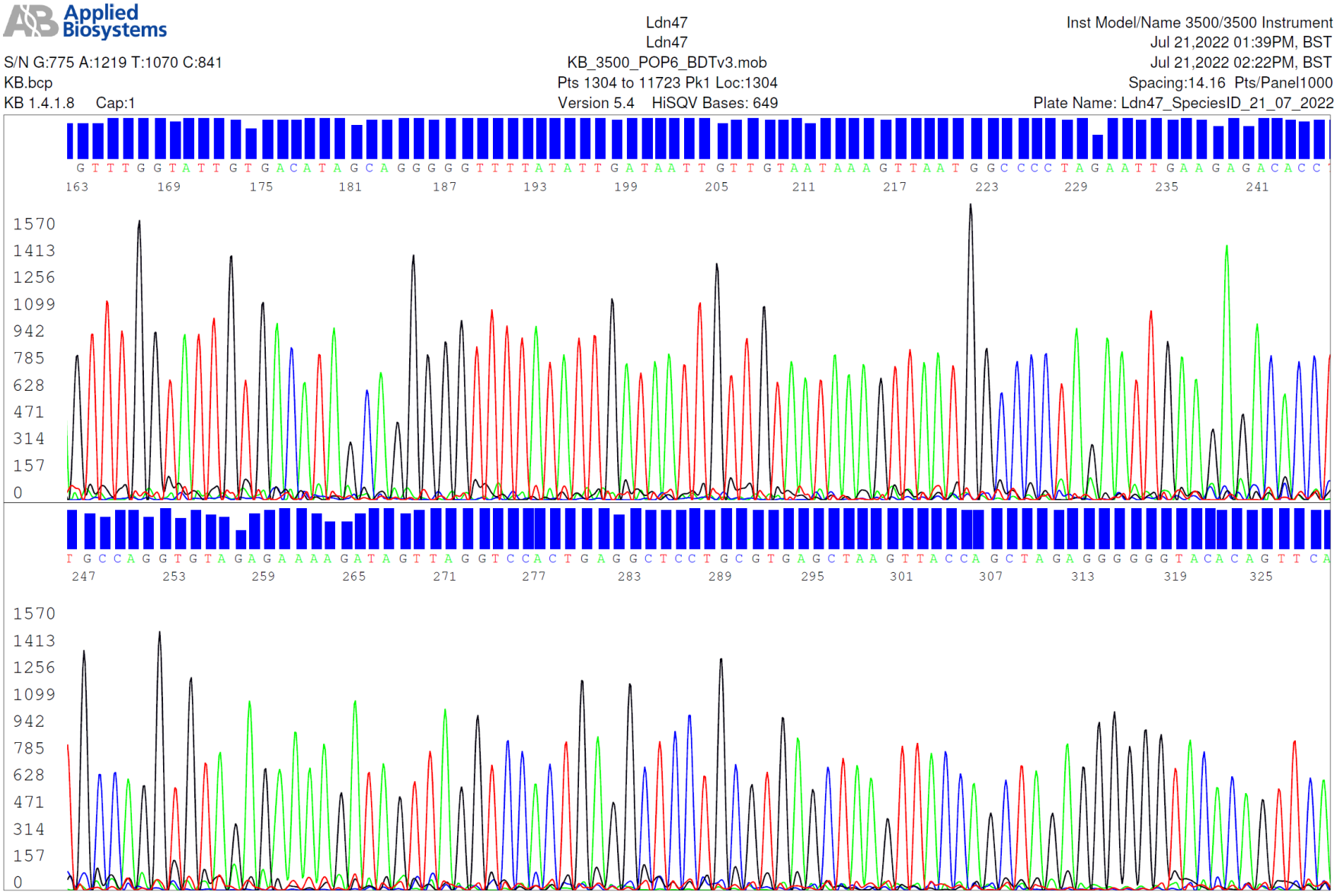


Figure 3.7: *A segment of the sequence electropherogram for sample Ldn47 a, showing high quality DNA and accurate base calls. Cycle sequencing performed following the BigDye™ Terminator v3.1 Cycle Sequencing Kit and sequenced via Sanger Sequencing on the ABI3500.*

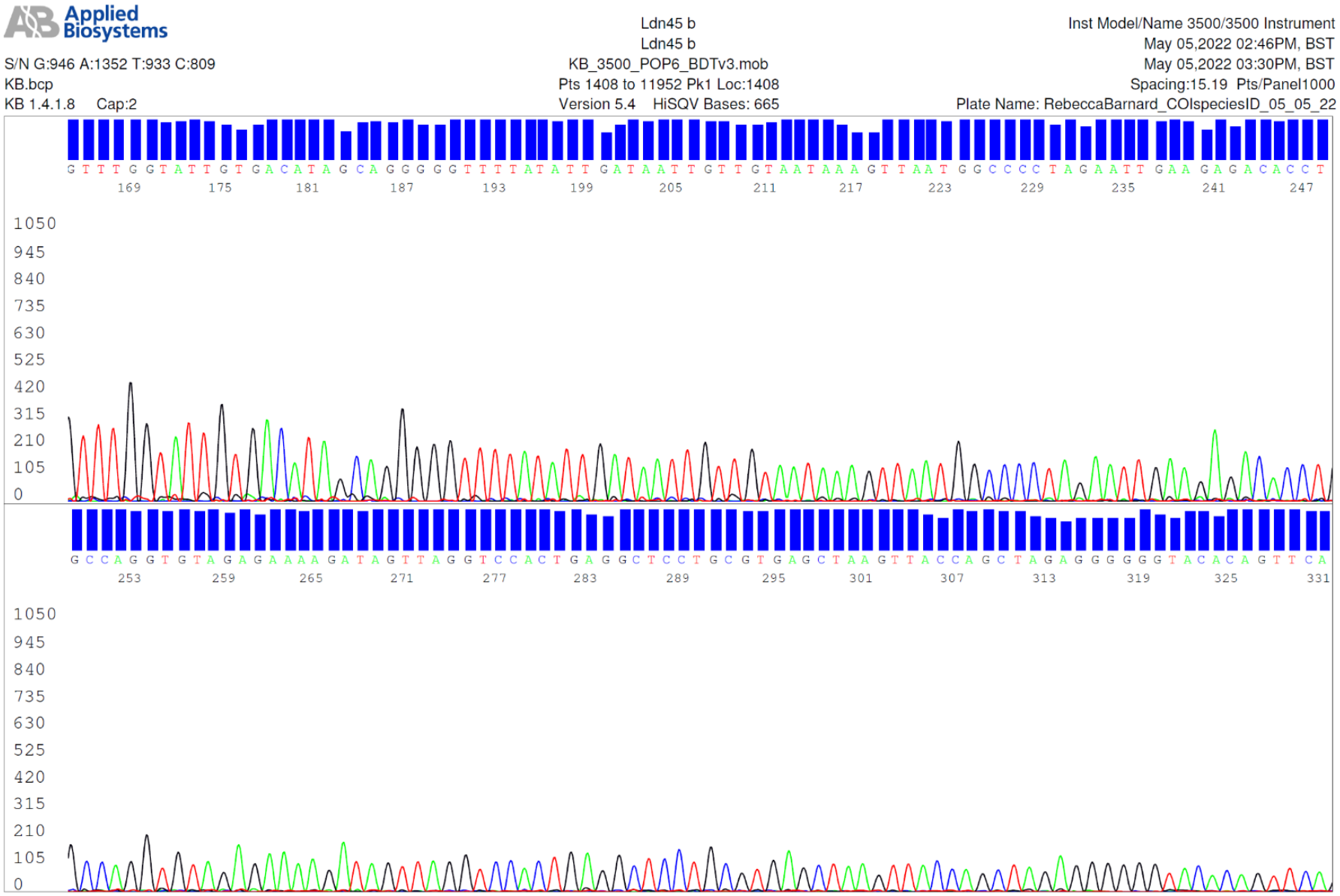


Figure 3.8: *A segment of the sequence electropherogram for sample Ldn45 b, showing high quality DNA and accurate base calls. Cycle sequencing performed following the BigDye™ Terminator v3.1 Cycle Sequencing Kit and sequenced via Sanger Sequencing on the ABI3500.*

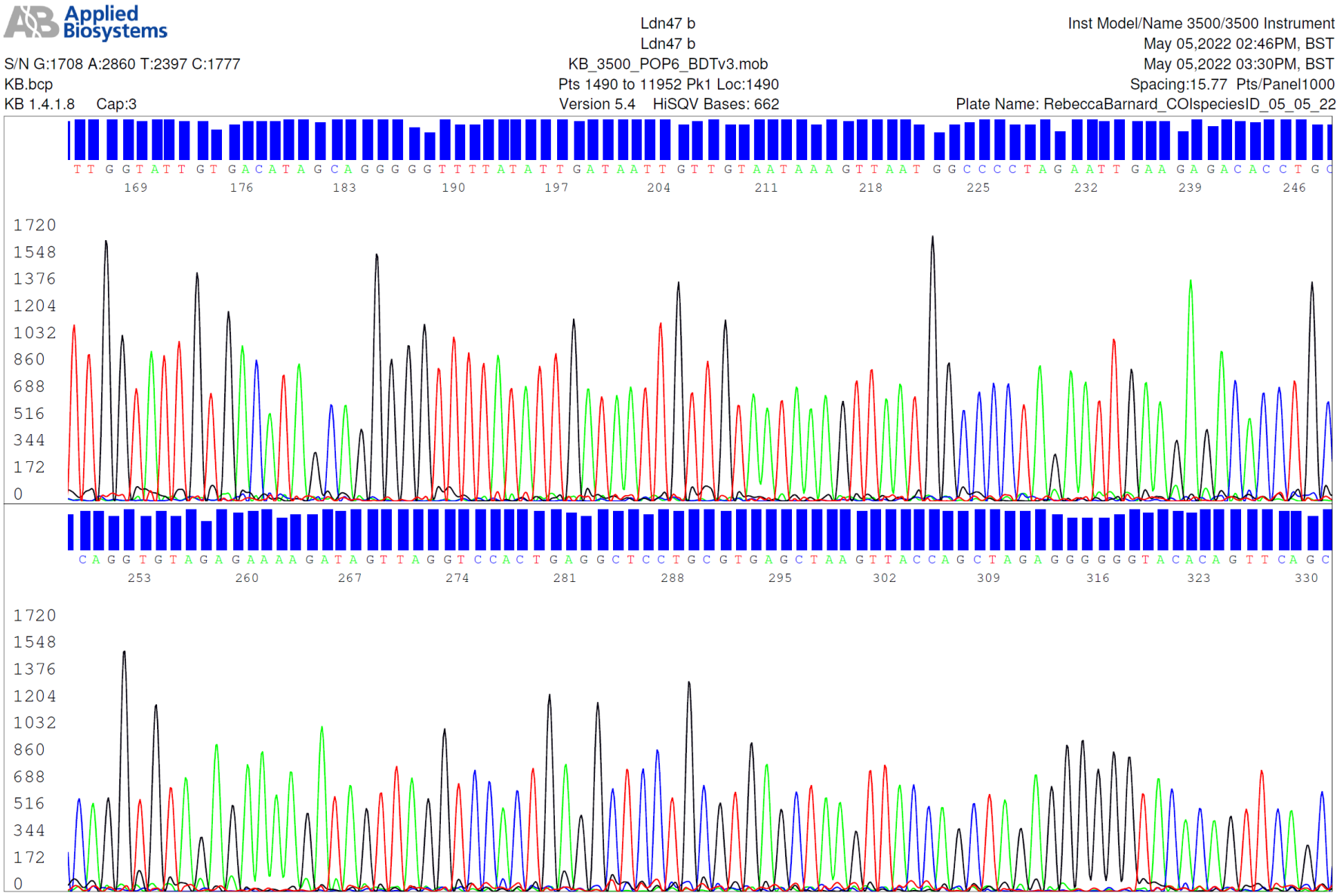


Figure 3.9: *A segment of the sequence electropherogram for sample Ldn47 b, showing high quality DNA and accurate base calls. Cycle sequencing performed following the BigDye™ Terminator v3.1 Cycle Sequencing Kit and sequenced via Sanger Sequencing on the ABI3500.*

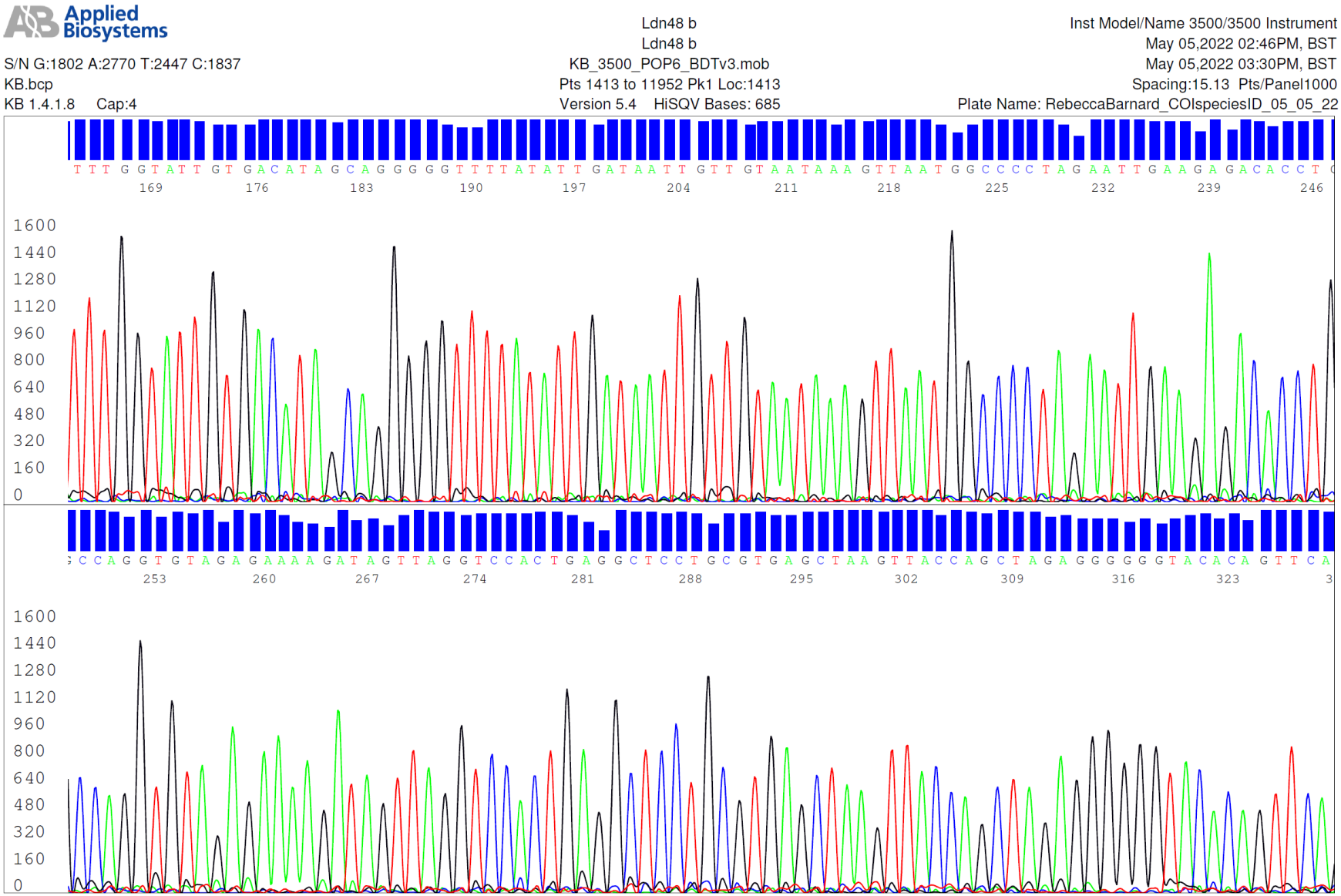


Figure 3.10: *A segment of the sequence electropherogram for sample Ldn48 b, showing high quality DNA and accurate base calls. Cycle sequencing performed following the BigDye™ Terminator v3.1 Cycle Sequencing Kit and sequenced via Sanger Sequencing on the ABI3500.*

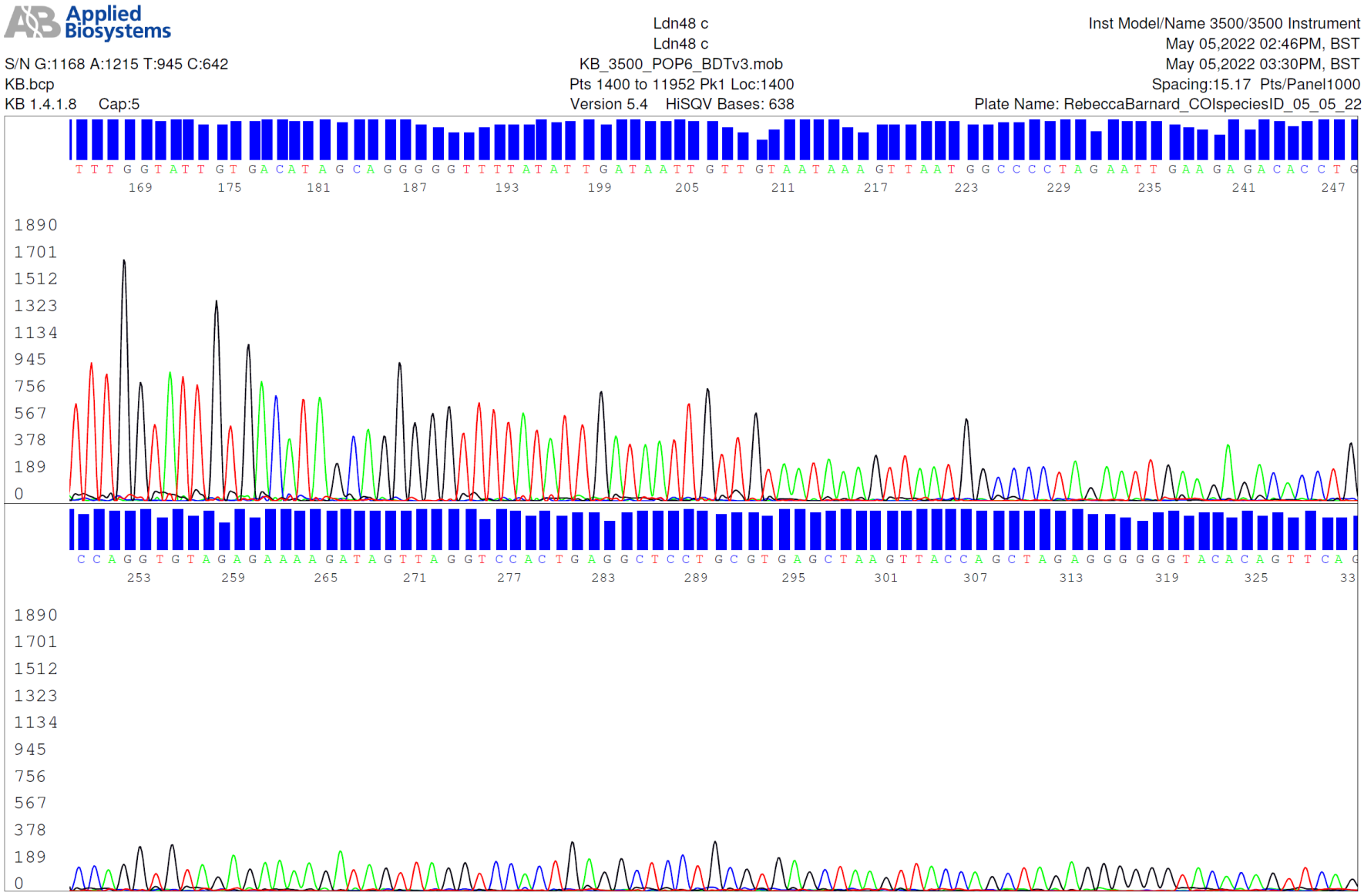


Figure 3.11: *A segment of the sequence electropherogram for sample Ldn48 c, showing high quality DNA and accurate base calls. Cycle sequencing performed following the BigDye™ Terminator v3.1 Cycle Sequencing Kit and sequenced via Sanger Sequencing on the ABI3500.*

3.3.3 Genome Assessment

Full results of the Genome sequencing including the sequence data can be found in the supplementary data file S.1_WGS.

3.3.3.1 Quality Control (QC)

All five samples were sent for QC at Edinburgh Genomics prior to PacBio WGS. This was done to ensure the best quality sample was sequenced to ensure the best possible result. Very HMW (high molecular weight) DNA was required with high DNA purity and a low percentage of RNA. After QC, the results showed that all samples had very high amounts of DNA, as expected from the DNA quantification previously done in house. The Nanodrop quantification had been repeated and showed high levels of pure DNA which was concordant with Qubit readings. There was still low A260/A230 ratios indicating some non-DNA contamination, but this was low enough to not be a concern. As shown by the TapeStation report given by Edinburgh genomics, the DNA samples do exhibit some damage/degradation however, as the HMW peak was so strong and so high this slight degradation was also no cause for concern. After analysis of the TapeStation reports, see figures 3.12 – 3.17, it was concluded that Sample 001 (Ldn47 a) should proceed to sequencing as this sample has the highest amount of intact HMW DNA.

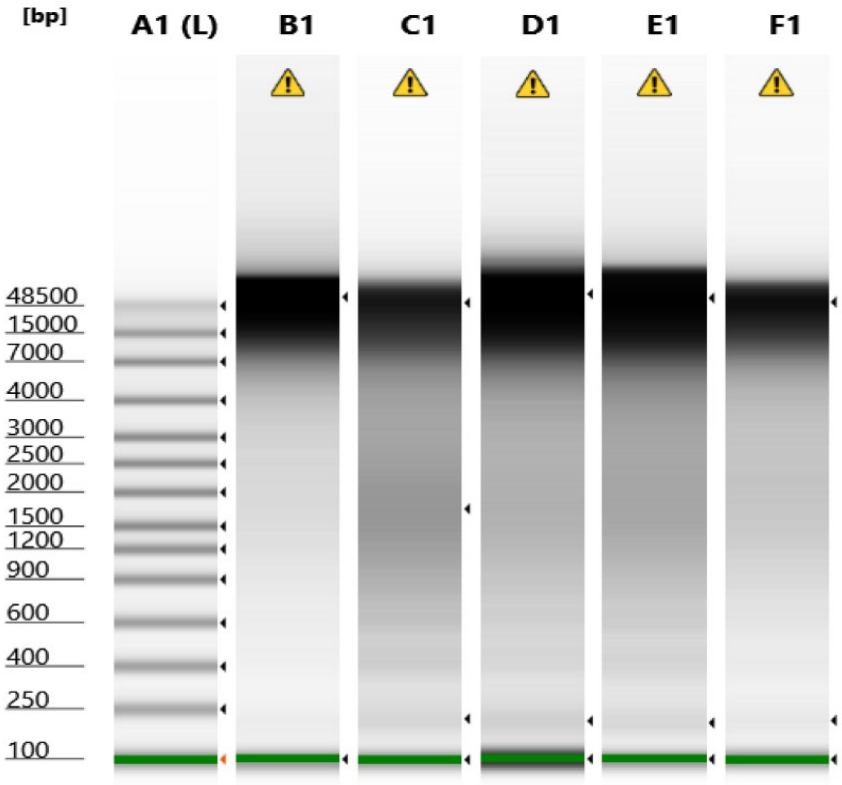
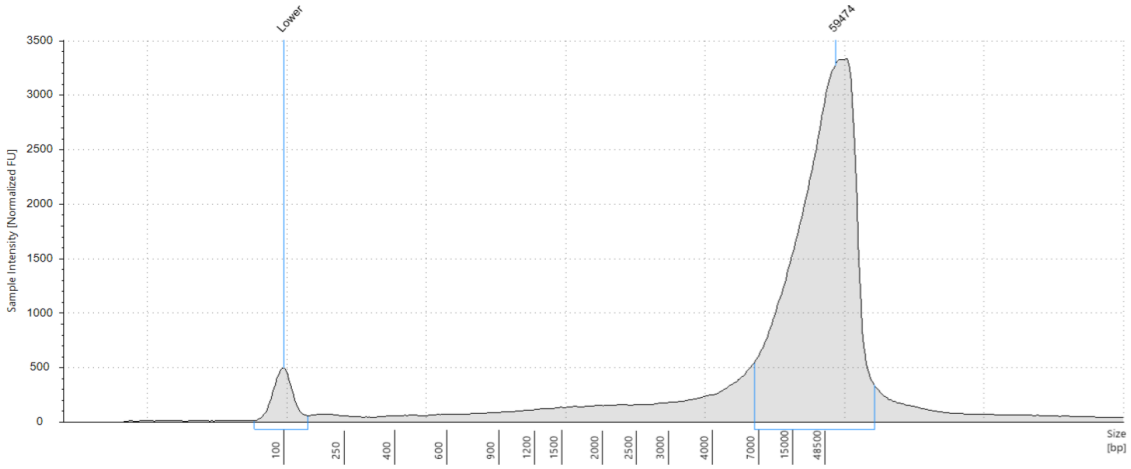


Figure 3.12: *TapeStation Gel image. From left to right A1: Ladder, B1: Ldn47 a, C1: Ldn45 b, D1: Ldn47 b, E1: Ldn48 b and F1: Ldn48 c.*

B1: 21751BR0001



Sample Table

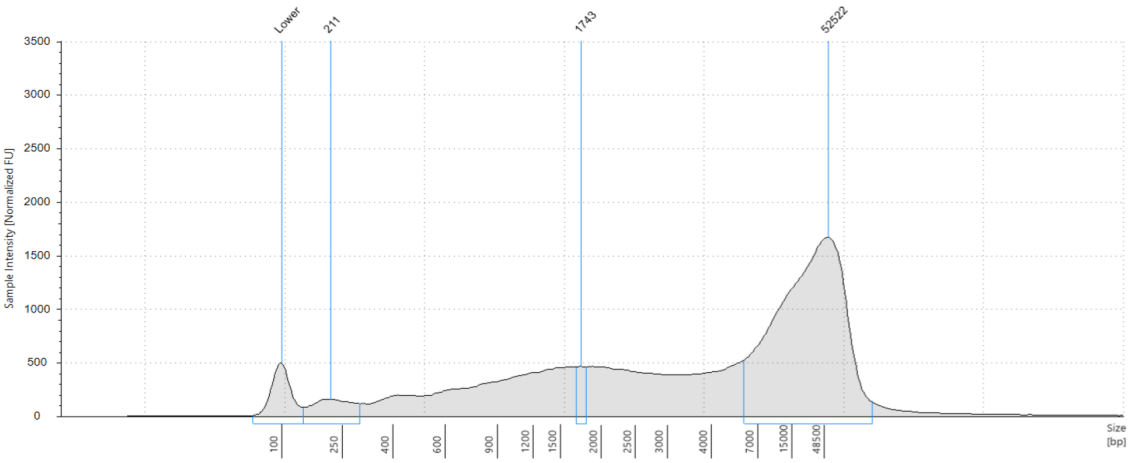
Well	DIN	Conc. [ng/µl]	Sample Description	Alert	Observations
B1	7.9	232	21751BR0001	⚠	Sample concentration outside recommended range

Peak Table

Size [bp]	Calibrated Conc. [ng/µl]	Assigned Conc. [ng/µl]	% Integrated Area	From [bp]	To [bp]	Peak Comment	Observations
100	8.50	8.50	-	63	144		Lower Marker
59474	170	-	93.38	6691	>60000		
-	-	-	-	-	-		Sample Well

Figure 3.13: *TapeStation gDNA quantification report for sample Ldn47 a.*

C1: 21751BR0002



Sample Table

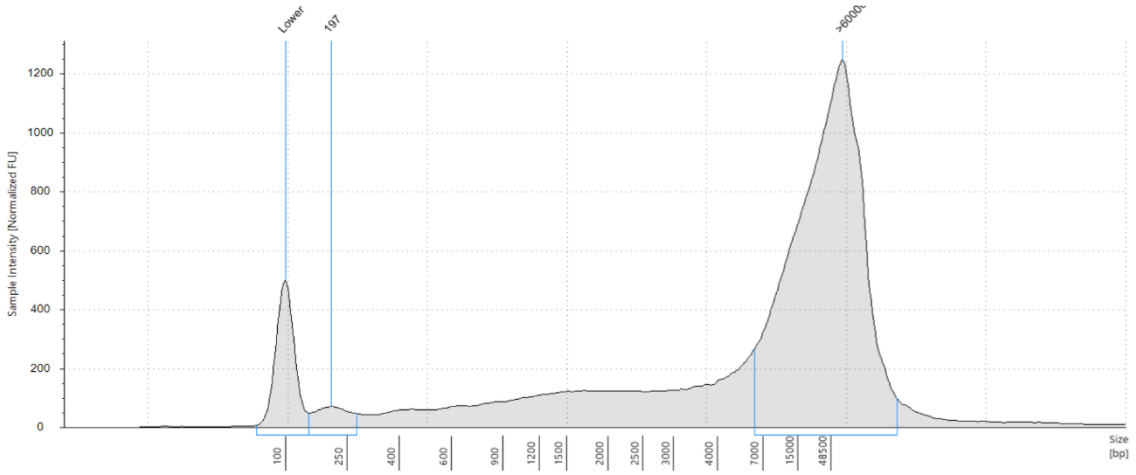
Well	DIN	Conc. [ng/µl]	Sample Description	Alert	Observations
C1	6.2	217	21751BR0002	⚠	Sample concentration outside recommended range

Peak Table

Size [bp]	Calibrated Conc. [ng/µl]	Assigned Conc. [ng/µl]	% Integrated Area	From [bp]	To [bp]	Peak Comment	Observations
100	8.50	8.50	-	65	139		Lower Marker
211	5.67	-	4.92	139	296		
1743	4.16	-	3.61	1675	1800		
52522	104	-	90.47	5963	>60000		
-	-	-	-	-	-		Sample Well

Figure 3.14: *TapeStation gDNA quantification report for sample Ldn45 b.*

D1 : 21751BR0003



Sample Table

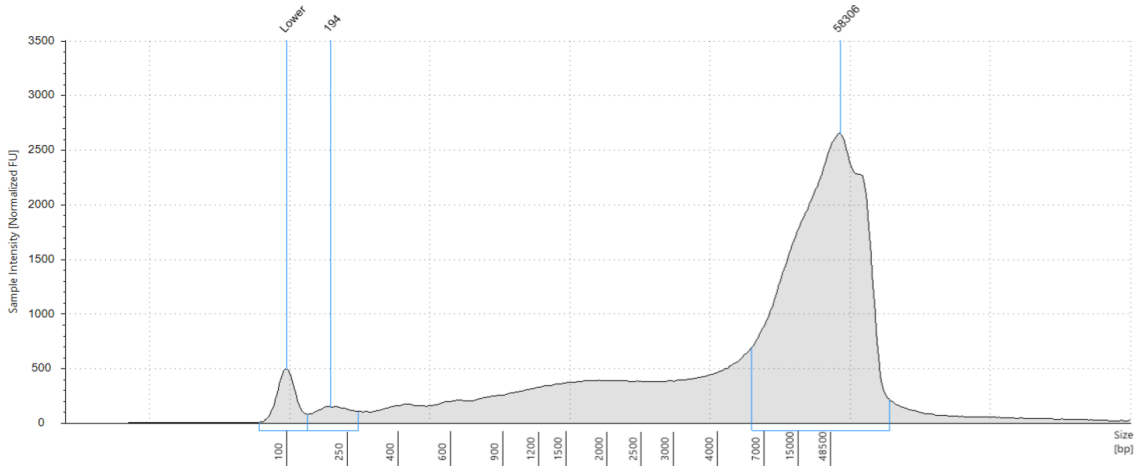
Well	DIN	Conc. [ng/µl]	Sample Description	Alert	Observations
B1	7.1	113	21751BR0003	⚠	Sample concentration outside recommended range

Peak Table

Size [bp]	Calibrated Conc. [ng/µl]	Assigned Conc. [ng/µl]	% Integrated Area	From [bp]	To [bp]	Peak Comment	Observations
100	8.50	8.50	-	65	143		Lower Marker
197	2.09	-	2.63	143	274		
>60000	75.4	-	94.83	6347	>60000		
-	-	-	-	-	-		Sample Well

Figure 3.15: TapeStation gDNA quantification report for sample Ldn47 b.

E1: 21751BR0004



Sample Table

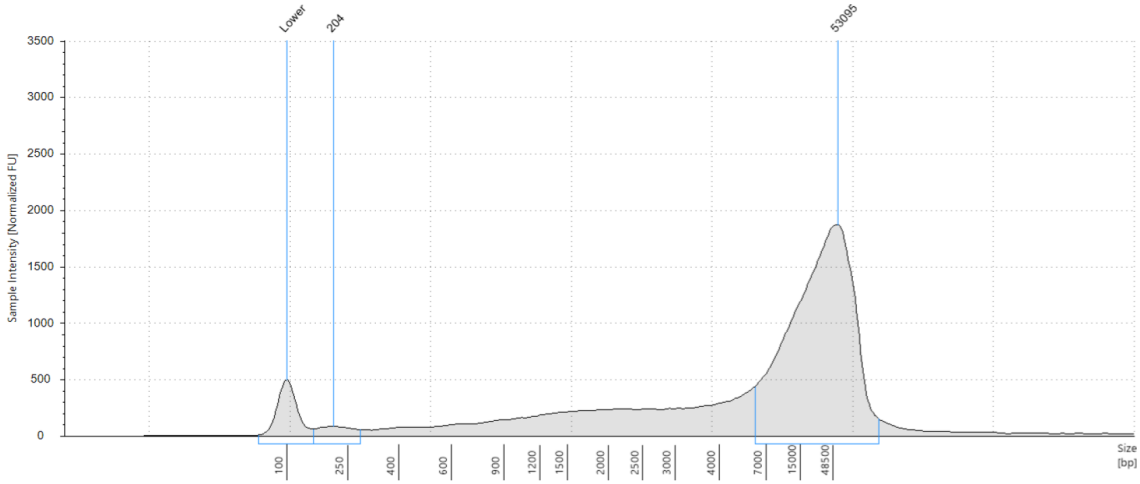
Well	DIN	Conc. [ng/µl]	Sample Description	Alert	Observations
E1	6.8	305	21751BR0004	⚠	Sample concentration outside functional range for DIN

Peak Table

Size [bp]	Calibrated Conc. [ng/µl]	Assigned Conc. [ng/µl]	% Integrated Area	From [bp]	To [bp]	Peak Comment	Observations
100	8.50	8.50	-	66	137		Lower Marker
194	4.68	-	2.32	137	277		
58306	193	-	95.82	6075	>60000		
-	-	-	-	-	-		Sample Well

Figure 3.16: TapeStation gDNA quantification report for sample Ldn48 b.

F1: 21751BR0005



Sample Table

Well	DIN	Conc. [ng/µl]	Sample Description	Alert	Observations
F1	6.8	167	21751BR0005	⚠	Sample concentration outside recommended range

Peak Table

Size [bp]	Calibrated Conc. [ng/µl]	Assigned Conc. [ng/µl]	% Integrated Area	From [bp]	To [bp]	Peak Comment	Observations
100	8.50	8.50	-	65	149		Lower Marker
204	2.45	-	2.25	149	281		
53095	102	-	93.49	6164	>60000		
-	-	-	-	-	-		Sample Well

Figure 3.17: TapeStation gDNA quantification report for sample Ldn48 c.

As can be observed from the TapeStation gel image and traces (see figures 3.12 – 3.17), sample Ldn47 a had the highest HMW peak at 59,474bp with a concentration of 170ng/ul at this level. There is also only one other peak at 100bp with a concentraion of 8.5ng/ul which is a very slight indication of some degraded DNA. However, that said, the majority of the sample is intact and of excellant quality, ideal for PacBio sequencing. Table 3.3 summerises the Quality control data provided by Edinburough genomics, further suporting Ldn47 a as the sample of choice. Compared to the other four samples which have multiple peaks registered indicating more degraded DNA within the sample and higher amounts of fragmented low molecular weight DNA, Ldn47 a had the highest chance of producing a highly intact and high quality whole genome sequence.

Table 3.3: *Quality Control report summarising the results from the DNA quantification assessment for samples Ldn47 a, Ldn45 b, Ldn47 b, Ldn48 b and Ldn48 c. Results support Ldn47 a as the most favourable for Whole Genome Sequencing.*

Internal Sample ID	External Sample ID	Qubit 1 (ng/µl)	Qubit 2 (ng/µl)	Qubit 3 (ng/µl)	Av Conc. (ng/µl)	Vol. (µl)	µg received	Qubit RNA 1 (ng/µl)	Qubit RNA 2 (ng/µl)	% RNA	Nanodrop A260/ A280	Nanodrop A260/ A230	Nanodrop (ng/µl)
21751BR0001	Ldn47 a	578.0	678.0	632.0	629.3	330	207.7	29.0	28.4	4.6	1.93	1.47	940.9
21751BR0002	Ldn45 b	646.0	742.0	646.0	678.0	330	223.7	11.9	11.6	1.7	1.93	2.02	1041.4
21751BR0003	Ldn47 b	696.0	724.0	764.0	728.0	330	240.2	19.0	18.0	2.5	1.89	1.69	1632.5
21751BR0004	Ldn48 b	352.0	432.0	352.0	378.7	350	132.5	11.1	10.6	2.9	1.89	1.66	513.7
21751BR0005	Ldn48 c	436.0	488.0	440.0	454.7	330	150.0	11.8	11.8	2.6	1.88	1.65	581.6

3.3.3.2 Library Preparation

During the library preparation process for PacBio Long-Read sequencing, using the SMRTbell barcoded adapter on the Sequel IIe system, the DNA must be sheared. At Edinburgh Genomics, DNA is sheared and the results are measured using the FEMTO Pulse system which is an automated pulsed-field capillary electrophoresis system. The FEMTO Pulse traces indicated that the sample Ldn47 a over sheared, see figures Figure 3.18, Figure 3.19 and Figure 3.20. The recommended shearing speeds on the Megaruptor for a 15-21kb product are 29 and 30. When sample 01 was sheared at 29 then 30, the product size produced was ~12.5kb. Lower speeds of 27 and 28 were attempted on the Megaruptor, however the same result of ~12.5kb was achieved. This over shearing potentially could be due to the way the DNA was extracted, or this could be typical of the way deer DNA behaves when being sheared. The Blue Pippin pulsed-field size selector was used to remove smaller fragments 9kb, however due to the lower-than-expected product size the final yield was also slightly reduced.

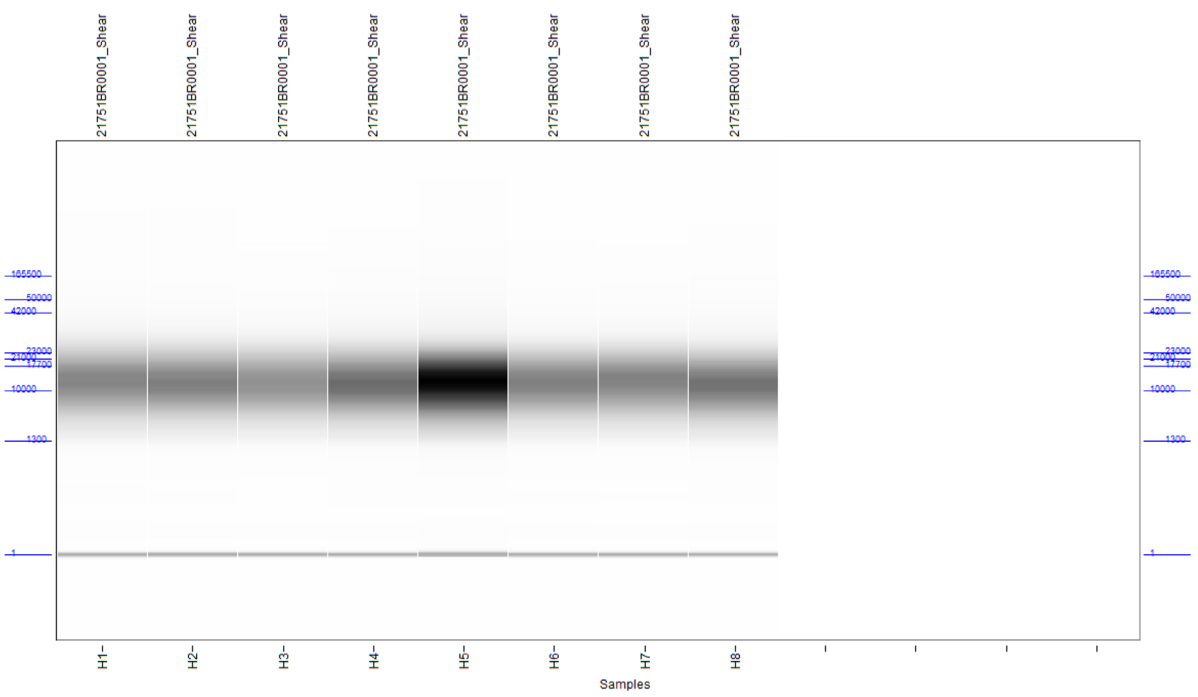


Figure 3.18: *Gel image from the FEMTO Pulse system post DNA shearing of sample Ldn47 a.*

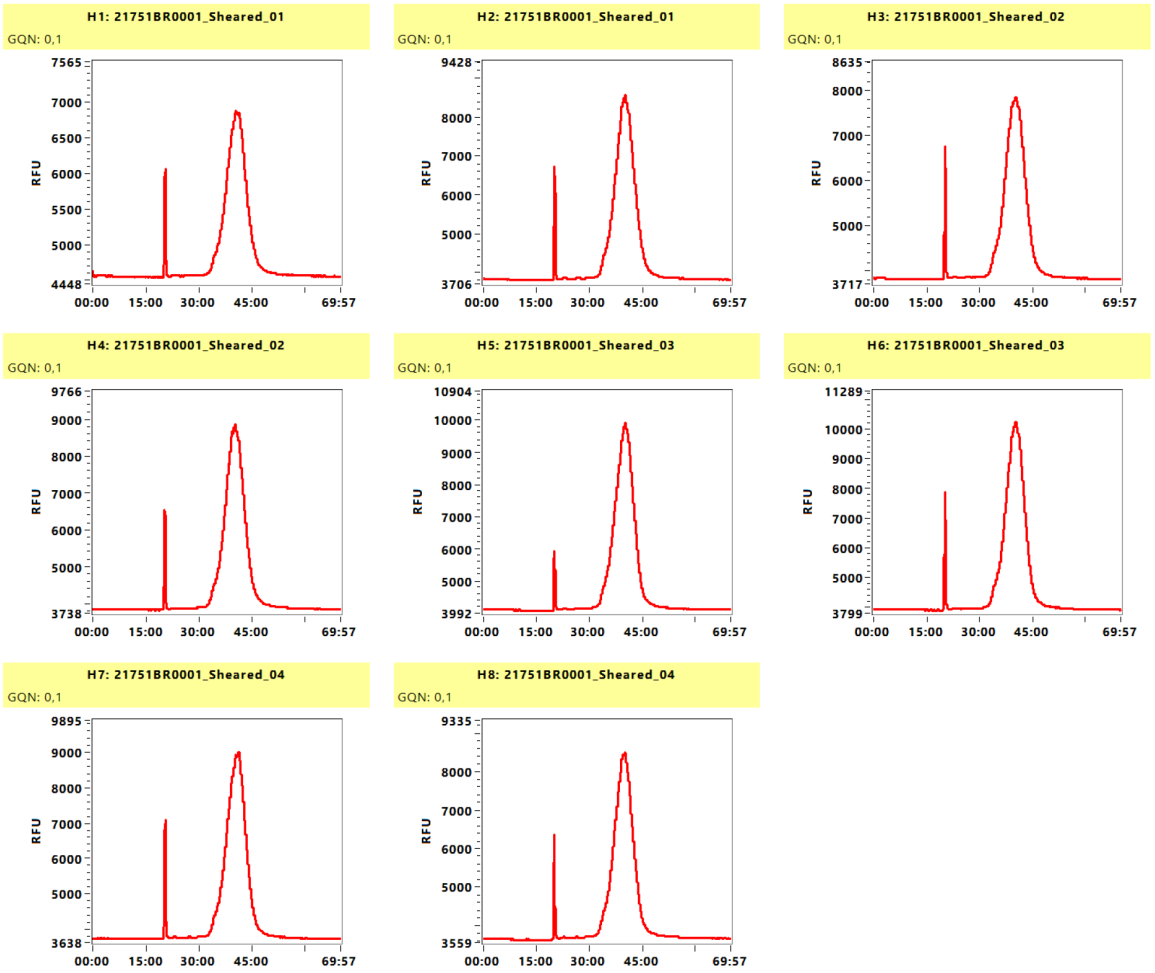
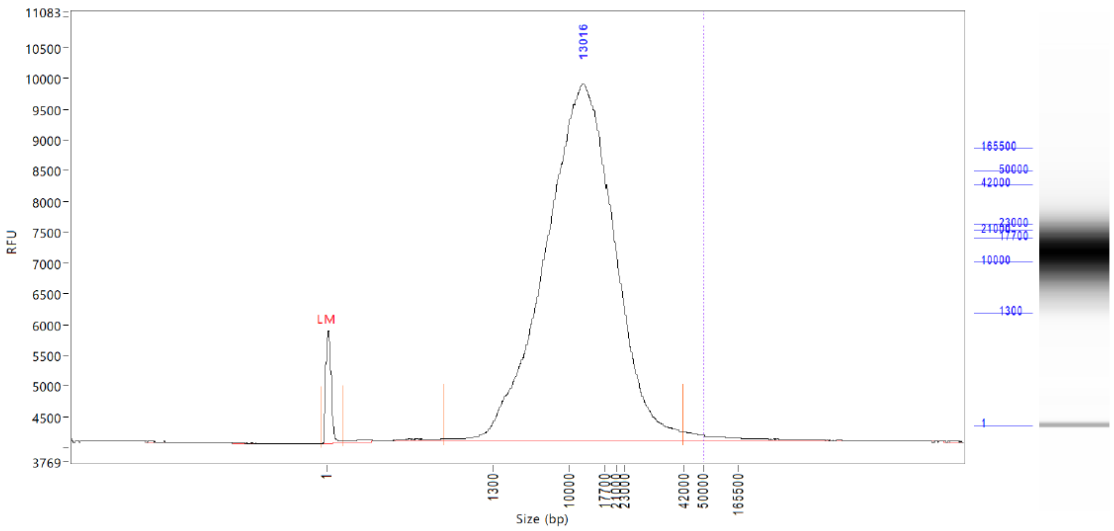


Figure 3.19: *Graphical representation of the sheared sample Ldn47 a on the FEMTO Pulse system.*

Sample: 21751BR0001_Sheared_03
Well Location: H5
Created: 26/05/2022 16:41



Peak	Size (bp)	Conc. (pg/uL)	From (bp)	To (bp)	Avg. Size (bp)	CV%	RFU	Corr. Peak Area
1	1 (LM)	1,9468	0	112	2	974,01	1827	42,912
2	13016	416,0950	912	41727	12712	51,35	5809	917,162

TIC:	416,0950	pg/uL
TIM:	53,8884	pmole/L
Total Conc.:	425,0168	pg/uL
GQN:	0,1	
Threshold:	50000	

```
Sample Peak Width (sec): 20   Sample Min Peak Height: 50    Sample Baseline V to V?: Y    Sample Baseline V to V pts: 3
Sample Filter: Binomial       # of Pts for Filter: 3       Sample Start Region (min): 0  Sample End Region (min): 70
Manual Baseline Start (min): 25     Manual Baseline End (min): 70
Marker Peak Width (sec): 8    Marker Min Peak Height: 500   Marker Baseline V to V?: N    Marker Baseline V to V pts: 3
Lower Marker Selection: First Peak > 500 RFU                Upper Marker Selection: Last Peak > 500 RFU
Ladder Size (bp)1, 1300, 10000, 17700, 21000, 23000, 42000, 50000, 165500
Quantification Using: Ladder          Final Concentration (ng/uL): 0,0200          Dilution Factor: 10,0
Min. RFU for Data Processing: 1      Size Threshold (b.p.): 50000
```

Figure 3.20: *Well H5 insight. H5 had the highest fragment size for the sheared sample Ldn47 a.*

As shown by the gel image in Figure 3.18, well H5 holds the longest fragment. Figure 3.19 and Figure 3.20 show this fragment to be 13,016 bp. This fragment is strong and of a high concentration. Therefore, there is little concern that this sample can still produce a good genome sequence. See the Genome report for the final results of the sequencing on the PacBio Long-Read Sequel IIe system.

3.3.3.3 Genome Assessment Report

Due to over shearing during the library preparation process, the yield of the four SMRT cells was reduced. Initially, around 60x was anticipated after sequencing, however this was reduced to 36x. However, 36x is still enough coverage for an accurate genome assembly.

The genome was assembled using HiFiasm (version 0.16.1-r375) (Cheng *et al.*, 2021). QUAST (version 5.0.2) was used to generate the assembly validation metrics (Gurevich *et al.*, 2013) and BUSCO (version 5.3.0) was used on the assembly generated in order to assess their quality in terms of gene completeness based on the mammalian lineage (Simão *et al.*, 2015). The final genome assembly has a total length of 3,108Mb in a total of 471 contigs which was assembled into 72 chromosomal scaffolds, see Table 3.4 for the full genome data. The final assembly also had a BUSCO completeness score of 92.7%.

The annotated genome was found to consist of 22,616 genes (Table 3.4), which was made up of 15,943 protein-coding genes and 6,676 non-coding genes. The final BUSCO v4.1.4 completeness score for the annotation was 72%.

Gene prediction analysis identified almost 16,000 genes with average transcript length of 49,000. On average each gene has 10 exons. Overall, the number of identified genes appeared lower than that expected for mammals (typically 20K-30K), however, these results can be expected due to the lack of additional RNA-seq data at this time, therefore some of the non-coding genes have yet to be annotated.

72 scaffolds were assigned to 35 Chromosomes, representing 33 autosomes as well as 2 allosomes (X and Y).

Table 3.4: *Genome data for Dama dama, Ldn47.*

Project data accession numbers	
Assembly identifier	Ldn47
Species	*Dama dama*
Specimen	Ldn47
NCBI taxonomy ID	ID_30532
BioProject	PRJNA905851
BioSample ID	SAMN31886425
Isolate information	Male, Muscle Tissue
Raw data accession numbers	
PacBio SEQUEL IIe	SRR24718831
Genome assembly	
Assembly accession	JASJUW000000000
Span (Mb)	3,108
Number of contigs	471
Contig N50 length (Mb)	68
Number of scaffolds	471
Scaffold N50 length (Mb)	68
Longest scaffold (Mb)	132
BUSCO genome score	C:96.4%[S:93.0%,D:3.4%],F:0.9%,M:2.7%,n:9226
Genome annotation	
Number of genes	22,619
Number of protein-coding genes	6,676
Average length of gene (bp)	49,099
Average number of exons per transcript	10.7
Average exon size (bp)	172.83
Average intron size (bp)	4,871
BUSCO annotation score	C:72.1%[S:70.5%,D:1.6%],F:3.0%,M:24.9%,n:13335

C= complete [S= single copy, D=duplicated], F=fragmented, M=missing, n=number of orthologues in comparison.

3.3.3.4 Genome Comparison

OrthoFinder (version 2.2.6) was used to do phylogenetic-orthology inference for comparative genomics (total of eight genomes including *Dama dama*) and mapped to a phylogenetic tree by FigTree (v1.4.3). The seven genomes used to compare to the fallow deer genome were: elk (*Cervus canadensis*), red deer (*Cervus elaphus*), white-tailed deer (*Odocoileus virginianus texanus*), Reeves' muntjac (*Muntiacus reevesi*), Central European red deer (*Cervus elaphus hippelaphus*), Yarkand deer (*Cervus hanglu yarkandensis*) and muntjac, (*Muntiacus muntjak*). Results of the genome comparison phylogenetic tree can be seen in Figure 3.21.

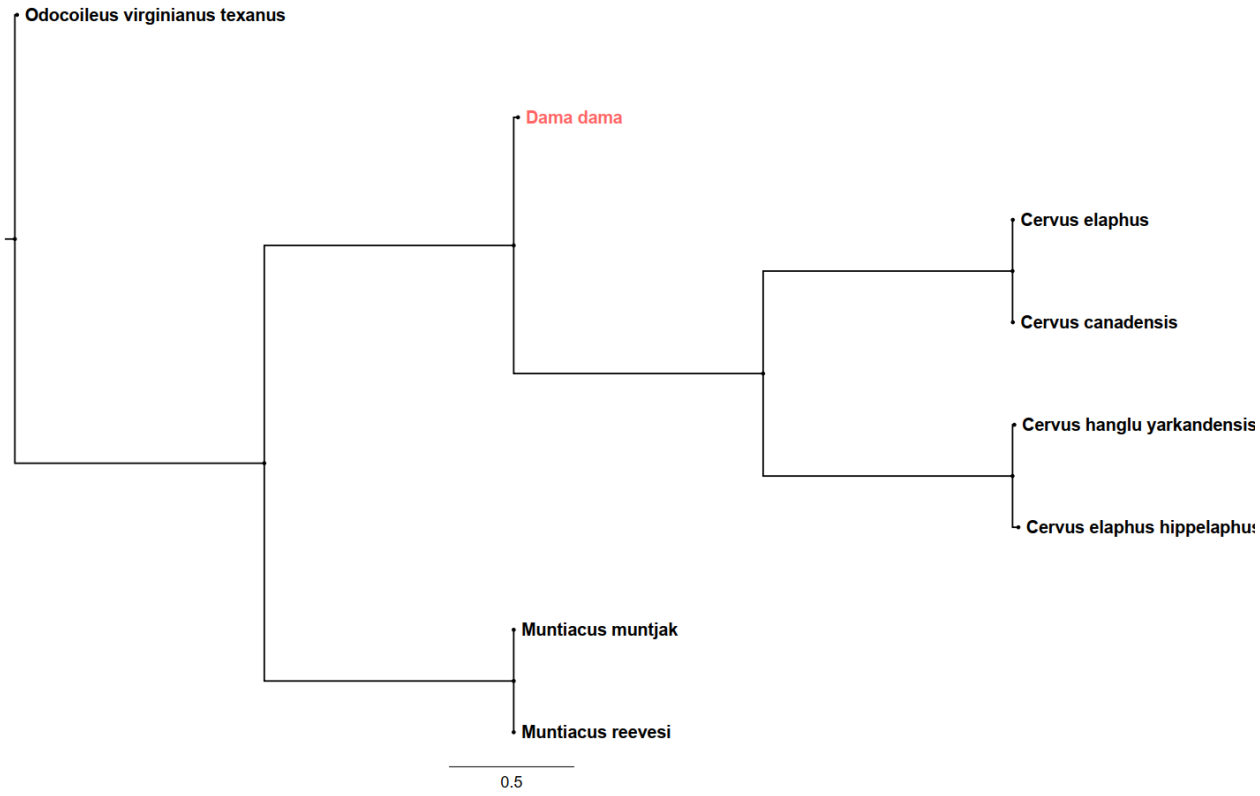


Figure 3.21: *Phylogenetic tree depicting the results of the phylogenetic-orthology inference of the fallow deer genome along with seven other species.*

3.4 Discussion

For WGS using a long read sequencing platform, high quality DNA is required. Typically purity must be A260/A280 1.8-2 and A260/A230 must be around 2 to signify pure DNA as anything lower than this will affect downstream analysis for WGS. Sample Ldn47 used had a purity of A260/A280 1.84 and A260/A230 1.46. This notably affected the library prep for WGS. As the integrity of the DNA was reduced, as depicted by the reduced DIN score of 7.9. Ideally, the DIN score would have been 8.5 or higher, 10 being fully intact DNA (Kong *et al.*, 2014). Low integrity of DNA caused the DNA to over shear, so smaller fragments of the DNA were produced. This resulted in the yield from each flow cell being slightly reduced (Kong *et al.*, 2014). The addition of DTT during extraction may have also caused DNA to weaken, which may have been a factor in the over shearing. DTT has shown to cause nicks within the DNA strand which can cause strands to weaken and in some cases unwind (Fjelstrup *et al.*, 2017). Despite DTT helping to increase purity of DNA, it may have caused a decrease in the integrity. In an attempt to reduce the over shearing, Edinburgh Genomics reduced the speed of the Megaruptor however this had no effect on how the DNA sheared. This however could just be the way fallow deer DNA shears; this was mentioned by Edinburgh genomics but there is no research to back this statement up. This provides a potential future research topic, to look into different species and effect of shearing on DNA with the possibility to adapt the methodology depending on the DNA being sheard to not affect the quality of preparation for sequencing.

With PacBio Sequel II sequencing, highly accurate long read sequences can be achieved. The more cells you run on the Sequel II the higher the sequencing depth however this also increases the cost. On average it costs around £2000 per cell. For this project, four cells were used. If DNA is undisrupted and of highest quality after the sequencing preparation process, 60x sequencing depth can be achieved using four cells. Another consequence of the over shearing is that it can lead to a reduced sequencing depth. Overall, 36X sequencing depth was achieved with a genome completeness score of 96.4%, which is comparable to the completeness of the Red deer genome which had been sequenced using PacBio and HIFI Illumina which achieved 96.2%. Despite not having any HiFi data for this sequencing, unlike the red deer genome, the completeness score is excellent. For this project, ~20x sequencing depth would have sufficed, the final genome achieved here is above and beyond our expectations and is an excellent example of a first reference genome for the fallow deer species.

In comparison to the red deer genome, which was sequenced, in 2021, by Edinburgh genomics and included Hi-C data and higher coverage due to the addition of Illumina short-read sequencing as well as PacBio Long read sequencing (Pemberton *et al.*, 2021), the fallow deer genome is extremely similar in its quality as well as other attributes. The genome coverage is almost the same despite having lower sequencing coverage, this is likely due to the capability of PacBio long read sequencing, as mentioned in section 1.5 and 3.1, and the high-quality DNA sample which was sequenced. Further aspects to note about the fallow deer genome include the length, structural assembly, and annotation. The length of the genome is as expected for a mammal at around 3Gb, 3,108Mb to be precise, which is slightly longer than the red deer genome, which was 2,887Mb. The genome was assembled into 471 contigs, which were assembled into 72 scaffolds and conformed into 35 Chromosomes, including X and Y allosomes. The red deer genome was female and consisted of 145 scaffolds conformed into 34 Chromosomes including the X allosome (Pemberton *et al.*, 2021). For annotation, the fallow deer genomes comprises of 22,619 genes, all of which at this time are protein coding. The Red deer genome by comparison has 32,781 genes, 22,941 of which are protein coding (Pemberton *et al.*, 2021). For comparisons with further published deer genomes, see Table 3.5. The annotation BUSCO completeness score for the fallow deer genomes is 72%, this is lower due to the unannotated non-coding genes. These have not yet been annotated due to the lack of RNA data. Further work is yet to be completed on this to complete the annotation. Software such as tRNAScan-SE can be used to predict tRNA, rRNAmmer may be used to predict rRNA, and mirScan to search for microRNA. This will then provide data to annotate the non-protein coding genes on the genome.

Table 3.5: *Comparison of genome characteristics amongst other published Cervus species.*

Species	Genome Size (Gb)	No. Genes	Avg. Gene Size (bp)	Avg. No. Exons per Transcript	Reference
Dama dama	3.1	22,619	49,099	10.7	(Barnard *et al.*, 2023)
Odocoileus virginianus texanus	2.4	20,651	-	-	(London *et al.*, 2022)
Cervus elaphus	2.8	32,781	37,644	11.26	(Pemberton *et al.*, 2021)
Cervus nippon	2.5	21,449	39,398	9.29	(Xing *et al.*, 2022)

Cervus elaphus yarkandensis	2.6	20,604	44,594	8.14	(Ba *et al.*, 2020)
Muntiacus muntjak	2.6	25,753	-	7.83	(Mudd *et al.*, 2019)
Muntiacus reevesi	2.6	26,054	-	7.77	(Mudd *et al.*, 2019)
Rangifer tarandus	2.7	24,814	39,409	7.62	(Li *et al.*, 2017)
Axis porcinus	2.7	22,473	34,537	8.61	(Wang *et al.*, 2019)
Elaphurus davidianus	2.5	20,125	46,124	9.76	(Zhang *et al.*, 2017)

The fallow deer genome was previously sequenced as part of research looking into the gene, MC1R, which is associated with the white coat in the species (Reiner *et al.*, 2020). However, The individual sequenced was a female, so no Y chromosome would have been analysed, the coverage was very low, and the genome was mapped to Bovine genome as at the time the *Cervus elaphus* wasn't available (Reiner *et al.*, 2020). Less than 10x sequencing depth was achieved which is very low and a subsequently less data was produced due to the low coverage, this causes an increase in gaps and missing data. Furthermore, the genome was sequenced using Illumina short read sequencing which is not ideal for achieving long intact sequences which can be assembled to produce a complete genome, as discussed in section 1.5. Short read sequencing is not ideal when searching for genetic variation, as there may be gaps and missing data at the loci and may result in a loss of key sites within the genome. Additionally, the genome was mounted using a bovine genome, despite there being some relation between fallow deer and bovine (same order), this is not as accurate as mounting the genome to the red deer (same family) as there is a higher chance of shared sequence producing a better more accurate assembly. Only 85% of the sequences could be mapped (Reiner *et al.*, 2020), whereas 100% of the current genome was successfully mounted to the red deer genome. 10x genome coverage was sufficient for the previous study as the interest focused on SNP variation (Reiner *et al.*, 2020), whereas this genome was produced to not only benefit this project but to provide a reference genome which can be utilised by the scientific community to further research into the fallow deer as expanded upon in section 3.4.1. Overall, there is no comparison with the genome achieved in this study compared to

previous work, this genome is of much higher quality and exhibits an excellent example of a reference genome.

Some genome comparison work has been done in shape of phylogenetic-orthology inference mapped to a phylogenetic tree, results of which can be found in section 3.3.3. The fallow deer genome was compared to seven other species of deer including, elk (*Cervus canadensis*), red deer (*Cervus elaphus*), white-tailed deer (Odocoileus *virginianus texanus*), Reeves' muntjac (*Muntiacus reevesi*), Central European red deer (*Cervus elaphus hippelaphus*), Yarkand deer (*Cervus hanglu yarkandensis*) and muntjac, (*Muntiacus muntjak*). As expected, the deer species within the genus *Cervus* are very closely related, situated on the same node of the tree and inline on the same short branch. As are the two species of muntjac which share a branch albeit on a separate node from the rest of the species. This suggests the inter genus evolutionary distance is low for these species showing a strong relationship between them. The White-tailed deer is the most unrelated, situated on its own longer branch near the root of the tree, this suggests this species has the largest evolutionary distance between it and the other genomes.

Fallow deer is showing to be genetically distinct to the other species, situated on its own branch. It is not surprising that the fallow deer species is genetically distinct when compared to other species due to the effect of the neolithic period ice age and the resultant genetic bottle neck, see section 1.2.1.1 of the introduction for further details of this (Chapman and Chapman, 1997; The British Deer Society (BDS), 2016; Baker *et al.*, 2017). This caused the species to be independent to other species of deer, and it is only more recently, after the 11th century, when fallow deer was reintroduced to other species of deer within the UK (Chapman and Chapman, 1997; The British Deer Society (BDS), 2016; Baker *et al.*, 2017). However, it wasn't until the 15th century when fallow deer entered the wild, before then they were kept in private parks and estates (Chapman and Chapman, 1997; The British Deer Society (BDS), 2016; Baker *et al.*, 2017). Therefore, compared to the other species of deer, which have had more time to interbreed, evolve together and inherit genetics from each other, the fallow deer is independent of that. However, fallow does present some genetic linkage to the four *Cervus* species, with the node of the tree branching down to *Cervus*, which is interesting as fallow deer live closely in Britain with red deer, *Cervus elaphus*. As depicted by the phylogenetic tree, the evolutionary distance between fallow and the other analysed species is high, but not as high as with the white tailed deer. As previous studies suggested (Pemberton and Smith, 1985; Hartl, Schleger and Slowak, 1986; Randi and Apollonio, 1988), fallow deer have a greater evolutionary divergence

to other species of deer, this comparison study supports their findings but show a new evolutionary relationship to the *Cervus* genus.

3.4.1 Future of the Project and beyond

In the future, further projects could work on increasing the quality of the genome sequencing. To do this would include the incorporation of Hi-C data to fill in gaps and missing data for a more accurate chromosome formation. Hi-C data consist of a genome wide three-dimensional capture of interactions between DNA loci (DeMaere & Darling, 2021). It is generated via a sample preparation method for NGS called Hi-C (DeMaere & Darling, 2021). This was not done at this stage due to budget restraints but easily achievable in the future with the data already available. Furthermore, to increase genome quality to 60x coverage or more, additional PacBio Sequel II cells could be run, a further 4 cells with a high-quality sample could double the current coverage thus increasing the completeness score of the genome. As before, this would require further funding to complete and not necessary for this project.

Due to the high quality of the data already achieved from the genome sequencing, this provides the groundwork for many other research projects in the future. As mentioned in section 1.2.1.1, a plethora of research has already been conducted on fallow deer, namely their place of origin, their distribution and their conservation (Baker *et al.*, 2017; Baker *et al.*, 2021). Using the genome now available, full genetic assessments can be conducted investigating the genetic diversity, genetic conservation and heredity between closely related species and the sub-species *Dama mesopotamica* (Persian fallow deer). There is important work to be done with the Persian fallow deer sub-species as they are facing extinction, finding genetic similarities with fallow deer could help with the preservation of this species. Research conducted into the genetic diversity of fallow deer is now made easier with the availability of the whole genome, full genome wide scans can be conducted rather than using a limited set of known shared loci (Pemberton & Smith, 1985). This will hopefully show that the fallow deer has a greater degree of genetic diversity than previously shown, this can be used to further increase genetic diversity of herds kept in captivity. The fallow deer project was a large-scale project with many influencers which ran for 4 years from 2011 to 2015 (Miller, 2013). This project was predominantly focused on distribution of fallow deer and where the species originated from. The project looked at the effect of the Roman period to the spread of fallow deer across Europe. The fallow deer genome can aid with geographical assessment by

searching genomes for conserved regions within different geographical populations. Furthermore, the fallow deer genome can play an important role in research for disease treatment and prevention. Common diseases associated with fallow deer are chronic wasting disease and lymes disease (Hamir *et al.*, 2011; Gandy *et al.*, 2021). The genome can help with the detection of how these diseases are spread, how they affect individuals and the development of medicinal treatments.

As discussed, the achievement of this novel fallow deer genome has a significant impact (Barnard et al., 2023). It not only secures the future research of this current project but will continue to be useful in many other aspects of genomics research. This fallow deer genome presented within this study provides the source data to identify polymorphic loci which can be used for population genetic studies and potential individual identification. The future of genomics research on fallow deer species is just getting started and is now ready to evolve.

3.5 Conclusion

Overall, this study achieved the following:

- First to fully sequence the genome of a male fallow deer to reference genome level. The genome was sequenced via PacBio sequencing technology on the Sequel IIe system.
- High quality genome achieved, 96.4% Completeness BUSCO score.
- The Genome is 3.1 Mb in length with the sequence assembled in to 35 Chromosomes.
- The genome of the fallow deer shows evolutionary relationship with the Cervus Genus.
- Benefits the welfare of the fallow deer, aiding the effort to combat wildlife crime.

4 Study 2: Exploring the best approach for individual identification

4.1 Introduction

Forensic wildlife genetics involves the analysis of non-human DNA. This includes DNA from the trade and possession of a species, or products made from a species, which is contrary to legislation; in cases where crime is against a person or property; cases of animal cruelty; and on the other hand, cases where the animal is the offender (Linacre *et al.*, 2011). Forensic wildlife genetics, especially identity testing, is fast becoming common practice in resolving criminal cases. STR multiplexing is the key form of analysis for identifying individuals. Increasingly, STR multiplexing is being adopted by wildlife forensic geneticists for the identification of individuals of an ever-growing list of species (Iyengar, 2014). STRs can be utilised for individual identification in wildlife species in many cases including animal cruelty, theft, illegal trade, predator identification, and poaching (Iyengar, 2014) see examples in section 1.3.4.

In 2005, Lorenzi reported on the use of STRs in a case of animal cruelty in Italy, DNA extracted from a knife of a poacher was matched to a wild boar carcass (Lorenzini, 2005). Theft of animals is a common occurrence across the world, particularly involving livestock (Cassidy and Gonzales, 2005). STRs have been adopted to bring about prosecutions in cases involving cattle theft (van de Goor and van Haeringen, 2007; Iyengar, 2014). In cases involving birds of prey and tortoises, STR profiles of the stolen individuals are compared to those of the alleged parents, shell, or feathers (depends on what is available for comparison testing) in order to distinguish that the individual is that of the stolen property (Ogden, Dawnay and McEwing, 2009). Parentage analysis is common in cases where illegally obtained wild animals are sold on as domestic pets (Ogden, Dawnay and McEwing, 2009). Illegal trade of rhino horn is one of the most commonly traded animal products on the black market, STR multiplexing has proved influential in the fight against this (Harper *et al.*, 2013; (Harper *et al.*, 2018). The identification of livestock predators is important in deciding the best preventive plan in order to protect your animals, Caniglia *et al.* (2013) used STRs to identify wolves which were then linked to 13 sheep, it was found by matching saliva from the carcasses to the wolves that the wolves were responsible for killing the sheep (Caniglia *et al.*, 2013). Lastly, and the topic that this project is focused on is poaching, i.e. the illegal killing of wildlife. The first ever case which was brought to court in 2009 for trial against a suspected poacher of wolf, this was reported by Caniglia *et al.* (2010). STRs were used to identify the items on a

necklace as teeth from wolf and linked them to carcasses found in the local area (Caniglia *et al.*, 2010). This provided strong evidence to support the prosecution of the poacher in court (Caniglia *et al.*, 2010).

These are prime examples of how STR profiling has been utilised by wildlife forensic geneticists to provide crucial evidence in order to gain a successful prosecution at court, showing how genetics is being implemented in the fight against wildlife crime, further supporting the importance of this project.

Fallow deer (*Dama dama*), are one of the main deer species in the UK and are the most widely spread throughout England, Wales, Scotland, and Northern Ireland, followed by red deer and roe deer. Over 50,000 fallow deer are illegally killed each year due to their expensive cuts of meat, fashionable antlers and sought after furs making them prime targets to poachers. Furthermore, red deer and roe deer already have STR multiplexes available for implementation in forensic case work (Szabolcsi *et al.*, 2014; Morf *et al.*, 2021) – respectively. The characterisation of suitable STR loci is crucial for the development of individual identification tools, however, nothing is yet available for fallow deer. Therefore, this project was the first to investigate suitable STR loci for individual identification, providing an original contribution to the literature and more importantly to forensic wildlife genetics.

4.2 Method

In this study, a total of 362 fallow deer samples were collected from various locations across the UK in an attempt to achieve a representative sample pool of the entire UK fallow deer population. This study initially set out to develop a STR multiplex for the identification of fallow deer individuals, however, due to the lack of genetic diversity within the species it was realised that this would not be possible. Therefore, the study was pivoted to utilise Oxford Nanopore Sequencing in order to compare the best approach for finding STR polymorphisms within novel species. Due to the complex nature of this study and the large volume of work done, the flow chart shown in Figure 4.1 details out the order of the work.

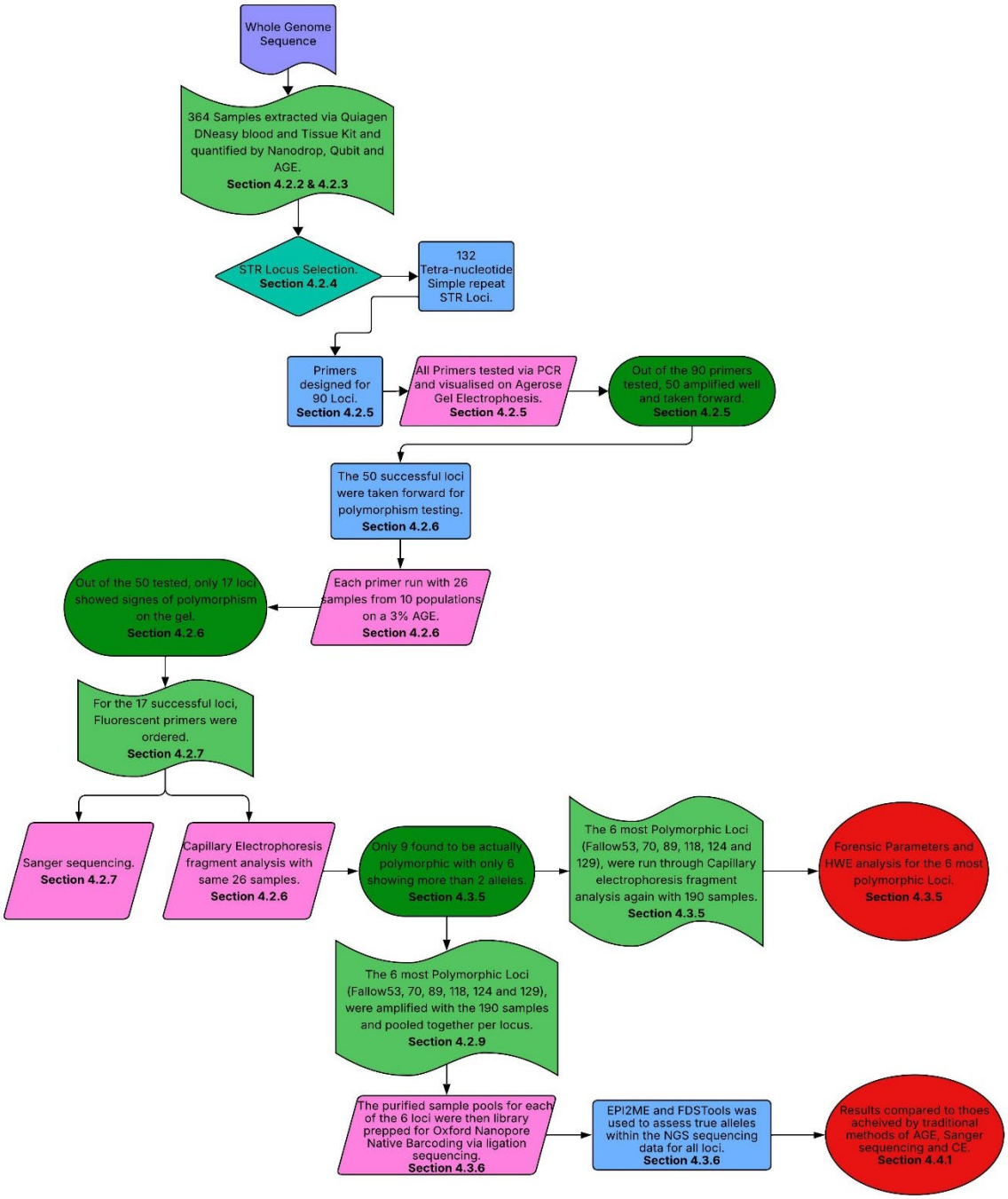


Figure 4.1: *Flow chart depicting the order of study 2.*

4.2.1 Sampling

For this study, a total of 362 fallow deer samples were collected from various locations across the UK (Figure 4.2), for the full list of samples collected for this project, see the sample database, within the supplementary data file S.9_Sample Data.

Figure 4.2: *A map to show the locations across the UK where fallow deer have been sampled. The red dots depict sample sites from the 1st round of sampling (2020-2021) and the blue dots from the 2nd round of sampling (2023-2024).*

4.2.2 DNA Extraction

A total of 362 samples were DNA extracted. Triplicate DNA extractions from whole blood was performed using the DNeasy Blood and Tissue Kit, following the Protocol: Purification of Total DNA from Animal Blood (Spin-Column Protocol) found on page 25 of the DNeasy® Blood & Tissue Handbook (QIAGEN, Germany), using the amendments listed below. A negative control was included with every batch of extractions.

Step 2: Incubation time increased to 10 mins.
Step 3: Incubation time increased to 30 mins on a thermomixer.
Step 8: Incubation time increased to 10 mins.

4.2.3 DNA quantification

Once DNA was extracted, it was quantified. Three methods of DNA quantification were used to assess sample quality and quantity.

Firstly, DNA was quantified using the NanoDrop 2000/2000c Spectrophotometer (Thermo Fisher Scientific, USA) following the manufactures guidelines, found in section 3, page 1-7 of the NanoDrop 2000/2000c Spectrophotometer user manual version 1.0. Each DNA extract was run three times on the NanoDrop 2000/2000c Spectrophotometer, and an average was made for an accurate quantification result. See results section 4.3.1.

Secondly, DNA was quantified using the Invitrogen™ Qubit™ 3 Fluorometer using the Qubit™ 1X dsDNA HS Assay Kit (Thermo Fisher Scientific, USA) following the manufactures guidelines, found on page 3-4 of the user guide. Each DNA extract was run three times on the Invitrogen™ Qubit™ 3 Fluorometer, and an average was made for an accurate quantification result. See results section 4.3.1.

Thirdly, Agarose Gel Electrophoresis (AGE) was used to Quantify DNA on a 1% 1x TAE agarose gel stained with gel red (Biotium) and visualised using a BioRad Molecular Imager® Gel Doc™ XR+ with image lab software. Comparison of band brightness was compared to known concentrations of Control Lambda DNA (Thermo Fisher Scientific, USA). 6 dilutions of control DNA with a final volume of 100µl were set up to 100ng/µl, 50ng/µl, 25ng/µl, 10ng/µl, 5ng/µl and 1ng/µl concentrations. 2µl of tracking dye is added to 4µl of the 6 dilutions, mixed and then 6µl of each can be loaded onto the gel. If DNA was present a band was visible on the gel, if DNA extraction had not been successful then no band was visible, and DNA extraction was repeated. See section 4.3.1 for results.

4.2.4 Locus Selection

The genome, sequenced in study 1, was mined for suitable STR loci via the online bioinformatics platform, Galaxy, using a STR detection tool (Fungtammasan *et al.*, 2015). Di-nucleotide, Tri-nucleotide, Tetra-nucleotide, Penta-nucleotide and Hexa-nucleotide repeats were all mined, but for the purposes of this study, only the tetranucleotide loci were utilised. A total of 15,589 tetra-nucleotide loci were identified in the fallow deer genome. 90 of which were selected for primer design. Potential polymorphic loci were selected for primer design based upon the repeat motif (Simple tetranucleotide repeats), length of repeat (between allele 5 and 15) and their location along the genome (two or three potential loci were selected per chromosome).

4.2.5 Primer Design and Selection

Primer sets used in this study are shown in section 10.1 along with the expected band size, location, and repeat motif. A split of Simple, Simple complex and hypervariable complex loci were chosen for the study which had either tetra-nucleotide or penta-nucleotide repeat motifs. This was to ensure the multiplex contained highly discriminatory loci. Furthermore, all primers were designed with a similar annealing temperature in mind, to ensure they would work together in a multiplex. 90 Primers were ordered and tested, if successful, primers were taken further for polymorphism (Section 4.2.6).

Each primer set was run through PCR. Amplification reactions consisted of a 15µl volume PCR containing 1x of Platinum® PCR Multiplex Master Mix (Life Technologies, USA), 3mM of MgCl2 (Fisher Scientific, UK) 0.1uM of both forward and reverse primer and 6ng DNA extract. For negative controls, nuclease free water was used instead of DNA extract. Samples were run on a gradient PCR testing 5 different annealing temperatures.

Amplifications were performed on the Applied Biosystems® Veriti® 96-Well Thermal Cycler (Fisher Scientific, UK), under the following conditions: five-minute denaturation step at 95°C followed by 32 cycles of 30 seconds at 94°C, 30 seconds at 54°C/56°C/58°C/60°C/62°C, 30 seconds at 68°C and a final extension step for ten minutes at 68°C.

PCR success was assessed by running 2µl of each reaction on a 1.5% agarose gel which was prepared as follows: For a 17 well gel, 1.5g of agarose (Promega, Spain)

was dissolved in 100ml of 1x TAE (homemade, Fisher Scientific, UK). Once the solution had cooled 1µl of gel red (10,000x in DMSO) (Biotium, USA) was added. The gel red was thoroughly mixed into the solution before pouring into the casting tray and allowing the gel to set. To make 1x TAE 20ml of 50x TAE (homemade, Fisher Scientific, UK) was mixed with 980ml of distilled water.

Once the gel was set it was placed into the electrophoresis tank and submerged in 1x TAE (homemade, Fisher Scientific, UK). 2µl of each sample, the negative control and a 1 in 10 100bp ladder (GeneCraft, Germany) were mixed with 2µl of tracking dye and loaded into separate wells on the gel. The gel ran at 100v for 30 minutes or until the tracking dye had moved just over halfway through the gel. The gel was visualised using a BioRad Molecular imager® Gel DocTM XR+ with Image LabTM Software. See section 4.3.2 for results. If the PCR had failed, no bands were shown, and primer set was re-tested. If the primer set successfully amplified in fallow deer, a band of expected size would be visible and therefore this primer set is taken forward to the next stage and is tested to see if it is polymorphic see section 4.2.6.

4.2.6 Polymorphism Testing

Loci which exhibited amplification in all 15 tested samples of fallow deer, were tested to see whether they were polymorphic. Firstly, a PCR was set up using the 50 primer sets which amplified successfully. Each primer set was run with 26 fallow deer DNA samples from 10 populations across the country.

Amplification reactions consisted of a 15µl volume PCR containing 1x of Platinum® PCR Multiplex Master Mix (Life Technologies, USA), 3mM of MgCl2 (Fisher Scientific, UK) 0.1uM of both forward and reverse primer and 6ng DNA extract. Samples were run on a gradient PCR testing 5 different annealing temperatures.

The samples were placed into The Applied Biosystems® Veriti® 96-Well Thermal Cycler (Fisher Scientific, UK). The PCR was run under the following conditions: five-minute activation step at 95°C followed by 32 cycles of 30 seconds at 94°C, 30 seconds at 62°C, 30 seconds at 68°C and a final extension step for ten minutes at 68°C.

PCR success was assessed by running a large 3% 28 well gel which was prepared using the following protocol. 12g of agarose (Fisher Scientific, UK) was added to 400ml 1X TAE (homemade, Fisher Scientific, UK) in a microwave proof glass bottle and heated in a microwave for 5-10 min, until the agarose had dissolved. Unlike the smaller 1% gels this 3% gel was not allowed to cool before adding the gel red as the solution would begin to set. 10µl of gel red (10,000x in DMSO)

(Biotium, USA) was added to the agarose solution and poured into the casting tray. Once the gel had set it was placed into the electrophoresis tank and submerged in 1x TAE (homemade, Fisher Scientific, UK). Each sample was loaded into each well and run for 6 hours and at 80V. The gel was visualised using a BioRad Molecular imager® Gel DocTM XR+ with Image LabTM Software (Bio-Rad, USA).

4.2.7 Fragment Analysis

The loci which appear to be polymorphic on the gel had fluorescently labeled primers ordered for them and are run on the ABI genetic analyser 3500 to analyse the fragments.

1µl of amplified sample was added to 11µl of HiDi formamide with 0.5µl Genescan Liz500 size standard and placed into a 96 well plate. Once it had been assured that no bubbles were present in the wells, the septum was placed on top and the samples denatured using a PCR machine set to 95°C for 5 minutes. Following denaturation, the plate was immediately placed in the freezer for 5 minutes prior to running on the ABI3500 via fragment analysis.

4.2.8 Sanger Sequencing

4.2.8.1 DNA Purification for Cycle Sequencing

PCR products were purified using The DNA Clean & Concentrator-5 PCR purification kit (Zymo Research, USA) following manufacturers guidelines on page 4 of the DNA Clean & Concentrator-5 handbook.

4.2.8.2 Cycle Sequencing

Cycle sequencing was performed using the purified amplified DNA. Cycle sequencing was done using the BigDye™ Terminator v3.1 Cycle Sequencing Kit following the manufacturers protocol with the modification of using half volume reactions.

4.2.8.3 Post Cycle Sequencing DNA Purification

After cycle sequencing the DNA was purified using the in-house protocol for PCR purification. Per sample reaction, 1µl of 3M NaoAc, 1µl of 20µg/µl Glycogen, 1µl of 100mM EDTA and 30µl of cold 100% proof ethanol was added. This was vortexed briefly and left overnight to incubate. Following the incubation, the reactions were centrifuged at top speed for 30 minutes at 3°C. The supernatant was then removed, leaving the pellet in place and washed with fresh 70% ethanol. This was then centrifuged at top speed for 15 min. This wash step was repeated

twice. Next, to air dry the pellet, the tubes were placed with lids open on a PCR machine set at 50°C for 10 min. When ready to run on the ABI 3500, the DNA was suspended in 13µl of HiDi formamide.

4.2.8.4 DNA Sequencing Analysis

Once DNA had been purified after cycle sequencing, the samples were prepared and sequences were analysed via capillary electrophoresis on the Applied Biosystems (ABI) 3500. Pop 6 polymer was used on the following instrument protocol: StdSeq50_POP6_Z. The base calling protocol used was the following: BDTv3.1_PA_Protocol-POP6. The injection time was 8 seconds and run time was 5000s.

4.2.9 Rapid Polymorphism screening via Oxford Nanopore

Taking primers Fallow53, Fallow70, Fallow89, Fallow118, Fallow124 and Fallow129, 190 samples were amplified with each primer set. For the full list of samples collected for this project, see the sample database, within the supplementary data file S.9_Sample Data. PCR conditions were used as set out in section 4.2.4, using their associated optimised annealing temperature. For Sample details and PCR set up, please see supplementary data file S.8_Nanopore Sequencing Study's_D-Loop & STR Nanopore Studies Set up.xlsx.

Following PCR amplification, each sample, per locus, was pooled together. In order to not incur any sequencing bias, all samples were pooled together in equal concentration. Once samples had been pooled together, the sample pools were purified using The DNA Clean & Concentrator-5 PCR purification kit (Zymo Research, USA) following manufacturers guidelines on page 4 of the DNA Clean & Concentrator-5 handbook.

The purified sample pools for each of the 6 loci were then library prepped for Oxford Nanopore Native Barcoding via ligation sequencing. The Native Barcoding Kit 24 V14 kit by Oxford Nanopore was used following manufacturers guidelines. Samples were loaded onto the flow cell and run until the flow cell was exhausted. Basecalling was carried out using MinKNOW and further alignment and amplicon variation analysis was conducted using EPI2ME software.

To assess true alleles within the NGS sequencing data for all loci, FDSTools was used. FDSTools, using the TSSV function can link raw reads from the FASTq files to known markers i.e. known primer sequences and counts the number of reads for each unique sequence (Hoogenboom *et al.*, 2017; Magoč & Salzberg,

2011). FDSTools is operated on command line via Python. All details on FDSTools can be found on their website (https://www.fdstools.nl/tools.html).

For the TSSV tool, the following command lines were used to produce the sequence data output followed by the named allele output (on all unmerged fasta.gz files):

```
for infile in *.fastq.gz; do fdstools tssv --report tssv-report-
    ${infile%.fastq.gz}.txt deer.ini $infile tssv-output-
                ${infile%.fastq.gz}.txt; done
```

```
for infile in *.fastq.gz; do fdstools tssv --sequence-format
    allelename deernaming.ini $infile tssv-output-named-
                ${infile%.fastq.gz}.txt; done
```

After Assigning the raw reads to the known markers, the true alleles were visualised via FDSTools using SampleVis (Hoogenboom *et al.*, 2017). The following command line was used to compute the results in html format for web browser viewing:

```
for infile in *tssv-output-named-*.txt; do fdstools vis sample
    $infile samplevis-${infile%.fastq.gz}.html; done
```

Alternatively, in order to run FDSTools on all sequence reads at once and produce a single results file output, the following command was used to merge all fasta.gz files into one:

```
cat *.gz > merged.fasta.gz
```

After merging files, the following command lines were used to compute the sequence data output followed by the named allele output:

```
fdstools tssv --report tssv1-report.txt deer.ini  merged.fasta.gz
                        tssv1-output.txt
```

```
fdstools tssv --sequence-format allelename deernaming.ini
            merged.fasta.gz  tssv-output-named.txt
```

Followed by the visualisation command:

```
fdstools vis sample tssv-output-named.txt tssv-samplevis.html
```

4.3 Results and Analysis

4.3.1 DNA Quantification

Results of the DNA Quantification for all samples extracted for this study, in triplicate, can be found in the supplementary data file S.2_DNA Quantification. This includes the quantification conducted using NanoDrop™ Spectrometry, Qubit™ Fluorometry and Agarose Gel Electrophoresis. All samples were taken forward with the assurance that the negative control had passed.

4.3.2 Locus Selection and Primer Design

Table 4.1 represents the total number of STR's found in each chromosome. 537,466 STRs were found in total when searching for the minimum allele of 5.

Table 4.1: *Total number of STR's found in each chromosome when searched for the minimum repeat of allele 5.*

Fallow Deer	Number of STR Repeats				
	Di-Nucleotide	Tri-Nucleotide	Tetra-Nucleotide	Penta-Nucleotide	Hexa-Nucleotide
Chromosome 1	17,024	1,903	748	287	95
Chromosome 2	10,338	1,092	511	136	135
Chromosome 3	10,753	1,292	440	170	174
Chromosome 4	14,201	1,527	603	157	55
Chromosome 5	27,995	3,202	1,177	343	166
Chromosome 6	10,250	1,331	504	203	86
Chromosome 7	11,054	1,211	412	142	19
Chromosome 8	10,618	1,439	493	207	26
Chromosome 9	20,990	2,590	776	295	19
Chromosome 10	8,551	979	436	108	43
Chromosome 11	20,711	2,498	826	326	72
Chromosome 12	19,062	2,327	741	288	319
Chromosome 13	14,100	1,596	531	224	232
Chromosome 14	15,818	1,801	648	246	187
Chromosome 15	18,843	2,177	706	348	77
Chromosome 16	9,582	1,042	369	147	64
Chromosome 17	190	64	7	6	0
Chromosome 18	22,479	2,566	725	311	93
Chromosome 19	18,413	2,250	758	321	23
Chromosome 20	8,662	1,024	441	144	12
Chromosome 21	16,089	1,894	445	230	24
Chromosome 22	11,829	22,410	366	170	7
Chromosome 23	8,589	857	252	130	4
Chromosome 24	11,815	1,251	508	121	82

Chromosome 25	14,923	1,702	527	210	53
Chromosome 26	1,042	89	36	24	1
Chromosome 27	13,525	1,389	454	174	60
Chromosome 28	12,722	1,448	456	223	8
Chromosome 29	12,582	1,409	432	190	26
Chromosome 30	8,905	254	29	0	0
Chromosome 31	5,176	131	27	0	0
Chromosome 32	15,535	934	323	125	41
Chromosome 33	14,906	1,728	630	272	19
Total:	**420,248**	**67,504**	**15,589**	**5,991**	**2,127**

132 tetra-nucleotide loci were shortlisted due to their optimal size, genomic location and repeat motif, and primers were designed. For full details of the designed primers see supplementary data file S.1_WGS_STR's_Shortlisted Loci_ Fallow Deer - Shortlisted STR's. Due to time restraints, out of the 132 tetra-nucleotide primers designed, 90 primers tested in fallow deer (Table 10.1 in section 10.1 of the appendix). 50 primer pairs worked well and were taken forward for polymorphism testing. See Figure 4.3, Figure 4.4 and Figure 4.5 for a few examples of these, the rest can be found in the supplementary data file, S.3_Primer Screening. Most primers worked optimally at 62°C, the optimal temperature shown by the results of the gradient PCR was chosen for the polymorphism testing PCR.

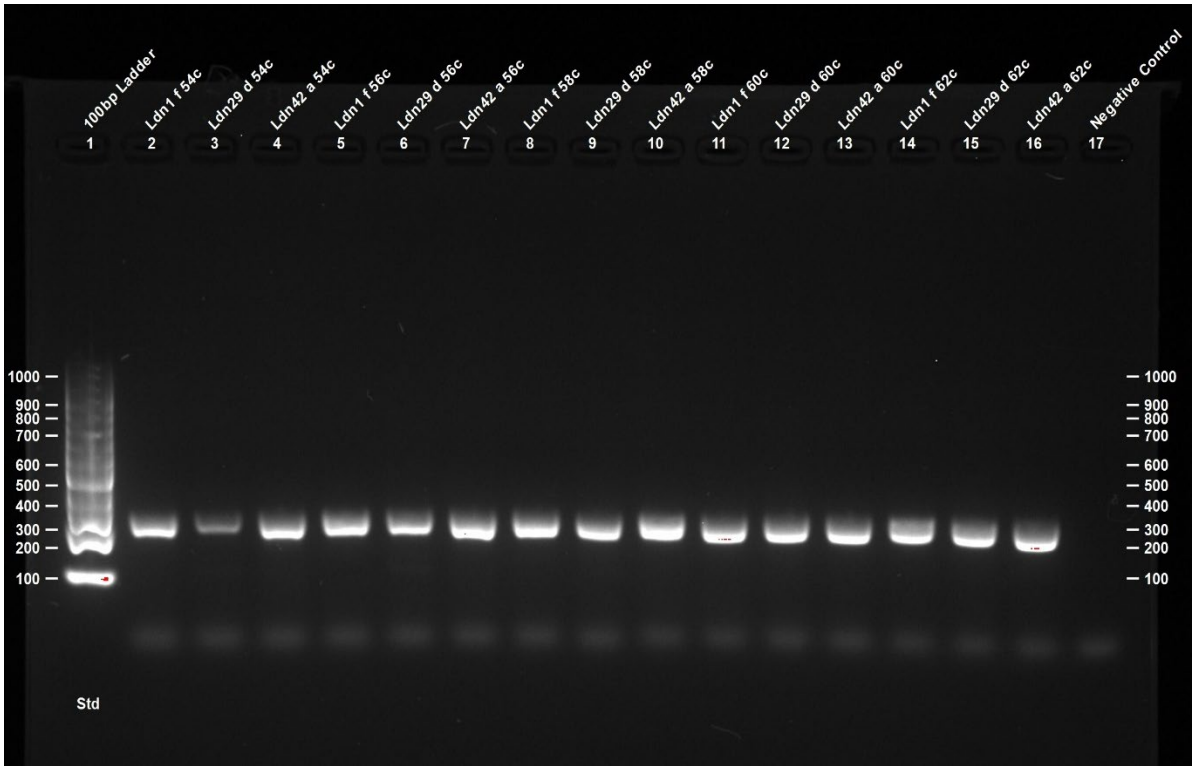


Figure 4.3: *Electrophoresis gel image of primer pair Fallow4 gradient PCR Result. Image shows successful amplification of 15 samples run on a 1.5% agarose gel.*

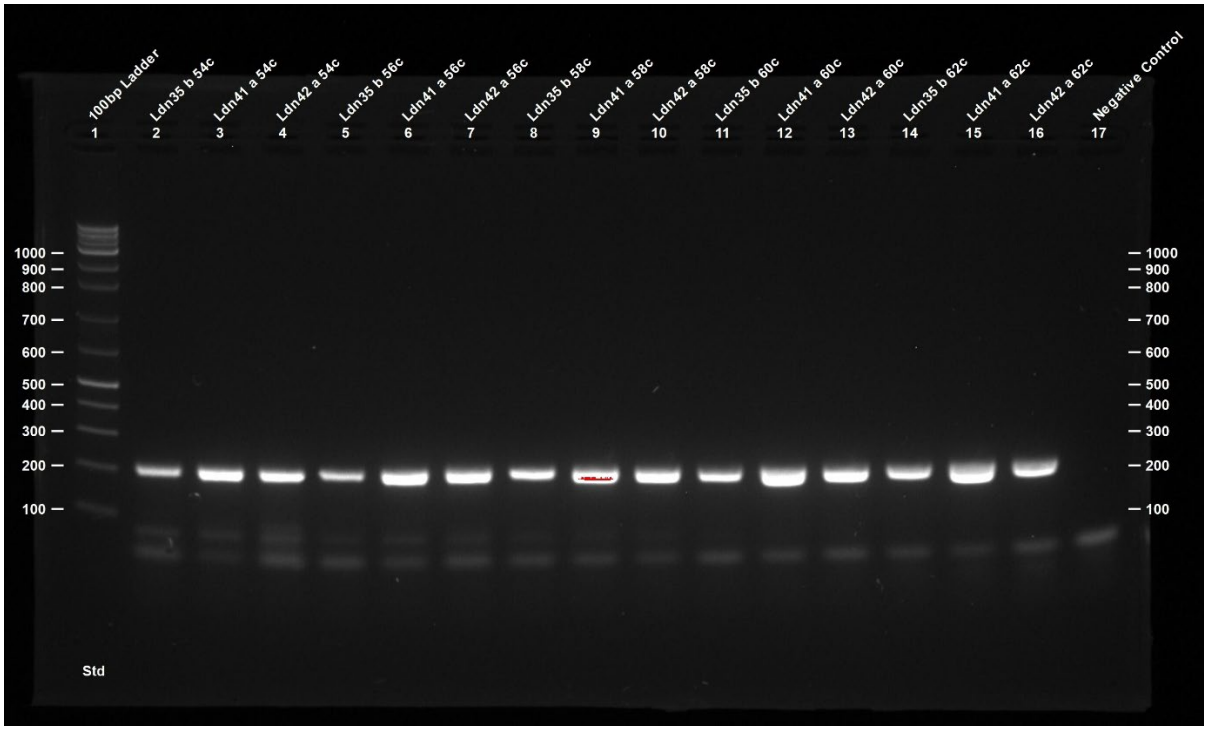


Figure 4.4: *Electrophoresis gel image of primer pair Fallow120 gradient PCR Result. Image shows successful amplification of 15 samples run on a 1.5% agarose gel.*

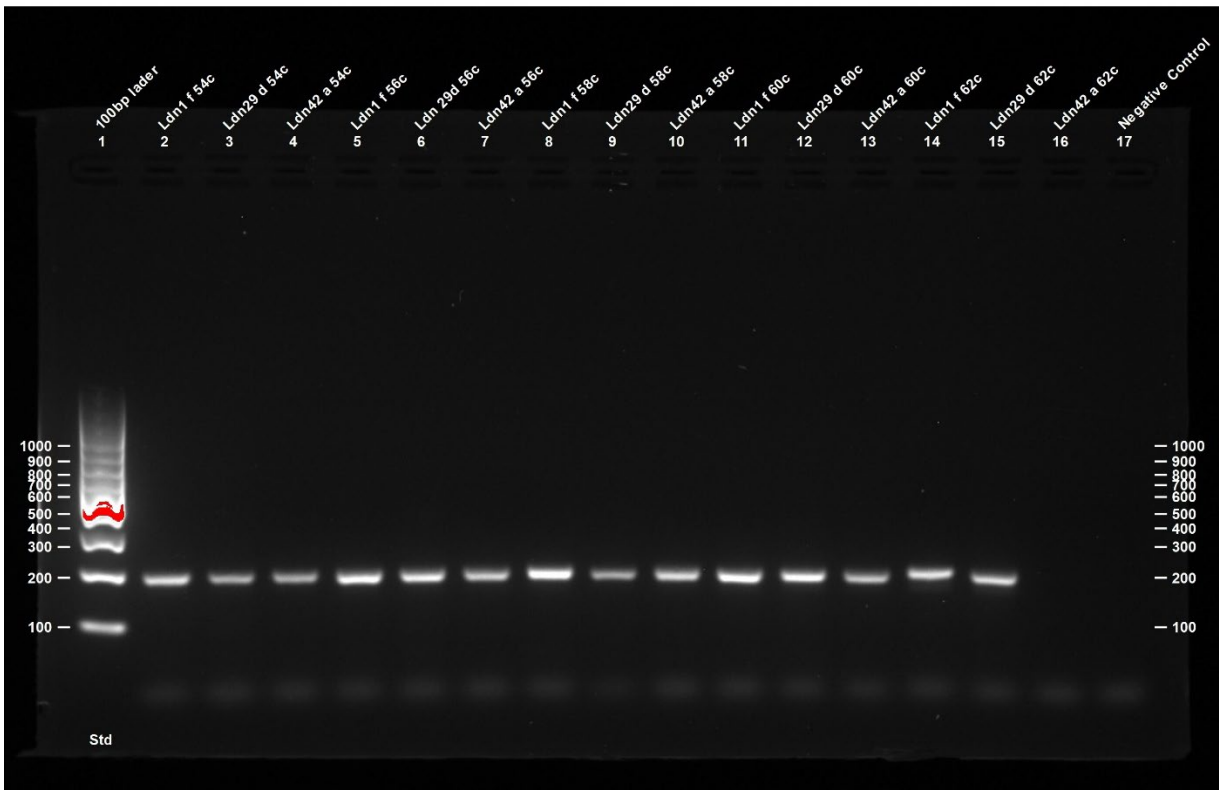


Figure 4.5: *Electrophoresis gel image of primer pair Fallow125 gradient PCR Result. Image shows successful amplification of 15 samples run on a 1.5% agarose gel.*

4.3.3 Polymorphism Testing

50 primers were tested across 26 samples for polymorphism. Out of these, 17 were found to show polymorphism (Table 4.2), see Figures 4.6 - 4.9 for examples of these. Full results of the polymorphism testing can be found in the supplementary data file, S.4_Polymorphism Testing.

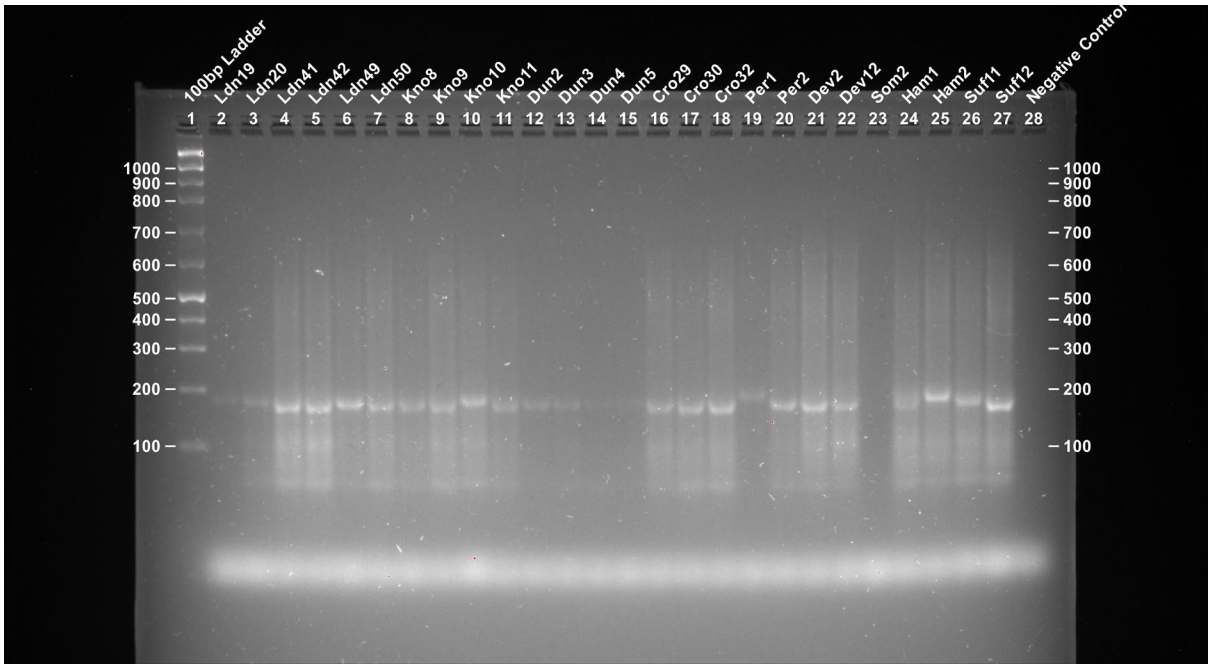


Figure 4.6: *Fallow4 polymorphism gel image. Successful amplification of 26 samples run on a 3% agarose gel.*

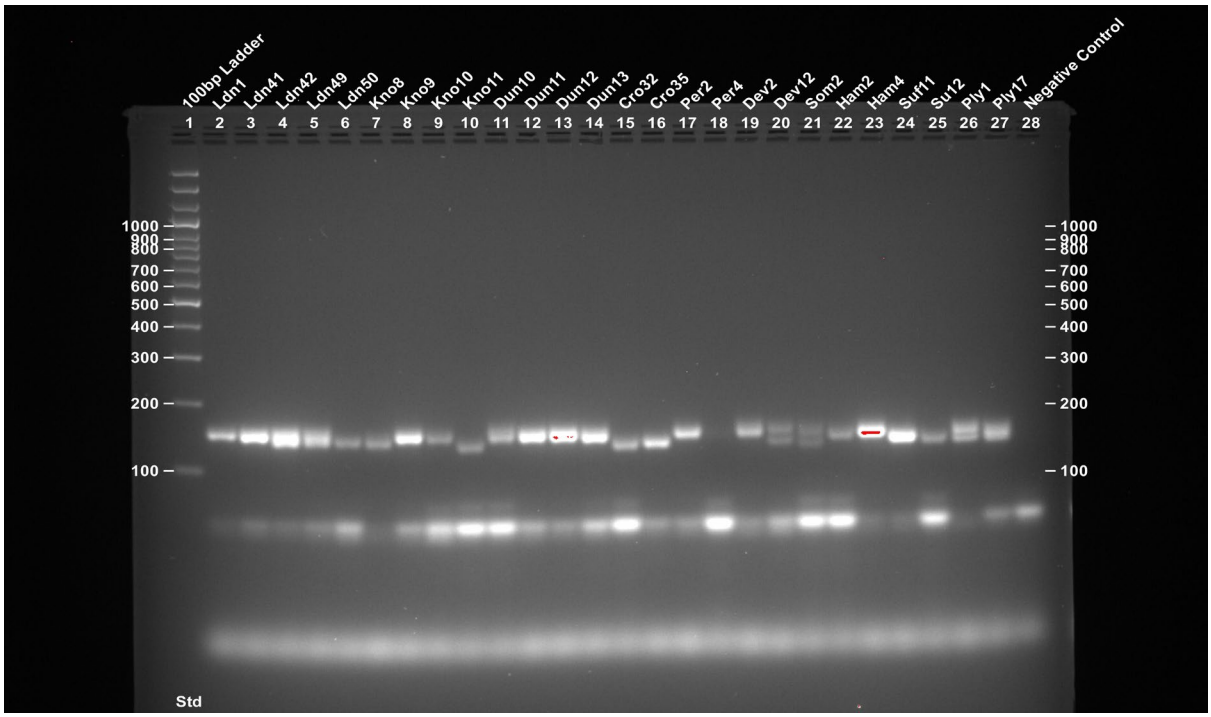


Figure 4.7: *Fallow53 polymorphism gel image. Successful amplification of 26 samples run on a 3% agarose gel.*

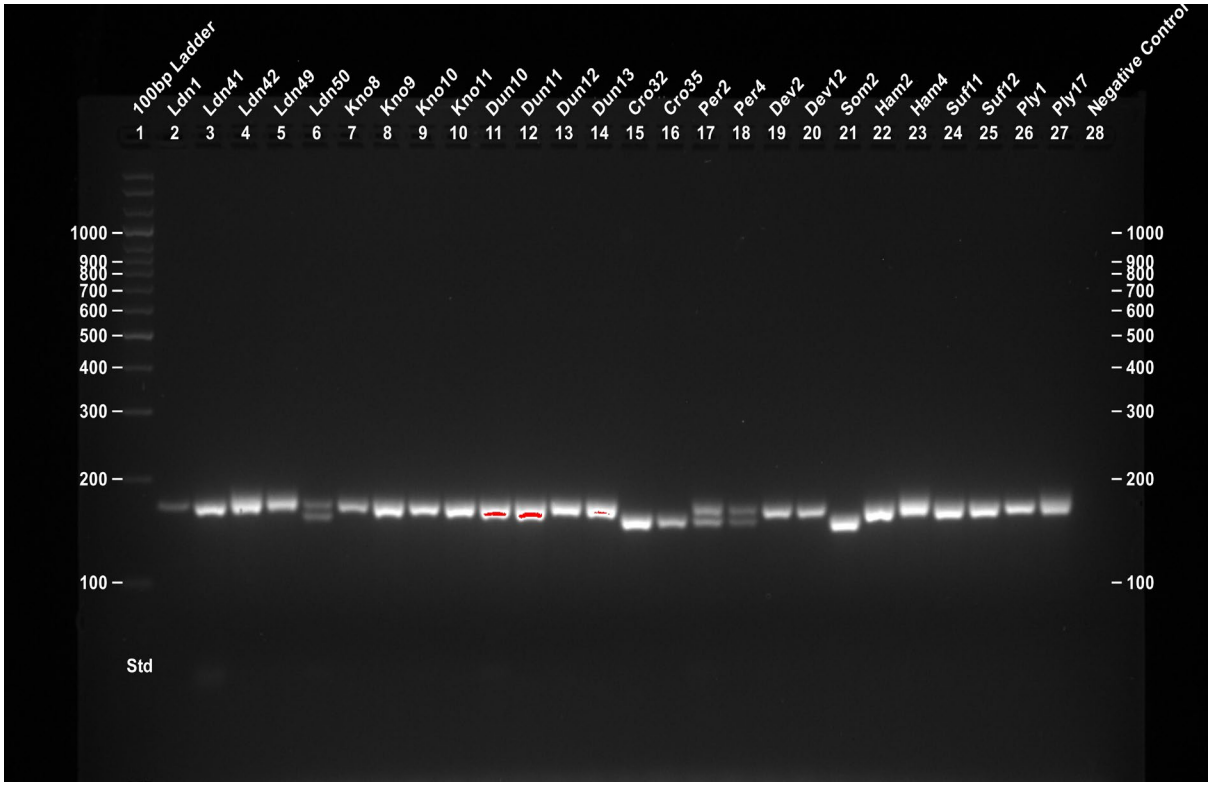


Figure 4.8: *Fallow70 polymorphism gel image. Successful amplification of 26 samples run on a 3% agarose gel. Despite some samples showing a tri-allelic pattern, this was not supported by the CE results, despite running repeats.*

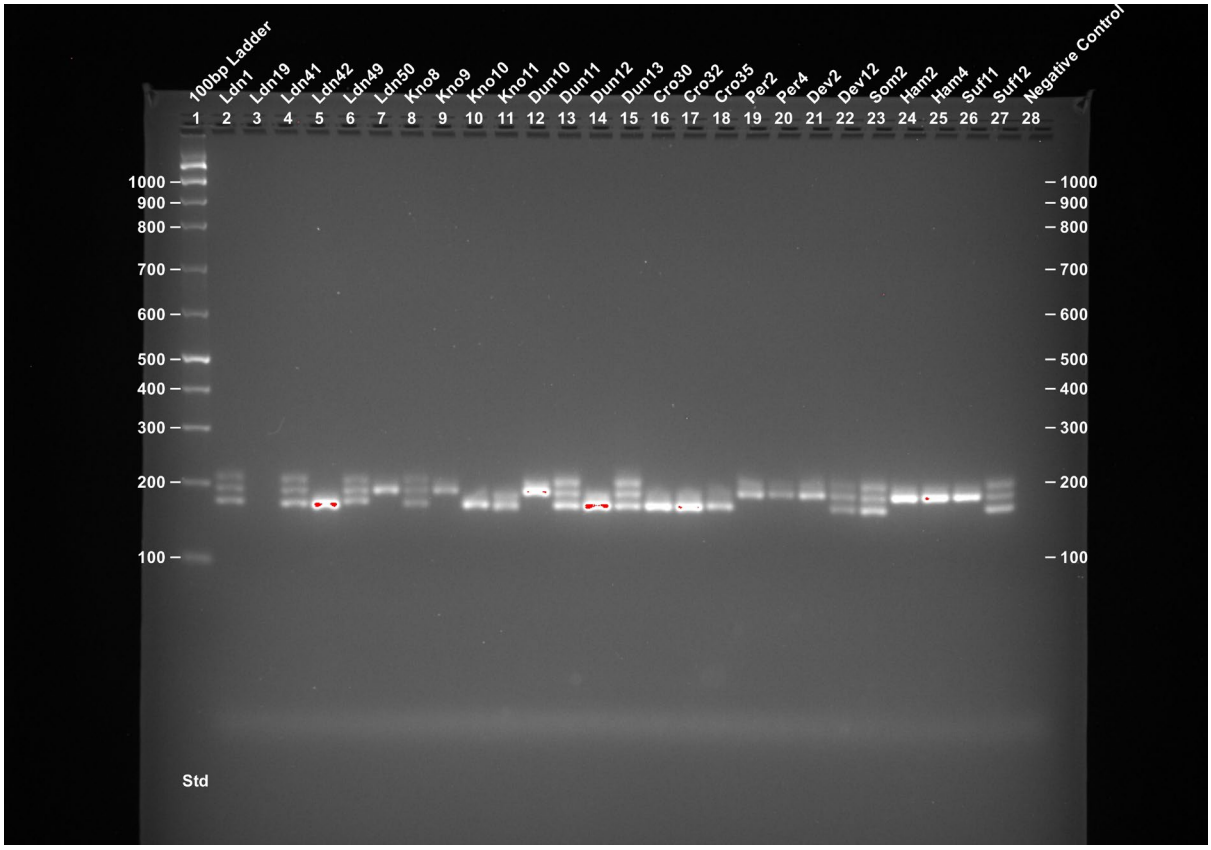


Figure 4.9: *Fallow118 polymorphism gel image. Successful amplification of 26 samples run on a 3% agarose gel.*

4.3.4 Fragment analysis

17 Primers which showed signs of polymorphism on the 3% Agarose gels were taken forward for fragment analysis using fluorescent primers to confirm allele size and determine if the locus is truly polymorphic, see Table 4.2 for a summary of the fragment analysis results. Further results of the fragment analysis testing can be found in the supplementary data file, S.5_Fragment Analysis.

Table 4.2: Summary of fragment analysis results.

Locus	Expected Fragment Length	Actual Fragment Length	Number of Alleles Present
Fallow4	186	183 – 200	3
Fallow6	158	152 – 157	2
Fallow10	153	153 – 157	2
Fallow12	212	212	1
Fallow17	189	187	1
Fallow26	231	221 – 229	2
Fallow37	202	196 – 200	2
Fallow53	142	139 – 156	7
Fallow64	221	219	1
Fallow70	174	155 – 171	3
Fallow89	115	99 – 111	3
Fallow94	150	146 – 159	3
Fallow101	124	115 – 120	3
Fallow109	180	173 – 177	2
Fallow118	200	173 – 197	7
Fallow124	184	182 – 190	3
Fallow129	220	210 – 223	3

Out of the 17 Primers showing polymorphism on the gels, only 9 actually presented polymorphism on the ABI 3500 fragment analysis (Table 4.2). Overall number of alleles found at each locus tested via fragment analysis on the ABI3500 can be seen in the supplementary data file S.5_Fragment Analysis.

4.3.5 Forensic Parameter's and Hardy-Weinberg Equilibrium Analysis

After the initial fragment analysis described in section 4.3.4, the most polymorphic loci were found to be Fallow53, 70, 89, 118, 124 and 129. Fragment analysis for these 6 most polymorphic loci were increased from 26 samples to 190. See Table 10.2 for the results (found in section 10.1 of the appendix). Based on the 190 samples, STRAF was used to compute forensic parameters, including HWE (Table 4.3, Figure 4.10, Figure 4.11 and Figure 4.12). Following fragment analysis, each homozygous polymorphic allele were sequenced via sanger sequencing, as detailed in section 4.2.8. See supplementary data file S.6_HWE Fragment Analysis for the full results including the sequencing and statistical analysis.

Table 4.3: *Summary of forensic parameters calculated using STRAF 2.1.5: STR Analysis for Forensics (Gouy & Zieger, 2017).*

locus	N	Nall	GD	PIC	PM	PD	Hobs	PE	TPI
Fallow118	50.00	7.00	0.60	0.50	0.25	0.75	0.44	0.14	0.89
Fallow124	50.00	3.00	0.36	0.30	0.51	0.49	0.24	0.04	0.66
Fallow129	50.00	3.00	0.38	0.32	0.47	0.53	0.48	0.17	0.96
Fallow53	52.00	7.00	0.79	0.74	0.18	0.82	0.15	0.02	0.59
Fallow70	52.00	3.00	0.50	0.43	0.43	0.57	0.15	0.02	0.59
Fallow89	52.00	3.00	0.54	0.44	0.44	0.56	0.12	0.01	0.57

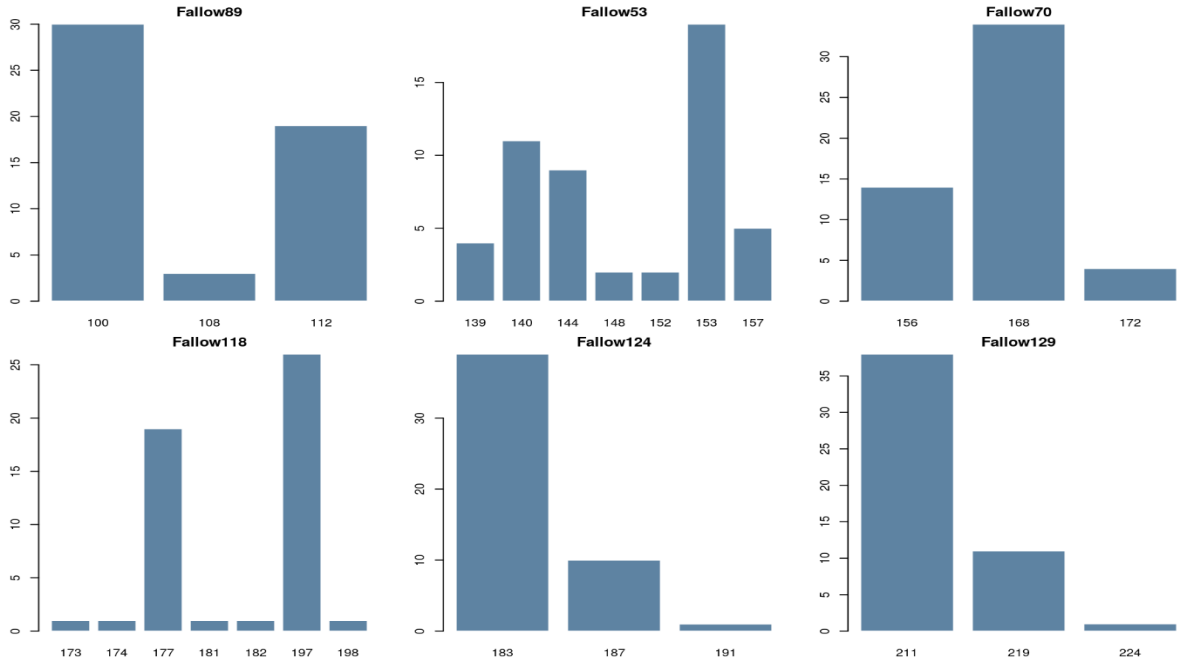


Figure 4.10: *Graphical representation of the abundance of alleles at each locus, developed using STRAF 2.1.5: STR Analysis for Forensics (Gouy & Zieger, 2017).*

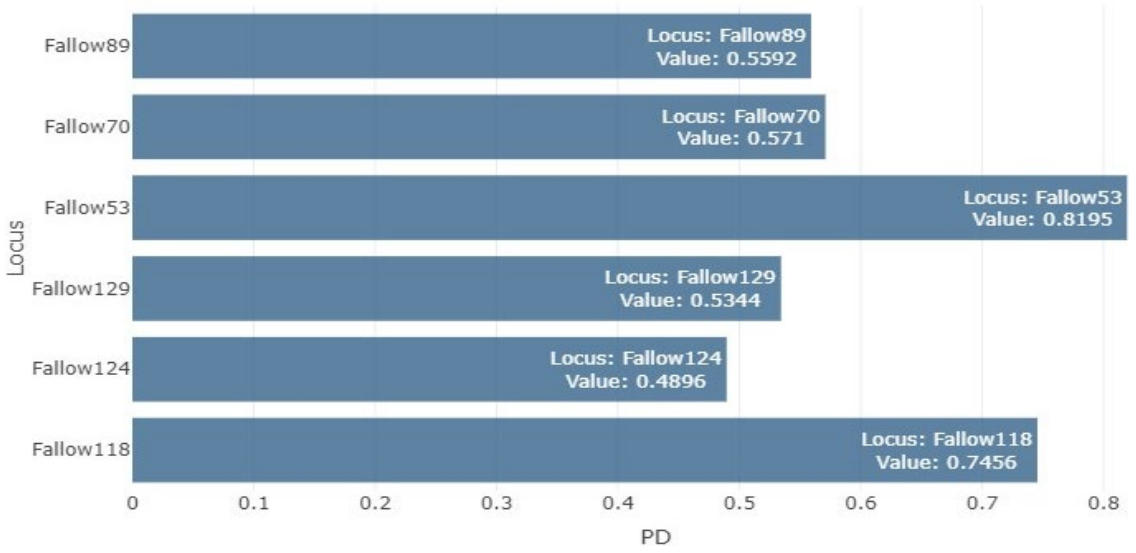


Figure 4.11: *Graphical representation of the power of discrimination (PD) at each locus, developed using STRAF 2.1.5: STR Analysis for Forensics (Gouy & Zieger, 2017).*

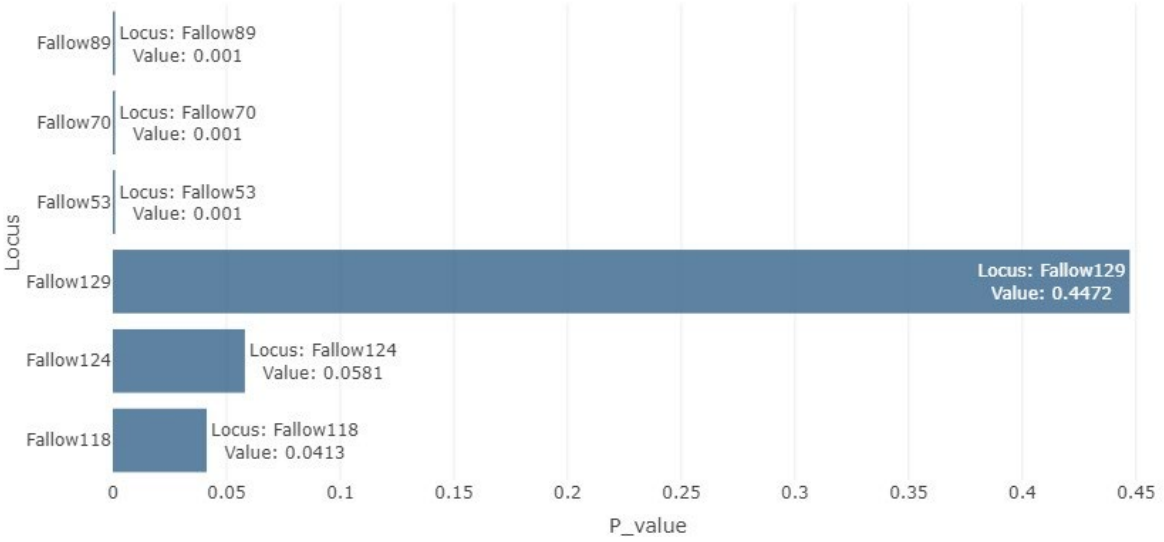


Figure 4.12*: Graphical representation of the P-values for each locus, developed using STRAF 2.1.5: STR Analysis for Forensics (Gouy & Zieger, 2017).*

4.3.6 Rapid Polymorphism screening via Oxford Nanopore

4.3.6.1 EPI2ME Analysis Results

Results of the Amplicon analysis conducted on EPI2ME can be viewed in Table 4.4. Alignments were assessed for variation in the form of SNVs and Indels, IGV alignment outputs can be viewed in Figures 4.13 – 4.18, these are generated as part of the EPI2ME alignment pipeline. Unfortunately, this software is limited to SNVs and Indels and cannot assess non-human STR variation. For the full results from the EPI2ME analysis, see supplementary data file, S.82_STR_NGS Study.

Table 4.4: *Summary of the EPI2ME analysis for all 6 loci sequenced by the ONT GridION.*

Locus	Reads	Mean Length	SNVs	Indels
Fallow53	191,459	144.7	2	2
Fallow70	3,245,760	167.3	4	1
Fallow89	608,239	109.9	6	1
Fallow118	373,121	175.9	2	1
Falllow124	72,168	174	3	0
Fallow129	148,587	194.5	0	1

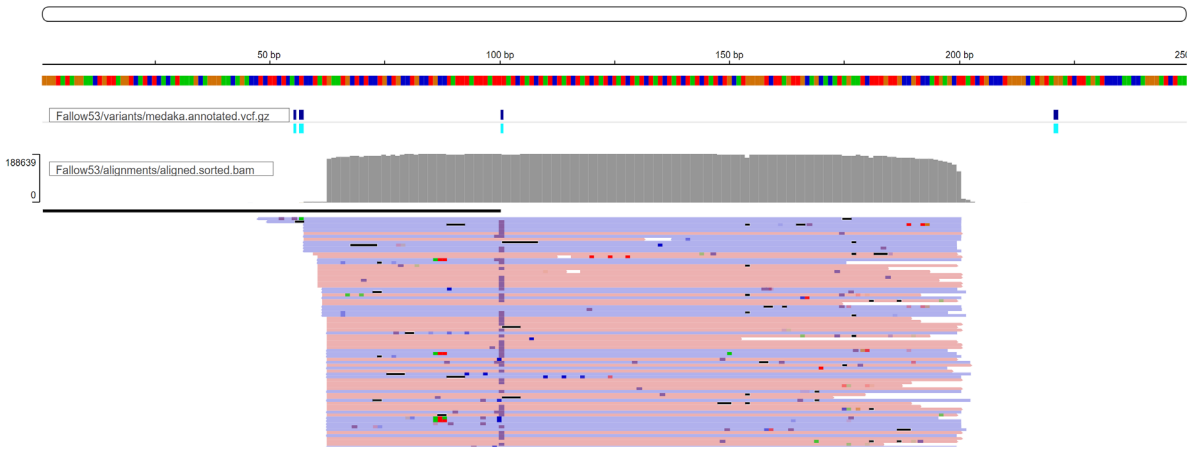


Figure 4.13: *Integrative Genomics Viewer (IGV) output for the alignment and variation analysis of Fallow53, produced my EPI2ME. Dark shading and colour blocks depict the SNVs and the Indels variation.*

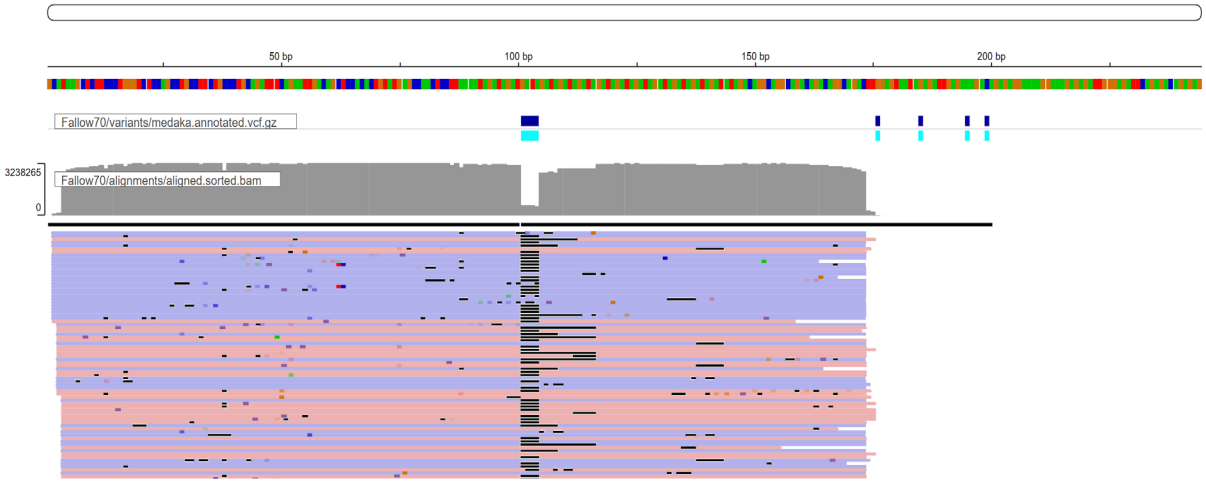


Figure 4.14: *Integrative Genomics Viewer (IGV) output for the alignment and variation analysis of Fallow70, produced my EPI2ME. Dark shading and colour blocks depict the SNVs and the Indels variation.*

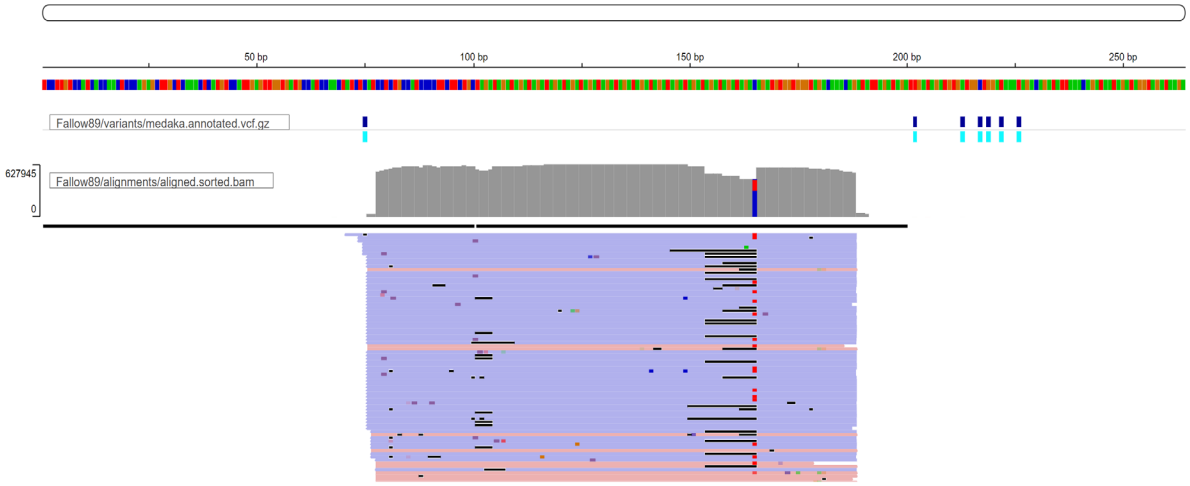


Figure 4.15: *Integrative Genomics Viewer (IGV) output for the alignment and variation analysis of Fallow89, produced my EPI2ME. Dark shading and colour blocks depict the SNVs and the Indels variation.*

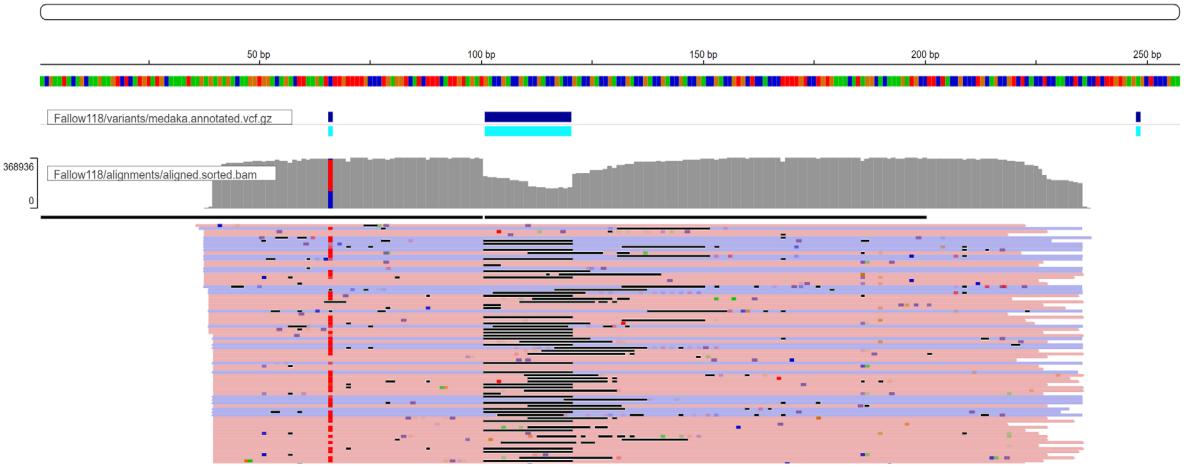


Figure 4.16: *Integrative Genomics Viewer (IGV) output for the alignment and variation analysis of Fallow118, produced my EPI2ME. Dark shading and colour blocks depict the SNVs and the Indels variation.*

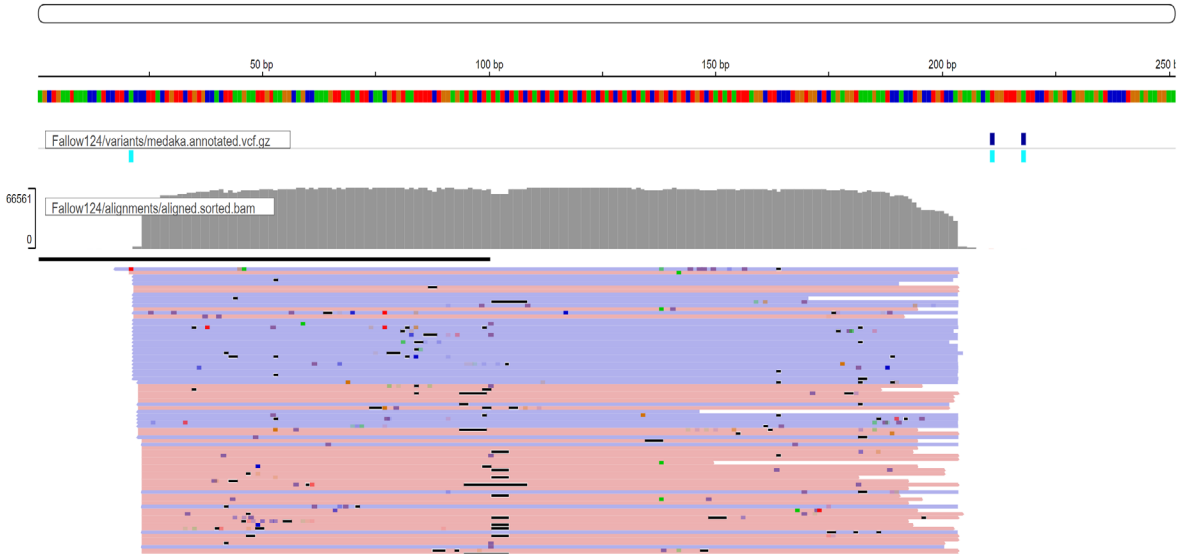


Figure 4.17: *Integrative Genomics Viewer (IGV) output for the alignment and variation analysis of Fallow124, produced my EPI2ME. Dark shading and colour blocks depict the SNVs and the Indels variation.*

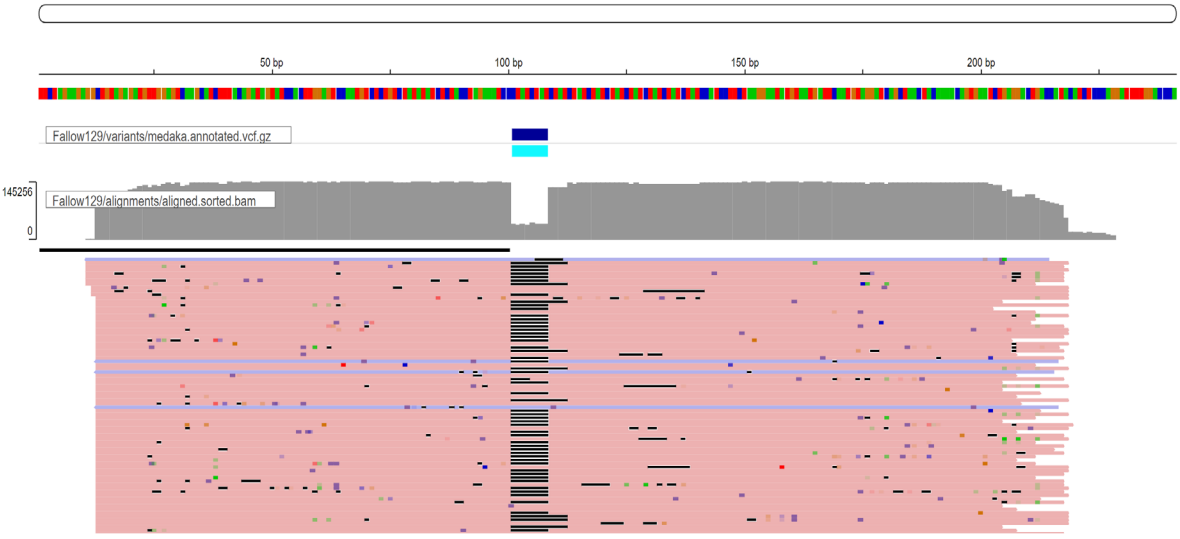
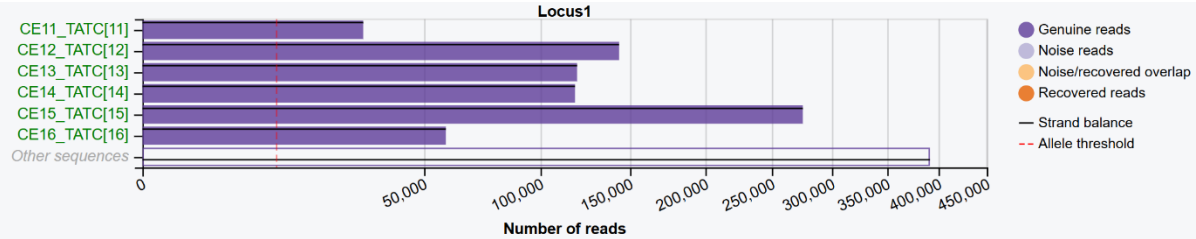


Figure 4.18: *Integrative Genomics Viewer (IGV) output for the alignment and variation analysis of Fallow129, produced my EPI2ME. Dark shading and colour blocks depict the SNVs and the Indels variation.*

4.3.6.2 FDSTools Analysis Results

FDSTools identified 100s of potential alleles, some of which were only represent by only 1 or 2 reads. Most of the under-represented alleles could be a result of sequencing noise generated by the Oxford Nanopore GridION. By incorporating thresholds into FDSTools, the top 1% of sequences can be isolated, see Figures 4.19 – 4.24. Comparison between alleles identified by CE sequencing and NGS sequencing is shown by Table 4.5.

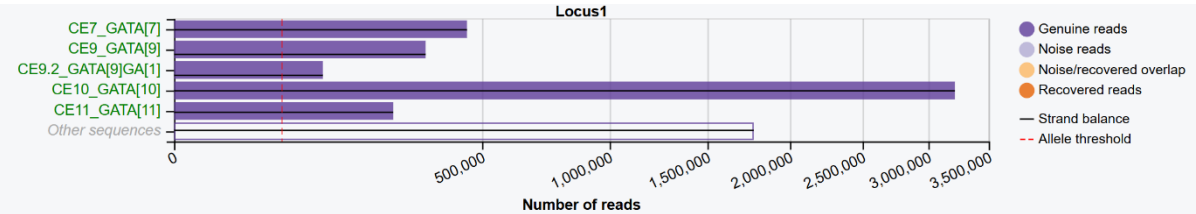
Fallow53



Marker	Allele	Reads	PctOfHighest	PctOfMarker	Correction	Recovery	Noise	PctForward	Notes
Locus1	CE11_TATC[11]	30734	11.17%	2.71%	+0 (+0.00%)	0 (0.00%)	0 (0.00%)	89.84%	
Locus1	CE12_TATC[12]	143284	52.07%	12.62%	+0 (+0.00%)	0 (0.00%)	0 (0.00%)	85.03%	
Locus1	CE13_TATC[13]	119192	43.32%	10.49%	+0 (+0.00%)	0 (0.00%)	0 (0.00%)	85.31%	
Locus1	CE14_TATC[14]	118068	42.91%	10.40%	+0 (+0.00%)	0 (0.00%)	0 (0.00%)	88.09%	
Locus1	CE15_TATC[15]	275152	100.00%	24.23%	+0 (+0.00%)	0 (0.00%)	0 (0.00%)	83.36%	
Locus1	CE16_TATC[16]	58053	21.10%	5.11%	+0 (+0.00%)	0 (0.00%)	0 (0.00%)	86.01%	

Figure 4.19: *FDSTools allele naming results, representing the top 99% of alleles for Fallow53.*

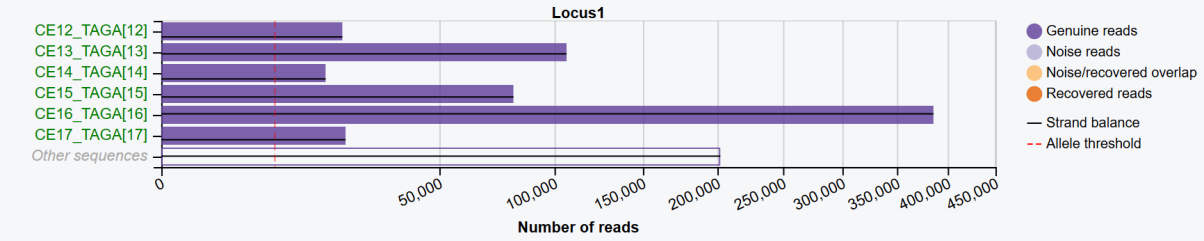
Fallow70



Marker	Allele	Reads	PctOfHighest	PctOfMarker	Correction	Recovery	Noise	PctForward	Notes
Locus1	CE7_GATA[7]	452014	14.06%	7.36%	+0 (+0.00%)	0 (0.00%)	0 (0.00%)	46.82%	
Locus1	CE9_GATA[9]	333278	10.37%	5.43%	+0 (+0.00%)	0 (0.00%)	0 (0.00%)	22.74%	
Locus1	CE9.2_GATA[9]GA[1]	116584	3.63%	1.90%	+0 (+0.00%)	0 (0.00%)	0 (0.00%)	20.07%	
Locus1	CE10_GATA[10]	3214777	100.00%	52.34%	+0 (+0.00%)	0 (0.00%)	0 (0.00%)	47.54%	
Locus1	CE11_GATA[11]	253218	7.88%	4.12%	+0 (+0.00%)	0 (0.00%)	0 (0.00%)	41.60%	

Figure 4.20: *FDSTools allele naming results, representing the top 99% of alleles for Fallow70.*

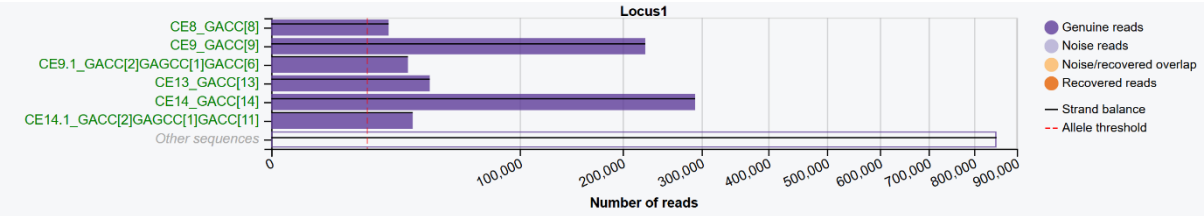
Fallow89



Marker	Allele	Reads	PctOfHighest	PctOfMarker	Correction	Recovery	Noise	PctForward	Notes
Locus1	CE12_TAGA[12]	21202	5.50%	2.54%	+0 (+0.00%)	0 (0.00%)	0 (0.00%)	19.06%	
Locus1	CE13_TAGA[13]	106225	27.55%	12.73%	+0 (+0.00%)	0 (0.00%)	0 (0.00%)	40.81%	
Locus1	CE14_TAGA[14]	17450	4.53%	2.09%	+0 (+0.00%)	0 (0.00%)	0 (0.00%)	17.62%	
Locus1	CE15_TAGA[15]	80180	20.79%	9.61%	+0 (+0.00%)	0 (0.00%)	0 (0.00%)	33.39%	
Locus1	CE16_TAGA[16]	385577	100.00%	46.19%	+0 (+0.00%)	0 (0.00%)	0 (0.00%)	55.51%	
Locus1	CE17_TAGA[17]	21957	5.69%	2.63%	+0 (+0.00%)	0 (0.00%)	0 (0.00%)	27.93%	

Figure 4.21: *FDSTools allele naming results, representing the top 99% of alleles for Fallow89.*

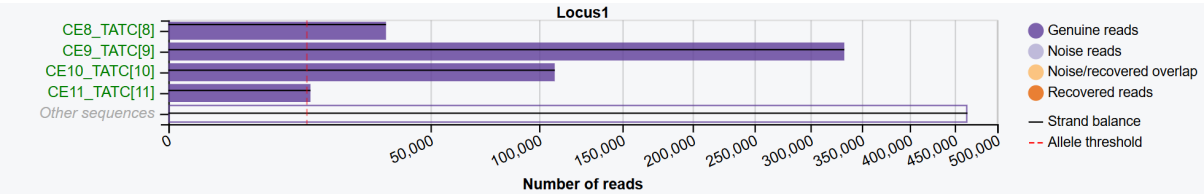
Fallow118



Marker	Allele	Reads	PctOfHighest	PctOfMarker	Correction	Recovery	Noise	PctForward	Notes
Locus1	CE8_GACC[8]	22297	7.66%	1.49%	+0 (+0.00%)	0 (0.00%)	0 (0.00%)	72.02%	
Locus1	CE9_GACC[9]	226673	77.89%	15.16%	+0 (+0.00%)	0 (0.00%)	0 (0.00%)	63.71%	
Locus1	CE9.1_GACC[2]GAGCC[1]GACC[6]	30224	10.39%	2.02%	+0 (+0.00%)	0 (0.00%)	0 (0.00%)	100.00%	
Locus1	CE13_GACC[13]	40641	13.96%	2.72%	+0 (+0.00%)	0 (0.00%)	0 (0.00%)	76.15%	
Locus1	CE14_GACC[14]	291022	100.00%	19.46%	+0 (+0.00%)	0 (0.00%)	0 (0.00%)	70.79%	
Locus1	CE14.1_GACC[2]GAGCC[1]GACC[11]	32421	11.14%	2.17%	+0 (+0.00%)	0 (0.00%)	0 (0.00%)	100.00%	

Figure 4.22: *FDSTools allele naming results, representing the top 99% of alleles for Fallow118.*

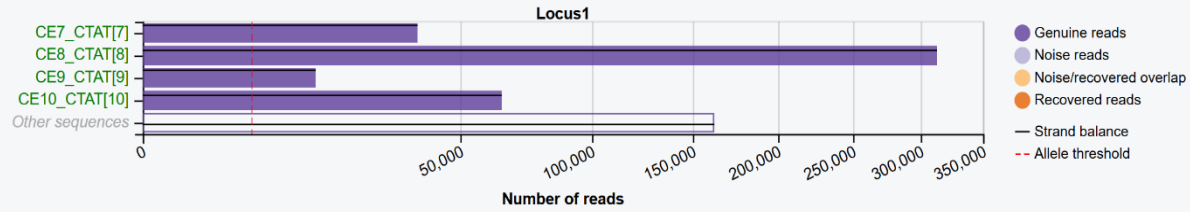
Fallow124



Marker	Allele	Reads	PctOfHighest	PctOfMarker	Correction	Recovery	Noise	PctForward	Notes
Locus1	CE8_TATC[8]	34473	10.37%	3.61%	+0 (+0.00%)	0 (0.00%)	0 (0.00%)	84.42%	
Locus1	CE9_TATC[9]	332512	100.00%	34.82%	+0 (+0.00%)	0 (0.00%)	0 (0.00%)	60.15%	
Locus1	CE10_TATC[10]	108630	32.67%	11.37%	+0 (+0.00%)	0 (0.00%)	0 (0.00%)	61.31%	
Locus1	CE11_TATC[11]	14692	4.42%	1.54%	+0 (+0.00%)	0 (0.00%)	0 (0.00%)	64.18%	

Figure 4.23: *FDSTools allele naming results, representing the top 99% of alleles for Fallow124.*

Fallow129



Marker	Allele	Reads	PctOfHighest	PctOfMarker	Correction	Recovery	Noise	PctForward	Notes
Locus1	CE7_CTAT[7]	37388	11.96%	6.33%	+0 (+0.00%)	0 (0.00%)	0 (0.00%)	89.12%	
Locus1	CE8_CTAT[8]	312549	100.00%	52.94%	+0 (+0.00%)	0 (0.00%)	0 (0.00%)	75.65%	
Locus1	CE9_CTAT[9]	14789	4.73%	2.51%	+0 (+0.00%)	0 (0.00%)	0 (0.00%)	88.85%	
Locus1	CE10_CTAT[10]	63834	20.42%	10.81%	+0 (+0.00%)	0 (0.00%)	0 (0.00%)	75.14%	

Figure 4.24: *FDSTools allele naming results, representing the top 99% of alleles for Fallow129.*

Comparison between alleles identified by CE sequencing and NGS sequencing is shown by Table 4.5.

Table 4.5: *Table comparing the identified Alleles by Sanger and Next Generation Sequencing.*

Loci	Alleles identified by: Sanger Sequencing	Alleles identified by: NGS
Fallow53	11	11
	12	12
	13	13
	14	14
	15	15
	16	16
Fallow70	8.3	7
		9
	11.3	9.2
		10
		11
Fallow89	13	12
		13
		14
		15
		16
		17
Fallow118	8	8
		9
	9	9.1
		13
	14	14
		14.1
Fallow124	10.2	8
		9
	11.2	10
		11
Fallow129	8	7
		8
	10	9
		10

4.4 Discussion

Study 2, set out to design and optimise an STR multiplex for the identification of fallow deer individuals for forensic purposes. Using the genome sequenced during Study 1, primers were designed and tested, optimised and used to identify polymorphic loci suitable for STR multiplexing. However, due to a lack of polymorphic loci this original aim was not possible and the discussions below outline how this study paved the way for the rest of the project.

Firstly, the whole genome was scanned for STR's. A total of 537,466 STR's were found, when searching for a minimum allele of 5 and a hammering threshold of 20, to include some complex repeats (Fungtammasan *et al.*, 2015). There is a high abundance of STR loci found throughout the genome, as shown by Table 4.1. This number compared to red deer is similar but slightly lower overall, despite the genome being larger. Total number of STR loci found in the red deer genome, when searching using the same parameters, was 564,673. A graphical representation of the comparison of total STR's per genome, can be seen in Figure 4.25. As expected, Di-nucleotide markers are the most abundant, with Hexa-nucleotide markers being the least abundant (Subramanian *et al.*, 2003). For fallow in particular, Chromosome 17 and 26 exhibited very low STR loci for all markers. This could be due to these chromosomes having fewer protein coding genes, and as such less non-coding variation associated, but as of yet no one really knows what causes gene distribution within the human genome, let alone a species with new genomic data (Subramanian *et al.*, 2003; Fan & Chu, 2007; Pray, 2008). It is hard to determine the true functions within each chromosome without first delving into each structure and purpose. This provides scope to future projects. Despite there being an expected amount of STR loci across the whole genome, polymorphism was not naturally followed. Using the STR loci detected, primers were developed for STR loci of interest, as explained below.

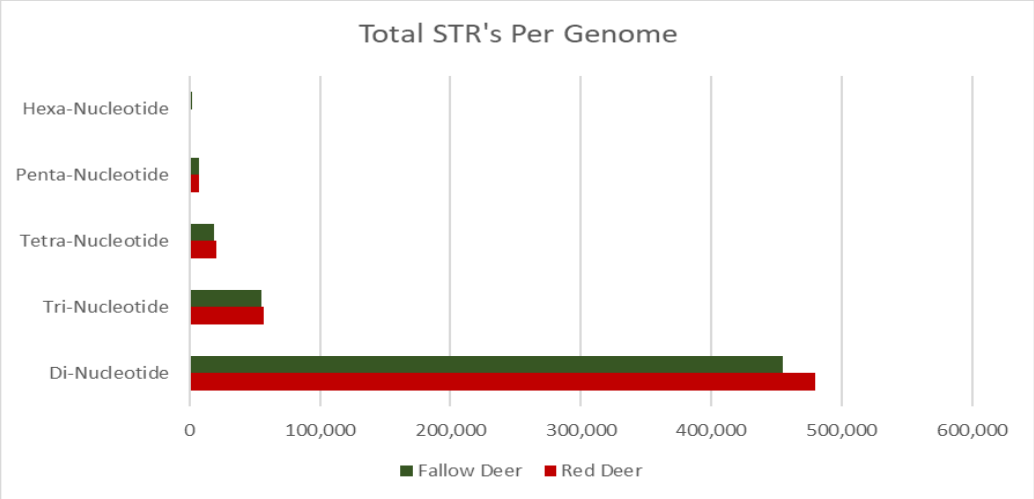


Figure 4.25: *Graph to show comparison of total STR's per genome in red deer and fallow deer.*

Utilising the STR loci discovered from the genome, primers were designed for 132 tetra-nucleotide and 132 penta-nucleotide loci. Tetra-nucleotide and penta-nucleotide loci were chosen as markers of choice due to their characteristics for DNA profiling. Tetra-nucleotide markers in particular are most commonly used for multiplexing due to their four base pair repeat structure allowing for closely spaced heterozygous peaks to be resolved easily via capillary electrophoresis, the same goes for penta-nucleotide repeats (Watherston & Ward, 2023). Further advantages of tetra-nucleotide markers include narrow allelic size ranges which is favourable for multiplexing and protects against drop out as a result of preferential amplification of small alleles (Butler, 2005; Watherston & Ward, 2023). Tetra-nucleotide markers are less effected by stutter formation and are beneficial when amplifying DNA from degraded samples due to the possibility of generating small PCR products, of less than 500bp (Butler, 2005; Watherston & Ward, 2023). Designing new primers from the genome sequence data overcame previous challenges, identified during a previous project, when trying to amplify STR loci using primers developed for closely linked species (Barnard *et al.*, n.p.). Due to the low genetic variation and the history of the fallow deer lineage their DNA is highly conserved and therefore cross-species amplification was not possible (Barnard *et al.*, n.p.). Primers were designed with specific criteria in mind in order to have STR loci ideal for multiplexing. Criteria included searching for STR loci larger than allele 5 repeats, this was to increase the probability of polymorphism at the site, STR loci which had a good GC ratio, loci which were separated in order to not incur any genetic linkage; therefore several loci were selected from every chromosome (Butler, 2005; Goodwin *et al.*, 2011). Careful selection of STR loci lends well to STR multiplexing, reduced stutter formation, amplification of distinguishable alleles and robust loci (Butler, 2005; Goodwin *et al.*, 2011). Primers specifically were designed with a length of 20-24bp, GC ratio of 40-60% and a melting temperature of 54-56˚C. Out of the 92 primers tested during the gradient PCR stage, 42 primers were not taken further for polymorphism testing due to either no amplification, non-specific binding or amplification in only some populations. It was expected to lose some at this stage and perhaps with some further optimisation those with non-specific binding could be rectified, however, for the purpose of this project, the 49 which had worked well was enough to continue with. By running each loci across three samples at 6 different temperatures, each increasing by 2˚C (52-62˚C), enabled the quick and effective screening and optimisation of the primers. Most Loci amplified the best at 56˚C or 62˚C, as expected considering the primer design criteria. 49 primers worked out of the 92 tested, this is a 45% increase in success rate compared to a previous project where, out of 117 primers tested, only 12 amplified. This can be

confidently attributed to the design of species-specific primers, via the whole genome sequence.

During the polymorphism testing, some loci exhibited polymorphism on the gel, such as Fallow4 and Fallow70 (see Figure 4.6 and Figure 4.8, respectively) however, once these loci were run through CE fragment analysis, only 2 or 3 different alleles were identified with the rest of the fragments being all the same. Fallow12, Fallow17 and Fallow64 all presented the possibility of some slight polymorphism on the gel, but once run through CE, only one allele was found, all individuals had the same fragment size. Fallow118 was very interesting on the gel, see Figure 4.9. It presented a high level of polymorphism and some tri-allelic patterns within some individuals. However, once run on CE, no tri-allelic patterns were noted and only 7 different alleles were found out of 26 individuals. The cause for such differences could be linked to a potential lack of sensitivity during CE preventing the detection of more alleles, this is rare, but a potential fault with our machine. CE has been linked to lack sensitivity, if using the incorrect polymer, when dealing with low level polymorphism, which could be the issue here (Krothapalli *et al.*, 2012). POP-4 is best for fragment analysis and is known to be able to differentiate between fragments as little as 1bp difference and due to its reduced viscosity, it is best for smaller fragments (Krothapalli *et al.*, 2012; Connon *et al.*, 2016). However, Pop-6, which is what is fitted in our machine, has high viscosity, which does have better resolution but is better for sequencing, when used for bigger fragment analysis (Connon *et al.*, 2016). Furthermore, our ABI3500 is fitted with a 50cm capillary which is best for larger fragments, Human Identification (HID) uses a 36cm capillary. Shorter capillaries have been associated with better resolution (Connon *et al.*, 2016). As the difference between alleles is very small, running fragment analysis with non-optimal conditions on such samples could explain why some different sized alleles have been missed.

The size range between the different fragments is small, for Fallow118, just 25bp between them, which when analyzing tetra-nucleotide STR's is only 6 repeats. This narrow range would be too narrow for multiplexing; the discriminatory value would be too low to be significant enough for forensic purposes. If the best five loci (Fallow53, Fallow89, Fallow118, Fallow124 and Fallow129), with the most polymorphism were currently taken and included into a multiplex the PD (power of discrimination) would be expected to be around 0.61. However, for a strong forensic multi-plex, such as GlobalFiler, the expected PD value is around 0.9. For other forensic wildlife STR multiplex's such as, SkydancerPlex (van Hoppe *et al.*, 2016) and Deer-plex (Szabolcsi *et al.*, 2014), the PD values are 0.85 and 0.99, respectively. This is much higher than what could ever be achieved in fallow deer. For full details of forensic parameters calculated using STRAF 2.1.5: STR Analysis for Forensics (Gouy & Zieger, 2017) for these 6 loci, see Figure 4.10,

Figure 4.11, Figure 4.12 and Table 4.3. The closer the PD value is to 1, then the greater the power of discrimination and thus the stronger the multiplex. This means that due to a lack of different alleles at a locus, there is a greater chance of a profile generated using the fallow deer hypothetical multiplex matching that of someone else other than the intended target. A PD of 0.61 is not forensically significant, therefore as a multiplex for forensic use this wouldn't be acceptable. This is significantly low when compared to other deer species individual identification kits such as STRoe, for roe deer, which has expected match probabilities ranging from 1.4×10^{-13} to 2.5×10^{-5} and Deer Plex, for red deer, with an expected match probability of 2.6736×10^{-15}. Ultimately, a population genetics test was done to assess Hardy-Weinberg equilibrium, see Figure 4.12. For all loci, the P-value was less than 0.5, with the lowest being 0.001 and the highest being 0.4. Therefore, all loci are not in Hardy-Weinberg equilibrium.

To increase reliability of the HWE results, the number of samples analysed for the 6 loci was increased from 26 to 190. This was to see if any further alleles were present. The results were fairly concordant with the previous analysis, with no drastic increase in the number of alleles detected at each locus, see Table 4.6. Fallow 53 decreased by 1 allele, this could be due to an allele drop out or a previous allele being recorded as true but was actually a result of a PCR artefact, such as stutter, which hasn't been repeated.

Table 4.6: *Comparison between number of alleles detected during first round of fragment analysis with 26 samples, and the second round with 190 samples.*

Locus	Fallow53	Fallow70	Fallow89	Fallow118	Fallow124	Fallow129
No. of Alleles (26 samples)	7	3	3	7	3	3
No. of Alleles (190 samples)	6	5	6	8	5	5

Furthermore, the population genetics and HWE analysis was repeated, see Table 4.7. For all loci, the P-value was less than 0.5, with the lowest being 0.000000358 and the highest being 0.005. For Fallow129 and Fallow53, the P-value was too low to be reported and is indicated by the '<' symbol. Therefore, all loci are not in Hardy-Weinberg equilibrium. Despite the increase in sample pool, P-values are greatly reduced compared to the initial study. These findings further support the statement that fallow deer have low genetic diversity and that a process, such as inbreeding, is having an effect on the allele distribution at each locus (Gouy &

Zieger, 2017). This further strengthens the point of view that an STR multiplex for fallow deer which is forensically significant is unlikely.

Table 4.7: Summary of population genetics analysis for 190 samples calculated using STRAF 2.1.5: STR Analysis for Forensics (Gouy & Zieger, 2017).

Locus	N	Nall	GD	Hobs	Fst	Ht	Fis	P_value
Fallow118	326	10	0.62516	0.31288	0.20496	0.6474	0.3344	0.008105
Fallow124	336	6	0.71	0.32143	0.38932	0.7014	0.2569	0.0000000676
Fallow129	328	5	0.67614	0.29268	0.44034	0.651	0.2129	<
Fallow53	290	11	0.747	0.25517	0.36417	0.7397	0.5798	<
Fallow70	346	5	0.36078	0.14451	0.47482	0.3859	0.4424	0.000000358
Fallow89	336	6	0.5293	0.30357	0.168	0.5483	0.1604	0.005462

No more primers were tested after the initial 92, despite more being developed, because it was a long, time-consuming laborious process which yielded very few useful results. Therefore, it was decided that no further primer screening would take place. Due to the acquisition of an Oxford Nanopore GridION at the university during the project, it was decided this equipment could be utilized to mass screen loci for polymorphism in the hope that the increased sensitivity could uncover genetic variation where it had previously potentially been missed, see study 3. Furthermore, we didn't move on to test any Penta-nucleotide markers because DNA is largely more conserved in the larger STR markers and would be even less likely to reveal polymorphism (Butler, 2005; Goodwin *et al.*, 2011). Additionally, di-nucleotide & tri-nucleotide markers were not tested due to their unfavorable characteristics for a forensic STR multiplex. Smaller markers are harder to distinguish between small allele size ranges and lend themselves to high level of stutter (Butler, 2005; Goodwin *et al.*, 2011). Smaller STR markers also tend to be less discriminatory due to their simpler repeats (Butler, 2005; Goodwin *et al.*, 2011).

However, despite the findings of this study being disappointing, it is also not unexpected or unexplainable. Depending on the CE results, the lack of polymorphism is easily explained. As discussed in section 1.2.1.1, we know that fallow deer have low genetic diversity as proven by previous projects (Pemberton and Smith, 1985; Hartl, Schleger and Slowak, 1986; Randi and Apollonio, 1988). This was due to a genetic bottle neck during the neolithic period and the captivation of this species for many decades (Baker *et al.*, 2017; Baker *et al.*, 2023). More recent studies are now suggesting that unless the genetic diversity in fallow deer is increased, there is a risk of this species facing conservation

vulnerability (Baker *et al.*, 2023). Genetic diversity could be increased by conducting more informed cull seasons and incorporating genetic strategies within hurds (Baker *et al.*, 2023). Further studies within this project are looking to help provide information to make this long-term goal viable, see study 3 & 4.

In conclusion, this study has been informative and insightful, however, has changed the overall aim of the project. The project will now no longer be focusing on the design of an STR multiplex for forensic purposes, instead it looks at the use of Oxford nanopore for rapid polymorphism screening (see section 4.3.6 for results and 4.4.1 for discussion) and Oxford Nanopore sequencing for mitochondrial analysis of the British fallow deer population. The changed focus of the project is just as important as the first and just as useful for the conservation of fallow deer, if not arguably more so. Findings can, and will, help increase the genetic diversity of fallow deer, protecting them from future conservation vulnerability.

4.4.1 Rapid Polymorphism screening via Oxford Nanopore

Results of the STR sequencing via NGS was very successful with over 11 million reads produced across all loci and full coverage across all loci. During Sanger sequencing, Fallow70 and Fallow89, in particular, were particularly hard to sequence, often resulting in no base calls, leaving sequences blank. This was not the case with NGS, all loci were sequenced to at least 100x sequencing depth. This is vastly more data than what could ever be achieved using Sanger sequencing. It is immediately evident that due to the increased sensitivity of NGS that more alleles were identified using NGS compared to Sanger. Using FDSTools to identify and name alleles (Hoogenboom *et al.*, 2021), it became apparent that thresholds which were representative of the data were needed in order to reduce complicating results with sequencing noise. Oxford Nanopore is traditionally used for long read sequencing and is known for its high sensitivity, with short reads, such as the loci analysed in this project, a lot of background sequencing noise was generated. This can cause complications when trying to determine true alleles from stutter alleles and background noise. For example, with loci Fallow53, when no thresholds were implemented on FDSTools, 1000s of potential allele repeats were identified, including complex and hypervariable repeats. When just simple repeats are selected, there are 24 potential alleles for Fallow53, see Table 4.8. However, many of these alleles are exhibited with very low reads and represent less than 1% of the total reads for the locus. Therefore, it is likely that these are not true alleles and are the result of sequencing noise or stutter. With

representative thresholds imputed into FDSTools, the bottom 1% of reads can be removed to leave behind, what is suspected to be the true alleles, see Figure 4.19 for Fallow53, Figure 4.20 for Fallow70, Figure 4.21 for Fallow89, Figure 4.22 for Fallow118, Figure 4.23 for Fallow124 and Figure 4.24 for Fallow129, Table 4.6 summarises these results and compares them to the alleles determined after Sanger sequencing. Once thresholds were in place, only six alleles remained for Fallow53, compared to the previous 24. These results clearly show the importance of representative thresholds in order to remove unwanted data and to retain the true alleles. However, it must be noted that there is the risk of losing extremely rare alleles, due to the crude nature of the threshold cut off.

Table 4.8: *FDSTools allele naming results, representing all simple repeat alleles for Fallow53.*

Marker	Allele	Reads	PctOfHighest	PctOfMarker	Correction	Recovery	Noise	PctForward
Locus1	CE1_TATC[1]	8	0.00%	0.00%	+0 (+0.00%)	0 (0.00%)	0 (0.00%)	50.00%
Locus1	CE2_TATC[2]	13	0.00%	0.00%	+0 (+0.00%)	0 (0.00%)	0 (0.00%)	76.92%
Locus1	CE3_TATC[3]	45	0.02%	0.00%	+0 (+0.00%)	0 (0.00%)	0 (0.00%)	91.11%
Locus1	CE4_TATC[4]	112	0.04%	0.01%	+0 (+0.00%)	0 (0.00%)	0 (0.00%)	90.18%
Locus1	CE5_TATC[5]	225	0.08%	0.02%	+0 (+0.00%)	0 (0.00%)	0 (0.00%)	82.67%
Locus1	CE7_TATC[7]	325	0.12%	0.03%	+0 (+0.00%)	0 (0.00%)	0 (0.00%)	81.23%
Locus1	CE8_TATC[8]	490	0.18%	0.04%	+0 (+0.00%)	0 (0.00%)	0 (0.00%)	79.39%
Locus1	CE9_TATC[9]	1083	0.39%	0.10%	+0 (+0.00%)	0 (0.00%)	0 (0.00%)	79.41%
Locus1	CE10_TATC[10]	4838	1.76%	0.43%	+0 (+0.00%)	0 (0.00%)	0 (0.00%)	86.63%
Locus1	CE11_TATC[11]	30734	11.17%	2.71%	+0 (+0.00%)	0 (0.00%)	0 (0.00%)	89.84%
Locus1	CE12_TATC[12]	143284	52.07%	12.62%	+0 (+0.00%)	0 (0.00%)	0 (0.00%)	85.03%
Locus1	CE13_TATC[13]	119192	43.32%	10.49%	+0 (+0.00%)	0 (0.00%)	0 (0.00%)	85.31%
Locus1	CE14_TATC[14]	118068	42.91%	10.40%	+0 (+0.00%)	0 (0.00%)	0 (0.00%)	88.09%
Locus1	CE15_TATC[15]	275152	100.00%	24.23%	+0 (+0.00%)	0 (0.00%)	0 (0.00%)	83.36%
Locus1	CE16_TATC[16]	58053	21.10%	5.11%	+0 (+0.00%)	0 (0.00%)	0 (0.00%)	86.01%
Locus1	CE17_TATC[17]	3338	1.21%	0.29%	+0 (+0.00%)	0 (0.00%)	0 (0.00%)	95.42%
Locus1	CE18_TATC[18]	175	0.06%	0.02%	+0 (+0.00%)	0 (0.00%)	0 (0.00%)	98.86%
Locus1	CE19_TATC[19]	27	0.01%	0.00%	+0 (+0.00%)	0 (0.00%)	0 (0.00%)	96.30%
Locus1	CE20_TATC[20]	8	0.00%	0.00%	+0 (+0.00%)	0 (0.00%)	0 (0.00%)	87.50%
Locus1	CE21_TATC[21]	9	0.00%	0.00%	+0 (+0.00%)	0 (0.00%)	0 (0.00%)	77.78%
Locus1	CE22_TATC[22]	4	0.00%	0.00%	+0 (+0.00%)	0 (0.00%)	0 (0.00%)	100.00%
Locus1	CE23_TATC[23]	6	0.00%	0.00%	+0 (+0.00%)	0 (0.00%)	0 (0.00%)	83.33%
Locus1	CE24_TATC[24]	2	0.00%	0.00%	+0 (+0.00%)	0 (0.00%)	0 (0.00%)	100.00%
Locus1	CE25_TATC[25]	2	0.00%	0.00%	+0 (+0.00%)	0 (0.00%)	0 (0.00%)	100.00%

Stutter alleles are more complicated to determine when sequencing with NGS. Stutter alleles are one repeat smaller or larger than the true allele, with NGS it is likely that stutter alleles are represented by lower repeat counts compared to the true alleles, but no formal documentation for guidelines has been released on this. Therefore, it is hard to quantify at what point is an allele a true allele or is

it a stutter. Traditionally, stutter alleles are determined by peak height represented by less than 10% of the true allele peak height on CE. Whereas when sequencing we do not get this data, therefore six repeats represented by NGS sequencing may turn out to only five when the locus is analysed by CE. However, when analysing loci by CE, it would take a lot more time to analyse the same number of samples, 223 in this case, than compared to pooling samples and running on one NGS run (Kokotas *et al.*, 2025). Therefore, NGS still poses as a faster way of scanning large pools of samples with a high level of data return for polymorphism as a pre-screening process before moving to CE for further fragment peak analysis (Kokotas *et al.*, 2025).

As a preliminary screening tool NGS is excellent at screening large pools of samples quickly and effectively. Our new proposed method of pooling samples required the concentration of each sample to be the same prior to pooling to remove sequencing bias (Whitford *et al.*, 2022). This can take some time to get right, but once achieved the overall method is very quick. In this study, 223 samples were pooled per loci prior to library preparation and then each locus was barcoded individually. This means that samples could not be analysed individually but instead added to the abundance of sequencing data for that locus without individual identifiers. As, in this case, it was important to ascertain level of polymorphism within the locus, individual sample analysis was not important. This method of sample pooling enables for many samples across a range of geographical locations to be analysed simultaneously, providing highly accurate and robust population data for a single locus for rapid screening. When evaluating a new or alternative methodology, cost is a considerable factor, especially in forensic science when budgets are low. On average, per sample to screen new loci for polymorphism by running sanger sequencing costs, £15 per sample. This is excluding the reagents required to run the ABI3500, the full cost is much higher. Whereas, with the new proposed method of sample pooling, as described in this study, cost per sample is £3. There is no other reagents needed for the running of the GridION sequencer. This figure is based on 200 samples pooled, per barcode. With the new method, the more samples pooled, the cheaper it becomes. NGS is considerably cheaper for a huge amount of data in return.

Sensitivity of the methodology is very important to forensic science. Often in forensic science DNA is highly degraded and low template, which is why methodology needs to be highly sensitive and to only need a small amount of input DNA. Unlike Sanger sequencing, NGS is highly sensitive and works well on degraded and low template DNA samples, with only 200fmol of DNA needed per

sample. Nanopore sequencing is one of the most sensitive sequencing technologies currently available, compared to other NGS platforms, but if too much DNA is loaded, it comes at the disadvantage of high sequencing noise, which can cause identification of false alleles. However, with good analysis software, such as FDSTools, back ground noise can be filtered out with good thresholds.

Historically, when identifying new polymorphic loci, fragment size identified by CE can be compared to the sequence obtained by sanger sequencing in order to link up fragment size and number of repeats, this is to ensure an accurate allele number is represented for the size of fragment. However, during this project it was found to be a difficult process, flanking regions contained many discrepancies and mutations across individuals which made it difficult in ascertaining allele number from fragment size (Carbó-Ramírez *et al.*, 2025; Gettings *et al.*, 2015; Hoogenboom *et al.*, 2021). However, with complete high accuracy NGS sequencing data, it makes determining allele number directly from complete sequences much easier with a much higher degree of accuracy (Carbó-Ramírez *et al.*, 2025; Gettings *et al.*, 2015; Hoogenboom *et al.*, 2021). However, NGS is still relatively new, compared to traditional DNA profiling methods, and therefore nomenclature for the naming of alleles from NGS data is lacking and there are very few guidelines currently available which have been universally approved and validated by the ISFG and forensic network (Carbó-Ramírez *et al.*, 2025; Gettings *et al.*, 2024; Hoogenboom *et al.*, 2021). There is currently a lot of researchers asking for improved and more robust nomenclature for allele naming from NGS data, this is necessary for NGS to be incorporated into the forensic work flow and for it to replace current CE methodologies.

There are several forensic MPS kits for human DNA profiling on the market now, IDseek®, PowerSeq® and ForenSeq®, to name a few. Most Kits for MPS incorporate STR loci previously amplified in traditional profiling kits, as well as mitochondrial markers and SNPs. Using MPS, it is now possible to incorporate STRs, SNPs, InDels, mtDNA, and methylation markers, into a single work flow (de Bruin *et al.*, 2025). Therefore, using NGS it is possible to combine forensic genetic and epigenetic analysis, into one sequencing run (de Bruin *et al.*, 2025; Ferreira *et al.*, 2025). Furthermore, there is an abundance of genomic data available now for the human genome, therefore there is the possibility for whole genome long read sequencing in forensics, especially with Oxford Nanopore technology and the ability to determine DNA methylation in real-time, making this tool highly beneficial in cases involving closely related family, including twins (de Bruin *et al.*, 2025; Ferreira *et al.*, 2025). The possibilities for NGS in human forensics is

vast and will continue to expand. However, for wildlife and other animal species concerned with forensic cases, such as poaching, it is a lot harder to move analysis methodologies forward due to the lack of genomic sequencing data (Cardinali *et al.*, 2023). Due to the abundance of genomic data available for humans, downstream analysis is easier, software tools can have kits preloaded and genomic data read on file for comparative analysis, such as FDSTools – STRNaming (Hoogenboom *et al.*, 2021) and STRaitRazor (King *et al.*, 2021). There is already several projects underway to increase genomic data available for all known species, such as the Wellcome Sanger Institute Tree of Life programme (Wellcome Sanger Institute , 2019), as mentioned in section 1.6. However, It will take many years of continuous work on projects, such as this PhD, to gain the same level of knowledge about species in the same way that there is with human genetic data. It is impossible to be able to have such a high level of knowledge for all species, but it is certainly possible to ascertain the relevant data and build databases for the species of most importance, such as rhino, elephant, tiger, leopard, deer etc. which are most effected by illegal activity, highly endangered and are of interest to forensics. The illegal wildlife trade has been connected with large, organised crime groups and terrorist organisations as a means for these organisations to fund their criminal activities, therefore there is increasing demand for more research into wildlife at the forefront of wildlife crime and illegal trade (INTERPOL, 2023; McNeish, 2014).

4.5 Conclusion

Overall, this study achieved the following:

- The Fallow deer genome was successfully mined for STR variation. A total of 537,466 STRs were found when searching for the minimum allele of 5. Within these, 21,580 tetra and penta nucleotide STRs were located and 264 primers were developed. Out of 132 designed primers for tetranucleotide STRs, 90 were tested and only 9 found to be polymorphic.
- The lack of polymorphic loci discovered can be attributed to the low genetic diversity of the fallow deer species. This lack of genetic diversity which is still observed today implicates the likelihood of a forensically significant STR multiplex for fallow deer individual identification.
- Oxford Nanopore Sequencing via the Native Barcoding kit was used to develop a novel method for rapidly screening samples for polymorphism. This method is quicker and overall more cost effective because more loci can be screened at once against an increased number of samples.

5 Study 3: The complete mitochondrial genome of *Dama dama*, and the assessment of genetic divergence between the species and *Dama mesopotamica*.

5.1 Introduction

Mitochondrial genomes in mammals are described as small, circular in shape and are highly conserved in length and gene content (Simon *et al.*, 1994; Boore, 1999; Sarvani *et al.*, 2018). Mitochondrial genomes are often approximately 16,000bp in length which contains 37 genes that encompasses 22 transfer RNA (tRNA) genes, 13 protein-coding genes (PCGs) two ribosomal RNA (rRNA) genes and an AT rich control region (CR) (Simon *et al.*, 1994; Boore, 1999; Sarvani *et al.*, 2018; Huo *et al.*, 2023). Mitochondrial genomes are highly abundant in each cell with high evolutionary rates, along with low levels of recombination and maternal inheritance (Curole & Kocher, 1999; Sarvani *et al.*, 2018; Huo *et al.*, 2023). Due to these characteristics, the mitogenome offers excellent capabilities to help understand genetic diversity of a species, including evolutionary relationships, population genetics and phylogenetic relationships (Simon *et al.*, 2006; Cameron, 2014; Sarvani *et al.*, 2018; Huo *et al.*, 2023).

Fallow deer (*Dama dama*) are wide spread across the globe; the species is considered wild and domestic whilst simultaneously being invasive and endangered. Several studies have been conducted into the European fallow deer population and comparisons between modern and archaeological samples which has revealed two clades of European *Dama dama* in the Balkans and Anatolia (Baker *et al.*, 2023). These areas represent the regions where fallow deer originated from, including the middle east. Despite their large wide spread population, the species has very low genetic diversity, this is causing concerns to arise that they have conservation vulnerabilities (Baker *et al.*, 2023). It has been suggested that the biocultural history of the species should impact how fallow deer are managed today (Baker *et al.*, 2023). The lack of genetic diversity was caused by a genetic bottle neck which affected the species population size and triggered a spike in inbreeding during the Mesolithic and Neolithic times (Baker *et al.*, 2017). Due to the glacial period, the population size was reduced and squeezed into a small geographical area, they almost became extinct (Baker *et al.*, 2017). Since this time the population of fallow has increased. Studies show that the Romans were responsible for the translocation of fallow from mainland Europe to Britain (Baker *et al.*, 2017; Baker *et al.*, 2023). However, Fallow were considered a symbol of wealth and often kept in captivity on big estates. Fallow deer re-entered the wild

after escaping captivity when private parks and estates were allowed to fall into disrepair during the 15th Century (Chapman and Chapman, 1997; The British Deer Society (BDS), 2016; Baker *et al.*, 2017). This is what lead to the population we see today in Britain, and why their genetic diversity is still so low, despite their large populace.

The sample used to sequence the fallow deer mitochondrial genome was muscle tissue collected from a male fallow deer from the Richmond Royal Park, London. Our sample can be considered a true fallow; there are no reported cases of hybridisation of the species with other species of deer in the UK. Due to their different genus, they are highly unlikely to be able to mate with other species of deer (de Queiroz, 2005), especially those which also inhabit Richmond park (sika, red and roe deer). Interbreeding is common between red, roe and sika however, due to these species sharing the same Genus, *Cervidae* (de Queiroz, 2005; Adavoudi & Pilot, 2021). The fallow herd on Richmond park dates back to 1529 and is home to over 330 fallow deer (The Royal Parks, 2023). Around 200 deer are born each year at Richmond park, so the exact number of fallow is not documented (The Royal Parks, 2023).

The mitogenome presented here is of great importance. Up until now there has been a lack of genomic sequence data available for fallow deer. Furthermore, the full characterisation of the complete mitochondrial genome had not been completed before. Despite fallow deer being classified as least concern (LC) on the ICUN red list, the species has a profound lack of genetic diversity which is causing the species to be at risk of conservation vulnerability as well as the effect of poaching on the population. Mitochondrial sequences are commonly used to resolve species phylogenetic positions and there evolutionary relationships, hence why this report is highly important to better understand the fallow deer genetic diversity, especially within the British population. This report also acts as the ground work for future studies researching the evolutionary genetics of *Dama dama*. Furthermore, phylogenetic studies involving fallow deer would help provide genetic data to support the re-designing of modern management strategies for the species, in particular, conservation breeding programs.

5.2 Method

5.2.1 DNA Extraction and Genome sequencing

The sample used to sequence the fallow deer Mitochondrial genome was muscle tissue collected from a male fallow deer, from the Richmond Royal Park in London UK (Latitude 51.443225, longitude −0.27042112) (Barnard *et al.*, 2023). DNA was extracted using a modified, optimised, method for Gentra Puregene (Barnard *et al.*, 2023). The Mitochondrial genome was sequenced during WGS of the fallow deer genome using PacBio® Long-Read sequencing, on the SMRTBell™ barcoded adapter on the Sequel IIe system (Barnard *et al.*, 2023).

The Genome data is publicly accessible on the NCBI found at the following URL, https://www.ncbi.nlm.nih.gov/datasets/genome/GCA_033118175.1/ (Accession Number: ASM3311817v1)(Barnard *et al.*, 2023). For the assembly of the mitochondrial genome, a previously sequenced fallow deer Mitogenome was downloaded from the NCBI and used (https://www.ncbi.nlm.nih.gov/nucleotide/NC_020700.1).

5.2.2 Mitochondrial genome annotation and analysis

The Mito sequence from this project was aligned with NC_020700.1 from NCBI on Bio-Edit v7.7 (Hall, 1999) and trimmed. The genome was annotated using Mitos WebServer (Bernt *et al.*, 2013) and MitoFish (Iwasaki *et al.*, 2013). A complete mitogenome gene map was generated using MitoAnnotator (Iwasaki *et al.*, 2013) for *Dama dama*. tRNAscan-SE software (Lowe & Eddy, 1997; Chan *et al.*, 2021) and Mitos2 on Galaxy.org (Bernt *et al.*, 2013; Iwasaki *et al.*, 2013; Afgan *et al.* , 2022) was used to predict proteins and transfer RNA (t-RNA) along with their secondary structures. Any t-RNAs which wasn't identified using the previous software, were identified by aligning the mitogenome with the only other previously published *Dama dama* Mito data, downloaded from the NCBI (accession: NC_020700.1).

5.2.3 Phylogenetic analysis

For comparison to other species, mitochondrial genome sequences were downloaded from the NCBI from Artiodactyla of the same family, Cervidae, see Table 5.1. Sequences were aligned with ClustalW (Thompson *et al.*, 1994) in Mega7 (Kumar *et al.*, 2016). For the phylogenetic tree, maximum likelihood (ML) analysis was conducted using MEGA 7 (Kumar *et al.*, 2016) on the alignment of the sequences in table 30. Maximum likelihood was assessed using the Tamura-

Nei model for substitution (Tamura & Nei, 1993; Kumar *et al.*, 2016). Sequences from species used as an out group were African black Rhino (*Diceros bicornis*, NC_012682.1) and the Bengal tiger (*Panthera tigris tigris*, KF892541.1).

Species divergence was calculated for *Dama dama* and sub-species *Dama* mesopotamica on MEGA7 (Kumar *et al.*, 2016), using the RelTime method (Tamura *et al.*, 2018) for estimating divergence times to compute a TimeTree.

Table 5.1: *Details of the 27 Mitochondrial genome sequences used for comparative and phylogenetic analysis, downloaded from the NCBI.*

Species	Common Name	Accession Number	Reference
Diceros bicornis	Black rhinoceros	NC_012682.1	(Willerslev *et al.*, 2009)
Panthera tigris tigris	Bengal tiger	KF892541.1	(Si, 2013)
Cervus elaphus	Red deer	NC_007704.2	(Wada *et al.*, 2006)
Capreolus capreolus	Roe deer	MN485773.1	(Hua, 2019)
Cervus nippon centralis	Honshū Sika deer	AB211429.1	(Wada *et al.*, 2007)
Cervus canadensis	Elk	NC_050863.1	(Kim *et al.*, 2020)
Muntiacus muntjak	Southern red muntjac	NC_004563.1	(Zhang *et al.*, 2004)
Odocoileus virginianus	White-tailed deer	NC_015247.1	(Seabury *et al.*, 2011)
Rangifer tarandus	Reindeer	NC_007703.1	(Wada *et al.*, 2010)
Alces alces	Moose	NC_020677.1	(Hassanin *et al.*, 2012)
Odocoileus hemionus	Mule deer	NC_020729.1	(Hassanin *et al.*, 2012)
Hydropotes inermis	Water deer	NC_011821.1	(Li *et al.*, 2020)
Axis axis	Chital	NC_020680.1	(Hassanin *et al.*, 2012)
Elaphurus davidianus	Père David's deer	NC_018358.1	(Zhang *et al.*, 2017)
Elaphodus cephalophus	Tufted deer	NC_008749.1	(Pang *et al.*, 2008)

Rucervus duvaucelii	Barasingha	NC_020743.1	(Hassanin *et al.*, 2012)
Rusa timorensis	Javan rusa	NC_020745.1	(Hassanin *et al.*, 2012)
Rucervus eldii	Eld's deer	KU133959.1	(Tabasum *et al.*, 2017)
Axis porcinus	Indian hog deer	NC_020681.1	(Hassanin *et al.*, 2012)
Hippocamelus antisensis	Taruca	NC_020711.1	(Hassanin *et al.*, 2012)
Cervus albirostris	Thorold's deer	MF966595.1	(Zhao *et al.*, 2017)
Megaloceros giganteus	Irish elk	MW802558.1	(Rey-Iglesia *et al.*, 2021)
Antilocapra americana	Pronghorn	NC_020679.1	(Hassanin *et al.*, 2012)
Rusa alfredi	Visayan spotted deer	NC_020744.1	(Hassanin *et al.*, 2012)
Mazama americana	Red brocket	NC_020719.1	(Hassanin *et al.*, 2012)
Ozotoceros bezoarticus	Pampas deer	JN632681.2	(Hassanin *et al.*, 2012)
Dama mesopotamica	Persian fallow deer	JN632630.1	(Hassanin *et al.*, 2012)

5.3 Results and Analysis

Full annotation results from MitoFish and Mitos 2 (Bernt *et al.*, 2013; Iwasaki *et al.*, 2013; Afgan *et al.* , 2022), can be found in the supplementary data file, S.7_Mitochondrial Genome Study.

The complete structure of the mitochondrial genome for *Dama dama* is presented in Figure 5.1. The mitochondrial genome for *Dama dama* is 16,322bp in length, consisting of 13 protein coding genes, 2 rRNAs and 22 tRNAs, see Table 5.2. This is similar length as the previously published *Dama dama* Mito data (16, 330bp) (accession: NC_020700.1). The notable difference with the mitogenome for *Dama dama* is that it is the smallest compared to all other mitochondrial genomes analysed in this study.

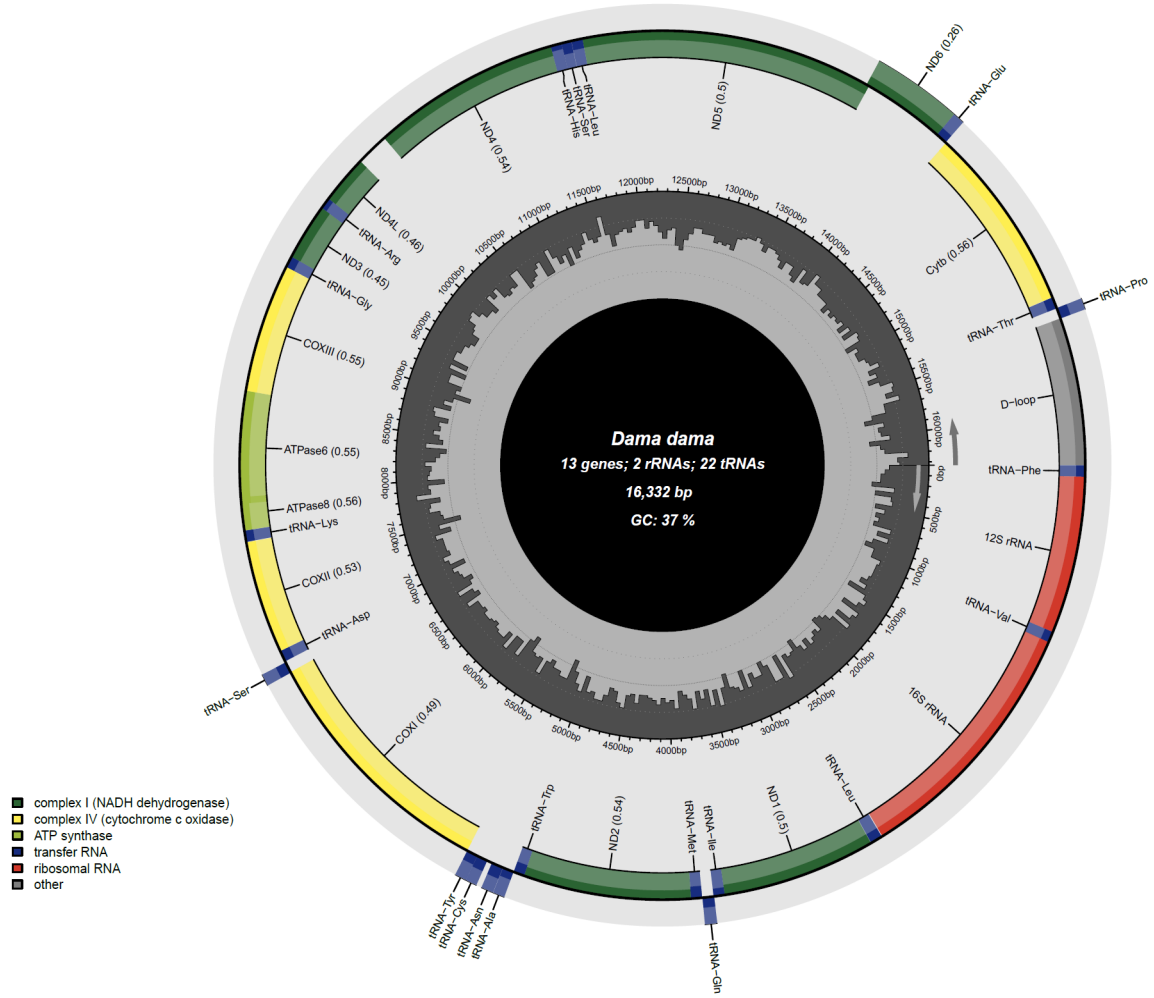


Figure 5.1: *Diagram depicting the circular structure of the mitochondrial genome, including the arrangement of genes for Dama dama. The figure was developed using the MitoAnnotator software (Iwasaki et al., 2013).*

All genes have been mapped to the mitochondrial genome and their arrangement is detailed in Table 5.2. There is marginal discrepancies between the arrangement of the genes in fallow deer and that of other deer species. Between species base pair location of genes differs by only a few bases and there is minor distinctions with the tRNA genes. Overall, the arrangement is almost identical.

Table 5.2: *Genes and their arrangement, for the Fallow deer (Dama dama) mitochondrial genome.*

Name	Site	Length (bp)	Intergenic nucleotide*	Strand	Start Codon	Stop Codon	Anti-Codon
$tRNA^{Phe}$	1 - 69	69	1	H			
12S rRNA	70 - 1024	955	0	H			
$tRNA^{Val}$	1025 - 1091	67	0	H			
16S rRNA	1091 - 2660	1570	-1	H			
$tRNA^{Leu}$	2666 - 2740	75	5	H			UAA
ND1	2743 - 3698	956	2	H	ATG	ATA	
$tRNA^{Ile}$	3699 - 3767	71	0	H			
$tRNA^{Gln}$	3765 - 3836	72	-1	L			
$tRNA^{Met}$	3839 - 3907	69	2	H			
ND2	3908 - 4949	1042	0	H	ATA	AGT	
$tRNA^{Trp}$	4950 - 5017	68	0	H			
$tRNA^{Ala}$	5019 - 5087	69	1	L			
$tRNA^{Asn}$	5089 - 5161	73	1	L			
$tRNA^{Cys}$	5194 - 5260	67	31	L			
$tRNA^{Tyr}$	5261 - 5329	69	0	L			
COX1	5331 - 6875	1545	1	H	ATG	TAA	
$tRNA^{Ser}$	6873 - 6941	69	-1	L			UGA
$tRNA^{Asp}$	6949 - 7016	68	7	H			
COX2	7018 - 7701	683	1	H	ATG	TAA	
$tRNA^{Lys}$	7705 - 7772	68	3	H			
ATP8	7774 - 7974	201	1	H	ATG	TAA	
ATP6	7935 - 8614	678	-37	H	ATG	TAA	
COX3	8615 - 9398	784	0	H	ATG	AT	
$tRNA^{Gly}$	9399 - 9467	69	0	H			
ND3	9468 - 9814	347	0	H	ATA	AAT	
$tRNA^{Arg}$	9815 - 9883	69	0	H			
ND4L	9884 - 10180	297	0	H	ATG	TAA	
ND4	10383 - 11553	1171	2	H	ATG	ATT	
$tRNA^{His}$	11554 - 11622	69	0	H			
$tRNA^{Ser}$	11623 -11683	61	0	L			GCU
$tRNA^{Leu}$	11685 - 11754	70	1	H			UAG
ND5	11755 - 13575	1821	0	H	ATA	TAA	
ND6	13559 - 14086	528	-14	L	TTA	AT	
$tRNA^{Glu}$	14087 - 14155	69	0	L			
CYTB	14160 - 15299	1140	4	H	ATG	AGA	
$tRNA^{Thr}$	15302 - 15371	70	2	H			
$tRNA^{Pro}$	15371 - 15436	66	-1	L			
Control Region	15437 - 16331	895	0	H			

** Overlapping adjacent genes are indicated by negative numbers.*

As part of mapping the genes on the mitochondrial genome, the protein coding regions have been represented in a graphical format, see Figure 5.2, based upon their location and the quality of the base-calling. Protein-coding genes present are identical to that of other *Cervidae* species (Frank *et al.*, 2016), just differing slightly in position, as mentioned previously. The 13 protein-coding genes amounts to 11,193bp in length, which is slightly shorter than other deer species (Wada *et al.*, 2010; Ju *et al.*, 2015; Frank *et al.*, 2016; Tabasum *et al.*, 2017; Sarvani *et al.*, 2018), which is to be expected as the whole genome is the smallest compared to other *Cervidae* species analysed in this study. The longest gene is ND5, at 1,821bp, and the shortest is the ATP8 gene, at 201bp.

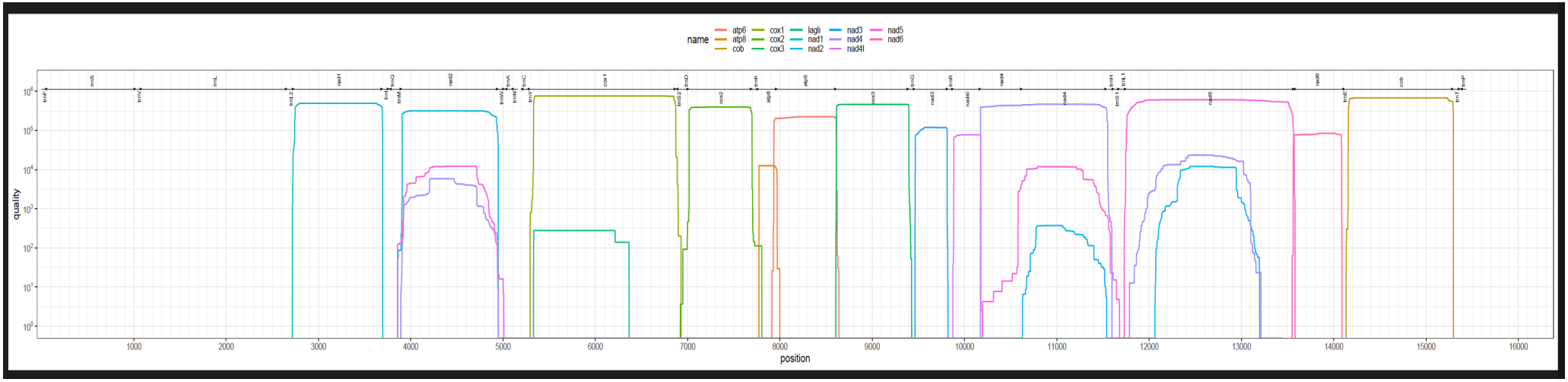
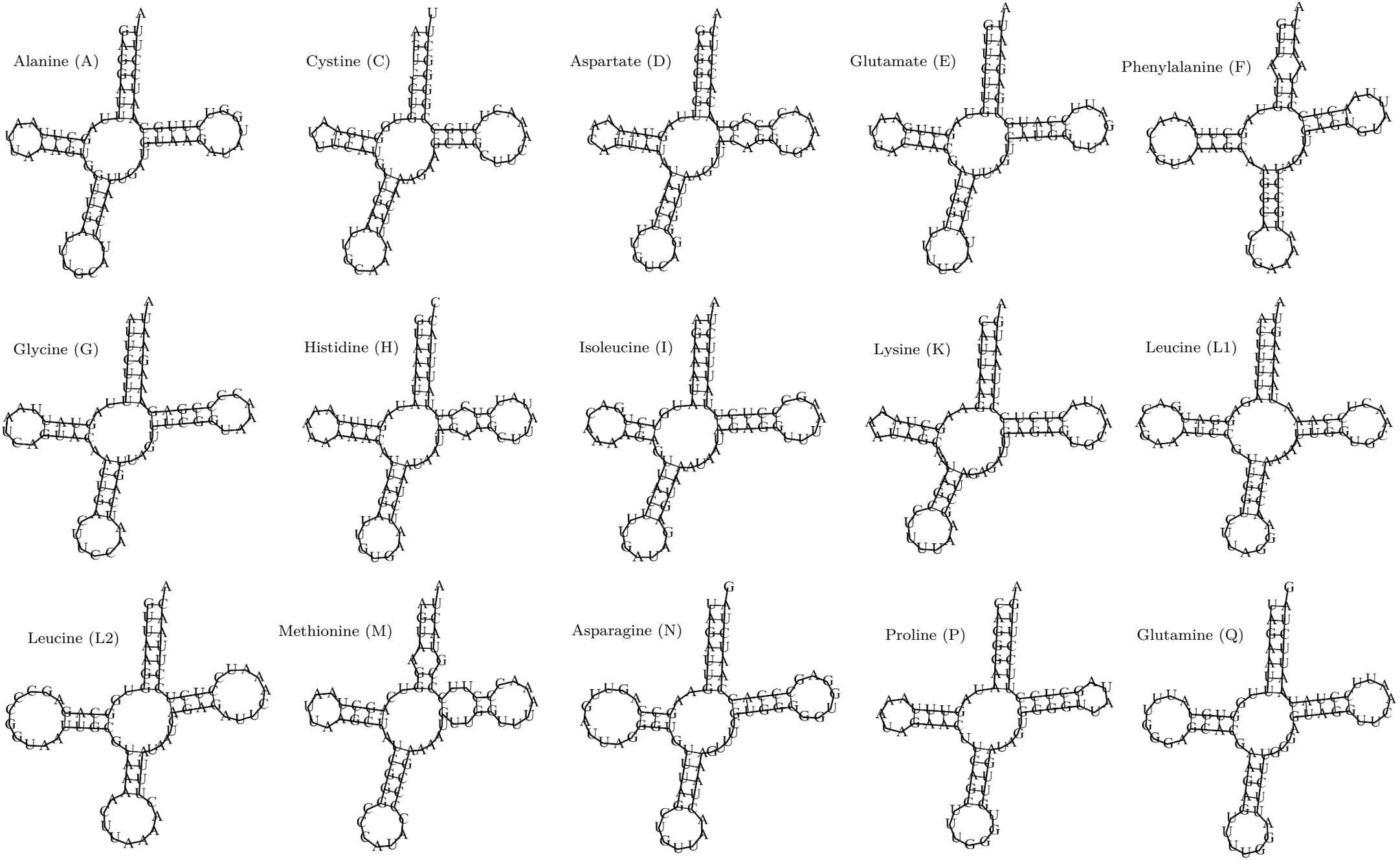


Figure 5.2: *Graphical depiction of the protein coding regions and their location along the mitochondrial genome (Bernt et al., 2013; Iwasaki et al., 2013; The Galaxy Community , 2022).*


Alanine (A)
Cystine (C)
Aspartate (D)
Glutamate (E)
Phenylalanine (F)
Glycine (G)
Histidine (H)
Isoleucine (I)
Lysine (K)
Leucine (L1)
Leucine (L2)
Methionine (M)
Asparagine (N)
Proline (P)
Glutamine (Q)

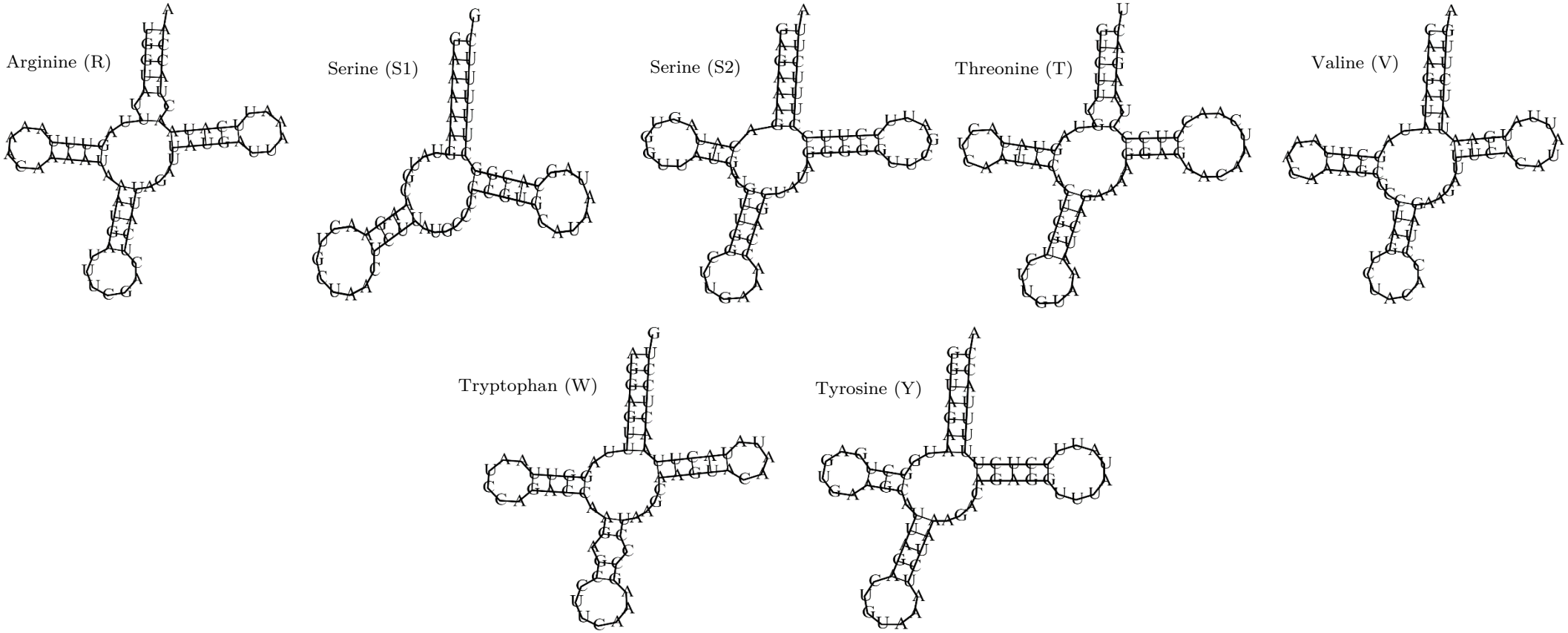


Figure 5.3: *The 22 tRNA genes of the Dama dama mitochondrial genome, and their secondary structures.*

Along with the 13 protein-coding genes, 22 tRNA genes have been identified, with a total length of 1,517bp. The size of the tRNA genes range from 61-75bp, the smallest being $tRNA^{Ser}$ and the largest being $tRNA^{Leu}$. Transfer RNA genes have a specific secondary structure, see Figure 5.3, apart from $tRNA^{Ser}$ (Serine S1) which lacks the dihydrouridine arm, the rest resemble a clover leaf secondary structure (Okimoto & Wolstenholme, 1990; Frank *et al.*, 2016). As with other Mammalia, there are 2 rRNA genes present on the mitochondrial genome, located between $tRNA^{Phe}$ and $tRNA^{Val}$ is the 12S rRNA and located between $tRNA^{Val}$ and $tRNA^{Leu}$ is the 16S rRNA. 12S rRNA in the *Dama dama* mitochondrial genome is 955 bps and the 16S rRNA is 1,570 bps in length, making it the larger of the two rRNA genes.

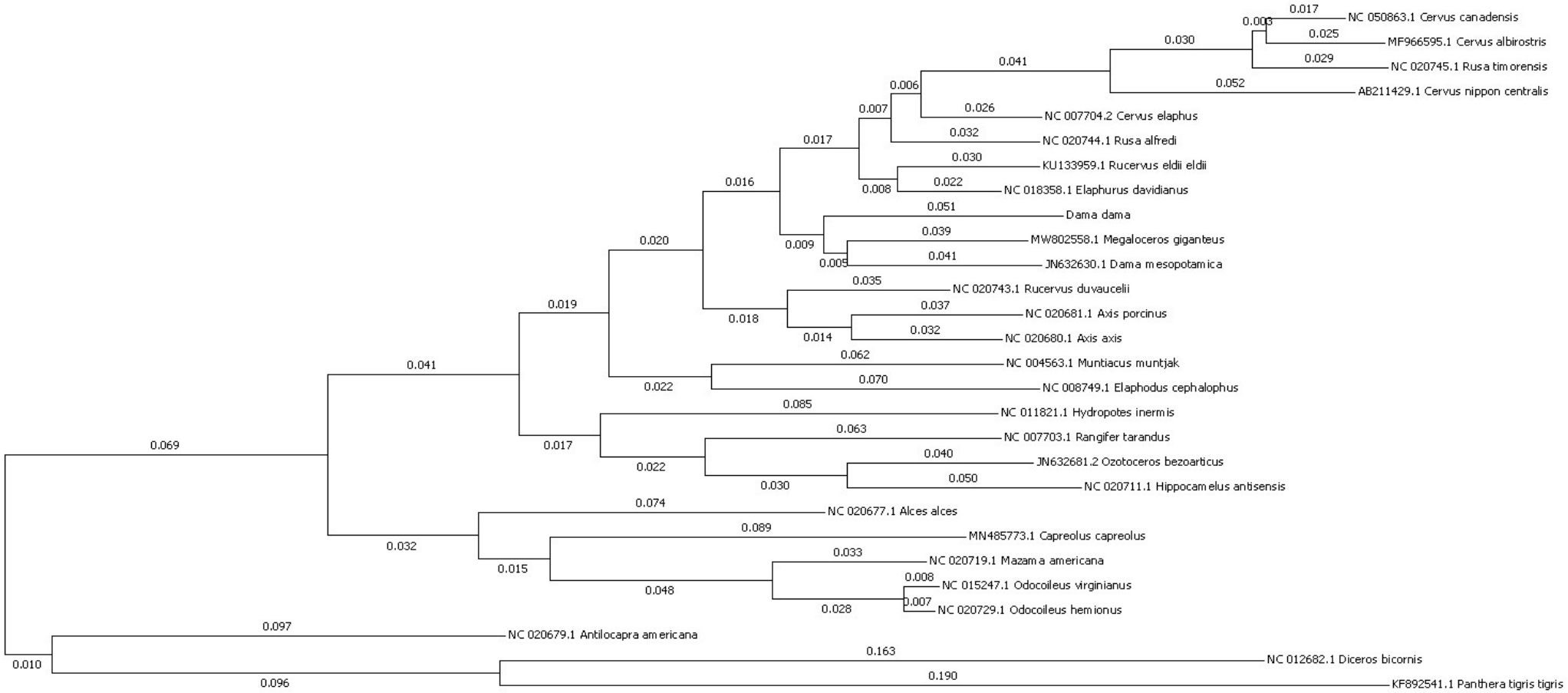


Figure 5.4: *Molecular Phylogenetic analysis by Maximum Likelihood method.*

The evolutionary history was inferred by using the Maximum Likelihood method based on the Tamura-Nei model (Tamura & Nei, 1993). The tree with the highest log likelihood (-73461.31) is shown in Figure 5.4. Initial trees for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach and then selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site (next to the branches). The analysis involved 28 nucleotide sequences. All positions containing gaps and missing data were eliminated. There were a total of 7358 positions in the final dataset. Evolutionary analyses were conducted in MEGA7 (Kumar *et al.*, 2016).

A TimeTree, see Figure 5.5, was inferred using the RelTime method (Tamura *et al.*, 2012) and estimates of branch lengths inferred using the Neighbor-Joining method (Saitou & Nei, 1987). The analysis involved 28 nucleotide sequences. All positions containing gaps and missing data were eliminated. There were a total of 7358 positions in the final dataset. Evolutionary analyses were conducted in MEGA7 (Kumar *et al.*, 2016).

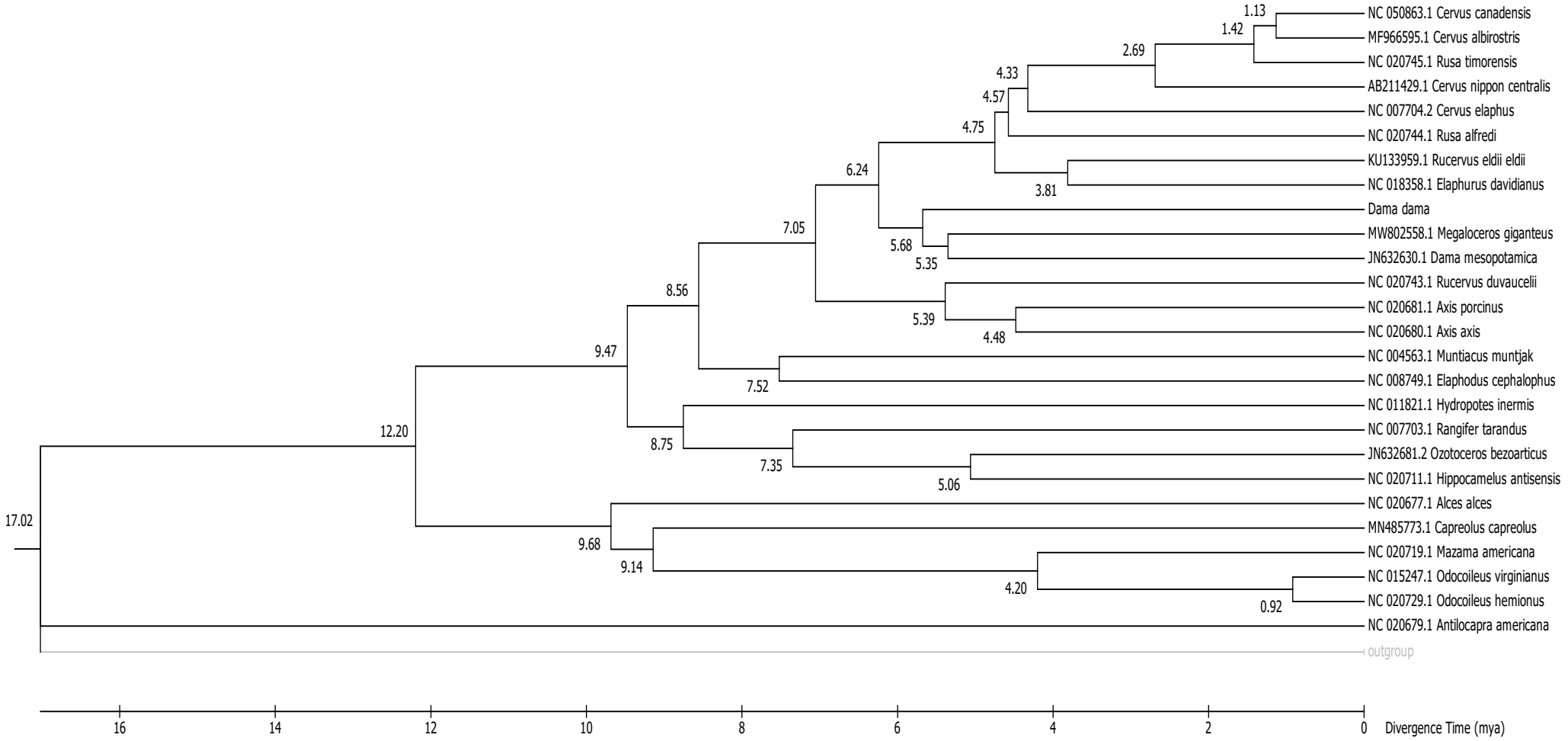


Figure 5.5: *A TimeTree inferred using the RelTime method to determine the evolutionary relationships of taxa (Tamura et al., 2012).*

5.4 Discussion

Over the years the genetic variation of fallow deer has been of increasing interest (Chapman and Chapmen, 1980; Pemberton and Smith, 1985; Hartl, Schleger and Slowak, 1986; Randi and Apollonio, 1988; Arslangündoğdu *et al.*, 2010, Ludwig *et al.*, 2011; Baker *et al.*, 2017; Baker *et al.*, 2024a). With the profound lack of genetic diversity with the species, many researchers have turned to mitochondrial DNA to help answer questions of phylogeny and heredity (Baker *et al.*, 2017; Baker *et al.*, 2024). Most current genetics research into *Dama dama* focusses on a known section of the Control region (Baker *et al.*, 2017; Baker *et al.*, 2024a), however, whole genome data provides the opportunity for a more detailed and robust analysis of phylogenetic relationships. This study provides an insight to the full length mtDNA genome for *Dama dama* and delivers analysis on their phylogenetic status and divergence time in relation to other Cervidae. As this project is the first to do this, the findings from this study help to better understand the species evolutionary status and relationship with other Cervidae to help unpiece why this species has such low genetic variation.

It is well documented that different molecular markers tend to have different evolutionary rates across taxa (Huang, 2008; Frank *et al.*, 2016). Mutation rates are important when reconstructing phylogenies using single genes or short sequences of genes as it may cause errors or inaccuracies when producing phylogenetic trees (Huang, 2008; Frank *et al.*, 2016). Using the complete mitochondrial genome for analysis, the results are more robust and a greater degree of support is provided for the study of evolutionary relationships between individuals and between species. The protein coding genes identified within the *Dama dama* mitochondrial genome, have the similar positions and lengths to that of other Cervidae species (Wada *et al.*, 2010; Ju *et al.*, 2015; Frank *et al.*, 2016; Tabasum *et al.*, 2017; Sarvani *et al.*, 2018). The start and stop codon details can be found in Table 5.2. For the most part, the start codons are either ATG or ATA which are methionine start codons, apart from ND6 which has a TTA start codon. Furthermore, the stop codons are mainly all complete with only two truncated to AT (COX3 and ND6), truncation of stop codons is common in animal mitochondrial genomes (Frank *et al.*, 2016). The secondary structures of the 22 tRNA genes located within the *Dama dama* mitochondrial genome can be found in Figure 5.3. Looking at the structure of tRNAser (Serine S1) closely it can be seen that there is a missing DHU arm, this has also been reported in other deer species (Frank *et al.*, 2016). The sequencing of the mitochondrial genome of the fallow deer using NGS has provided highly reliable genome data which

includes the full control region. This non-coding region is important for in-depth population studies due to the high mutation rate and conserved regions (Bronstein *et al.*, 2018).

A Phylogenetic tree was drawn using the maximum likelihood method, see Figure 5.4. Several species of Cervidae have grouped together within a clade, therefore all those arranged on the same node share a common ancestor (Baum, 2008). For example, all deer from the Cervus genus are all grouped together on the same node of the tree. On the same outer node as the Cervus genus is *Dama dama*. *Dama dama* shares a node with two other species which have formed a separate clade, *Dama mesopotamica* and Megaloceros giganteus. It is documented that *Dama mesopotamica*, a critically endangered species, is a sub-species of *Dama dama* (Werner *et al.*, 2015; Baker *et al.*, 2017). The findings from the phylogenetic tree presented in this study support this claim. With *Dama dama* being on the same node but on its own branch, not part of the *Dama mesopotamica* clade. From this, it can be concluded that there is enough genetic variation between these two mitochondrial genomes that they are no longer the same species and that genetic divergence has occurred. This has most likely been caused by the founder effect, where a population of *Dama dama* became separated and formed a new gene pool, forming *Dama mesopotamica* (Baker *et al.*, 2024b). Furthermore, as *Dama mesopotamica* is now critically endangered and is only found naturally in the middle-eastern areas, Natural selection could also have a part to play. Genetic drift and natural selection often both occur simultaneously to cause genetic divergence and result in the occurrence of a new sub-species. It is also interesting to note the close presence of the *Megaloceros giganteus* species with *Dama dama* and in particular *Dama mesopotamica*. *Megaloceros giganteus* shares a clade with *Dama mesopotamica* therefore are closely related, even more so than to *Dama dama. Megaloceros giganteus* is a much older species, dating back approximately 450,000 years ago, it is known as Irish elk or Giant deer, however it has also been called, the Giant Fallow deer, and with the findings from this study to support, it is likely that the *Megaloceros giganteus* is the common ancestor of *Dama dama* and *Dama mesopotamica* and the origin species of these two sub-species.

Estimation of divergence times was inferred using the RelTime method (Tamura *et al.*, 2012) and the results can be viewed in Figure 5.5. According to the time tree, the origin of the Cervidae sample group used in this study is estimated to be a mean value of 17.02 mya. This indicates that the Cervidae family dates back to later than previous evidence has indicated. The Cervus genus node, diverges

from the Dama family node at 6.24 mya. The largest divergence time is 17.02 mya for Antilocapra americana and the smallest is 0.92 mya for the Odocoileus clade. Another more recent divergence time is 1.13 mya for the Rusa/Cervus clade. This suggests that Rusa timorensis, formerly known as *Cervus timorensis* (Pangau-Adam *et al.*, 2022), is still a closely related species of the *Cervus* genus and may not belong to two separate species as the name suggests. With regards to *Dama dama*, the species divergence time from *Dama mesopotamica* and *Megaloceros giganteus* is 5.68 mya. Whereas the divergence time between *Dama mesopotamica* and *Megaloceros giganteus* is 5.35 mya. This suggests that there has been a divergence from *Dama dama*, which happened prior to the divergence of *Dama mesopotamica* and *Megaloceros giganteus*. This provides support for the distinction between *Dama mesopotamica* and *Dama dama* as sub-species (Baker *et al.*, 2017). The divergence time between *Dama mesopotamica* and *Megaloceros giganteus* is small, compared to that of the Cervidae family, this provides evidence for the theory that they are less of two separate species but more of the same species, much more closely related than some previously documented and providing support for a sister-group relationship with a shared common ancestor (Lister *et al.*, 2005). It suggests that as Megaloceros giganteus is the older of the Dama species, *Dama mesopotamica* has held onto more of the ancestral genetic sequence than *Dama dama*, and that it is *Dama dama* that diverged first and has mutated away from the ancestral lineage.

Overall, the phylogenetic and divergence time analysis has provided support for previously published theories but has also given a new perspective on the divergence of the *Dama dama* species and its sub-species *Dama mesopotamica* which has been a popular topic over the years due to the rapid decline of the *Dama mesopotamica* species. The overall divergence times of the Cervidae family studies in this report are concordant with that reported by Frank. *et al* (2016). However, Frank. *et al* (2016) notes the complexities around divergence estimation for the Cervidae family, as the figures reported here, as within their paper, are higher than previously reported and gathered from fossil data. Fossil records for *Dama dama* date back to the Mesolithic/neolithic times, which was 9,000-4,000 years ago (Breda & Lister, 2013; Baker *et al.*, 2017). The divergence times for *Dama dama* from other Cervidae species supports this, at 6.24 mya. This study has also provided a new perspective on the ancestral origin of the Dama genus, which can be further investigated using the *Dama dama* mitochondrial genome presented in this report.

5.5 Conclusion

Overall, this study achieved the following:

- This is the first study to fully analyse the complete mitochondrial genome of the fallow deer and carry out in depth phylogenetic relationship and species divergence analysis in respect to other Cervidae.
- The mitochondrial genome presented here is 16,332 bp which is comprised of 13 genes, 2 rRNAs and 22 tRNAs.
- The phylogenetic analysis conducted in this study compared the mitochondrial sequences from 25 different Cervidae species. Findings suggest that *Dama dama*, compared to other Cervidae, is most closely related to *Dama mesopotamica* and *Megaloceros giganteus*.
- *Dama dama*, divergence time from *Dama mesopotamica* and *Megaloceros giganteus* is 5.68 mya. Whereas the divergence time between *Dama mesopotamica* and *Megaloceros giganteus* is 5.35 mya.
- This study provides strong support for the distinction between *Dama dama* and *Dama mesopotamica* as a sub-species and a close evolutionary relationship between *Dama mesopotamica* and *Megaloceros giganteus*. Supporting previous reports of a sister-group relationship with a shared common ancestor.

6 Study 4: Haplotype diversity of the British Fallow deer population

6.1 Introduction

Mitochondrial haplotypes are specific regions of mtDNA which group with other mtDNA sequences to show phylogenetic origins of maternal lineages (Lee *et al.*, 2017). Mitochondrial haplotypes can be associated with phenotypes and disease (Lee *et al.*, 2017). Individuals which share the same variants, are said to belong to the same haplogroup. Haplogroups represent the major points of the branches on a phylogenetic tree, because of this, haplotyping can be used to determine the geographical origin of an individual as well as their ancestry (Syndercombe Court, 2021). For haplotype sequencing the displacement loop, also known as the D-loop or control region, of the mtDNA is utilised. This region is 1.1kb in size and is a non-coding region which has an important role in controlling molecule transcription and replication (Sinha *et al.*, 2020). In the human mitochondrial genome the D-loop is divided into three short regions, the hypervariable HVS-I, HVS-II and HVS-III. These regions are aptly named due to the, highly variable, highly muting, nature of the sequence compared to the rest of the genome (Sinha *et al.*, 2020). Despite this however, the mtDNA is highly preserved along the maternal linage, it recombines with similar copies of itself. This makes it highly beneficial for tracking down missing individuals, piecing together remains and making geographical assessments, especially in forensics (Sinha *et al.*, 2020).

In forensics, there are several applications for mtDNA, importantly for wildlife forensics applications include: maternal lineage identification of highly decayed remains, identification of ancient DNA, identification of trophies i.e. ivory, species ID and phylogenetic studies (Sinha *et al.*, 2020). Under the umbrella term phylogenetic studies, mitochondrial haplotyping is utilised for identification of individuals via haplogroup frequency assessment (Syndercombe Court, 2021). How rare a haplotype is, is simply established by counting the number of times the sequence is detected within a sample set/database or population (Syndercombe Court, 2021). As there is currently no haplogroup frequency database for fallow deer, this project will develop one by sequencing approximately 500 individual known D-loop regions and determining each haplotype and grouping them into haplogroups and forming the database. Where polymorphism variation is lacking, mtDNA D-loop haplotyping could provide an excellent second option, providing a similarly beneficial tool for helping fight crime against this species. Understanding the haplotype of a confiscated or seized

piece of evidence suspected to be fallow deer, can link the suspect to the geographical location of the crime and potentially the poached individual itself.

Some haplotype assessments have already been conducted on fallow deer, predominantly those from Germany, Greece and wider Europe (Baker *et al.*, 2017; Ludwig *et al.*, 2011; Masseti *et al.*, 2008). Figure 6.1 shows the sample locations from the Baker *et al* (2017) study, which is the largest study to date on fallow haplotypes involving 364 fallow deer individuals from 10 European countries, including 57 samples from England.

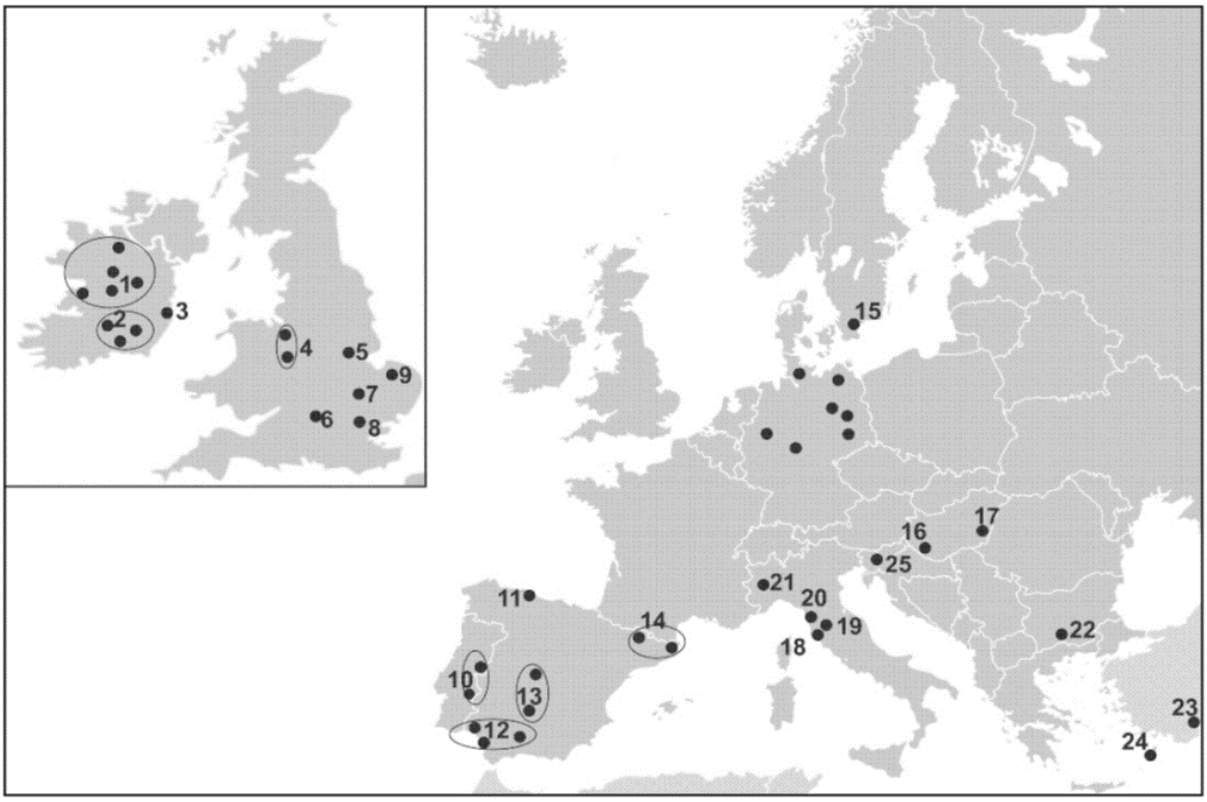


Figure 6.1: *Map depicting sample sites for the 364 fallow deer individuals included in the Baker et al (2017) study.*

The study concluded some interesting findings. The 57 samples tested from England exhibited 15 haplotypes and showed strong links to the population in Ireland, Sweden and Canada, so it is likely that the British fallow deer population founded these three further populations (Baker *et al.*, 2017). Baker *et al* (2017) also discovered that the diversity pattern amongst the 57 samples from England was complex, representing HWE deviations, suggesting a Wahlund effect with a structure of multiple clusters. It must be noted here that 57 samples is a very small sample size, and these finding will be interesting to compare what this study uncovers when looking at a much larger population set. The same primers are being used in this study as was used in the Baker *et al* (2017) study, this shall

provide the capability to conduct direct comparisons. Finally, Baker *et al* (2017) discovered that the Greek Rhodian population is unique, unlike other European populations, with their haplotype sequences presenting with a unique indel, not present in other populations. Therefore, it is suggested this population must be treated with current IUCN recommendations in order to preserve this potential sub-species (Baker *et al.*, 2017; Masseti & Mertzanidou, 2008). A lot of data is already available for European fallow deer haplotypes, and there is keen interest to delve into this further. Such as, to determine founding British fallow deer populations and to gain a clearer understanding of British fallow diversity, all of which can be achieved by this study via assessing ~500 British fallow individuals.

This project has a strong forensic use case, but it also has wider applicability. This project will also benefit the conservation of the fallow deer species in the UK. By undertaking an in-depth study of 352 individuals, a strong haplogroup frequency table can be created. Utilising this, not only for forensics, can provide an insight into the genetic diversity of separate populations/herds within Britain. Thus, providing better knowledge of genetic variation via haplotype density. Carrying out haplotype assessments of individual populations would identify those with decreased variation, occupying more frequent haplotypes. These individuals would be candidates for the cull season. By carrying out cull seasons based upon genetics, an increase in genetic diversity amongst herds and ultimately in the British fallow population as a whole could be achieved. Increasing genetic diversity is crucial in terms of disease protection, should a disease effect a population with low genetic diversity it is more likely to cause that species to become extinct compared to disease in a highly diverse species (Lively, 2010). Furthermore, increased genetic diversity increases the chances of a species being able to adapt to environmental changes and avoids inbreeding (Kardos *et al.*, 2021). Therefore, this project has the potential to have a significant impact.

6.2 Method

6.2.1 Sampling

For the population analysis, sampling was increased. A total of 352 samples have been collected from 15 locations across the UK, see Figure 4.1. For this study, 223 samples from the 352 were selected due to their quality, needed for NGS. These are the same 233 samples used in study 2 NGS analysis.

Samples of blood were collected by rangers within minutes of the deer being culled. The culling season for fallow is between the months of April to August for bucks and November to March for does. Blood is kept in EDTA vacutainer tubes. The blood and samples were kept in the freezer. Once a batch of samples (20 – 30) were collected, they were shipped or collected and brought to the lab for extraction. This process insures the samples are kept fresh and DNA quality is optimal.

6.2.2 Extraction

DNA extraction from whole blood was performed using the DNeasy Blood and Tissue Kit, following the Protocol: Purification of Total DNA from Animal Blood (Spin-Column Protocol) found on page 25 of the DNeasy® Blood & Tissue Handbook (QIAGEN, Germany), using the amendments listed below:

- Step 2: Incubation time increased to 10 mins.
- Step 3: Incubation time increased to 30 mins on a thermomixer.
- Step 8: Incubation time increased to 10 mins.

A negative control was included with every batch of extractions.

Following DNA extraction, DNA was quantified using the NanoDrop One Spectrophotometer (Thermo Fisher Scientific, USA) following the manufactures guidelines. Each DNA extract was run three times on the NanoDrop One Spectrophotometer, and an average was made for an accurate quantification result. See results section 6.3.1.

6.2.3 PCR

Primers used to isolate the full ~1000bp region of the mtDNA D-Loop for haplotype sequencing were developed with the help of previously published universal Control region primers from a former study investigating the European fallow deer population genetic variation (Baker *et al.*, 2017). The primers used by Baker *et al.* 2017 were originally developed by Randi *et al.* (1998). Using these primers, sequences were aligned with our newly sequenced Mitogenome and new, specific *Dama dama* primers, were developed:

CRDamaF 5′- GAGCTATAACCCCACTATCAACAC - 3′

CRDamaR 5′- CTCATCTAGGCATTTTCAGTGCCTT - 3′

Amplification reactions consisted of a 15µl volume PCR containing 1x of Platinum® PCR Multiplex Master Mix (Life Technologies, USA), 3mM of MgCl2 (Fisher Scientific, UK) 0.1uM of both forward and reverse primer and 6ng DNA extract. Nuclease free water was used as a negative PCR control, instead of DNA extract.

Cycle conditions were run as follows: 95 °C for 5 min; 32 cycles at 95 °C for 30 s, 62 °C for 30 s and 68 °C for 2 min; and 68 °C for 10 min.

To assess the success of PCR before sequencing, 2ul of product was run on 2% agarose gels.

6.2.4 Sanger sequencing

Six samples, each from a different geographical location was selected for Sanger sequencing, see Table 6.1. Each sample was amplified as stated in section 6.2.3.

Table 6.1: *Samples used for sanger sequencing of the D-loop.*

Sanger Sequencing		
No.	**Sample Name**	**Population**
1	Ldn46	London
2	Dun10	Dunham Massey
3	Dev25	Devon
4	Ham15	Hampshire
5	Gal4	Galloway
6	Ply5	Plymouth

6.2.4.1 DNA Purification for Cycle Sequencing

PCR products were purified using The DNA Clean & Concentrator-5 PCR purification kit (Zymo Research, USA) following manufacturers guidelines on page 4 of the DNA Clean & Concentrator-5 handbook.

6.2.4.2 Cycle Sequencing

Cycle sequencing was performed using the purified amplified DNA. Cycle sequencing was done using the BigDye™ Terminator v3.1 Cycle Sequencing Kit following the manufacturers protocol with the modification of using half volume reactions.

6.2.4.3 Post Cycle Sequencing DNA Purification

After cycle sequencing the DNA was purified using the in-house protocol for PCR purification.

Per sample reaction, 1µl of 3M NaoAc, 1µl of 20µg/µl Glycogen, 1µl of 100mM EDTA and 30µl of cold 100% proof ethanol was added. This was vortexed briefly and left overnight to incubate. Following the incubation, the reactions were centrifuged at top speed for 30 minutes at 3°C. The supernatant was then removed, leaving the pellet in place and washed with fresh 70% ethanol. This was then centrifuged at top speed for 15 min. This wash step was repeated twice. Next, to air dry the pellet, the tubes were placed with lids open on a PCR machine set at 50°C for 10 min. When ready to run on the ABI 3500, the DNA was suspended in 13µl of HiDi formamide.

6.2.4.4 DNA Sequencing Analysis

Once DNA had been purified after cycle sequencing, the samples were prepared and sequences were analysed via capillary electrophoresis on the Applied Biosystems (ABI) 3500. Pop 6 polymer was used on the following instrument protocol: StdSeq50_POP6_Z. The base calling protocol used was the following: BDTv3.1_PA_Protocol-POP6. The injection time was 8 seconds and run time was 5000s.

6.2.5 Nanopore Sequencing

Taking the primers developed and optimised in section 6.2.3 for the full fallow deer D-loop, 223 samples were amplified. PCR conditions were used as set out in section 6.2.3, using their associated optimised annealing temperature. For Sample details and PCR set up, please see supplementary data file S.8_Nanopore Sequencing Study's_D-Loop & STR Nanopore Studies Set up.xlsx.

Following PCR amplification, each sample, per population, was pooled together. In order to not incur any sequencing bias, all samples were pooled together in equal concentration. Once samples had been pooled together, the sample pools were purified using The DNA Clean & Concentrator-5 PCR purification kit (Zymo Research, USA) following manufacturers guidelines on page 4 of the DNA Clean & Concentrator-5 handbook.

The purified sample pools for each of the 15 populations were then library prepped for Oxford Nanopore Native Barcoding via ligation sequencing. The Native Barcoding Kit 24 V14 kit by Oxford Nanopore was used following manufacturers guidelines. 15 barcodes were used, one for each population pool. Samples were loaded onto the flow cell and run until the flow cell was exhausted.

6.2.6 Bioinformatics

D-loop sequences analysed via sanger sequencing were aligned on Bio-Edit to the reference sequence to identify sites of variation. Haplotypes were matched to those previously identified (Baker *et al.*, 2017) by aligning sequences with 34 identified haplotype sequences available on the NCBI, from the Baker *et al* (2017) previous project (Haplotype 1-34). Mega7 was used to produce a phylogenetic tree depicting evolutionary relationships between the haplotype sequences and the D-loop sequences produced in this study (Kumar *et al.*, 2016).

For D-Loop sequences achieved from the Oxford Nanopore sequencing, basecalling was carried out using MinKNOW on the GridION, based on high accuracy basecalling. Further alignment and amplicon variation analysis was conducted using EPI2ME software. Following the EPI2ME analysis, FDSTools was used to analyse sequences containing sites of variation compared to the reference fallow deer D-Loop sequence (Hoogenboom *et al.*, 2017; Magoč & Salzberg, 2011). Sites of SNPs and InDels were identified by analysing the CLUSTAL W sequence alignment of the sequences generated by FDSTools. The alignment was also performed on Mega 12 software to assess genetic divergence between the potential haplotypes and the reference sequence (Kumar *et al.*, 2024). Estimates of evolutionary divergence between sequences were conducted using the Maximum Composite Likelihood model (Tamura *et al.*, 2004). A phylogenetic tree was built using MEGA7 (Kumar *et al.*, 2024). The evolutionary history was inferred using the Neighbor-Joining method (Saitou & Nei, 1987). The evolutionary distances were computed using the p-distance method (Nei & Kumar, 2000) and are in the units of the number of base differences per site. The pairwise deletion option was applied to all ambiguous positions for each sequence pair resulting in a final data set comprising 1,012 positions.

6.3 Results and Analysis

6.3.1 DNA Quantification

Results of the DNA Quantification for all samples extracted for this study, can be found in the supplementary data file S.2_DNA Quantification. DNA quantification was conducted using NanoDrop™ Spectrometry. All samples were taken forward with the assurance that the negative control had passed.

6.3.2 Sanger Sequencing and Haplotype Analysis

The six individual D-loop sequences were aligned with the full mitochondrial genome, sequenced in study 1 of this thesis. Only eight sites of variation were identified amongst them. See Figure 6.2 for the details of each site of variation.

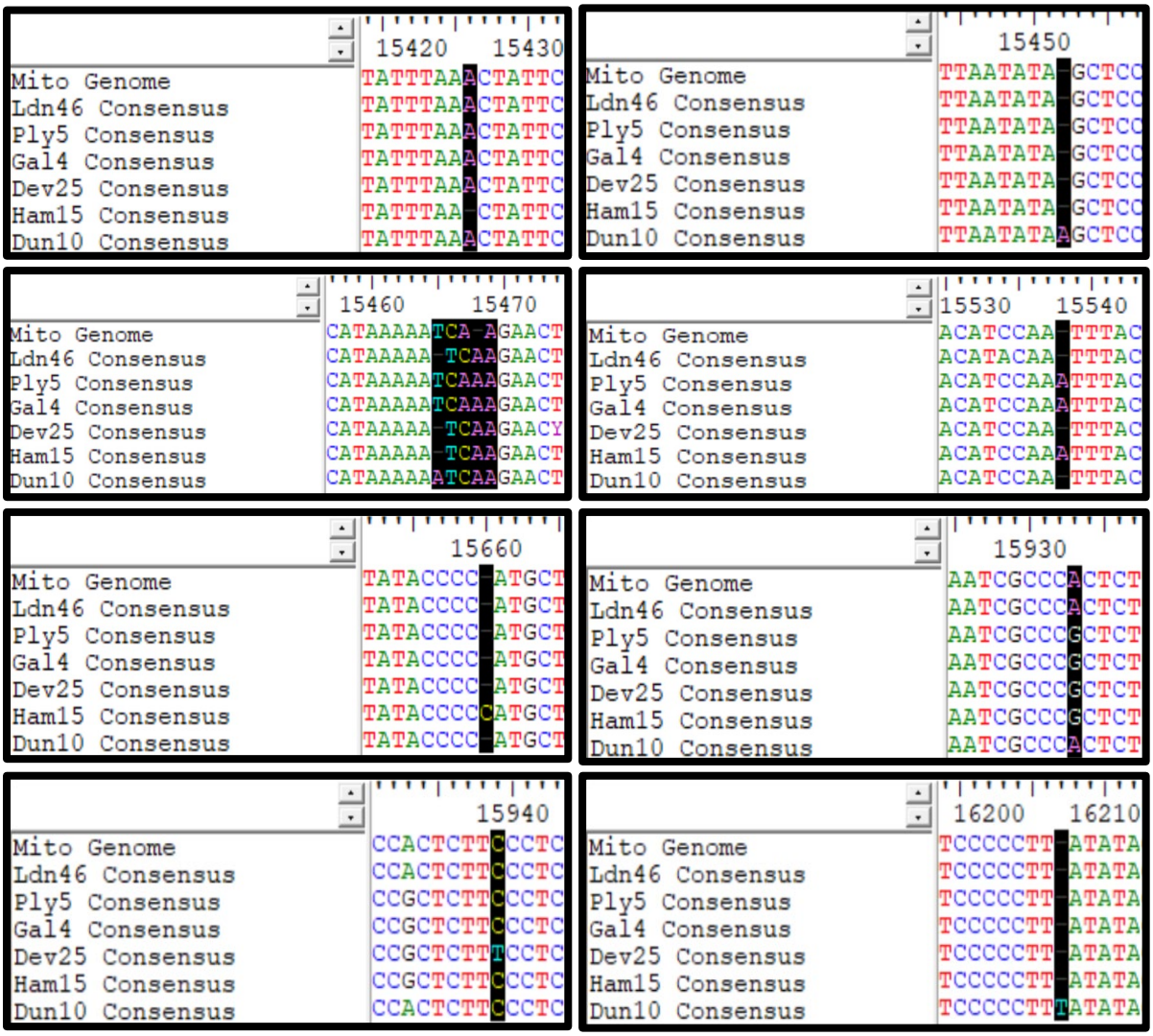


Figure 6.2: *Sections of the genetic code depicting the eight identified sites of variation within the Dama dama D-loop, from Sanger Sequencing.*

Following sequence alignment with the 34 known haplotypes previously identified by Baker *et al.* (2017), it was identified that Ply5, Gal4 and Ham15 are all Haplotype 3. Dun 10 and Ldn46 are both Haplotype 4, and Dev25 is Haplotype 8. Based upon the sequence alignment, a Maximum Likelihood phylogenetic tree was computed, see Figure 6.3. The evolutionary history was inferred by using the Maximum Likelihood method based on the Tamura-Nei model (Tamura & Nei, 1993). The tree with the highest log likelihood (-1354.19) is shown. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 0.7226)). The rate variation model allowed for some sites to be evolutionarily invariable ([+I], 76.93% sites). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 40 nucleotide sequences. Evolutionary analyses were conducted in MEGA7 (Kumar *et al.*, 2016). A larger full resolution copy of this tree is available in the supplementary data file: "S.8_Nanopore Sequencing Study's\S.81_Control Region_ Study\Sanger Sequencing\Haplotype Tree.mtsx"

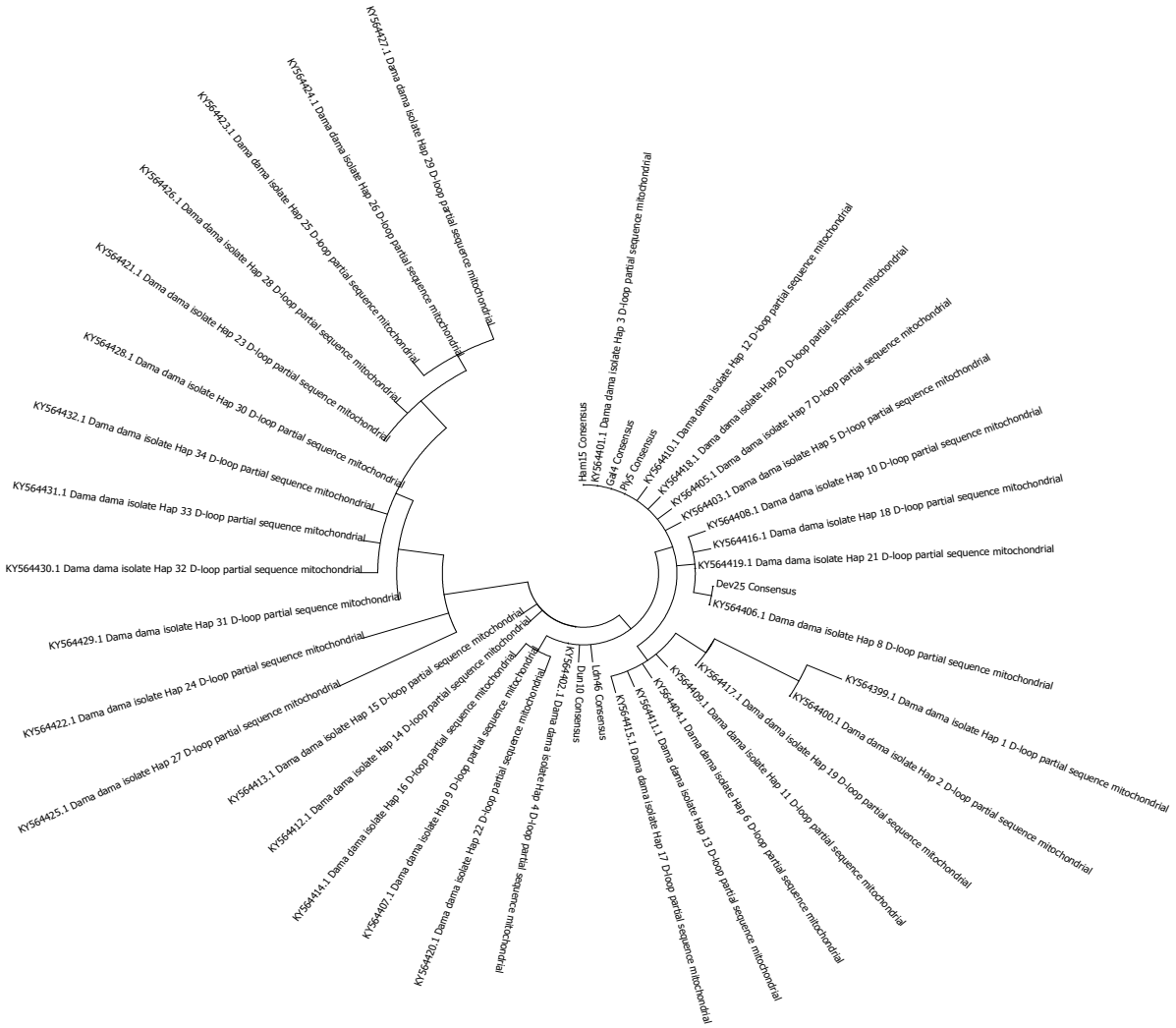


Figure 6.3: *Molecular Phylogenetic analysis by Maximum Likelihood method for the identification of haplotype relationships.*

6.3.3 Nanopore Sequencing

In looking for haplotypes, variation sites including SNPs and Insertion-deletion (InDels) were investigated. EPI2ME calls sites of variation with at least 10 reads whereas FDSTools calls sites with as little as 1 read. Figure 6.4 shows the 6 SNPs identifies by EPI2ME. Table 6.2 details all sites of SNP and InDel variation identified by EPI2ME and FDSTools sequence alignments.

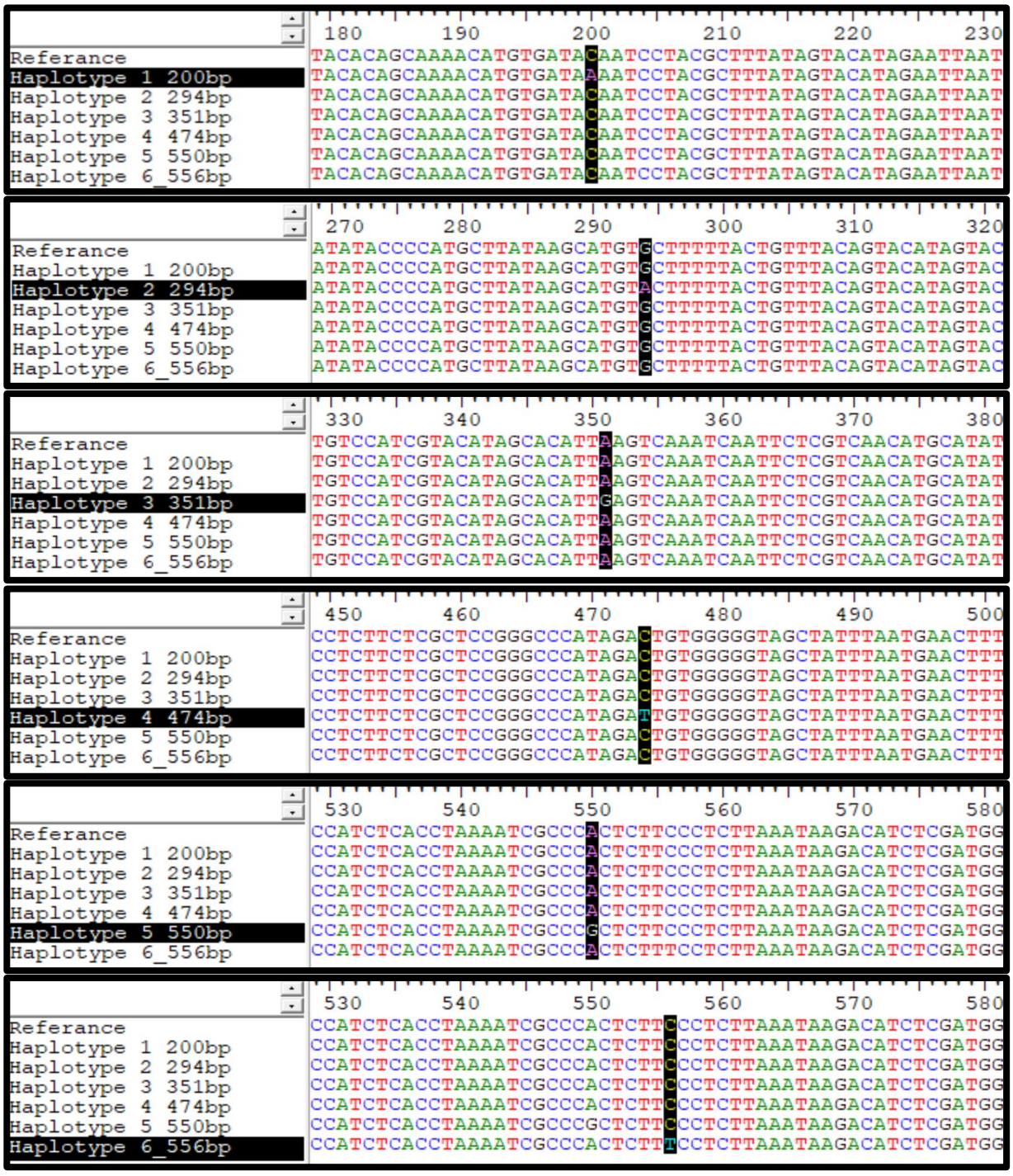


Figure 6.4: Sections of the genetic code depicting six SNP sites of variation within the Dama dama D-loop, identified from Nanopore sequencing on EPI2ME Amplicon workflow.

Table 6.2: *Details of the site variation identified using the EPI2ME Amplicon workflow from nanopore sequencing and FDSTools Sequencing viewer, including their associated geographical locations. Green lettering depicts the sites of variation also picked up by the Baker et al. (2017) study.*

Site	Reference Location	SNP/InDels	Barcode Location	Associated Population(s)
1	200bp	C → A	Barcode 3	Knole Park
2	203bp	T → C	Barcode 5	Devon
3	278bp	Insertion → C	Barcode 2	N-Ireland
4	279bp	T → C	Barcode 5	Devon
5	294bp	G → A	Barcode 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 & 15	Dunham Massey, N-Ireland, Knole Park, London, Devon, Hampshire, Suffolk, Plymouth, New Forest, Basingstoke, Somerset, Galloway, Republic of Ireland, Wiltshire & Perthshire
6	300bp	T → Deletion	Barcode 5	Devon
7	301bp	Insertion → T	Barcode 2, 5, 8 & 13	N-Ireland, Devon, Plymouth & Republic of Ireland
8	317bp	A → C/G	Barcode 5	Devon
9	349bp	T → C	Barcode 15	Perthshire
10	351bp	A → G	Barcode 5	Devon
11	369bp	T → C	Barcode 7	Suffolk
12	442bp	A → Deletion	Barcode 2	N-Ireland
13	443bp	G → A	Barcode 3	Knole Park
14	455bp	Insertion → T	Barcode 5	Devon
15	457bp	C → T	Barcode 3, 10 & 11	Knole Park, Basingstoke & Somerset
16	458bp	G → A	Barcode 2	N-Ireland
17	465bp	G → Deletion	Barcode 2	N-Ireland
18	474bp	C → T	Barcode 9, 10 & 14	New Forest, Basingstoke & Wiltshire
19	475bp	T → C	Barcode 2	N-Ireland
20	523bp	T → Deletion	Barcode 5	Devon
21	525bp	A → G	Barcode 2	N-Ireland
22	550bp	A → G	Barcode 2, 3, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 & 15	N-Ireland, Knole Park, Devon, Hampshire, Suffolk, Plymouth, New Forest, Basingstoke,

				Somerset, Galloway, Republic of Ireland, Wilshire & Perthshire
23	556bp	C → T	Barcode 5	Devon
24	557bp	Insertion → T	Barcode 5	Devon
25	557bp	Insertion → C	Barcode 2, 8, 9, 10, 11 & 13	N-Ireland, Plymouth, New Forest, Basingstoke, Somerset & Republic of Ireland
26	558bp	C → Deletion	Barcode 2, 8 & 9	N-Ireland, Plymouth& New Forest
27	595bp	C → G	Barcode 5	Devon
28	654bp	G → Deletion	Barcode 3, 5, 9 11, 13 & 15	Knole Park, Devon, New Forest, Somerset, Republic of Ireland & Perthshire
29	661bp	Insertion → T	Barcode 2	N-Ireland
30	680bp	T → G	Barcode 13	Republic of Ireland
31	682bp	A → G	Barcode 5	Devon
32	689bp	G → A	Barcode 2	N-Ireland
33	724bp	A → G	Barcode 12	Galloway
34	753bp	G → C	Barcode 5	Devon
35	761bp	T → C	Barcode 3	Knole Park
36	804bp	C → Deletion	Barcode 2, 3, 5, 8, 9, 10, 11, 13 & 15	N-Ireland, Knole Park, Devon, Plymouth, New Forest, Basingstoke, Somerset, Republic of Ireland & Perthshire
37	805bp	Insertion → C	Barcode 3, 5 & 11	Knole Park, Devon & Somerset
38	821bp	C → Deletion	Barcode 2	N-Ireland
39	821bp	C → T	Barcode 13	Republic of Ireland
40	821bp	Insertion → C	Barcode 5	Devon
41	822bp	T → Deletion	Barcode 13	Republic of Ireland
42	853bp	Deletion → TT	Barcode 5 & 11	Devon & Somerset
43	855bp	Insertion → T	Barcode 2	N-Ireland
44	870bp	T → Deletion	Barcode 2	N-Ireland
45	938bp	T → C	Barcode 2	N-Ireland
46	976bp	A → G	Barcode 10	Basingstoke
47	999bp	A → Deletion	Barcode 1	Dunham Massey

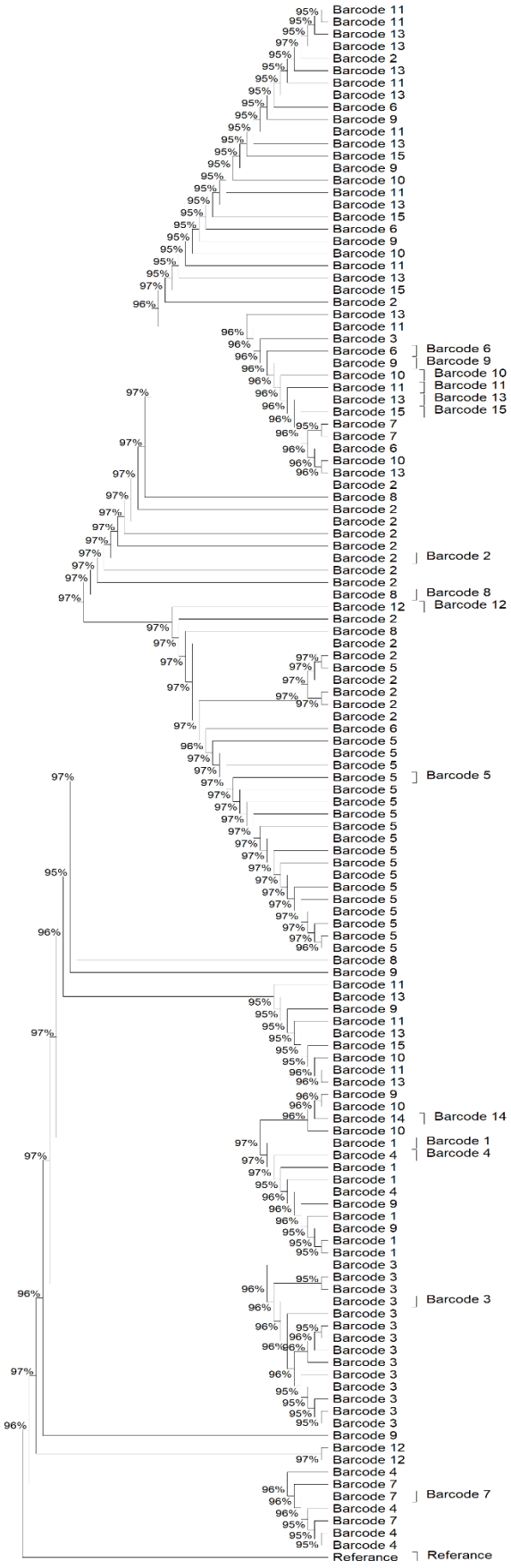


Figure 6.5: *Phylogenetic tree displaying the evolutionary history inferred using the Neighbor-Joining method (Saitou & Nei, 1987).*

The evolutionary history was inferred using the Neighbor-Joining method, see Figure 6.5 (Saitou & Nei, 1987). The optimal tree with the sum of branch length = 0.028 is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (100 replicates) are shown (Felsenstein, 1985). The evolutionary distances were computed using the p-distance method (Nei & Kumar, 2000) and are in the units of the number of base differences per site. The proportion of sites where at least 1 unambiguous base is present in at least 1 sequence for each descendent clade is shown next to each internal node in the tree. The analytical procedure encompassed 128 nucleotide sequences. The pairwise deletion option was applied to all ambiguous positions for each sequence pair resulting in a final data set comprising 1,012 positions. Evolutionary analyses were conducted in MEGA12 (Kumar *et al.*, 2024) utilizing up to 6 parallel computing threads. Due to its size, the tree can be better visualised in the supplementary data file "D:\S.8_Nanopore Sequencing Study's\S.81_Control Region_ Study\Nanopore Sequencing\Mega12 & Alignments\FDStools Haplotypes_By Barcode_Alignment_Grouped_Tree.mtsx"

For the results of the estimates of evolutionary divergence between sequences, performed on MEGA12, see supplementary data file "S.8_Nanopore Sequencing Study's\S.81_Control Region_ Study\Nanopore Sequencing\FDStools\Assessment of genetic divergence.xls".

The abundance of variation at each sample site from Table 6.2 was translated into pie charts and positioned at each geographical location on a map in Figure 6.6.

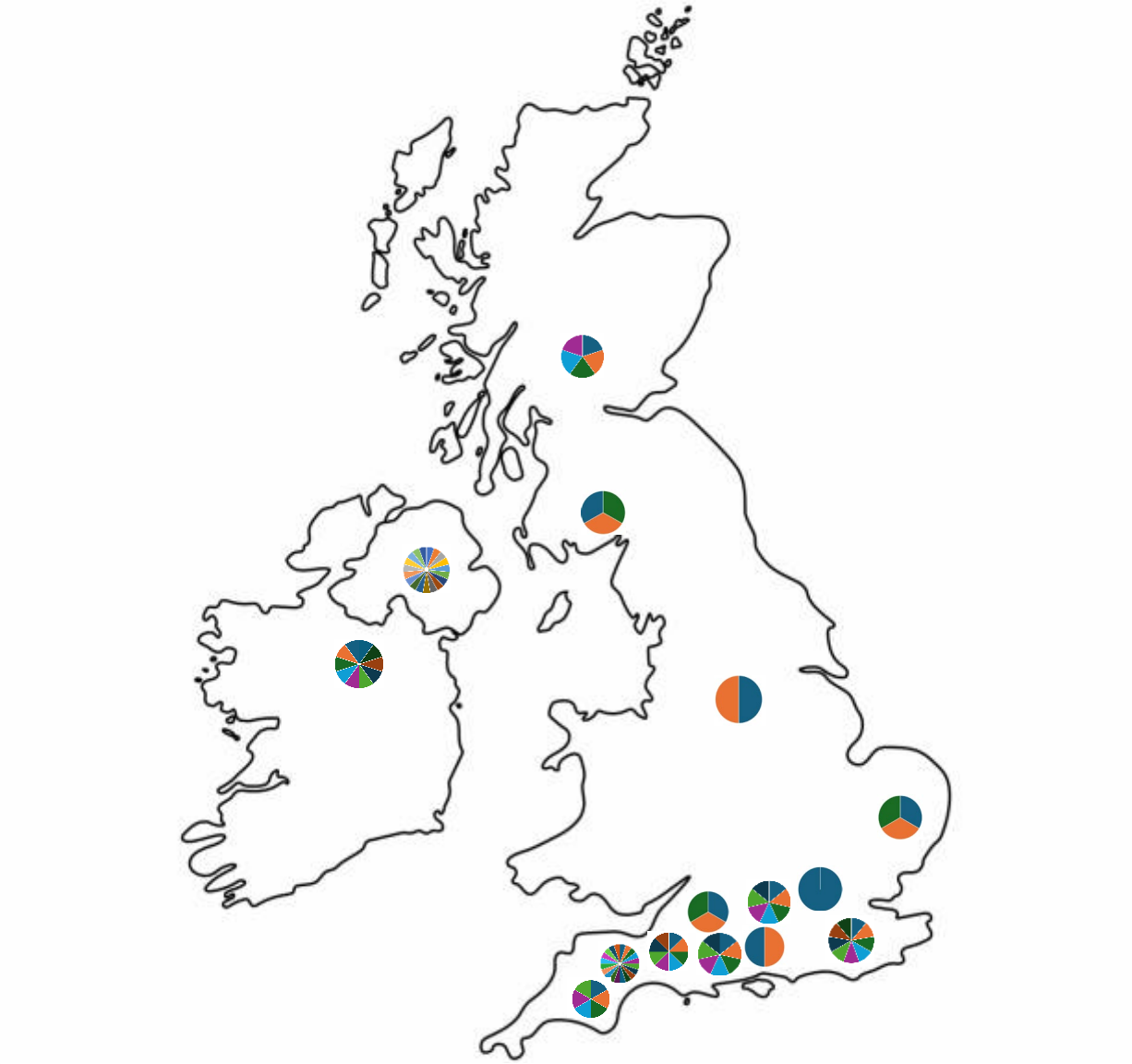


Figure 6.6: *Map of UK and Ireland to show abundance of variation at each sample site, inferred from table 5.2. Each colour represents how many different sites of variation were exhibited at each location.*

6.4 Discussion

This study was a full deep dive into the mitochondrial genetic diversity of the fallow deer within the complete D-Loop sequence. As mentioned throughout this work, fallow deer are known to have low genetic diversity, which is why it wasn't surprising when only 6 sites of variation were identified as a result of the Sanger Sequencing in section 6.3.2. However, this was latterly increased to 47 sites of variation as a result of the Nanopore Sequencing, as displayed in section 6.3.3. This increase in variation detected is likely due to the increased sequencing depth achieved with Oxford Nanopore Sequencing. Variation calling via Nanopore Sequencing was achieved at a 300x sequencing depth. This compared to the Sanger Sequencing depth of 1x, where only 1 read is produced per DNA fragment, the ability to be able to detect low-frequency variation is vastly reduced. Therefore, the application of Next Generation Sequencing, especially with samples from species where diversity is low and variation is at low-frequency, it is crucial in order to detect what may have been missed before.

Out of the 47 sites of variation found as a result of the NGS, site 5 and 22 were the most common, with site 5 exhibited in all 15 of the populations and haplotype 22 was found in 13 out of the 15 populations analysed (Table 6.2). Both these sites of variation were also picked up the Sanger Sequencing. It is likely that these sites are high-frequency and more easily identified. Variation abundance at each geographical location sampled in this study can be visualised by the map in Figure 6.6. Our study differs to ones previously published, such as Baker *et al.* (2017), as other studies looked as the British population of fallow deer as a whole, and compared to other European countries, rather than seeking out differences within the British population itself. However, comparisons can still be drawn within the sites of variation observed. Table 6.2 shows all 47 sites of variation observed within the 15 populations, the sites coloured green were also observed within the Baker *et al.* (2017) published haplotypes. Interestingly, the site of variation at 203 bp and 443bp observed solely in Devon and Knole park, respectively, was also observed in haplotypes 23-34 in the Baker *et al.* (2017) study which corresponded to Germany, Rhodes, Italy, Portugal and Spain. There was no link to the UK at these variation sites. Whereas site 10, found in the Devon population, was only observed in haplotype 8 which corresponded solely to England in the Baker *et al* (2017) study. All other sites coloured green in Table 6.2 were linked to England as well as other European countries. Sites of variation linked to one particular geographical location only could be indicative of population signature haplotypes, but without individual sequence data for each barcode, it currently can only be speculated.

In order to assess level of genetic diversity between populations and the reference sequence, evolutionary divergence was estimated via MEGA12 using the Maximum Composite Likelihood model, see supplementary data file "S.8_Nanopore Sequencing Study's\S.81_Control Region_ Study\Nanopore Sequencing\FDStools\Assessment of genetic divergence.xls". Due to the large amount of sequence data compared, the data is vast, but some interesting findings can be drawn. By highlighting all cells containing "0", barcodes which have no divergence from each other can be visualised. Barcode 4, London, is the closest to the reference, with zero divergence in most cases from the reference. This is likely because the original reference genome was sequenced from a sample from the London population. Furthermore, barcode 2 (Northern Ireland) stands out as an interesting population with a lot of conservation with barcodes 6 (Hampshire), 8 (Plymouth), 9 (New Forest), 10 (Basingstoke), 11 (Somerset), 12 (Galloway), 13 (Republic of Ireland) and 15 (Perthshire). Barcode 13 (Republic of Ireland) also shows similar traits. This could be due to the migration of fallow deer. As we know, the fallow deer species was initially introduced to England by the Romans (Baker *et al.*, 2024). However, the Romans never invaded Ireland, and for that reason fallow deer never made it to Ireland until much later, when the Norman's invaded in the 13th century (Beglane *et al.*, 2018). The population which the Angolo-Normans introduced to Ireland came from Britain, predominantly Southern and South-Western Britain (Beglane *et al.*, 2018). Due to the sampling for this study predominantly also being from the South of England, which are more likely to be a part of the founding populations, the lack of genetic divergence supports this claim.

The phylogenetic tree in Figure 6.5 displays the evolutionary history inferred using the Neighbor-Joining method. There are some patterns emerging here with barcodes, 2 (Northern Ireland), 3 (Knole Park) and 5 (Devon) individually clustered together within their populations/group. The other populations are spread out amongst each other. barcodes 2, 3 and 5 have the most variation within the population with 19, 9 and 20 sites of variation within each, respectively. Populations with the variation and that exhibit all the haplotypes clustered together are more genetically distinct, other haplotypes cluster across a number of different barcodes which may cause implications for breeding programmes to maximise mitotype diversity. By analysing the phylogenetic tree, it can also be deduced that Northern Ireland and Devon are amongst the most common with the reference sequence with the highest percentage coverage match, which suggests that they diverged away slightly later than other populations and have more in common with the fallow deer population thought to be brought

across to the UK in the 11th century by the Normans from Europe after the original Roman introduced population had become extinct (Baker *et al.*, 2024). This also explains the strong genetic links to European fallow, especially those from Eastern Europe.

Benefits of NGS over Sanger Sequencing has been demonstrated in this study. The level of variation observed in the NGS sequences compared to Sanger is a 783% increase. Bearing in mind this is with the D-Loop, just a 1000bp region of the full 16,322bp mitochondrial genome. Most previous studies look at a smaller ~500bp region of the D-Loop. NGS makes it much easier to sequence and analyse longer regions, especially with Nanopore Sequencing which specialises in long reads. Furthermore, due to the nature of the library preparation for NGS, it is much quicker to sequence a higher number of samples. With our methodology in particular, the novel approach to pool samples prior to barcoding by population meant a huge sample pool could be assessed as a mixture to reveal all possible variation across each population. This type of sample pooling is cheaper and faster than sequencing each samples individually. It provided an oversight into the level of variation in each population. In the case of this project, only 15 barcoded samples had to be ran, rather than running the total 223 samples individually which co-existed within the sample pool. The downside is that individual specific data is lost and sites of variation can only be attributed to population rather than individuals. In this instance it is not possible to reveal exact haplotype sequences, but due to the success of the initial screening, there is an application for future work here to investigate the same D-Loop region individually by sequencing each of the 223 samples via NGS. The initial screening allowed the assessment of this use case, rather than deep diving into individual sequencing first which would have cost almost 15x more, this in terms of nanopore costs would have been £1000's of pounds extra. Future work involving the individual sequencing of the 223 samples from the 15 locations would permit the assessment of individuals to reveal the specific haplotype sequences and to compare inter-population differentiation.

Overall, Long-read sequencing (LRS) has a growing demand in forensic science and particularly wildlife forensics where investigations still rely largely on species identification and geographical assessment. Oxford-nanopore revolutionised the LRS scene and now dominates with Ultra-LRS with the longest recorded sequenced complete DNA fragment is 4.2Mb (Jain, 2021; Ferreira *et al.*, 2025). This is highly beneficial for haplotype assessment as the longer the sequence of DNA fragments the more reliable the assembly can be and thus the more variation

is likely to be preserved (Ferreira *et al.*, 2025). With longer complete sequences there is increased potential for the detection of multiple variants which enables the ability to obtain haplotype phasing information (Ferreira *et al.*, 2025). This allows the merging of epigenetic modifications and variants which occur close together along the DNA sequence, to be assigned to the same haplotype (Ferreira *et al.*, 2025). This has been previously very difficult to do due to a PCR amplification step, this can lead to amplification bias and the presence of PCR artifacts (Ferreira *et al.*, 2025). LRS via Oxford Nanopore does not require PCR, therefore increasing reliability and robustness (Ferreira *et al.*, 2025). Having Phasing information along with epigenetic details in relation to haplotypes increases the power of mitochondrial evidence when building a case for court, especially concerning closely related individuals, such as fallow deer. For an overview of the main applications and benefits of LRS in forensic genetics, see Figure 6.7 for details (Ferreira *et al.*, 2025). The main draw back for the use of LRS in forensics is the amount of input DNA required, this is typically more than traditional forensic DNA workflows. However, LRS is a new methodology and is continually improving. With over 40 PubMed publication results for forensics and LRS, research is clearly pushing forward this field (Ferreira *et al.*, 2025). Going forward, It is likely LRS will be seen much more regularly in wildlife forensics and conservation work.

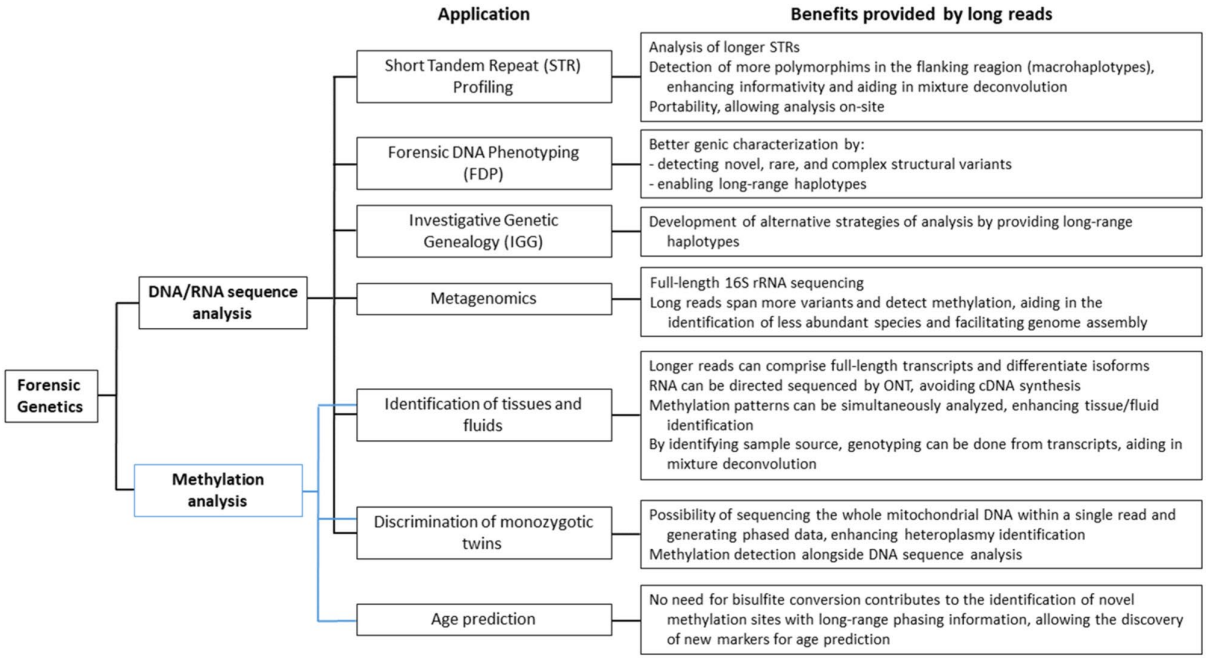


Figure 6.7: *Main applications of Long Read Sequencing (LRS) in forensic genetics and its associated benefits (Ferreira et al., 2025).*

6.5 Conclusion

Overall, this study achieved the following:

- This study was the first to utilise the full sequence of the D-Loop to assess population diversity and analyse population signature haplotypes.
- This study compared the use of Sanger sequencing to NGS via Oxford Nanopore for the application of SNP and InDel detection.
- Only 6 sites of variation were identified as a result of the sanger sequencing compared to 47 sites of variation detected as a result of the Nanopore sequencing.
- Variation calling via nanopore sequencing was achieved at a 300x sequencing depth compared to 1x with Sanger Sequencing. The level of variation observed in the NGS sequences compared to Sanger equates to a staggering 783% increase.

7 Overall Discussion and Future Directions

The project initially set out to develop an individual identification tool for fallow deer, harnessing polymorphic loci mined from the genome sequence. The Whole genome sequence was achieved by NGS on the Pac-Bio Sequel II. In order to obtain the high level of sequencing that was gained by this project the DNA extraction method had to be perfect in order to preserve as much genomic, unfragmented, high quality and high yield DNA as possible. This is to ensure as much of the genome is sequenced at a high % coverage. The genome achieved in this project made reference genome standard with a 96.4% coverage BUSCO score. Having such high quality was crucial for the identification of variation and the mining of STR loci, because any errors in the DNA sequence could cause the miss-identification of variants or potential variation being lost. In this case, a total of 21,580 tetra and penta nucleotide STRs were located within the genome and 264 primers were developed. Out of the 132 deigned primers for tetranucleotide markers, 90 were tested and 9 were found to be polymorphic. Due to the lack of polymorphism found within the STR loci and the time consuming nature of the work involved it was decided that time was better spent investigating this lack of genetic diversity rather than trying to find something which wasn't there.

The new phase of the project embarked on utilising NGS as a tool to mass screen for polymorphism within the fallow deer. This new way of screening for polymorphism is fast and effective and unearths genetic data, which is previously lost with traditional methods, such as Sanger Sequencing. The new methodology used to pool and screen across large data sets using Oxford Nanopore presented by this project offers a new way to screen new species which currently lack published genetic data. Following the success of screening autosomal STRs with NGS, and despite slightly more polymorphism being found, it was still not enough to be considered as forensically significant in terms of individual identification therefore attentions turned to the mitochondrial genome and furthermore specifically the hypervariable D-Loop region. The mitochondrial genome was isolated from the initial whole genome sequencing and annotated. The mitochondrial genome for *Dama dama* is 16,322bp in length, consisting of 13 protein coding genes, 2 rRNAs and 22 tRNAs. Interesting finding came out of in depth phylogenetic analysis when comparing to 25 other *Cervidae* species. Whilst most of the findings were in favour of previously published work, new discoveries suggest that as *Megaloceros giganteus* is the older of the *Dama* species, *Dama mesopotamica* has held onto more ancestral genetic sequence than *Dama dama*,

and that it is *Dama dama* that diverged first and has mutated away from the ancestral lineage, around 5.35 million years ago. Finally, the final part of this project used NGS to mass screen across 15 British populations of fallow deer to assess potential haplotype diversity. Despite not being able to assess and call individual haplotypes due to the lack of individual data, the study provides a population wide assessment of diversity and any variation is indicative of population signature haplotypes only. However, The variation found provides scope to further the NGS sequencing, to pull apart the pooled individuals and repeat the work on each sample separately. The new methodology presented in this study enables labs to assess whether such work and funds to support it is worth it or not. Despite this PhD project ending up in a different place then where it started, it is to only be seen as a triumph of new discovery, new possibilities and a new perspective for the future of fallow deer in the UK.

One of the main successes of this project was achieving a high sample pool of British fallow deer samples from across the UK. A total of 352 samples were collected spanning 15 individual populations; this is the highest number of modern fallow deer samples collected and analysed from the UK for genetic diversity purposes to date. However, the majority of the sampling regions are based around the south of England (Figure 4.1), this was due to the responses I got back from park rangers volunteering to help collect samples. Other national and private parks were contacted but not everyone responded and only from the 15 sampled populations was a response received from the ranger with samples. It also must be noted that the majority of the UK population exists in the south of England with numbers reducing as you go north, therefore despite this study having fewer samples from northern populations, it does include a representative sample set from the UK fallow deer distribution (Figure 1.7). If future projects were to take place based off of the research presented in this study, it would be ideal to include more samples from Ireland, Scotland and the North of England to compare more widely with the southern populations. A further recommendation would be to include samples from Europe; this would provide genetic data from potential founding fallow deer populations which could provide useful in sight to the origins of the samples analysed in this project. The inclusion of European samples would also enable further comparison with previous work completed by Baker *et al* (2017; 2024). As for this project, the lack of European samples was due to England going through Brexit alterations and changing border control legislation. It was becoming increasingly difficult to transport animal tissue/blood samples from Europe to the UK. Furthermore, due to differences in Europe surrounding deer

culling, it was difficult to find someone to collect fallow samples in a humane way to not incur any ethical scrutiny and additionally, due to language barriers it was difficult to explain the importance of my project and reassure people of the legitimacy of my work. Now the Brexit transition has settled, with more time and support on this it is more likely to be able to get this set up in the future.

Fallow deer in the UK are often overlooked due to the increasing herd numbers in the south causing over populated areas effecting farmland and natural habitats (Potts *et al.*, 2014). Most rangers and landowners want to increase culling in order to manage the growing numbers. However, with the profound lack of genetic diversity detailed out in this study, there is a cause for concern surrounding issues of conservation vulnerability. Low genetic diversity within a population can lead to evolutionary bias and puts the species under stress from disease and potential mass extinction events (Kardos, 2021). Whilst the fallow deer species in the UK is often rolled under the same umbrella as red and roe deer, termed invasive by many due to the large population size, fallow deer deserve to be individualised, this project adds to previous published justifications for fallow deer to have their own conservation protection (Baker *et al.*, 2017; Baker *et al.*, 2024a; Baker *et al.*, 2024b). This study provides support for population wide genetic diversity assessments in order to better understand which populations are more vulnerable than others. This genetic data can then help advise modern management strategies, including future breeding programs. For example, unrelated individuals with known increased diversity could be moved into areas/populations with lower diversity where natural movement corridors are not possible to increase gene flow, this is known as 'genetic rescue' (Kardos, 2021). For example, this type of management program would be most effective in private parks rather than wild herds where natural movement is not possible, however it could still be beneficial in wider re-locations such as between wild Scottish herds and those from Southern England or Ireland rather than within Southern England itself. A further possibility in order to reduce population size but increase genetic diversity could be a gene dive to reduce fertility in populations lacking genetic diversity and are more susceptible to decline as a result of natural selection (Chen *et al.*, 2024). This would involve genetic engineering and CRISPR homing gene drives (Chen *et al.*, 2024). However this approach could be difficult to manage, potentially less specific depending on the deployment method used and may risk losing genetic diversity further if done incorrectly which could be costly, therefore harder to justify in the current climate.

This study is just the beginning of a journey to learn more about the population genetics of the British fallow deer. Future work stemming from this project has already been discussed with focus on increasing genetic diversity of the fallow deer population, this is paramount and should be a clear focus of future research on the species. However, there is potential further work to be done within all aspects of fallow deer welfare and management which can now be made more achievable by the results and publications on the back of this project. For example, future whole genome sequencing is made easier due to the availability of the reference genome, sequenced in study 1. Future genomes can be easily obtained at lower sequencing depths and assembled due to alignment with the reference genome, which makes it cheaper and more accessible to researchers across the world. Furthermore, the availability of whole genome data lends itself to disease treatment and prevention research which was previously not possible. Common diseases associated with fallow deer are chronic wasting disease and lymes disease (Hamir *et al.*, 2011; Gandy *et al.*, 2021). The genome can help with the detection of how these diseases are spread, how they affect individuals and the development of medicinal treatments. This project has begun work assessing genetic divergence with comparisons with other Cervidae, this could be furthered with more in-depth investigations with the *Dama mesopotamica* (Persian fallow deer) with the aim for reintroduction of this critically endangered sub-species into the wild. This work could utilise STR loci discovered in Study 2 of this project or the Hypervariable D-Loop, also sequenced in depth as part of this project in study 3. There is scope to further work into the STR loci. Utilising the new locus screening methodology via Oxford Nanopore, more STR loci can be screened for polymorphism. A total of 21,580 tetra and penta nucleotide STRs were located within the genome and due to time and budget restraints only 90 were tested. Now the new methodology has been proven successful in rapid screening and detection a future project could utilise this to investigate these loci further. This project was only able to analyse tetra-nucleotide loci, it may be that Di-nucleotide or Tri-nucleotide markers house more variation, which despite their draw backs as markers for forensic multiplexing due to increased PCR artefacts associated, they may be the way to go for the fallow deer species due to the distinct lack of polymorphism within the tetra-nucleotide markers. Where possible, future work should encompass the Persian fallow deer species, in a bid to protect the current most endangered deer species in the world.

Technology available for DNA analysis is rapidly evolving and is intrinsic to the future of this project and the continuation of research into the genetics of fallow

deer. For example, new developments in sequencing technology over the last 20 years, with the introduction of NGS, enabled the identification of the autosomal and mitochondrial variation detailed in this project which previous studies had missed due to the legacy analysis technology used. There have been recent developments within the field of DNA analysis which is important to forensic wildlife conservation.

New to the scene of species identification is SpelID (Ghemrawi *et al.*, 2021). SpelID is a pyrosequencing-based assay utilizing the 12S rRNA gene. The fragment targeted is small, only 35 bases, and is able to distinguish between 32 commonly known species as pets. The assay is able to process up to 48 samples within 48 hours (Ghemrawi *et al.*, 2021). The short target fragment and high sensitivity (down to 5pg DNA) makes this an ideal assay for forensic applications where DNA is often degraded. This assay is ideal for forensic cases which have occurred in and around the home, but less applicable currently to major wildlife crime. The assay is fully validated for forensic use in accordance with the Scientific Working Group on DNA Analysis Methods (SWGDAM), specifically looking at the sensitivity, specificity, reproducibility, stability of the assay and run-on casework samples (Ghemrawi *et al.*, 2021). The assay uses relatively new technology and requires a small bench top pyrosequencer which makes it fairly portable and easily obtainable. For some closely related species, this assay however remains a presumptive test, as it cannot distinguish the difference between them, for example fish (Ghemrawi *et al.*, 2021). In these cases, other confirmatory analysis would be required, such as Massive Parallel sequencing (Ghemrawi *et al.*, 2021). Overall, this assay is quick and effective for common forensic cases, especially when it is needed to differentiate between human and non-human samples, however it requires further development if it is to be wider used to help tackle wildlife crime but the foundations for this are there.

Other advances are happening across the field from standard PCR amplification to LAMP and the use of NGS with CRISPR-Cas9 methodology (Nagarajan *et al.*, 2024; Yugovich *et al.*, 2025). Figure 7.1 details recent advances in the field of species ID and how NGS is being utilised. There is a growing demand for NGS in forensic wildlife conservation due to the higher throughput, increased sensitivities and ability to achieve full target sequences from low input DNA. NGS has the capability to sequence multiple target regions simultaneously and new developments from Oxford Nanopore enable work flows to be undertaken in the field on their MinION™ machine (Yugovich *et al.*, 2025). NGS also opens up the possibility to gain identifications from difficult samples such as horn, teeth, ivory and bone (Meiklejohn *et al.*, 2021). NGS is also used in the detection of

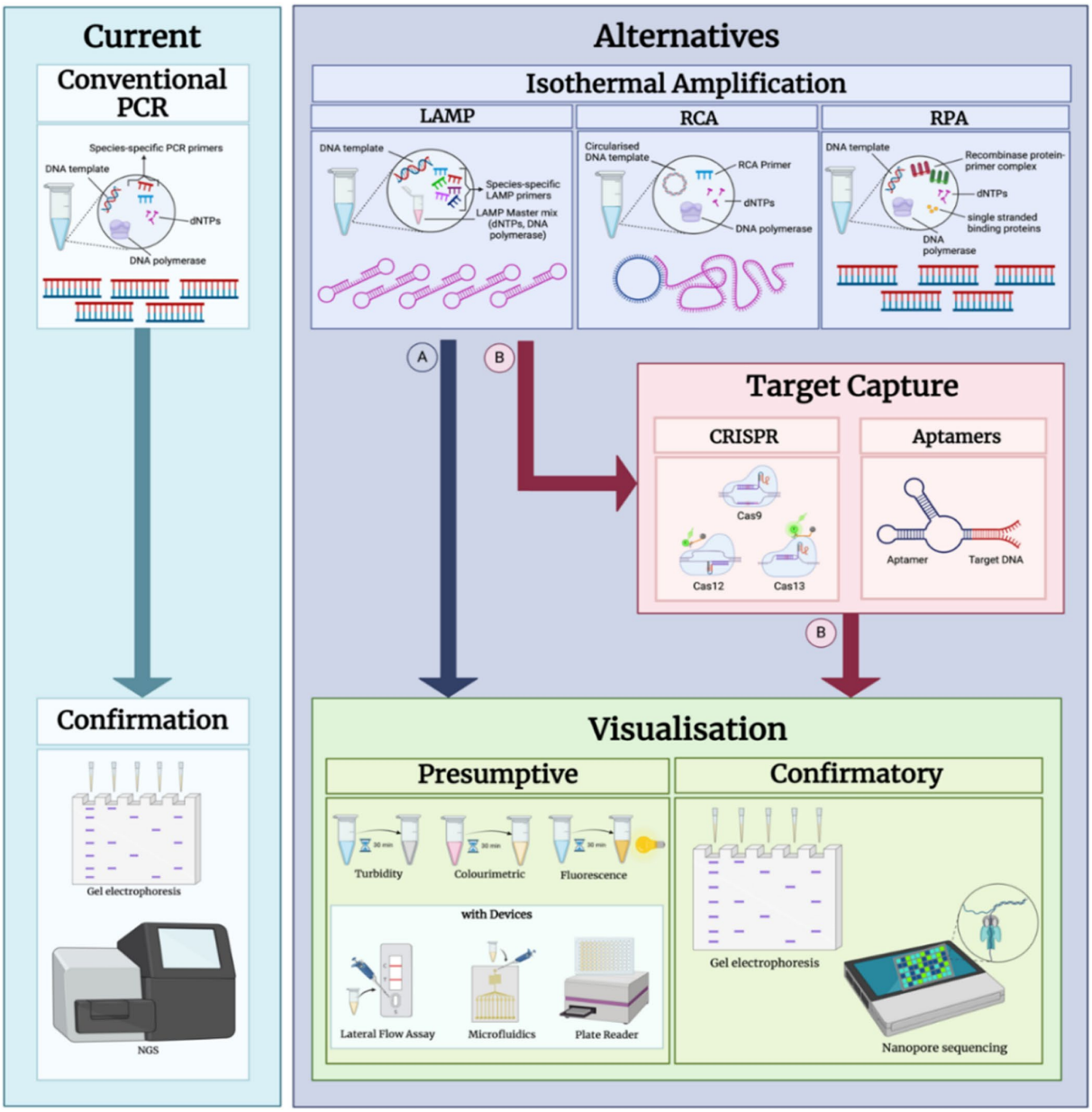


Figure 7.1: *Overview of the recent advancements in Species ID DNA analysis with the utilisation of NGS (Yugovich et al., 2025).*

mixtures and can distinguish taxa present as low as 1% (Meiklejohn *et al.*, 2021). This is beneficial in many environmental cases and those investigating food mislabelling violations. However, despite the many advantages of implementing NGS into a wildlife forensic context, many advances currently published from research are not validated to an ISFG or SWGDAM standard so are not yet applicable to active casework. Furthermore, the data generated from NGS, albeit advantageous for many aspects of wildlife conservation analysis, comes at the challenge of storage. Single runs can generate in excess of 15GB worth of data, laboratories will require their own dedicated data storage platform in order to perform necessary downstream data processing and analysis in house (Meiklejohn *et al.*, 2021). It is likely as NGS developments continue and future research pushes

the boundaries of what this technology can do for the field of forensic wildlife conservation genetics, the more we will see the barriers, which are preventing laboratories from currently utilising it, lowered.

Forensic genetics has developed over the years, historically known as forensic biology, later changing to forensic haemogenetics and now forensic genetics (Carracedo, 2025). More recently the term forensic wildlife genetics has been seen, separating human and wildlife forensic genetics disciplines. In 2024 we saw the introduction of the journal Forensic Science International: Animals and Environments, the first journal specifically dedicated to wildlife forensics. There have been recent calls to have the term forensic genetics changed again to forensic genomics due to the advancements in forens-omics with growing interest in the microbiome, proteomics, transcriptomics and the use of DNA methylation for forensic purposes (Carracedo, 2025). As far as the future of wildlife forensics is concerned, it is residual on the continuation of novel research and technology as mentioned, to address continually evolving trends in wildlife crime. In order to properly tackle wildlife crime, it is paramount that genetic tools developed for wildlife are validated to current standards for use in active casework (Ogden *et al.*, 2009; Johnson *et al.*, 2014). This step is unfortunately often lacking in research due to the rigorous and time consuming nature of the work involved and the lack of fixed guidelines for NGS data in wildlife species. Furthermore, in order to progress the field of wildlife genetics, NGS must be utilised to bolster reference data available. A project such as the Darwin Tree of Life is influential in achieving this by conducting sequencing of 70,000 species genomes from Britain and Ireland, however this needs to go further to cover species around the world especially in wildlife crime hotspot areas. Another development, which would be beneficial to the future of wildlife forensic conservation genetics, would be the increase of dedicated accredited wildlife DNA laboratories, such as RhODIS® (Ogden *et al.*, 2009; Harper *et al.*, 2013; Johnson *et al.*, 2014). The future of pure conservation genetics is also bright, with developments with NGS making whole genome sequencing possible for less than $1000, presents new opportunities for researchers across the world to harness, combining genetics and conservation (Kardos, 2021). As presented in this project, genome sequencing enables detailed understanding on population structure, evolutionary divergence, relationships within and amongst populations and relationships with closely related taxa. As with fallow deer, this future work is crucial for understanding conservation vulnerabilities in species, which is currently a growing concern due to increased climate change risk and other human inflicted conflicts, in order to preserve populations with low genetic diversity and adaptation rate which we may see become extinct in the not so distant future if nothing is done to protect them.

8 Conclusion

In Conclusion, the findings presented in this thesis collectively support previous work on fallow deer, displaying a profound lack of genetic diversity. This project went one step further and presents an in depth assessment of the current British fallow deer population and their relationships to other European fallow deer (previously sequenced in the Baker *et al* (2017) study) as well as wider Cervidae species. Variation discovered in the Devon and Knole Park populations were found to be linked to German, Rhodian, Italian, Portuguese and Spanish populations. Whereas other variation found within the Devon population corresponded solely to England in the Baker *et al* (2017) study. This shines a light on the ancestral origins of these populations within Britain. It can also be used to identify which populations are more genetically diverse than others, information which can help shape modern management and breeding programs. The data presented within this project should be treated as a warning for the future of the species, currently fallow deer are listed as least concern (LC) on the IUCN due to their vast population size in many areas, however, due to their severe lack of genetic diversity there is growing concern for the species and their low adaptation rate causing conservation vulnerabilities. Through an increase of awareness for the effects of low genetic diversity and the implications on a species, regardless of the population size, it is possible to change this outlook and increase diversity with fallow deer. With no change, the species is at risk of mass extinction due to disease or strain from climate change events. Due to current attitudes towards deer and the archaic hunting culture associated in the UK, it will be a journey to change perceptions. Especially with, park rangers, deer stalkers and hunters who are typically bound by tradition. These individuals are on the front line when it comes to deer management, therefore they are in a position to make the biggest impact. However, it is time to rethink attitudes towards animals given the global biodiversity crisis, increasing wildlife conservation research and continuing to spread awareness of species which were previously over looked, will in time shift perceptions and create an awareness for the creatures we share our planet with.

9 References

Abe, H., Hayano, A., Inoue-Murayama, M., 2012. Forensic species identification of large macaws using DNA barcodes and microsatellite profiles. Molecular Biology Reports 39, 693–699. doi:10.1007/s11033-011-0787-1

Adavoudi, R., Pilot, M., 2021. Consequences of hybridization in mammals: A systematic review. Genes 13, 50. doi:10.3390/genes13010050

Afgan, E., Nekrutenko, A., Grüning, Bjórn A, Blankenberg, D., Goecks, J., Schatz, M.C., Ostrovsky, A.E., Mahmoud, A., Lonie, A.J., Syme, A., Fouilloux, A., Bretaudeau, A., Nekrutenko, A., Kumar, A., Eschenlauer, A.C., DeSanto, A.D., Guerler, A., Serrano-Solano, B., Batut, B., Grüning, Björn A, Langhorst, B.W., Carr, B., Raubenolt, B.A., Hyde, C.J., Bromhead, C.J., Barnett, C.B., Royaux, C., Gallardo, C., Blankenberg, D., Fornika, D.J., Baker, D., Bouvier, D., Clements, D., de Lima Morais, D.A., Tabernero, D.L., Lariviere, D., Nasr, E., Afgan, E., Zambelli, F., Heyl, F., Psomopoulos, F., Coppens, F., Price, G.R., Cuccuru, G., Corguillé, G.L., Von Kuster, G., Akbulut, G.G., Rasche, H., Hotz, H.-R., Eguinoa, I., Makunin, I., Ranawaka, I.J., Taylor, J.P., Joshi, J., Hillman-Jackson, J., Goecks, J., Chilton, J.M., Kamali, K., Suderman, K., Poterlowicz, K., Yvan, L.B., Lopez-Delisle, L., Sargent, L., Bassetti, M.E., Tangaro, M.A., van den Beek, M., Čech, M., Bernt, M., Fahrner, M., Tekman, M., Föll, M.C., Schatz, M.C., Crusoe, M.R., Roncoroni, M., Kucher, N., Coraor, N., Stoler, N., Rhodes, N., Soranzo, N., Pinter, N., Goonasekera, N.A., Moreno, P.A., Videm, P., Melanie, P., Mandreoli, P., Jagtap, P.D., Gu, Q., Weber, R.J., Lazarus, R., Vorderman, R.H., Hiltemann, S., Golitsynskiy, S., Garg, S., Bray, S.A., Gladman, S.L., Leo, S., Mehta, S.P., Griffin, T.J., Jalili, V., Yves, V., Wen, V., Nagampalli, V.K., Bacon, W.A., de Koning, W., Maier, W., Briggs, P.J., 2022. The Galaxy Platform for Accessible, reproducible and collaborative biomedical analyses: 2022 update. Nucleic Acids Research 50. doi:10.1093/nar/gkac247

Al-Senaidi R, A.-M.A., 2015. DNA Barcoding of Panulirus homarus from Oman and Yemen. Journal of Aquaculture & Marine Biology 2, 1–6. doi:10.15406/jamb.2015.02.00043

Alberts, B., Johnson, A., Lewis, J., Martin, M., Roberts, K., Walter, P., 2002. Molecular biology of the cell, 4th ed. Garland Science, New York.

Alonge, M., Lebeigle, L., Kirsche, M., Aganezov, S., Wang, X., Lippman, Z.B., Schatz, M.C., Soyk, S., 2021. Automated Assembly scaffolding elevates a new

tomato system for high-throughput genome editing. BioRxiv The PrePrint Server for Biology. doi:10.1101/2021.11.18.469135

An, J., Lee, M., Min, M.-S., Lee, M.-H., Lee, H., 2007. A molecular genetic approach for species identification of mammals and sex determination of birds in a forensic case of poaching from South Korea. Forensic Science International 167, 59–61. doi:10.1016/j.forsciint.2005.12.031

Arslangündoğdu, Z., Kasparek, M., Sarbaçak, H., Kaçar, M.S., Yöntem, O., Şahin, M.T., 2010. Development of the population of the European Fallow Deer, *Dama dama dama* (Linnaeus, 1758), in Turkey. Zoology in the Middle East 49, 3–12. doi:10.1080/09397140.2010.10638383

Ayres, K.L., Overall, A.D.J., 1999. Allowing for within-subpopulation inbreeding in forensic match probabilities. Forensic Science International 103, 207–216. doi:10.1016/s0379-0738(99)00087-0

Ba, H., Cai, Z., Gao, H., Qin, T., Liu, W., Xie, L., Zhang, Y., Jing, B., Wang, D., Li, C., 2020. Chromosome-level genome assembly of Tarim Red Deer, cervus elaphus yarkandensis. Scientific Data 7. doi:10.1038/s41597-020-0537-0

Baker, C.S., Steel, D., Choi, Y., Lee, H., Kim, K.S., Choi, S.K., Ma, Y.-U., Hambleton, C., Psihoyos, L., Brownell, R.L., Funahashi, N., 2010. Genetic evidence of illegal trade in protected whales links Japan with the US and South Korea. Biology Letters 6, 647–650. doi:10.1098/rsbl.2010.0239

Baker, K. H., Gray, H.W., Lister, A.M., Spassov, N., Welch, A.J., Trantalidou, K., De Cupere, B., Bonillas, E., De Jong, M., Çakırlar, C., Sykes, N., Hoelzel, A.R., 2024. Ancient and modern DNA track temporal and spatial population dynamics in the European fallow deer since the Eemian interglacial. Scientific Reports 14. doi:10.1038/s41598-023-48112-6

Baker, K.H., Gray, H.W., Pekşen, Ç.A., Hoelzel, A.R., 2021. Conservation genetics of the European fallow deer: A reply to Marchesini *et al.*. Mammalian Biology 101, 313–319. doi:10.1007/s42991-021-00121-6

Baker, K.H., Gray, H.W., Ramovs, V., Mertzanidou, D., Akın Pekşen, Bilgin, C.C., Sykes, N., Hoelzel, A.R., 2017. Strong population structure in a species manipulated by humans since the neolithic: The European fallow deer (*Dama dama dama*). Heredity 119, 16–26. doi:10.1038/hdy.2017.11

Baker, Karis H., Miller, H., Doherty, S., Gray, H.W., Daujat, J., Çakırlar, C., Spassov, N., Trantalidou, K., Madgwick, R., Lamb, A.L., Ameen, C., Atici, L.,

Baker, P., Beglane, F., Benkert, H., Bendrey, R., Binois-Roman, A., Carden, R.F., Curci, A., De Cupere, B., Detry, C., Gál, E., Genies, C., Kunst, G.K., Liddiard, R., Nicholson, R., Perdikaris, S., Peters, J., Pigière, F., Pluskowski, A.G., Sadler, P., Sicard, S., Strid, L., Sudds, J., Symmons, R., Tardio, K., Valenzuela, A., van Veen, M., Vuković, S., Weinstock, J., Wilkens, B., Wilson, R.J., Evans, J.A., Hoelzel, A.R., Sykes, N., 2024. The 10,000-year biocultural history of fallow deer and its implications for conservation policy. Proceedings of the National Academy of Sciences 121. doi:10.1073/pnas.2310051121

Balding, D.J., 1995. Estimating Products in Forensic Identification Using DNA Profiles. Journal of the American Statistical Association 90, 839–844. doi:10.1080/01621459.1995.10476582

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 Science International 64, 125–140. doi:10.1016/0379-0738(94)90222-4

Ball, M.C., Finnegan, L.A., Nette, T., Broders, H.G., Wilson, P.J., 2011. Wildlife forensics: "Supervised" assignment testing can complicate the association of suspect cases to source populations. Forensic Science International: Genetics 5, 50–56. doi:10.1016/j.fsigen.2010.02.002

Banos-Ruiz, I.B.R., 2017. Europe, a silent hub of illegal wildlife trade [WWW Document]. DW.COM. URL https://www.dw.com/en/europe-a-silent-hub-of-illegal-wildlife-trade/a-37183459 (accessed 5.14.20).

Bao, W., Kojima, K.K., Kohany, O., 2015. Repbase update, a database of repetitive elements in eukaryotic genomes. Mobile DNA 6. doi:10.1186/s13100-015-0041-9

Barbanera, F., Guerrini, M., Beccani, C., Forcina, G., Anayiotos, P., Panayides, P., 2012. Conservation of endemic and threatened wildlife: Molecular forensic DNA against poaching of the Cypriot mouflon (Ovis orientalis ophion, Bovidae). Forensic Science International: Genetics 6, 671–675. doi:10.1016/j.fsigen.2011.12.001

Barbisin, M., Fang, R., O'Shea, C.E., Calandro, L.M., Furtado, M.R., Shewale, J.G., 2009. Developmental Validation of the Quantifiler®Duo DNA Quantification Kit for Simultaneous Quantification of Total Human and Human Male DNA and Detection of PCR Inhibitors in Biological Samples. Journal of Forensic Sciences 54, 305–319. doi:10.1111/j.1556-4029.2008.00951.x

Barnard, R., Jackson, R., Gee, O., De Boer, J., Iyengar, I., n.d. . Work Towards the Development of an STR Multiplex For The Identification of Fallow Deer Individuals.

Barnard, R.K., Smith, J.A., Yuan, N., Liu, F., Hadi, S.S., 2023. An announcement of a new genome sequence available for *Dama dama* (fallow deer). Forensic Science International: Animals and Environments 4, 100074. doi:10.1016/j.fsiae.2023.100074

Baum, D., 2008. Reading a Phylogenetic Tree: The Meaning of Monophyletic Groups. Nature Education 1.

Beglane, F., Baker, K., Carden, R.F., Hoelzel, A.R., Lamb, A.L., Fhionnghaile, R.M., Miller, H., Sykes, N., 2018. Ireland's Fallow Deer: Their historical, Archaeological and Biomolecular Records. Proceedings of the Royal Irish Academy: Archaeology, Culture, History, Literature 118C, 141. doi:10.3318/priac.2018.118.01

Berger, B., Berger, C., Hecht, W., Hellmann, A., Rohleder, U., Schleenbecker, U., Parson, W., 2014. Validation of two canine STR multiplex-assays following the ISFG recommendations for non-human DNA analysis. Forensic Science International: Genetics 8, 90–100. doi:10.1016/j.fsigen.2013.07.002

Bernt, M., Donath, A., Jühling, F., Externbrink, F., Florentz, C., Fritzsch, G., Pütz, J., Middendorf, M., Stadler, P.F., 2013. Mitos: Improved de novo metazoan mitochondrial genome annotation. Molecular Phylogenetics and Evolution 69, 313–319. doi:10.1016/j.ympev.2012.08.023

Bingpeng, X., Heshan, L., Zhilan, Z., Chunguang, W., Yanguo, W., Jianjun, W., 2018. DNA barcoding for identification of fish species in the Taiwan Strait. PLOS ONE 13, e0198109. doi:10.1371/journal.pone.0198109

Boore, J.L., 1999. Animal mitochondrial genomes. Nucleic Acids Research 27, 1767–1780. doi:10.1093/nar/27.8.1767

Brandies, P., Peel, E., Hogg, C., Belov, K., 2019. The Value of Reference Genomes in the Conservation of Threatened Species. Genes 10, 846. doi:10.3390/genes10110846

Breda, M., Lister, A.M., 2013. Dama roberti, a new species of deer from the early middle pleistocene of Europe, and the origins of modern Fallow Deer. Quaternary Science Reviews 69, 155–167. doi:10.1016/j.quascirev.2013.01.029

The British Association for Shooting and Conservation (BASC), 2013. , Poaching of Game, Deer and Fish, BASC Campaign Against Poaching. The British Association for Shooting and Conservation Ltd, Wrexham.

The British Deer Society (BDS), 2016. Fallow deer [WWW Document]. Bds.org.uk. URL https://www.bds.org.uk/index.php/advice-education/species/fallow-deer (accessed 4.30.20).

Bronstein, O., Kroh, A., Haring, E., 2018. Mind the gap! the mitochondrial control region and its power as a phylogenetic marker in echinoids. BMC Evolutionary Biology 18. doi:10.1186/s12862-018-1198-x

Brown, W.M., George, M., Wilson, A.C., 1979. Rapid evolution of animal mitochondrial DNA. Proceedings of the National Academy of Sciences 76, 1967–1971. doi:10.1073/pnas.76.4.1967

Buckleton, J.S., Bright, J.-A., Taylor, D., 2016. Forensic DNA Evidence Interpretation, 2nd ed. Chapman & Hall/CRC Pres, Boca Raton.

Buckleton, J.S., Triggs, C.M., Walsh, S.J., 2005. Forensic DNA evidence interpretation, 1st ed. CRC Press, Boca Raton.

Budowle, B., Garofano, P., Hellman, A., Ketchum, M., Kanthaswamy, S., Parson, W., van Haeringen, W., Fain, S., Broad, T., 2005. Recommendations for animal DNA forensic and identity testing. International Journal of Legal Medicine 119, 295–302. doi:10.1007/s00414-005-0545-9

Butler, J., 2005. Forensic DNA typing, 1st ed. Elsevier Academic Press, Amsterdam.

Butler, J.M., 2014. Advanced topics in forensic DNA typing: Interpretation, 1st ed. Academic Press, Oxford, England.

Butler, J.M., 2015. The future of forensic DNA analysis. Philosophical Transactions of the Royal Society B: Biological Sciences 370, 20140252. doi:10.1098/rstb.2014.0252

Butler, J.M., Willis, S., 2020. Interpol review of forensic biology and forensic DNA typing 2016-2019. Forensic Science International: Synergy. doi:10.1016/j.fsisyn.2019.12.002

Cameron, S.L., 2014. Insect mitochondrial genomics: Implications for evolution and Phylogeny. Annual Review of Entomology 59, 95–117. doi:10.1146/annurev-ento-011613-162007

Caniglia, R., Fabbri, E., Greco, C., Galaverni, M., Randi, E., 2010. Forensic DNA against wildlife poaching: Identification of a serial wolf killing in Italy. Forensic Science International: Genetics 4, 334–338. doi:10.1016/j.fsigen.2009.10.012

Caniglia, R., Fabbri, E., Mastrogiuseppe, L., Randi, E., 2013. Who is who? Identification of livestock predators using forensic genetic approaches. Forensic Science International: Genetics 7, 397–404. doi:10.1016/j.fsigen.2012.11.001

Carbó-Ramírez, S., Codoñer-Alejos, A., Reyes-Prieto, M., Bernat, C., Gil, A., Ruiz-Ramírez, J., Soriano-Chirona, V., Marco-Romero, G.D., Gonzalez-Candelas, F., Martinez-Priego, L., 2025. One sample, three genotypes: A flanking region deletion at the D19S433 locus causes genotyping discrepancies between CE and NGS Technologies. Forensic Science International: Genetics 78, 103301. doi:10.1016/j.fsigen.2025.103301

Cardinali, I., Tancredi, D., Lancioni, H., 2023. The Revolution of Animal Genomics in Forensic Sciences. International Journal of Molecular Sciences 24, 8821. doi:10.3390/ijms24108821

Carracedo, A., 2025. Forensic science international: Genetics: Past, present, and future of the journal and the Field. Forensic Science International: Genetics 79, 103255. doi:10.1016/j.fsigen.2025.103255

Carver, T., Harris, S.R., Berriman, M., Parkhill, J., McQuillan, J.A., 2011. Artemis: An integrated platform for visualization and analysis of high-throughput sequence-based experimental data. Bioinformatics 28, 464–469. doi:10.1093/bioinformatics/btr703

Cassidy, B.G., Gonzales, R.A., 2005. DNA Testing in Animal Forensics. Journal of Wildlife Management 69, 1454–1462. doi:10.2193/0022-541x(2005)69[1454:dtiaf]2.0.co;2

Ceballos, G., Ehrlich, P.R., Barnosky, A.D., García, A., Pringle, R.M., Palmer, T.M., 2015. Accelerated modern human–induced species losses: Entering the sixth mass extinction. Science Advances 1. doi:10.1126/sciadv.1400253

Chaisson, M.J.P., Wilson, R.K., Eichler, E.E., 2015. Genetic variation and the de novo assembly of human genomes. Nature Reviews Genetics 16, 627–640. doi:10.1038/nrg3933

Chan, P.P., Lin, B.Y., Mak, A.J., Lowe, T.M., 2021. TRNAscan-SE 2.0: Improved detection and functional classification of transfer RNA genes. Nucleic Acids Research 49, 9077–9096. doi:10.1093/nar/gkab688

Chapman, D., Chapman, N., 1997. Fallow Deer, 2nd ed. Coch-Y-Bonddu Books, Powys.

Chapman, N.G., Chapman, D.I., 1980. The distribution of fallow deer: a worldwide review. Mammal Review 10, 61–138. doi:10.1111/j.1365-2907.1980.tb00234.x

Chen, J.-W., Uboh, C.E., Soma, L.R., Li, X., Guan, F., You, Y., Liu, Y., 2010. Identification of racehorse and sample contamination by novel 24-plex STR system. Forensic Science International: Genetics 4, 158–167. doi:10.1016/j.fsigen.2009.08.001

Chen, W., Guo, J., Liu, Y., Champer, J., 2024. Population suppression by release of insects carrying a dominant sterile homing gene drive targeting doublesex in drosophila. Nature Communications 15. doi:10.1038/s41467-024-52473-5

Cheng, H., Concepcion, G.T., Feng, X., Zhang, H., Li, H., 2021. Haplotype-resolved de Novo Assembly using phased assembly graphs with hifiasm. Nature Methods 18, 170–175. doi:10.1038/s41592-020-01056-5

Chin Chin, T., Adibah, A.B., Danial Hariz, Z.A., Siti Azizah, M.N., 2016. Detection of mislabelled seafood products in Malaysia by DNA barcoding: Improving transparency in food market. Food Control 64, 247–256. doi:10.1016/j.foodcont.2015.11.042

Ciavaglia, S., Linacre, A., 2018. OzPythonPlex: An optimised forensic STR multiplex assay set for the Australasian carpet python (Morelia spilota). Forensic Science International: Genetics 34, 231–248. doi:10.1016/j.fsigen.2018.03.002

Ciavaglia, S.A., Henry, J.M., Linacre, A.M.T., 2013. Profiling pythons to combat common illegal wildlife activities. Forensic Science International: Genetics Supplement Series 4, e31–e32. doi:10.1016/j.fsigss.2013.10.015

CITES (The Convention on International Trade in Endangered Species of Wild Fauna and Flora), 2019. What is CITES? | CITES [WWW Document]. Cites.org. URL https://www.cites.org/eng/disc/what.php (accessed 4.23.20).

Clayton, T.M., Whitaker, J.P., Sparkes, R., Gill, P., 1998. Analysis and interpretation of mixed forensic stains using DNA STR profiling. Forensic Science International 91, 55–70. doi:10.1016/s0379-0738(97)00175-8

Cler, L., Bu, D., Lewis, C., Euhus, D., 2006. A comparison of five methods for extracting DNA from paucicellular clinical samples. Molecular and Cellular Probes 20, 191–196. doi:10.1016/j.mcp.2005.12.003

Clifford, S.L., Anthony, N.M., Bawe-Johnson, M., Abernethy, K.A., Tutin, C.E.G., White, L.J.T., Bermejo, M., Goldsmith, M.L., McFarland, K., Jeffery, K.J., Bruford, M.W., Wickings, E.J., 2004. Mitochondrial DNA phylogeography of western lowland gorillas (Gorilla gorilla gorilla). Molecular Ecology 13, 1551–1565. doi:10.1111/j.1365-294x.2004.02140.x

CMS, 2019. , 1st ed, Wildlife Crime. Convention on the Conservation of Migratory Species of Wild Animals, Bonn, Germany.

Cock, P.J.A., Fields, C.J., Goto, N., Heuer, M.L., Rice, P.M., 2009. The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. Nucleic Acids Research 38, 1767–1771. doi:10.1093/nar/gkp1137

Coghlan, M.L., White, N.E., Parkinson, L., Haile, J., Spencer, P.B.S., Bunce, M., 2012. Egg forensics: An appraisal of DNA sequencing to assist in species identification of illegally smuggled eggs. Forensic Science International: Genetics 6, 268–273. doi:10.1016/j.fsigen.2011.06.006

Connon, C.C., LeFebvre, A.K., Benjamin, R.C., 2016. Validation of alternative capillary electrophoresis detection of strs using pop-6 polymer and a 22 cm array on a 3130 XL genetic analyzer. Forensic Science International: Genetics 22, 113–127. doi:10.1016/j.fsigen.2016.02.006

Coyle, H.M., van de Goor, L.H.P., van Haeringen, W.A., 2007. Identification of stolen cattle using 22 microsatellite markers, in: Nonhuman DNA Typing: Theory and Casework Applications. CRC Press, Boca Raton, pp. 122–123.

Criminal justice response to wildlife crime in Malaysia, 2017. Combating Wildlife and Forest Crime. United Nations.

Curole, J.P., Kocher, T.D., 1999. Mitogenomics: Digging deeper with complete mitochondrial genomes. Trends in Ecology & amp; Evolution 14, 394–398. doi:10.1016/s0169-5347(99)01660-2

Dalton, D.L., Kotze, A., 2011. DNA barcoding as a tool for species identification in three forensic wildlife cases in South Africa. Forensic Science International 207, e51–e54. doi:10.1016/j.forsciint.2010.12.017

Davoren, J., Vanek, D., Konjhodzić, R., Crews, J., Huffine, E., Parsons, T.J., 2007. Highly Effective DNA Extraction Method for Nuclear Short Tandem Repeat Testing of Skeletal Remains from Mass Graves. Croatian Medical Journal 48, 478–485.

Dawnay, N., Ogden, R., McEwing, R., Carvalho, G.R., Thorpe, R.S., 2007. Validation of the barcoding gene COI for use in forensic genetic species identification. Forensic Science International 173, 1–6. doi:10.1016/j.forsciint.2006.09.013

Dawnay, N., Ogden, R., Thorpe, R.S., Pope, L.C., Dawson, D.A., McEwing, R., 2008. A forensic STR profiling system for the Eurasian badger: A framework for developing profiling systems for wildlife species. Forensic Science International: Genetics 2, 47–53. doi:10.1016/j.fsigen.2007.08.006

Dawnay, N., Ogden, R., Wetton, J.H., Thorpe, R.S., McEwing, R., 2009. Genetic data from 28 STR loci for forensic individual identification and parentage analyses in 6 bird of prey species. Forensic Science International: Genetics 3, e63–e69. doi:10.1016/j.fsigen.2008.07.001

de Bruin, D.D.S.H., Haagmans, M.A., van der Gaag, K.J., Hoogenboom, J., Weiler, N.E.C., Tesi, N., Salazar, A., Zhang, Y., Holstege, H., Reinders, M., M'charek, A.A., Sijen, T., Henneman, P., 2025. Exploring nanopore direct sequencing performance of forensic STRs, SNPs, InDels, and DNA methylation markers in a single assay. Forensic Science International: Genetics 74, 103154. doi:10.1016/j.fsigen.2024.103154

de Knijff, P., 2019. From next generation sequencing to now generation sequencing in forensics. Forensic Science International: Genetics 38, 175–180. doi:10.1016/j.fsigen.2018.10.017

de Queiroz, K., 2005. Ernst Mayr and the modern concept of species. Proceedings of the National Academy of Sciences 102, 6600–6607. doi:10.1073/pnas.0502030102

Deer Act, 1963, 1963.

Deer Act, 1991, 2020.

DeMaere, M.Z., Darling, A.E., 2021. QC3C: Reference-free quality control for hi-C sequencing data. PLOS Computational Biology 17. doi:10.1371/journal.pcbi.1008839

Di Giovanni, G., 2020. Fallow deer DNA marker research [WWW Document]. Australian Deer Association. URL https://www.austdeer.com.au/2020/05/19/1380477/fallow-deer-dna-marker-research (accessed 9.17.23).

DNeasy Blood & Tissue Handbook - QIAGEN [WWW Document], 2020. . Qiagen.com. URL https://www.qiagen.com/br/resources/resourcedetail?id=68f29296-5a9f-40fa-8b3d-1c148d0b3030&lang=en (accessed 7.23.20).

Doukakis, P., Pikitch, E.K., Rothschild, A., DeSalle, R., Amato, G., Kolokotronis, S.-O., 2012. Testing the Effectiveness of an International Conservation Agreement: Marketplace Forensics and CITES Caviar Trade Regulation. PLoS ONE 7, e40907. doi:10.1371/journal.pone.0040907

Eaton, M.J., Meyers, G.L., Kolokotronis, S.-O., Leslie, M.S., Martin, A.P., Amato, G., 2009. Barcoding bushmeat: molecular identification of Central African and South American harvested vertebrates. Conservation Genetics 11, 1389–1404. doi:10.1007/s10592-009-9967-0

Eichmann, C., Berger, B., Parson, W., 2004. A proposed nomenclature for 15 canine-specific polymorphic STR loci for forensic purposes. International Journal of Legal Medicine 118, 249–266. doi:10.1007/s00414-004-0452-5

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., Kuse, R., Lacroix, Y., Lin, S., Lundquist, P., Ma, 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. doi:10.1126/science.1162986

Eiken, H.G., Andreassen, R.J., Kopatz, A., Bjervamoen, S.G., Wartiainen, I., Tobiassen, C., Knappskog, P.M., Aspholm, P.E., Smith, M.E., Aspi, J., 2009. Population data for 12 STR loci in Northern European brown bear (Ursus arctos) and application of DNA profiles for forensic casework. Forensic Science International: Genetics Supplement Series 2, 273–274. doi:10.1016/j.fsigss.2009.07.007

Endrullat, C., Glökler, J., Franke, P., Frohme, M., 2016. Standardization and quality management in next-generation sequencing. Applied & Translational Genomics 10, 2–9. doi:10.1016/j.atg.2016.06.001

Evett, I.W., Jackson, G., Lambert, J.A., 2000. More on the hierarchy of propositions: exploring the distinction between explanations and propositions. Science & Justice 40, 3–10. doi:10.1016/s1355-0306(00)71926-5

Evett, I.W., Weir, B.S., 1998. Interpreting DNA evidence. Sinauer, Sunderland, Mass.

Excoffier, L., Lischer, H.E.L., 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Molecular Ecology Resources 10, 564–567. doi:10.1111/j.1755-0998.2010.02847.x

Faircloth, B.C., 2008. msatcommander: detection of microsatellite repeat arrays and automated, locus-specific primer design. Molecular Ecology Resources 8, 92–94. doi:10.1111/j.1471-8286.2007.01884.x

Fan, H., Chu, J.-Y., 2007. A brief review of short tandem repeat mutation. Genomics, Proteomics & Bioinformatics 5, 7–14. doi:10.1016/s1672-0229(07)60009-6

Farquharson, K.A., Hogg, C.J., Grueber, C.E., 2019. A case for genetic parentage assignment in captive group housing. Conservation Genetics 20, 1187–1193. doi:10.1007/s10592-019-01198-w

Felsenstein, J., 1985. Confidence limits on phylogenies: An approach using the bootstrap. Evolution 39, 783. doi:10.2307/2408678

Ferreira, M.R., Carratto, T.M., Frontanilla, T.S., Bonadio, R.S., Jain, M., de Oliveira, S.F., Castelli, E.C., Mendes-Junior, C.T., 2025. Advances in forensic genetics: Exploring the potential of long read sequencing. Forensic Science International: Genetics 74, 103156. doi:10.1016/j.fsigen.2024.103156

Filonzi, L., Chiesa, S., Vaghi, M., Nonnis Marzano, F., 2010. Molecular barcoding reveals mislabelling of commercial fish products in Italy. Food Research International 43, 1383–1388. doi:10.1016/j.foodres.2010.04.016

Finlay, J., Sutherland, N., Sturge, G., 2019. , Wildlife Crime Data Pack, Number: CDP 2019/0070. The House of Commons Library, London.

FitzSimmons, N., Moritz, C., Moore, S., 1995. Conservation and dynamics of microsatellite loci over 300 million years of marine turtle evolution. Molecular Biology and Evolution 12, 432–440. doi:10.1093/oxfordjournals.molbev.a040218

Fjelstrup, S., Andersen, M., Thomsen, J., Wang, J., Stougaard, M., Pedersen, F., Ho, Y.-P., Hede, M., Knudsen, B., 2017. The effects of dithiothreitol on DNA. Sensors 17, 1201. doi:10.3390/s17061201

Folmer, O., Black, M., Hoeh, w, Lutz, R., Vrijenhoek, R., 1994. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. Molecular Marine Biology and Biotechnology 3, 294–299.

Fox, J.C., Cave, C.A., Schumm, J.W., 2003. Development, Characterization, and Validation of a Sensitive Primate-Specific Quantification Assay for Forensic Analysis. BioTechniques 34, 314–322. doi:10.2144/03342rr02

Frank, K., Barta, E., Bana, N., Nagy, J., Horn, P., Orosz, L., Stéger, V., 2016. Complete mitochondrial genome sequence of a Hungarian red deer (Cervus Elaphus Hippelaphus) from high-throughput sequencing data and its phylogenetic position within the family Cervidae. Acta Biologica Hungarica 67, 133–147. doi:10.1556/018.67.2016.2.2

Fuentes-López, A., Ruiz, C., Galián, J., Romera, E., 2020. Molecular identification of forensically important fly species in Spain using COI barcodes. Science & Justice 60, 293–302. doi:10.1016/j.scijus.2019.12.003

Fuentes-Pardo, A.P., Ruzzante, D.E., 2017. Whole-genome sequencing approaches for conservation biology: Advantages, limitations and practical recommendations. Molecular Ecology 26, 5369–5406. doi:10.1111/mec.14264

Fungtammasan, A., Ananda, G., Hile, S.E., Su, M.S.-W., 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 Research 25, 736–749. doi:10.1101/gr.185892.114

Gandy, S., Kilbride, E., Biek, R., Millins, C., Gilbert, L., 2021. Experimental evidence for opposing effects of high deer density on tick-borne pathogen prevalence and hazard. Parasites & Vectors 14. doi:10.1186/s13071-021-05000-0

Gaur, A., Singh, C.S., Sreenivas, A., Singh, L., 2012. DNA-based identification of a snake in a wine bottle using universal primers: A case of mistaken identity. Forensic Science International 214, e51–e53. doi:10.1016/j.forsciint.2011.07.045

Gentra Puregene Blood Kit - QIAGEN Online Shop [WWW Document], 2010. . Qiagen.com. URL https://www.qiagen.com/gb/products/discovery-and-translational-research/dna-rna-purification/dna-purification/genomic-dna/gentra-puregene-blood-kit/?clear=true#orderinginformation (accessed 2.21.21).

Gettings, K.B., Aponte, R.A., Vallone, P.M., Butler, J.M., 2015. STR allele sequence variation: Current knowledge and future issues. Forensic Science International: Genetics 18, 118–130. doi:10.1016/j.fsigen.2015.06.005

Gettings, K.B., Bodner, M., Borsuk, L.A., King, J.L., Ballard, D., Parson, W., Benschop, C.C.G., Børsting, C., Budowle, B., Butler, J.M., van der Gaag, K.J., Gill, P., Gusmão, L., Hares, D.R., Hoogenboom, J., Irwin, J., Prieto, L., Schneider, P.M., Vennemann, M., Phillips, C., 2024. Recommendations of the DNA commission of the International Society for Forensic Genetics (ISFG) on short tandem repeat sequence nomenclature. Forensic Science International: Genetics 68, 102946. doi:10.1016/j.fsigen.2023.102946

Gettings, K.B., Borsuk, L.A., Ballard, D., Bodner, M., Budowle, B., Devesse, L., King, J., Parson, W., Phillips, C., Vallone, P.M., 2017. STRSeq: A catalog of sequence diversity at human identification Short Tandem Repeat loci. Forensic Science International: Genetics 31, 111–117. doi:10.1016/j.fsigen.2017.08.017

Ghemrawi, M., Fischinger, F., Duncan, G., Dukes, M.J., Guilliano, M., McCord, B., 2021. Developmental validation of speid: A pyrosequencing-based assay for species identification. Forensic Science International: Genetics 55, 102560. doi:10.1016/j.fsigen.2021.102560

Ghiani, M.E., Mameli, A., Vecchio, C., Francalacci, P., Robledo, R., Calò, C.M., 2021. Estimating population genetics and forensic efficiency of the globalfiler PCR amplification kit in the population of Sardinia (Italy). Gene 794, 145775. doi:10.1016/j.gene.2021.145775

Ghobrial, L., Lankester, F., Kiyang, J.A., Akih, A.E., de Vries, S., Fotso, R., Gadsby, E.L., Jenkins, P.D., Gonder, M.K., 2010. Tracing the origins of rescued chimpanzees reveals widespread chimpanzee hunting in Cameroon. BMC Ecology 10. doi:10.1186/1472-6785-10-2

Gill, P., 2001. Application of low copy number DNA profiling. Croatian Medical Journal 42, 229–232.

Gill, P., Kimpton, C., D'Aloja, E., Andersen, J.F., Bar, W., Brinkmann, B., Holgersson, S., Johnsson, V., Kloosterman, A.D., Lareu, M.V., Nellemann, L., Pfitzinger, H., Phillips, C.P., Schmitter, H., Schneider, P.M., Stenersen, M., 1994. Report of the European DNA profiling group (EDNAP) — towards standardisation of short tandem repeat (STR) loci. Forensic Science International 65, 51–59. doi:10.1016/0379-0738(94)90299-2

Gill, R.M.A., Morgan, G., 2009. The effects of varying deer density on natural regeneration in woodlands in lowland Britain. Forestry 83, 53–63. doi:10.1093/forestry/cpp031

Gomes, C., Martínez-Gómez, J., Díez-Juárez, L., Díaz-Sánchez, S., Palomo-Díez, S., Arroyo-Pardo, E., Cano-López, M., Fernández-Serrano, J., 2017. Prep-n-GoTM: A new and fast extraction method for forensic blood samples. Forensic Science International: Genetics Supplement Series 6, e265–e266. doi:10.1016/j.fsigss.2017.09.089

Goodwin, W., Linacre, A., Hadi, S., 2011. An introduction to forensic genetics, 2nd ed. John Wiley & Sons, New York, NY.

Gooley, R., Hogg, C.J., Belov, K., Grueber, C.E., 2017. No evidence of inbreeding depression in a Tasmanian devil insurance population despite significant variation in inbreeding. Scientific Reports 7. doi:10.1038/s41598-017-02000-y

Gortari, M.J. de, Freking, B.A., Kappes, S.M., Leymaster, K.A., Stone, R.T., Beattie, C.W., Crawford, A.M., 1997. Extensive genomic conservation of cattle microsatellite heterozygosity in sheep. Animal Genetics 28, 274–290. doi:10.1111/j.1365-2052.1997.00153.x

Gouy, A., Zieger, M., 2017. STRAF—a convenient online tool for STR data evaluation in forensic genetics. Forensic Science International: Genetics 30, 148–151. doi:10.1016/j.fsigen.2017.07.007

Govender, A., Groeneveld, J., Singh, S., Willows-Munro, S., 2019. The design and testing of mini-barcode markers in marine lobsters. PLOS ONE 14, e0210492. doi:10.1371/journal.pone.0210492

Grahn, R.A., Alhaddad, H., Alves, P.C., Randi, E., Waly, N.E., Lyons, L.A., 2015. Feline mitochondrial DNA sampling for forensic analysis: When enough is enough! Forensic Science International: Genetics 16, 52–57. doi:10.1016/j.fsigen.2014.11.017

Green, R.L., Roinestad, I.C., Boland, C., Hennessy, L.K., 2005. Developmental Validation of the QuantifilerTM Real-Time PCR Kits for the Quantification of

Herper, M., 2013. Four lessons from IBM and Roche's failed DNA collaboration [WWW Document]. Forbes. URL https://www.forbes.com/sites/matthewherper/2013/04/23/four-lessons-from-ibm-and-roches-failed-dna-collaboration/ (accessed 8.5.25).

Human Nuclear DNA Samples. Journal of Forensic Sciences 50, 1–17. doi:10.1520/jfs2004478

Grueber, C.E., Peel, E., Wright, B., Hogg, C.J., Belov, K., 2019. A Tasmanian devil breeding program to support wild recovery. Reproduction, Fertility and Development 31, 1296. doi:10.1071/rd18152

Gupta, S.K., Bhagavatula, J., Thangaraj, K., Singh, L., 2011. Establishing the identity of the massacred tigress in a case of wildlife crime. Forensic Science International: Genetics 5, 74–75. doi:10.1016/j.fsigen.2010.05.004

Gupta, S.K., Kumar, A., Hussain, S.A., Vipin, Singh, L., 2013. Cytochrome b based genetic differentiation of Indian wild pig (Sus scrofa cristatus) and domestic pig (Sus scrofa domestica) and its use in wildlife forensics. Science & Justice 53, 220–222. doi:10.1016/j.scijus.2012.09.005

Gurevich, A., Saveliev, V., Vyahhi, N., Tesler, G., 2013. Quast: Quality Assessment Tool for Genome Assemblies. Bioinformatics 29, 1072–1075. doi:10.1093/bioinformatics/btt086

Haas, B.J., Salzberg, S.L., Zhu, W., Pertea, M., Allen, J.E., Orvis, J., White, O., Buell, C.R., Wortman, J.R., 2008. Automated eukaryotic gene structure annotation using EVidenceModeler and the program to assemble spliced alignments. Genome Biology 9. doi:10.1186/gb-2008-9-1-r7

Hall, A., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symposium Series 41, 95–98.

Hamir, A.N., Greenlee, J.J., Nicholson, E.M., Kunkle, R.A., Richt, J.A., Miller, J.M., Hall, M., 2011. Experimental transmission of chronic wasting disease (CWD) from elk and white-tailed deer to fallow deer by intracerebral route: Final report. Canadian Journal of Veterinary Research 75, 152–156.

Hardwick, S.A., Deveson, I.W., Mercer, T.R., 2017. Reference standards for next-generation sequencing. Nature Reviews Genetics 18, 473–484. doi:10.1038/nrg.2017.44

Hardy, G.H., 1908. Mendelian proportions in a mixed population. Zeitschrift für Induktive Abstammungs- und Vererbungslehre 1, 395–395. doi:10.1007/bf01990610

Harper, C., Ludwig, A., Clarke, A., Makgopela, K., Yurchenko, A., Guthrie, A., Dobrynin, P., Tamazian, G., Emslie, R., van Heerden, M., Hofmeyr, M., Potter, R., Roets, J., Beytell, P., Otiende, M., Kariuki, L., du Toit, R., Anderson, N., Okori, J., Antonik, A., Koepfli, K.-P., Thompson, P., O'Brien, S.J., 2018. Robust forensic matching of confiscated horns to individual poached African rhinoceros. Current Biology 28, R13–R14. doi:10.1016/j.cub.2017.11.005

Harper, C.K., Vermeulen, G.J., Clarke, A.B., de Wet, J.I., Guthrie, A.J., 2013. Extraction of nuclear DNA from rhinoceros horn and characterization of DNA profiling systems for white (Ceratotherium simum) and black (Diceros bicornis) rhinoceros. Forensic Science International: Genetics 7, 428–433. doi:10.1016/j.fsigen.2013.04.003

Hartl, G.B., Schleger, A., Slowak, M., 2009. Genetic variability in fallow deer, *Dama dama* L. Animal Genetics 17, 335–341. doi:10.1111/j.1365-2052.1986.tb00726.x

Hassanin, A., Delsuc, F., Ropiquet, A., Hammer, C., Jansen van Vuuren, B., Matthee, C., Ruiz-Garcia, M., Catzeflis, F., Areskoug, V., Nguyen, T.T., Couloux, A., 2012. Pattern and timing of diversification of Cetartiodactyla (Mammalia, Laurasiatheria), as revealed by a comprehensive analysis of mitochondrial genomes. Comptes Rendus Biologies 335, 32–50. doi:10.1016/j.crvi.2011.11.002

Haynes, E., Jimenez, E., Pardo, M.A., Helyar, S.J., 2019. The future of NGS (next generation sequencing) analysis in Testing Food Authenticity. Food Control 101, 134–143. doi:10.1016/j.foodcont.2019.02.010

Henderson, A., Lee, C.M., Mistry, V., Thomas, M.D., Iyengar, A., 2012. Reliable and Robust Molecular Sexing of the Hen Harrier (Circus cyaneus) Using PCR-RFLP Analysis of the CHD 1 Gene. Journal of Forensic Sciences 58, 491–494. doi:10.1111/1556-4029.12030

Hogg, C.J., Ivy, J.A., Srb, C., Hockley, J., Lees, C., Hibbard, C., Jones, M., 2015. Influence of genetic provenance and birth origin on productivity of the Tasmanian

devil insurance population. Conservation Genetics 16, 1465–1473. doi:10.1007/s10592-015-0754-9

Hogg, C.J., Lee, A.V., Srb, C., Hibbard, C., 2016. Metapopulation management of an Endangered species with limited genetic diversity in the presence of disease: the Tasmanian devilSarcophilus harrisii. International Zoo Yearbook 51, 137–153. doi:10.1111/izy.12144

Hogg, C.J., McLennan, E.A., Wise, P., Lee, A.V., Pemberton, D., Fox, S., Belov, K., Grueber, C.E., 2020. Preserving the demographic and genetic integrity of a single source population during multiple translocations. Biological Conservation 241, 108318. doi:10.1016/j.biocon.2019.108318

Hohenlohe, P.A., McCallum, H.I., Jones, M.E., Lawrance, M.F., Hamede, R.K., Storfer, A., 2019. Conserving adaptive potential: lessons from Tasmanian devils and their transmissible cancer. Conservation Genetics 20, 81–87. doi:10.1007/s10592-019-01157-5

Hoogenboom, J., Sijen, T., van der Gaag, K.J., 2021. STRNaming: Generating simple, informative names for sequenced STR alleles in a standardised and automated manner. Forensic Science International: Genetics 52, 102473. doi:10.1016/j.fsigen.2021.102473

Hoogenboom, J., van der Gaag, K.J., de Leeuw, R.H., Sijen, T., de Knijff, P., Laros, J.F.J., 2017. FDSTools: A software package for analysis of massively parallel sequencing data with the ability to recognise and correct STR stutter and other PCR or sequencing noise. Forensic Science International: Genetics 27, 27–40. doi:10.1016/j.fsigen.2016.11.007

Hua, Y., 2019. Capreolus capreolus mitochondrion, complete genome. College of Wildlife and Protected Area, Northeast Forestry University, No. 26, Hexing Road, Xiangfang District, Harbin, Heilongjiang Province 150040, China. doi:https://www.ncbi.nlm.nih.gov/nuccore/MN485773.1

Huang, S., 2008. The genetic equidistance result of Molecular Evolution is independent of mutation rates. Journal of Computer Science & Systems Biology 01. doi:10.4172/jcsb.1000009

Hume, S., Nelson, T.N., Speevak, M., McCready, E., Agatep, R., Feilotter, H., Parboosingh, J., Stavropoulos, D.J., Taylor, S., Stockley, T.L., 2019. CCMG practice guideline: laboratory guidelines for next-generation sequencing. Journal of Medical Genetics 56, 792–800. doi:10.1136/jmedgenet-2019-106152

Huo, C., Bao, F., Long, H., Qin, T., Zhang, S., 2023. The complete mitochondrial genome of Wellcomia Compar (spirurina: Oxyuridae) and its genome characterization and phylogenetic analysis. Scientific Reports 13. doi:10.1038/s41598-023-41638-9

Illes, A., 2016. , Wildlife Crime In The UK. Policy Department A: Economic And Scientific Policy, London.

Illumina, 2011. Quality Scores for Next-Generation Sequencing [WWW Document]. Illumina.com. URL https://www.illumina.com/documents/products/technotes/technote_Q-Scores.pdf (accessed 20).

Illumina, 2015. De Novo Sequencing | Assemble novel genomes [WWW Document]. Emea.illumina.com. URL https://emea.illumina.com/techniques/sequencing/dna-sequencing/whole-genome-sequencing/de-novo-sequencing.html#:~:text=De%20novo%20sequencing%20refers%20to,of%20gaps%20in%20the%20data). (accessed 8.22.20).

The International Consortium on Combating Wildlife Crime (ICCWC), 2016. [WWW Document]. Cites.org. URL https://cites.org/eng/prog/iccwc.php (accessed 20/11.2020).

International Criminal Police Organisation (INTERPOL), 2018. Wildlife crime [WWW Document]. Interpol.int. URL https://www.interpol.int/en/Crimes/Environmental-crime/Wildlife-crime (accessed 5.11.20).

International Organisation for Standardisation, 2017. ISO/IEC 17025 — Testing and calibration laboratories [WWW Document]. ISO. URL https://www.iso.org/ISO-IEC-17025-testing-and-calibration-laboratories.html (accessed 5.13.20).

INTERPOL, 2023. Illegal wildlife trade has become one of the “world’s largest criminal activities” [WWW Document]. Illegal wildlife trade has become one of the ‘world’s largest criminal activities.’ URL https://www.interpol.int/en/News-and-Events/News/2023/Illegal-wildlife-trade-has-become-one-of-the-world-s-largest-criminal-activities#:~:text=With%20the%20black%20market%20for,other%20forms%20of%20organized%20crime. (accessed 12.6.24).

Iwasaki, W., Fukunaga, T., Isagozawa, R., Yamada, K., Maeda, Y., Satoh, T.P., Sado, T., Mabuchi, K., Takeshima, H., Miya, M., Nishida, M., 2013. Mitofish and MitoAnnotator: A mitochondrial genome database of fish with an accurate and automatic annotation pipeline. Molecular Biology and Evolution 30, 2531–2540. doi:10.1093/molbev/mst141

Iyavoo, S., Hadi, S., Goodwin, W., 2013. Evaluation of five DNA extraction systems for recovery of DNA from bone. Forensic Science International: Genetics Supplement Series 4, e174–e175. doi:10.1016/j.fsigss.2013.10.090

Iyengar, A., 2014. Forensic DNA analysis for animal protection and biodiversity conservation: A review. Journal for Nature Conservation 22, 195–205. doi:10.1016/j.jnc.2013.12.001

Jain, M., 2021. From kilobases to "Whales": A short history of ultra-long reads and high-throughput genome sequencing [WWW Document]. Oxford Nanopore Technologies. URL https://nanoporetech.com/blog/news-blog-kilobases-whales-short-history-ultra-long-reads-and-high-throughput-genome (accessed 7.1.25).

Jain, M., Fiddes, I.T., Miga, K.H., Olsen, H.E., Paten, B., Akeson, M., 2015. Improved data analysis for the MinION nanopore sequencer. Nature Methods 12, 351–356. doi:10.1038/nmeth.3290

Jan, C., Fumagalli, L., 2016. Polymorphic DNA microsatellite markers for forensic individual identification and parentage analyses of seven threatened species of parrots (family Psittacidae). PeerJ 4, e2416. doi:10.7717/peerj.2416

Jobin, R.M., Patterson, D., Zhang, Y., 2008. DNA typing in populations of mule deer for forensic use in the Province of Alberta. Forensic Science International: Genetics 2, 190–197. doi:10.1016/j.fsigen.2008.01.003

Johnson, R.N., Wilson-Wilde, L., Linacre, A., 2014. Current and future directions of DNA in wildlife forensic science. Forensic Science International: Genetics 10, 1–11. doi:10.1016/j.fsigen.2013.12.007

Joint Nature Conservation Committee, 2017. Deer [WWW Document]. Jncc.defra.gov.uk. URL http://jncc.defra.gov.uk/page-3206 (accessed 4.29.20).

Jones, M.E., Paetkau, D., Geffen, E., Moritz, C., 2003. Microsatellites for the Tasmanian devil (Sarcophilus laniarius). Molecular Ecology Notes 3, 277–279. doi:10.1046/j.1471-8286.2003.00425.x

Ju, Y., Liu, H., Rong, M., Yang, Y., Wei, H., Shao, Y., Chen, X., Xing, X., 2015. Complete mitochondrial genome sequence of Aoluguya Reindeer (rangifer tarandus). Mitochondrial DNA 1–2. doi:10.3109/19401736.2014.984171

Kanthaswamy, S., Tom, B.K., Mattila, A.-M., Johnston, E., Dayton, M., Kinaga, J., Joy-Alise Erickson, B., Halverson, J., Fantin, D., DeNise, S., Kou, A., Malladi, V., Satkoski, J., Budowle, B., Glenn Smith, D., Koskinen, M.T., 2009. Canine Population Data Generated from a Multiplex STR Kit for Use in Forensic Casework. Journal of Forensic Sciences 54, 829–840. doi:10.1111/j.1556-4029.2009.01080.x

Kardos, M., 2021. Conservation genetics. Current Biology 31. doi:10.1016/j.cub.2021.08.047

Kardos, M., Armstrong, E.E., Fitzpatrick, S.W., Hauser, S., Hedrick, P.W., Miller, J.M., Tallmon, D.A., Funk, W.C., 2021. The crucial role of genome-wide genetic variation in conservation. Proceedings of the National Academy of Sciences 118. doi:10.1073/pnas.2104642118

Katugin, O.N., Chichvarkhina, O.V., Zolotova, A.O., Chichvarkhin, A.Yu., 2015. DNA barcoding for squids of the family Gonatidae (Cephalopoda: Teuthida) from the boreal North Pacific. Mitochondrial DNA Part A 28, 41–49. doi:10.3109/19401736.2015.1110792

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*. Journal of Forensic Sciences 56, 1457–1463. doi:10.1111/j.1556-4029.2011.01871.x

Kim, H.-J., Hwang, J.-Y., Park, K.-J., Park, H.-C., Kang, H.-E., Park, J., Sohn, H.-J., 2020. The complete mitochondrial genome of Cervus canadensis (Erxleben, 1777), as a model species of chronic wasting disease (CWD). Mitochondrial DNA Part B 5, 2621–2623. doi:10.1080/23802359.2020.1780983

Kim, M.-J., Kim, H.-Y., 2019. A fast multiplex real-time PCR assay for simultaneous detection of pork, chicken, and beef in commercial processed meat products. LWT 114, 108390. doi:10.1016/j.lwt.2019.108390

King, J.L., Woerner, A.E., Mandape, S.N., Kapema, K.B., Moura-Neto, R.S., Silva, R., Budowle, B., 2021. Strait Razor Online: An enhanced user interface to facilitate interpretation of mps data. Forensic Science International: Genetics 52, 102463. doi:10.1016/j.fsigen.2021.102463

Kokotas, S., Budowle, B., Papatheodorou, A., Bolanaki, E., Kondili, A., Metheniti, A., Vouropoulou, M., Koukouvinos, G., Palaigeorgiou, E., Makras, P., 2025. Comparison of next generation sequencing (NGS) - (snps) and capillary electrophoresis (CE) - (strs) in the genetic analysis of human remains. Forensic Science International: Genetics 74, 103131. doi:10.1016/j.fsigen.2024.103131

Kolmogorov, M., Yuan, J., Lin, Y., Pevzner, P.A., 2019. Assembly of long, error-prone reads using repeat graphs. Nature Biotechnology 37, 540–546. doi:10.1038/s41587-019-0072-8

Kong, N., Ng, W., Cai, L., Leonardo, A., Weimer, B., Kelly, L., 2014. Integrating the DNA Integrity Number (DIN) to Assess Genomic DNA (gDNA) Quality Control Using the Agilent 2200 TapeStation System. Technical Note. doi:10.13140/RG.2.1.3616.8409

Koontz, D., Baecher, K., Amin, M., Nikolova, S., Gallagher, M., Dollard, S., 2015. Evaluation of DNA extraction methods for the detection of Cytomegalovirus in dried blood spots. Journal of Clinical Virology 66, 95–99. doi:10.1016/j.jcv.2015.03.015

Krems, M., Zwolak, M., Pershin, Y.V., Di Ventra, M., 2009. Effect of Noise on DNA Sequencing via Transverse Electronic Transport. Biophysical Journal 97, 1990–1996. doi:10.1016/j.bpj.2009.06.055

Krenke, B.E., Nassif, N., Sprecher, C.J., Knox, C., Schwandt, M., Storts, D.R., 2008. Developmental validation of a real-time PCR assay for the simultaneous quantification of total human and male DNA. Forensic Science International: Genetics 3, 14–21. doi:10.1016/j.fsigen.2008.07.004

Krivanek, O.L., Chisholm, M.F., Nicolosi, V., Pennycook, T.J., Corbin, G.J., Dellby, N., Murfitt, M.F., Own, C.S., Szilagyi, Z.S., Oxley, M.P., Pantelides, S.T., Pennycook, S.J., 2010. Atom-by-atom structural and chemical analysis by annular dark-field electron microscopy. Nature 464, 571–574. doi:10.1038/nature08879

Krothapalli, S., May, M.K., Hestekin, C.N., 2012. Capillary electrophoresis-single strand conformation polymorphism for the detection of multiple mutations leading to tuberculosis drug resistance. Journal of Microbiological Methods 91, 147–154. doi:10.1016/j.mimet.2012.07.021

Kumar Jha, D., Kumar Gupta, S., Kshetry, N.T., Panday, R., Pokharel, B.R., 2016. A Pioneer Case Study on Identification of Infant Rhinoceros Horn. Journal of Forensic Research 08. doi:10.4172/2157-7145.1000374

Kumar, S., Stecher, G., Suleski, M., Sanderford, M., Sharma, S., Tamura, K., 2024. MEGA12: Molecular evolutionary genetic analysis version 12 for adaptive and Green Computing. Molecular Biology and Evolution 41. doi:10.1093/molbev/msae263

Kumar, S., Stecher, G., Tamura, K., 2016. Mega7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. Molecular Biology and Evolution 33, 1870–1874. doi:10.1093/molbev/msw054

Küpper, C., Burke, T., Székely, T., Dawson, D.A., 2008. Enhanced cross-species utility of conserved microsatellite markers in shorebirds. BMC Genomics 9, 502. doi:10.1186/1471-2164-9-502

Laman Web Rasmi Jabatan Perlindungan Hidupan Liar & Taman Negara Semenanjung Malaysia [WWW Document], 2015. . wildlife.gov.my. URL https://www.wildlife.gov.my/ (accessed 5.13.20).

Lee, E., Lee, Y., Moon, S., Kim, N., Kim, S., Yang, M., Choi, D., Han, M., 2013. The identification of elephant ivory evidences of illegal trade with mitochondrial cytochrome b gene and hypervariable D-loop region. Journal of Forensic and Legal Medicine 20, 174–178. doi:10.1016/j.jflm.2012.06.014

Lee, J.C.-I., Tsai, L.-C., Kuan, Y.-Y., Chien, W.-H., Chang, K.-T., Wu, C.-H., Linacre, A., Hsieh, H.-M., 2007. Racing pigeon identification using STR and chromo-helicase DNA binding gene markers. ELECTROPHORESIS 28, 4274–4281. doi:10.1002/elps.200700063

Lee, S.B., McCord, B., Buel, E., 2014. Advances in forensic DNA quantification: A review. ELECTROPHORESIS 35, 3044–3052. doi:10.1002/elps.201400187

Lee, W.T., Sun, X., Tsai, T.-S., Johnson, J.L., Gould, J.A., Garama, D.J., Gough, D.J., McKenzie, M., Trounce, I.A., St. John, J.C., 2017. Mitochondrial DNA haplotypes induce differential patterns of DNA methylation that result in differential chromosomal gene expression patterns. Cell Death Discovery 3. doi:10.1038/cddiscovery.2017.62

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. doi:10.1126/science.1079700

Lewin, H.A., Robinson, G.E., Kress, W.J., Baker, W.J., Coddington, J., Crandall, K.A., Durbin, R., Edwards, S.V., Forest, F., Gilbert, M.T.P., Goldstein, M.M., Grigoriev, I.V., Hackett, K.J., Haussler, D., Jarvis, E.D., Johnson, W.E.,

Patrinos, A., Richards, S., Castilla-Rubio, J.C., van Sluys, M.-A., Soltis, P.S., Xu, X., Yang, H., Zhang, G., 2018. Earth BioGenome Project: Sequencing life for the future of life. Proceedings of the National Academy of Sciences 115, 4325–4333. doi:10.1073/pnas.1720115115

Li, Z., Lin, Z., Ba, H., Chen, L., Yang, Y., Wang, K., Qiu, Q., Wang, W., Li, G., 2017. Draft genome of the reindeer (rangifer tarandus). GigaScience 6. doi:10.1093/gigascience/gix102

Li, Z., Zhang, Z., Mi, S., Wu, J., Xu, T., Liu, Z., Teng, L., 2020. The complete mitochondrial genome of water deer in Liaoning, China. Mitochondrial DNA Part B 5, 922–923. doi:10.1080/23802359.2020.1719936

Linacre, A., Gusmão, L., Hecht, W., Hellmann, A.P., Mayr, W.R., Parson, W., Prinz, M., Schneider, P.M., Morling, N., 2011. ISFG: Recommendations regarding the use of non-human (animal) DNA in forensic genetic investigations. Forensic Science International: Genetics 5, 501–505. doi:10.1016/j.fsigen.2010.10.017

Linacre, A., Tobe, S.S., 2008. On the trial of tigers–tracking tiger in Traditional East Asian Medicine. Forensic Science International: Genetics Supplement Series 1, 603–604. doi:10.1016/j.fsigss.2007.10.112

Linacre, A., Tobe, S.S., 2011. An overview to the investigative approach to species testing in wildlife forensic science. Investigative Genetics 2, 2. doi:10.1186/2041-2223-2-2

Lister, A.M., Edwards, C.J., Nock, D.A., Bunce, M., van Pijlen, I.A., Bradley, D.G., Thomas, M.G., Barnes, I., 2005. The phylogenetic position of the 'Giant Deer' Megaloceros giganteus. Nature 438, 850–853. doi:10.1038/nature04134

Lively, C.M., 2010. The effect of host genetic diversity on disease spread. The American Naturalist 175. doi:10.1086/652430

London, E.W., Roca, A.L., Novakofski, J.E., Mateus-Pinilla, N.E., 2022. A de novo chromosome-level genome assembly of the white-tailed deer, odocoileus virginianus. Journal of Heredity 113, 479–489. doi:10.1093/jhered/esac022

Lopez-Oceja, A., Lekube, X., Ruiz, L., Mujika-Alustiza, J.A., De Pancorbo, M.M., 2019. CYT B L15601 and H15748 primers for genetic identification of cetacean species. Forensic Science International: Genetics Supplement Series 7, 771–772. doi:10.1016/j.fsigss.2019.10.171

Lorenzini, R., 2005. DNA forensics and the poaching of wildlife in Italy: A case study. Forensic Science International 153, 218–221. doi:10.1016/j.forsciint.2005.04.032

Lorenzini, R., Cabras, P., Fanelli, R., Carboni, G.L., 2011. Wildlife molecular forensics: Identification of the Sardinian mouflon using STR profiling and the Bayesian assignment test. Forensic Science International: Genetics 5, 345–349. doi:10.1016/j.fsigen.2011.01.012

Lowe, T.M., Eddy, S.R., 1997. TRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Research 25, 955–964. doi:10.1093/nar/25.5.955

Luan, B., Peng, H., Polonsky, S., Rossnagel, S., Stolovitzky, G., Martyna, G., 2010. Base-By-Base Ratcheting of Single Stranded DNA through a Solid-State Nanopore. Physical Review Letters 104. doi:10.1103/physrevlett.104.238103

Ludeman, M.J., Zhong, C., Mulero, J.J., Lagacé, R.E., Hennessy, L.K., Short, M.L., Wang, D.Y., 2018. Developmental validation of GlobalFilerTM PCR amplification kit: a 6-dye multiplex assay designed for amplification of casework samples. International Journal of Legal Medicine 132, 1555–1573. doi:10.1007/s00414-018-1817-5

Ludwig, A., Vernesi, C., Lieckfeldt, D., Lattenkamp, E.Z., Wiethölter, A., Lutz, W., 2011. Origin and patterns of genetic diversity of German fallow deer as inferred from mitochondrial DNA. European Journal of Wildlife Research 58, 495–501. doi:10.1007/s10344-011-0571-5

Lundquist, P.M., Zhong, C.F., Zhao, P., Tomaney, A.B., Peluso, P.S., Dixon, J., Bettman, B., Lacroix, Y., Kwo, D.P., McCullough, E., Maxham, M., Hester, K., McNitt, P., Grey, D.M., Henriquez, C., Foquet, M., Turner, S.W., Zaccarin, D., 2008. Parallel confocal detection of single molecules in real time. Optics Letters 33, 1026. doi:10.1364/ol.33.001026

MacKenzie, S., Jentoft, S., Xu, P., Jiang, Y., Xu, J., Li, J., Sun, X., 2016. Genomics in the common carp, in: Genomics in Aquaculture. Academic Press, pp. 247–274.

Magoč, T., Salzberg, S.L., 2011. Flash: Fast length adjustment of short reads to improve genome assemblies. Bioinformatics 27, 2957–2963. doi:10.1093/bioinformatics/btr507

Mahamdallie, S., Ruark, E., Yost, S., Münz, M., Renwick, A., Poyastro-Pearson, E., Strydom, A., Seal, S., Rahman, N., 2018. The Quality Sequencing Minimum (QSM): providing comprehensive, consistent, transparent next generation sequencing data quality assurance. Wellcome Open Research 3, 37. doi:10.12688/wellcomeopenres.14307.1

Majoros, W.H., Pertea, M., Salzberg, S.L., 2004. Tigrscan and GlimmerHMM: Two open source ab initio eukaryotic gene-finders. Bioinformatics 20, 2878–2879. doi:10.1093/bioinformatics/bth315

Mandrekar, M., Erickson, A., Kopp, K., Krenke, B., Mandrekar, P., Nelson, R., Peterson, K., Shultz, J., Tereba, A., Westphal, N., 2001. Development of a human DNA quantitation system. Croatian Medical Journal 42, 336–339.

Manni, M., Berkeley, M.R., Seppey, M., Simão, F.A., Zdobnov, E.M., 2021. Busco update: Novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. Molecular Biology and Evolution 38, 4647–4654. doi:10.1093/molbev/msab199

Mantelatto, F.L., Carvalho, F.L., Simões, S.M., Negri, M., Souza-Carvalho, E.A., Terossi, M., 2016. New primers for amplification of cytochrome c oxidase subunit I barcode region designed for species of Decapoda (Crustacea). Nauplius 24, 1–4. doi:10.1590/2358-2936e2016030

Margaret, M., M.,.S. and R., Ewing, Jonelle, M., Thompson, Robert, S., McLaren, Storts, D.R., 2014. The PowerQuantTM System: A New Quantification Assay for Determining DNA Concentration and Quality [WWW Document]. Promega.co.uk. URL https://www.promega.co.uk/resources/profiles-in-dna/2014/the-powerquant-system-a-new-quantification-assay-for-determining-dna-concentration-and-quality/ (accessed 7.24.20).

Margres, M.J., Jones, M.E., Epstein, B., Kerlin, D.H., Comte, S., Fox, S., Fraik, A.K., Hendricks, S.A., Huxtable, S., Lachish, S., Lazenby, B., O'Rourke, S.M., Stahlke, A.R., Wiench, C.G., Hamede, R., Schönfeld, B., McCallum, H., Miller, M.R., Hohenlohe, P.A., Storfer, A., 2018. Large-effect loci affect survival in Tasmanian devils (Sarcophilus harrisii) infected with a transmissible cancer. Molecular Ecology 27, 4189–4199. doi:10.1111/mec.14853

Marín, J.C., Saucedo, C.E., Corti, P., González, B.A., 2009. Application of DNA Forensic Techniques for Identifying Poached Guanacos (Lama guanicoe) in Chilean Patagonia. Journal of Forensic Sciences 54, 1073–1076. doi:10.1111/j.1556-4029.2009.01087.x

Masseti, M., Mertzanidou, D., 2008. Dama dama. The IUCN Red List of Threatened Species 2008. doi:10.2305/iucn.uk.2008.rlts.t42188a10656554.en

Masseti, M., Pecchioli, E., Vernesi, C., 2008. Phylogeography of the last surviving populations of Rhodian and Anatolian fallow deer (*Dama dama Dama* L., 1758). Biological Journal of the Linnean Society 93, 835–844. doi:10.1111/j.1095-8312.2007.00951.x

McNeish, H., 2014. $213bn illegal wildlife and charcoal trade "funding global terror groups" [WWW Document]. Our World. URL https://ourworld.unu.edu/en/213bn-illegal-wildlife-and-charcoal-trade-funding-global-terror-groups (accessed 12.6.24).

Meiklejohn, K.A., Burnham-Curtis, M.K., Straughan, D.J., Giles, J., Moore, M.K., 2021. Current methods, future directions and considerations of DNA-based taxonomic identification in wildlife forensics. Forensic Science International: Animals and Environments 1, 100030. doi:10.1016/j.fsiae.2021.100030

Menotti-Raymond, M., David, V.A., Agarwala, R., Schäffer, A.A., Stephens, R., O'Brien, S.J., Murphy, W.J., 2003. Radiation hybrid mapping of 304 novel microsatellites in the domestic cat genome. Cytogenetic and Genome Research 102, 272–276. doi:10.1159/000075762

Menotti-Raymond, M.A., David, V.A., Wachter, L.L., Butler, J.M., O'Brien, S.J., 2005. An STR Forensic Typing System for Genetic Individualization of Domestic Cat (Felis catus) Samples. Journal of Forensic Sciences 50, 1–10. doi:10.1520/jfs2004317

Meredith, E.P., Adkins, J.K., Rodzen, J.A., 2020. UrsaPlex: An STR multiplex for forensic identification of North American black bear (Ursus americanus). Forensic Science International: Genetics 44, 102161. doi:10.1016/j.fsigen.2019.102161

Merkel, A., Gemmell, N., 2008. Detecting short tandem repeats from genome data: opening the software black box. Briefings in Bioinformatics 9, 355–366. doi:10.1093/bib/bbn028

Miller, Holly, 2013. Dama International- The Fallow Deer Project [WWW Document]. Britgeopeople.blogspot.com. URL https://britgeopeople.blogspot.com/2013/04/dama-international-fallow-deer-project.html (accessed 6.24.20).

Miller, W., Hayes, V.M., Ratan, A., Petersen, D.C., Wittekindt, N.E., Miller, J., Walenz, B., Knight, J., Qi, J., Zhao, F., Wang, Q., Bedoya-Reina, O.C., Katiyar, N., Tomsho, L.P., Kasson, L.M., Hardie, R.-A., Woodbridge, P., Tindall, E.A., Bertelsen, M.F., Dixon, D., Pyecroft, S., Helgen, K.M., Lesk, A.M., Pringle, T.H., Patterson, N., Zhang, Y., Kreiss, A., Woods, G.M., Jones, M.E., Schuster, S.C., 2011. Genetic diversity and population structure of the endangered marsupial Sarcophilus harrisii (Tasmanian devil). Proceedings of the National Academy of Sciences 108, 12348–12353. doi:10.1073/pnas.1102838108

Milner, J.M., Bonenfant, C., Mysterud, A., Gaillard, J.-M., Csanyi, S., Stenseth, N.CHR., 2006. Temporal and spatial development of red deer harvesting in Europe: biological and cultural factors. Journal of Applied Ecology 43, 721–734. doi:10.1111/j.1365-2664.2006.01183.x

Mondol, S., Sridhar, V., Yadav, P., Gubbi, S., Ramakrishnan, U., 2014. Tracing the geographic origin of traded leopard body parts in the indian subcontinent with DNA-based assignment tests. Conservation Biology 29, 556–564. doi:10.1111/cobi.12393

Moore, S.S., Sargeant, L.L., King, T.J., Mattick, J.S., Georges, M., Hetzel, D.J.S., 1991. The conservation of dinucleotide microsatellites among mammalian genomes allows the use of heterologous PCR primer pairs in closely related species. Genomics 10, 654–660. doi:10.1016/0888-7543(91)90448-n

Morey, M., Fernández-Marmiesse, A., Castiñeiras, D., Fraga, J.M., Couce, M.L., Cocho, J.A., 2013. A glimpse into past, present, and future DNA sequencing. Molecular Genetics and Metabolism 110, 3–24. doi:10.1016/j.ymgme.2013.04.024

Morf, N.V., Kopps, A.M., Nater, A., Lendvay, B., Vasiljevic, N., Webster, L.M.I., Fautley, R.G., Ogden, R., Kratzer, A., 2021. Stroe Deer: A validated forensic STR profiling system for the European roe deer (capreolus capreolus). Forensic Science International: Animals and Environments 1, 100023. doi:10.1016/j.fsiae.2021.100023

Morling, N., Allen, R.W., Carracedo, A., Geada, H., Guidet, F., Hallenberg, C., Martin, W., Mayr, W.R., Olaisen, B., Pascali, V.L., Schneider, P.M., 2002. Paternity Testing Commission of the International Society of Forensic Genetics: recommendations on genetic investigations in paternity cases. Forensic Science International 129, 148–157. doi:10.1016/s0379-0738(02)00289-x

Mucci, N., Mengoni, C., Randi, E., 2014. Wildlife DNA forensics against crime: Resolution of a case of tortoise theft. Forensic Science International: Genetics 8, 200–202. doi:10.1016/j.fsigen.2013.10.001

Mudd, A.B., Bredeson, J.V., Baum, R., Hockemeyer, D., Rokhsar, D.S., 2019. . Muntjac chromosome evolution and architecture. doi:10.1101/772343

Murchison, E.P., Schulz-Trieglaff, O.B., Ning, Z., Alexandrov, L.B., Bauer, M.J., Fu, B., Hims, M., Ding, Z., Ivakhno, S., Stewart, C., Ng, B.L., Wong, W., Aken, B., White, S., Alsop, A., Becq, J., Bignell, G.R., Cheetham, R.K., Cheng, W., Connor, T.R., Cox, A.J., Feng, Z.-P., Gu, Y., Grocock, R.J., Harris, S.R., Khrebtukova, I., Kingsbury, Z., Kowarsky, M., Kreiss, A., Luo, S., Marshall, J., McBride, D.J., Murray, L., Pearse, A.-M., Raine, K., Rasolonjatovo, I., Shaw, R., Tedder, P., Tregidgo, C., Vilella, A.J., Wedge, D.C., Woods, G.M., Gormley, N., Humphray, S., Schroth, G., Smith, G., Hall, K., Searle, S.M.J., Carter, N.P., Papenfuss, A.T., Futreal, P.A., Campbell, P.J., Yang, F., Bentley, D.R., Evers, D.J., Stratton, M.R., 2012. Genome Sequencing and Analysis of the Tasmanian Devil and Its Transmissible Cancer. Cell 148, 780–791. doi:10.1016/j.cell.2011.11.065

Murugaiah, C., Noor, Z.M., Mastakim, M., Bilung, L.M., Selamat, J., Radu, S., 2009. Meat species identification and Halal authentication analysis using mitochondrial DNA. Meat Science 83, 57–61. doi:10.1016/j.meatsci.2009.03.015

Mármol, P., Gómez, B., Gomes, C., Romero, C., Baeza-Richer, C., Hernández-Cordero, A., Martín-Arrebola, M., Rosell-Herrera, R., López-Parra, A.M., Palomo-Díez, S., Labajo-González, E., Perea-Pérez, B., López-Matayoshi, C., Arroyo-Pardo, E., 2019. An innovative DNA extraction method: Water versus commercial buffers. Forensic Science International: Genetics Supplement Series 7, 282–284. doi:10.1016/j.fsigss.2019.09.111

Nagarajan, R.P., Sanders, L., Kolm, N., Perez, A., Senegal, T., Mahardja, B., Baerwald, M.R., Schreier, A.D., 2024. CRISPR-based environmental DNA detection for a rare endangered estuarine species. Environmental DNA 6. doi:10.1002/edn3.506

National Research Council, 1996. The evaluation of forensic DNA evidence. National Academy Press, Washington, D.C.

Nei, M., Kumar, S., 2000. . Molecular evolution and phylogenetics. doi:10.1093/oso/9780195135848.001.0001

Neme, L.A., Leakey, R., 2014. Animal investigators, 1st ed. Scribner, New York.

Nicklas, J.A., Noreault-Conti, T., Buel, E., 2011. Development of a Real-Time Method to Detect DNA Degradation in Forensic Samples*. Journal of Forensic Sciences 57, 466–471. doi:10.1111/j.1556-4029.2011.02001.x

Ogden, R., 2010. Forensic science, genetics and wildlife biology: getting the right mix for a wildlife DNA forensics lab. Forensic Science, Medicine, and Pathology 6, 172–179. doi:10.1007/s12024-010-9178-5

Ogden, R., Dawnay, N., McEwing, R., 2009. Wildlife DNA forensics—bridging the gap between conservation genetics and law enforcement. Endangered Species Research 9, 179–195. doi:10.3354/esr00144

Ogden, R., Linacre, A., 2015. Wildlife forensic science: A review of genetic geographic origin assignment. Forensic Science International: Genetics 18, 152–159. doi:10.1016/j.fsigen.2015.02.008

Okimoto, R., Wolstenholme, D.R., 1990. A set of trnas that lack either the T psi c arm or the dihydrouridine arm: Towards a minimal trna adaptor. The EMBO Journal 9, 3405–3411. doi:10.1002/j.1460-2075.1990.tb07542.x

Ottolini, B., Lall, G.M., Sacchini, F., Jobling, M.A., Wetton, J.H., 2017. Application of a mitochondrial DNA control region frequency database for UK domestic cats. Forensic Science International: Genetics 27, 149–155. doi:10.1016/j.fsigen.2016.12.008

Oxford Nanopore Technologies, 2017. Nanopore sequencing: The advantages of long reads for genome assembly [WWW Document]. Nanoporetech.com. URL https://nanoporetech.com/sites/default/files/s3/white-papers/WGS_Assembly_white_paper.pdf?submissionGuid=40a7546b-9e51-42e7-bde9-b5ddef3c3512#:~:text=Long%20read%20lengths%20are%20more,to%20span%20repetitive%20genomic%20regions. (accessed 20).

O'Brien, S.J., Womack, J.E., Lyons, L.A., Moore, K.J., Jenkins, N.A., Copeland, N.G., 1993. Anchored reference loci for comparative genome mapping in mammals. Nature Genetics 3, 103–112. doi:10.1038/ng0293-103

O'Mahony, P.J., 2013. Finding horse meat in beef products--a global problem. QJM 106, 595–597. doi:10.1093/qjmed/hct087

PacBio, 2022. Whole genome sequencing [WWW Document]. PacBio. URL https://www.pacb.com/products-and-services/applications/whole-genome-sequencing/ (accessed 4.12.22).

Pang, H., Pang, H., Liu, W., Pang, H., Liu, W., Chen, Y., Fang, L., Zhang, X., Cao, X., 2008. Identification of complete mitochondrial genome of the tufted deer. DNA Sequence 19, 411–417. doi:10.1080/19401730802389517

Pangau-Adam, M., Flassy, M., Trei, J., Waltert, M., Soofi, M., 2022. The role of the introduced Rusa Deer (cervus timorensis) for wildlife hunting in West Papua, Indonesia. Ecological Solutions and Evidence 3. doi:10.1002/2688-8319.12118

Parkes, C., Thornley, J., 2008. Deer: Law and Liabilities, 2nd ed. Quiller Press, Shrewsbury.

Parson, W., Ballard, D., Budowle, B., Butler, J.M., Gettings, K.B., Gill, P., Gusmão, L., Hares, D.R., Irwin, J.A., King, J.L., Knijff, P. de, Morling, N., Prinz, M., Schneider, P.M., Neste, C.V., Willuweit, S., Phillips, C., 2016. Massively parallel sequencing of forensic STRs: Considerations of the DNA commission of the International Society for Forensic Genetics (ISFG) on minimal nomenclature requirements. Forensic Science International: Genetics 22, 54–63. doi:10.1016/j.fsigen.2016.01.009

Peakall, R., Smouse, P.E., 2012. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research--an update. Bioinformatics 28, 2537–2539. doi:10.1093/bioinformatics/bts460

Peakall, R., Smouse, Peter.E., 2006. genalex 6: genetic analysis in Excel. Population genetic software for teaching and research. Molecular Ecology Notes 6, 288–295. doi:10.1111/j.1471-8286.2005.01155.x

Pemberton, J., Johnston, S.E., Fletcher, T.J., Darwin Tree of Life Barcoding collective, Wellcome Sanger Institute Tree of Life programme, Wellcome Sanger Institute Scientific Operations: DNA Pipelines collective, Tree of Life Core Informatics collective, Consortium, D.T. of L., 2021. The genome sequence of the red deer, Cervus Elaphus Linnaeus 1758 [WWW Document]. Wellcome Open Research. URL https://wellcomeopenresearch.org/articles/6-336 (accessed 9.22.22).

Pemberton, J.M., Smith, R.H., 1985. Lack of biochemical polymorphism in British fallow deer. Heredity 55, 199–207. doi:10.1038/hdy.1985.92

Peng, Q., Vijaya Satya, R., Lewis, M., Randad, P., Wang, Y., 2015. Reducing amplification artifacts in high multiplex amplicon sequencing by using molecular barcodes. BMC Genomics 16. doi:10.1186/s12864-015-1806-8

Peppin, L., McEwing, R., Carvalho, G.R., Ogden, R., 2008. A DNA-Based Approach for the Forensic Identification of Asiatic Black Bear (Ursus thibetanus) in a Traditional Asian Medicine*. Journal of Forensic Sciences 53, 1358–1362. doi:10.1111/j.1556-4029.2008.00857.x

Pettersson, E., Lundeberg, J., Ahmadian, A., 2009. Generations of sequencing technologies. Genomics 93, 105–111. doi:10.1016/j.ygeno.2008.10.003

Phillips, C., Gettings, K.B., King, J.L., Ballard, D., Bodner, M., Borsuk, L., Parson, W., 2018. "The devil's in the detail": Release of an expanded, enhanced and dynamically revised forensic STR Sequence Guide. Forensic Science International: Genetics 34, 162–169. doi:10.1016/j.fsigen.2018.02.017

Pineda, G.M., Montgomery, A.H., Thompson, R., Indest, B., Carroll, M., Sinha, S.K., 2014. Development and validation of InnoQuantTM, a sensitive human DNA quantitation and degradation assessment method for forensic samples using high copy number mobile elements Alu and SVA. Forensic Science International: Genetics 13, 224–235. doi:10.1016/j.fsigen.2014.08.007

Pink, G., White, R., Scanlon, J.E., Farroway, L., 2016. Organisational Consortiums: The International Consortium on Combating Wildlife Crime (ICCWC), in: Environmental Crime and Collaborative State Intervention. Palgrave Studies in Green Criminology. Palgrave Macmillan, London, pp. 77–99.

Poetsch, M., Seefeldt, S., Maschke, M., Lignitz, E., 2001. Analysis of microsatellite polymorphism in red deer, roe deer, and fallow deer — possible employment in forensic applications. Forensic Science International 116, 1–8. doi:10.1016/s0379-0738(00)00337-6

Potts, J.M., Beeton, N.J., Bowman, D.M., Williamson, G.J., Lefroy, E.C., Johnson, C.N., 2014. Predicting the future range and abundance of fallow deer in Tasmania, Australia. Wildlife Research 41, 633. doi:10.1071/wr13206

Pray, L., 2008. Eukaryotic Genome Complexity. Nature Education, 96 1.

Primrose, S., Woolfe, M., Rollinson, S., 2010. Food forensics: methods for determining the authenticity of foodstuffs. Trends in Food Science & Technology 21, 582–590. doi:10.1016/j.tifs.2010.09.006

Publications [WWW Document], 2020. . Swgdam.org/publications. URL https://www.swgdam.org/publications (accessed 6.29.20).

Pye, R., Patchett, A., McLennan, E., Thomson, R., Carver, S., Fox, S., Pemberton, D., Kreiss, A., Baz Morelli, A., Silva, A., Pearse, M.J., Corcoran, L.M., Belov, K., Hogg, C.J., Woods, G.M., Lyons, A.B., 2018. Immunization Strategies Producing a Humoral IgG Immune Response against Devil Facial Tumor Disease in the Majority of Tasmanian Devils Destined for Wild Release. Frontiers in Immunology 9. doi:10.3389/fimmu.2018.00259

Ramsey, K., 2014. Wildlife Crime: A guide to the use of forensic and specialist techniques in the investigation of wildlife crime. United Kingdom: Forensic Working Group (FWG) pp.46-47.

Randi, E., Apollonio, M., 1988. Low biochemical variability in European fallow deer (*Dama dama* L.): natural bottlenecks and the effects of domestication. Heredity 61, 405–410. doi:10.1038/hdy.1988.131

Reiner, G., Weber, T., Nietfeld, F., Fischer, D., Wurmser, C., Fries, R., Willems, H., 2020. A genome-wide scan study identifies a single nucleotide substitution in MC1R gene associated with white coat colour in fallow deer (*Dama dama*). BMC Genetics 21. doi:10.1186/s12863-020-00950-3

Rey-Iglesia, A., Lister, A.M., Campos, P.F., Brace, S., Mattiangeli, V., Daly, K.G., Teasdale, M.D., Bradley, D.G., Barnes, I., Hansen, A.J., 2021. Exploring the phylogeography and population dynamics of the giant deer (megaloceros giganteus) using late quaternary mitogenomes. Proceedings of the Royal Society B: Biological Sciences 288. doi:10.1098/rspb.2020.1864

RhODIS™ [WWW Document], 2010. . Erhodis.org. URL https://erhodis.org/ (accessed 5.13.20).

Robin, J.D., Ludlow, A.T., LaRanger, R., Wright, W.E., Shay, J.W., 2016. Comparison of DNA Quantification Methods for Next Generation Sequencing. Scientific Reports 6. doi:10.1038/srep24067

Rodrigues Filho, L.F. da S., Feitosa, L.M., Silva Nunes, J.L., Onodera Palmeira, A.R., Martins, A.P.B., Giarrizzo, T., Carvalho-Costa, L.F., Monteiro, I.L.P., Gemaque, R., Gomes, F., Souza, R.F.C., Sampaio, I., Sales, J.B. de L., 2020. Molecular identification of ray species traded along the Brazilian Amazon coast. Fisheries Research 223, 105407. doi:10.1016/j.fishres.2019.105407

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M.A., Huelsenbeck, J.P., 2012. MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. Systematic Biology 61, 539–542. doi:10.1093/sysbio/sys029

The Royal Parks, 2023. Deer in richmond park [WWW Document]. Deer in the Richmond Park. URL https://www.royalparks.org.uk/visit/parks/richmond-park/deer-richmond-park (accessed 11.22.23).

RSPCA, 2020. . Advice and welfare/wildlife/deer. URL https://www.rspca.org.uk/adviceandwelfare/wildlife/deer (accessed 1.14.26).

Saitou, N., Nei, M., 1987. The neighbor-joining method: A new method for reconstructing phylogenetic trees. Molecular Biology and Evolution 4, 406–425. doi:10.1093/oxfordjournals.molbev.a040454

Sambrook, J., Fritsch, E.F., Maniatis, T., 1989. Molecular cloning, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

Sanches, A., Perez, W.A.M., Figueiredo, M.G., Rossini, B.C., Cervini, M., Galetti, P.M., Galetti, M., 2010. Wildlife forensic DNA and lowland tapir (Tapirus terrestris) poaching. Conservation Genetics Resources 3, 189–193. doi:10.1007/s12686-010-9318-y

Sanders, J.G., Cribbs, J.E., Fienberg, H.G., Hulburd, G.C., Katz, L.S., Palumbi, S.R., 2008. The tip of the tail: molecular identification of seahorses for sale in apothecary shops and curio stores in California. Conservation Genetics 9, 65–71. doi:10.1007/s10592-007-9308-0

Sanger, F., Nicklen, S., Coulson, A.R., 1977. DNA sequencing with chain-terminating inhibitors. Proceedings of the National Academy of Sciences 74, 5463–5467. doi:10.1073/pnas.74.12.5463

Sarvani, R.K., Parmar, D.R., Tabasum, W., Thota, N., Sreenivas, A., Gaur, A., 2018. Characterization of the complete mitogenome of Indian mouse deer, Moschiola Indica (artiodactyla: Tragulidae) and its evolutionary significance. Scientific Reports 8. doi:10.1038/s41598-018-20946-5

Scanlon , J.E., Farroway , L., 2016. Organisational Consortiums: The International Consortium on Combating Wildlife Crime (ICCWC), in: Environmental Crime and Collaborative State Intervention. Palgrave Macmillan, Basingstoke, Hampshire, pp. 77–99.

Schadt, E.E., Turner, S., Kasarskis, A., 2010. A window into third-generation sequencing. Human Molecular Genetics 19, R227–R240. doi:10.1093/hmg/ddq416

Schneider, P.M., 2007. Scientific standards for studies in forensic genetics. Forensic Science International 165, 238–243. doi:10.1016/j.forsciint.2006.06.067

Schoch, C.L., Ciufo, S., Domrachev, M., Hotton, C.L., Kannan, S., Khovanskaya, R., Leipe, D., Mcveigh, R., O'Neill, K., Robbertse, B., Sharma, S., Soussov, V., Sullivan, J.P., Sun, L., Turner, S., Karsch-Mizrachi, I., 2020. NCBI taxonomy: A comprehensive update on curation, resources and Tools. Database (Oxford) 2020. doi:10.1093/database/baaa062

Schubart, C.D., Huber, M.G.J., 2006. Genetic Comparisons Of German Populations Of The Stone Crayfish, Austropotamobius Torrentium (Crustacea: Astacidae). Bulletin Français de la Pêche et de la Pisciculture 1019–1028. doi:10.1051/kmae:2006008

Seabury, C.M., Bhattarai, E.K., Taylor, J.F., Viswanathan, G.G., Cooper, S.M., Davis, D.S., Dowd, S.E., Lockwood, M.L., Seabury, P.M., 2011. Genome-wide polymorphism and comparative analyses in the white-tailed deer (Odocoileus virginianus): A model for conservation genomics. PLoS ONE 6. doi:10.1371/journal.pone.0015811

Service, R.F., 2006. Gene Sequencing: The Race for the $1000 Genome. Science 311, 1544–1546. doi:10.1126/science.311.5767.1544

Shaler, R.C., 2002. , in: Science Handbook. Prentice Hall, Upper Saddle River.

Shaw, C.N., Wilson, P.J., White, B.N., 2003. A Reliable Molecular Method of Gender Determination for Mammals. Journal of Mammalogy 84, 123–128. doi:10.1644/1545-1542(2003)084<0123:armmog>2.0.co;2

Shivji, M.S., Chapman, D.D., Pikitch, E.K., Raymond, P.W., 2005. Genetic profiling reveals illegal international trade in fins of the great white shark, Carcharodon carcharias. Conservation Genetics 6, 1035–1039. doi:10.1007/s10592-005-9082-9

Si, T., 2013. Complete mitochondrial genome of white Bengal tiger, Panthera tigris tigris. College of Biological Engineering, Jingchu University of Technology, Jingmen, Hubei Province 480000, China. doi:https://www.ncbi.nlm.nih.gov/nuccore/576896669

Simon, C., Buckley, T.R., Frati, F., Stewart, J.B., Beckenbach, A.T., 2006. Incorporating molecular evolution into phylogenetic analysis, and a new compilation of conserved polymerase chain reaction primers for animal mitochondrial DNA. Annual Review of Ecology, Evolution, and Systematics 37, 545–579. doi:10.1146/annurev.ecolsys.37.091305.110018

Simon, C., Frati, F., Beckenbach, A., Crespi, B., Liu, H., Flook, P., 1994. Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. Annals of the Entomological Society of America 87, 651–701. doi:10.1093/aesa/87.6.651

Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., Zdobnov, E.M., 2015. Busco: Assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31, 3210–3212. doi:10.1093/bioinformatics/btv351

Singer, V.L., Jones, L.J., Yue, S.T., Haugland, R.P., 1997. Characterization of PicoGreen Reagent and Development of a Fluorescence-Based Solution Assay for Double-Stranded DNA Quantitation. Analytical Biochemistry 249, 228–238. doi:10.1006/abio.1997.2177

Sinha, M., Rana, M., Kushwaha, P., 2020. Applications of Mitochondrial DNA in Forensic Science, in: Forensic DNA Typing: Principles, Applications and Advancements. Springer, Singapore, pp. 329–343.

Slate, J., Coltman, D.W., Goodman, S.J., MacLean, I., Pemberton, J.M., Williams, J.L., 1998. Bovine microsatellite loci are highly conserved in red deer (Cervus elaphus), sika deer (Cervus nippon) and Soay sheep (Ovis aries). Animal Genetics 29, 307–315. doi:10.1046/j.1365-2052.1998.00347.x

Slater, G.S., Birney, E., 2005. Automated generation of heuristics for biological sequence comparison. BMC Bioinformatics 6.

Smart, U., Cihlar, J.C., Budowle, B., 2021. International Wildlife Trafficking: A perspective on the challenges and potential Forensic Genetics Solutions. Forensic Science International: Genetics 54, 102551. doi:10.1016/j.fsigen.2021.102551

Smit, A., Hubley, R., Green, P., 2015. RepeatMasker [WWW Document]. RepeatMasker. URL http://www.repeatmasker.org/ (accessed 9.7.22).

Smith, P., DenDanto, D., Smith, K., Palman, D., Kornfield, I., 2002. Allele frequencies for three STR loci RT24, RT09, and BM1225 in northern New England white-tailed deer. Journal of Forensic Science 47, 673–675.

Stanke, M., Steinkamp, R., Waack, S., Morgenstern, B., 2004. Augustus: A web server for gene finding in Eukaryotes. Nucleic Acids Research 32. doi:10.1093/nar/gkh379

Stern, C., 1943. The Hardy–Weinberg law. Science 97, 137–138. doi:10.1126/science.97.2510.137

Subramanian, S., Mishra, R.K., Singh, L., 2003. Genome-wide analysis of microsatellite repeats in humans: their abundance and density in specific genomic regions. Genome Biology 4. doi:10.1186/gb-2003-4-2-r13

Swango, K.L., Hudlow, W.R., Timken, M.D., Buoncristiani, M.R., 2007. Developmental validation of a multiplex qPCR assay for assessing the quantity and quality of nuclear DNA in forensic samples. Forensic Science International 170, 35–45. doi:10.1016/j.forsciint.2006.09.002

SWGDAM, 2003. Revised validation guidelines [WWW Document]. Strbase.nist.gov. URL https://strbase.nist.gov/validation/SWGDAM_Validation.doc (accessed 6.29.20).

SWGDAM, 2012. Scientific Working Group on DNA Analysis Methods: Validation Guidelines for DNA Analysis Methods [WWW Document]. Media.wix.com. URL http://media.wix.com/ugd/4344b0_cbc27d16dcb64fd88cb36ab2a2a25e4c.pdf (accessed 6.29.20).

SWGDAM, 2019. Addendum to "SWGDAM Interpretation Guidelines for Autosomal STR Typing by Forensic DNA Testing Laboratories" to Address Next Generation Sequencing [WWW Document]. 1ecb9588-ea6f-4feb-971a-73265dbf079c.filesusr.com. URL https://1ecb9588-ea6f-4feb-971a-73265dbf079c.filesusr.com/ugd/4344b0_91f2b89538844575a9f51867def7be85.pdf (accessed 6.29.20).

Syndercombe Court, D., 2021. Mitochondrial DNA in forensic use. Emerging Topics in Life Sciences 5, 415–426. doi:10.1042/etls20210204

Szabolcsi, Z., Egyed, B., Zenke, P., Padar, Z., Borsy, A., Steger, V., Pasztor, E., Csanyi, S., Buzas, Z., Orosz, L., 2014. Constructing STR Multiplexes for Individual Identification of Hungarian Red Deer. Journal of Forensic Sciences 59, 1090–1099. doi:10.1111/1556-4029.12403

Szibor, R., Edelmann, J., Hering, S., Plate, I., Wittig, H., Roewer, L., Wiegand, P., Caĺı, F., Romano, V., Michael, M., 2003. Cell line DNA typing in forensic genetics—the necessity of reliable standards. Forensic Science International 138, 37–43. doi:10.1016/j.forsciint.2003.09.002

Tabasum, W., Parmar, D.R., Jayaraman, A., Mitra, S., Sreenivas, A., Kunteepuram, V., Gaur, A., 2017. The complete mitochondrial genome of Eld's deer (Rucervus eldii eldii) and its phylogenetic implications. Gene Reports 9, 98–107. doi:10.1016/j.genrep.2017.10.001

Tak, Y.K., Kim, W.Y., Kim, M.J., Han, E., Han, M.S., Kim, J.J., Kim, W., Lee, J.E., Song, J.M., 2012. Highly sensitive polymerase chain reaction-free quantum dot-based quantification of forensic genomic DNA. Analytica Chimica Acta 721, 85–91. doi:10.1016/j.aca.2012.01.056

Tamura, K., Battistuzzi, F.U., Billing-Ross, P., Murillo, O., Filipski, A., Kumar, S., 2012. Estimating divergence times in large molecular phylogenies. Proceedings of the National Academy of Sciences 109, 19333–19338. doi:10.1073/pnas.1213199109

Tamura, K., Nei, M., 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Molecular Biology and Evolution 10, 512–526. doi:10.1093/oxfordjournals.molbev.a040023

Tamura, K., Nei, M., Kumar, S., 2004. Prospects for inferring very large phylogenies by using the neighbor-joining method. Proceedings of the National Academy of Sciences 101, 11030–11035. doi:10.1073/pnas.0404206101

Tamura, K., Tao, Q., Kumar, S., 2018. Theoretical Foundation of the RELTIME method for estimating divergence times from variable evolutionary rates. Molecular Biology and Evolution 35, 1770–1782. doi:10.1093/molbev/msy044

Tao, R., Qi, W., Chen, C., Zhang, J., Yang, Z., Song, W., Zhang, S., Li, C., 2019. Pilot study for forensic evaluations of the Precision ID GlobalFilerTM NGS STR Panel v2 with the Ion S5TM system. Forensic Science International: Genetics 43, 102147. doi:10.1016/j.fsigen.2019.102147

Taylor, D., Bright, J.-A., Buckleton, J., 2014. The 'factor of two' issue in mixed DNA profiles. Journal of Theoretical Biology 363, 300–306. doi:10.1016/j.jtbi.2014.08.021

Terencio, M.L., Schneider, C.H., Gross, M.C., Feldberg, E., Porto, J.I.R., 2012. Structure and organization of the mitochondrial DNA control region with tandemly repeated sequence in the Amazon ornamental fish. Mitochondrial DNA 24, 74–82. doi:10.3109/19401736.2012.717934

The Evaluation of Forensic DNA Evidence, 1996. Committee on DNA Forensic Science: An Update. Washington, D.C.

Thomas, J.T., Berlin, R.M., Barker, J.M., Dawson Cruz, T., 2013. Qiagen's Investigator™ Quantiplex Kit as a Predictor of STR Amplification Success from Low-Yield DNA Samples,. Journal of Forensic Sciences 58, 1306–1309. doi:10.1111/1556-4029.12171

Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. Clustal W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Research 22, 4673–4680. doi:10.1093/nar/22.22.4673

Thompson, W.C., Schumann, E.L., 1987. Interpretation of statistical evidence in criminal trials: The prosecutor's fallacy and the defense attorney's fallacy. Law and Human Behavior 11, 167–187. doi:10.1007/bf01044641

Toole, K., Roffey, P., Young, E., Cho, K., Shaw, T., Smith, M., Blagojevic, N., 2019. Evaluation of commercial forensic DNA extraction kits for decontamination and extraction of DNA from biological samples contaminated with radionuclides. Forensic Science International 302, 109867. doi:10.1016/j.forsciint.2019.06.025

TRAFFIC, 2021. Wildlife Trade and the UK - traffic - The Wildlife Trade Monitoring Network [WWW Document]. wildlife trade in the UK. URL https://www.traffic.org/wildlife-trade-and-the-uk/#:~:text=Between%202013%20and%202020%20the,wood%20products%20and%20medicinal%20plants. (accessed 1.20.25).

UNODC (United Nations on Drugs and Crime), 2016. , World Wildlife Crime Report: Trafficking in protected species. UNODC, Vienna.

UNODC (United Nations on Drugs and Crime), 2020. , World Wildlife Crime Report, Trafficking in protected species. United Nations Office on Drugs and Crime, New York.

UNODC (United Nations on Drugs and Crime), 2024. , World Wildlife Crime Report. United Nations publications, Vienna.

van Dijk, E.L., Auger, H., Jaszczyszyn, Y., Thermes, C., 2014. Ten years of next-generation sequencing technology. Trends in Genetics 30, 418–426. doi:10.1016/j.tig.2014.07.001

van Dijk, E.L., Jaszczyszyn, Y., Naquin, D., Thermes, C., 2018. The Third Revolution in Sequencing Technology. Trends in Genetics 34, 666–681. doi:10.1016/j.tig.2018.05.008

van Hoppe, M.J.C., Dy, M.A.V., van den Einden, M., Iyengar, A., 2016. SkydancerPlex : A novel STR multiplex validated for forensic use in the hen harrier (Circus cyaneus). Forensic Science International: Genetics 22, 100–109. doi:10.1016/j.fsigen.2016.02.003

Vanek, D., Rihova, P., Ehler, E., Dalihodova, S., Stikarova, R., Vankova, L., Strnad, Z., 2019. STRAND: A Cloud expert system for non-human DNA analysis. Forensic Science International: Genetics Supplement Series 7, 147–149. doi:10.1016/j.fsigss.2019.09.057

Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A., Gocayne, J.D., Amanatides, P., Ballew, R.M., Huson, D.H., Wortman, J.R., Zhang, Q., Kodira, C.D., Zheng, X.H., Chen, L., Skupski, M., Subramanian, G., Thomas, P.D., Zhang, J., Gabor Miklos, G.L., Nelson, C., Broder, S., Clark, A.G., Nadeau, J., McKusick, V.A., Zinder, N., Levine, A.J., Roberts, R.J., Simon, M., Slayman, C., Hunkapiller, M., Bolanos, R., Delcher, A., Dew, I., Fasulo, D., Flanigan, M., Florea, L., Halpern, A., Hannenhalli, S., Kravitz, S., Levy, S., Mobarry, C., Reinert, K., Remington, K., Abu-Threideh, J., Beasley, E., Biddick, K., Bonazzi, V., Brandon, R., Cargill, M., Chandramouliswaran, I., Charlab, R., Chaturvedi, K., Deng, Z., Francesco, V.D., Dunn, P., Eilbeck, K., Evangelista, C., Gabrielian, A.E., Gan, W., Ge, W., Gong, F., Gu, Z., Guan, P., Heiman, T.J., Higgins, M.E., Ji, R.-R., Ke, Z., Ketchum, K.A., Lai, Z., Lei, Y., Li, Z., Li, J., Liang, Y., Lin, X., Lu, F., Merkulov, G.V., Milshina, N., Moore, H.M., Naik, A.K., Narayan, V.A., Neelam, B., Nusskern, D., Rusch, D.B., Salzberg, S., Shao, W., Shue, B., Sun, J., Wang, Z.Y., Wang, A., Wang, X., Wang, J., Wei, M.-H., Wides, R., Xiao, C., Yan, C., Yao, A., Ye, J., Zhan, M., Zhang, W., Zhang, H., Zhao, Q., Zheng, L., Zhong, F., Zhong, W., Zhu, S.C., Zhao, S., Gilbert, D., Baumhueter, S., Spier, G., Carter, C., Cravchik, A., Woodage, T., Ali, F., An, H., Awe, A., Baldwin, D., Baden, H., Barnstead, M., Barrow, I., Beeson, K., Busam, D., Carver, A., Center, A., Cheng, M.L., Curry, L., Danaher, S., Davenport, L., Desilets, R., Dietz, S., Dodson, K., Doup, L., Ferriera, S., Garg, N., Gluecksmann, A., Hart, B., Haynes, J., Haynes,

C., Heiner, C., Hladun, S., Hostin, D., Houck, J., Howland, T., Ibegwam, C., Johnson, J., Kalush, F., Kline, L., Koduru, S., Love, A., Mann, F., May, D., McCawley, S., McIntosh, T., McMullen, I., Moy, M., Moy, L., Murphy, B., Nelson, K., Pfannkoch, C., Pratts, E., Puri, V., Qureshi, H., Reardon, M., Rodriguez, R., Rogers, Y.-H., Romblad, D., Ruhfel, B., Scott, R., Sitter, C., Smallwood, M., Stewart, E., Strong, R., Suh, E., Thomas, R., Tint, N.N., Tse, S., Vech, C., Wang, G., Wetter, J., Williams, S., Williams, M., Windsor, S., Winn-Deen, E., Wolfe, K., Zaveri, J., Zaveri, K., Abril, J.F., Guigó, R., Campbell, M.J., Sjolander, K.V., Karlak, B., Kejariwal, A., Mi, H., Lazareva, B., Hatton, T., Narechania, A., Diemer, K., Muruganujan, A., Guo, N., Sato, S., Bafna, V., Istrail, S., Lippert, R., Schwartz, R., Walenz, B., Yooseph, S., Allen, D., Basu, A., Baxendale, J., Blick, L., Caminha, M., Carnes-Stine, J., Caulk, P., Chiang, Y.-H., Coyne, M., Dahlke, C., Mays, A.D., Dombroski, M., Donnelly, M., Ely, D., Esparham, S., Fosler, C., Gire, H., Glanowski, S., Glasser, K., Glodek, A., Gorokhov, M., Graham, K., Gropman, B., Harris, M., Heil, J., Henderson, S., Hoover, J., Jennings, D., Jordan, C., Jordan, J., Kasha, J., Kagan, L., Kraft, C., Levitsky, A., Lewis, M., Liu, X., Lopez, J., Ma, D., Majoros, W., McDaniel, J., Murphy, S., Newman, M., Nguyen, T., Nguyen, N., Nodell, M., Pan, S., Peck, J., Peterson, M., Rowe, W., Sanders, R., Scott, J., Simpson, M., Smith, T., Sprague, A., Stockwell, T., Turner, R., Venter, E., Wang, M., Wen, M., Wu, D., Wu, M., Xia, A., Zandieh, A., Zhu, X., 2001. The Sequence of the Human Genome. Science 291, 1304–1351. doi:10.1126/science.1058040

Vieira-Silva, C., Afonso-Costa, H., Ribeiro, T., Porto, M.J., Dias, M., Amorim, A., 2015. Quantifiler ® Trio DNA validation and usefulness in casework samples. Forensic Science International: Genetics Supplement Series 5, e246–e247. doi:10.1016/j.fsigss.2015.09.098

Vinogradov, A.E., 2003. DNA helix: The importance of being GC-rich. Nucleic Acids Research 31, 1838–1844. doi:10.1093/nar/gkg296

Wada, K., Nakamura, M., Nishibori, M., Yokohama, M., 2006. The complete nucleotide sequence of mitochondrial genome in the rein deer (Rangifer tarandus) and red deer (Cervus elaphus). National Center for Biotechnology Information, NIH, Bethesda, MD 20894, USA. doi:https://www.ncbi.nlm.nih.gov/nucleotide/NC_007704.2

Wada, K., Nishibori, M., Yokohama, M., 2007. The complete nucleotide sequence of mitochondrial genome in the Japanese sika deer (Cervus Nippon), and a

phylogenetic analysis between Cervidae and bovidae. Small Ruminant Research 69, 46–54. doi:10.1016/j.smallrumres.2005.12.002

Wada, K., Okumura, K., Nishibori, M., Kikkawa, Y., Yokohama, M., 2010. The complete mitochondrial genome of the domestic red deer (cervus elaphus) of New Zealand and its phylogenic position within the family Cervidae. Animal Science Journal 81, 551–557. doi:10.1111/j.1740-0929.2010.00799.x

Wahlund, S., 1928. Zusammensetzung von Population und Korrelationserscheinung vom Standpunkt der Vererbungslehre aus betrachtet. Hereditas 11, 65–106. doi:10.1111/j.1601-5223.1928.tb02483.x

Walsh, P.S., Varlaro, J., Reynolds, R., 1992. A rapid chemiluminescent method for quanititation of human DNA. Nucleic Acids Research 20, 5061–5065. doi:10.1093/nar/20.19.5061

Wang, W., Yan, H.-J., Chen, S.-Y., Li, Z.-Z., Yi, J., Niu, L.-L., Deng, J.-P., Chen, W.-G., Pu, Y., Jia, X., Qu, Y., Chen, A., Zhong, Y., Yu, X.-M., Pang, S., Huang, W.-L., Han, Y., Liu, G.-J., Yu, J.-Q., 2019. The sequence and de novo assembly of Hog Deer Genome. Scientific Data 6. doi:10.1038/sdata.2018.305

Wasser, S.K., Joseph Clark, W., Drori, O., Stephen Kisamo, E., Mailand, C., Mutyoba, B., Stephens, M., 2008. Combating the Illegal Trade in African Elephant Ivory with DNA Forensics. Conservation Biology 22, 1065–1071. doi:10.1111/j.1523-1739.2008.01012.x

Wasser, S.K., Mailand, C., Booth, R., Mutayoba, B., Kisamo, E., Clark, B., Stephens, M., 2007. Using DNA to track the origin of the largest ivory seizure since the 1989 trade ban. Proceedings of the National Academy of Sciences 104, 4228–4233. doi:10.1073/pnas.0609714104

Wasser, S.K., Shedlock, A.M., Comstock, K., Ostrander, E.A., Mutayoba, B., Stephens, M., 2004. Assigning African elephant DNA to geographic region of origin: Applications to the ivory trade. Proceedings of the National Academy of Sciences 101, 14847–14852. doi:10.1073/pnas.0403170101

Watherston, J., Ward, J., 2023. Autosomal short tandem repeat (STR) profiling of human skeletal remains. Forensic Genetic Approaches for Identification of Human Skeletal Remains 167–197. doi:10.1016/b978-0-12-815766-4.00008-x

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. doi:10.1016/j.fsigen.2013.10.015

Weir, B.S., 2012. Estimating F-Statistics: A Historical View. Philosophy of Science 79, 637–643. doi:10.1086/667904

Wellcome Sanger Institute , 2019. . Darwin Tree of Life Project. URL https://www.sanger.ac.uk/collaboration/darwin-tree-of-life-project/ (accessed 6.16.22).

Welton, L.J., Siler, C.D., Linkem, C.W., Diesmos, A.C., Diesmos, M.L., Sy, E., Brown, R.M., 2013. Dragons in our midst: Phyloforensics of illegally traded Southeast Asian monitor lizards. Biological Conservation 159, 7–15. doi:10.1016/j.biocon.2012.10.013

Werner, N.Y., Rabiei, A., Saltz, D., Daujat, J., Baker, K., 2015. Dama mesopotamica. The IUCN Red List of Threatened Species 2015 e.T6232A97672550. doi:10.2305

Wesselink, M., Kuiper, I., 2011. Individual identification of fox (Vulpes vulpes) in forensic wildlife investigations. Forensic Science International: Genetics Supplement Series 3, e214–e215. doi:10.1016/j.fsigss.2011.08.107

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 Science International 123, 215–223. doi:10.1016/s0379-0738(01)00557-6

White, N.E., Dawson, R., Coghlan, M.L., Tridico, S.R., Mawson, P.R., Haile, J., Bunce, M., 2012. Application of STR markers in wildlife forensic casework involving Australian black-cockatoos (Calyptorhynchus spp.). Forensic Science International: Genetics 6, 664–670. doi:10.1016/j.fsigen.2011.10.003

Whitford, W., Hawkins, V., Moodley, K.S., Grant, M.J., Lehnert, K., Snell, R.G., Jacobsen, J.C., 2022. Proof of concept for multiplex amplicon sequencing for mutation identification using the Minion Nanopore sequencer. Scientific Reports 12. doi:10.1038/s41598-022-12613-7

Wildlife and Countryside Link, 2019. Link Annual wildlife crime report 2018 [WWW Document]. Wildlife and Countryside Link. URL https://www.wcl.org.uk/docs/Link_Annual_Wildlife_Crime_Report_2018.pdf (accessed 5.23.22).

Wildlife and Countryside Link, 2022. WCL wildlife crime report 2021 [WWW Document]. Wildlife and Countryside Link. URL https://www.wcl.org.uk/docs/assets/uploads/WCL_Wildlife_Crime_Report_2021_29.11.22.pdf (accessed 1.20.25).

Wildlife and Countryside Link, 2023. WCL Annual wildlife crime report 2022 [WWW Document]. Wildlife and Countryside Link. URL https://www.wcl.org.uk/assets/uploads/0/Wildlife_Crime_Report_October_2023.pdf (accessed 1.20.25).

Wildlife and Countryside Link, 2024. WCL Annual wildlife crime report 2023 [WWW Document]. Wildlife and Countryside Link. URL https://www.wcl.org.uk/docs/2024/WCL_Wildlife_Crime_2023_Final.pdf (accessed 1.20.25).

Wildlife Crime in 2017, A report on the scale of wildlife crime in England and Wales. 2018. London.

Wildlife Forensics // Cellmark [WWW Document], 2015. . Cellmarkforensics.co.uk. URL https://www.cellmarkforensics.co.uk/services/wildlife-forensics/ (accessed 5.13.20).

Willerslev, E., Gilbert, M.T., Binladen, J., Ho, S.Y., Campos, P.F., Ratan, A., Tomsho, L.P., da Fonseca, R.R., Sher, A., Kuznetsova, T.V., Nowak-Kemp, M., Roth, T.L., Miller, W., Schuster, S.C., 2009. Analysis of complete mitochondrial genomes from extinct and extant rhinoceroses reveals lack of phylogenetic resolution. BMC Evolutionary Biology 9, 95. doi:10.1186/1471-2148-9-95

Wilson, J., Fuller, V., Benson, G., Juroske, D., Duvall, E., Fu, J., Pritchard, J., Allen, R.W., 2010. Molecular Assay for Screening and Quantifying DNA in Biological Evidence: The Modified Q-TAT Assay*. Journal of Forensic Sciences 55, 1050–1057. doi:10.1111/j.1556-4029.2010.01371.x

Wu, H., Wan, Q.-H., Fang, S.-G., Zhang, S.-Y., 2005. Application of mitochondrial DNA sequence analysis in the forensic identification of Chinese sika deer subspecies. Forensic Science International 148, 101–105. doi:10.1016/j.forsciint.2004.04.072

Wyatt, T., 2016. Victimless venison? Deer poaching and black market meat in the UK. Contemporary Justice Review 19, 188–200. doi:10.1080/10282580.2016.1169700

Xing, X., Ai, C., Wang, T., Li, Y., Liu, Huitao, Hu, P., Wang, G., Liu, Huamiao, Wang, H., Zhang, R., Zheng, J., Wang, Xiaobo, Wang, L., Chang, Y., Qian, Q., Yu, J., Tang, L., Wu, S., Shao, Xiujuan, Li, A., Cui, P., Zhan, W., Zhao, S., Wu, Z., Shao, Xiqun, Dong, Y., Rong, M., Tan, Y., Cui, X., Chang, S., Song, X., Yang, T., Sun, L., Ju, Y., Zhao, P., Fan, H., Liu, Y., Wang, Xinhui, Yang, W., Yang, M., Wei, T., Song, S., Xu, J., Yue, Z., Liang, Q., Li, C., Ruan, J., Yang, F., 2022. The first high-quality reference genome of sika deer provides insights into high-tannin adaptation. Genomics, Proteomics & Bioinformatics 21, 203–215. doi:10.1016/j.gpb.2022.05.008

Yang, Y., Xie, B., Yan, J., 2014. Application of Next-generation Sequencing Technology in Forensic Science. Genomics, Proteomics & Bioinformatics 12, 190–197. doi:10.1016/j.gpb.2014.09.001

Young, B., Faris, T., Armogida, L., 2019. A nomenclature for sequence-based forensic DNA analysis. Forensic Science International: Genetics 42, 14–20. doi:10.1016/j.fsigen.2019.06.001

Yugovich, O., Bunce, M., Harbison, SA., 2025. Point-of-need species identification using non-PCR DNA-based approaches to combat wildlife crime. Forensic Science International: Genetics 78, 103278. doi:10.1016/j.fsigen.2025.103278

Zhang , X., Zheng , A., Li , J., Zhang , H., Shi , Y., Shan , X., 2004. Sequence and organization of Muntiacus reevesi mitochondrial genome. Hereditas (Beijing) 6, 849–853. doi:PMID: 15640115

Zhang, C., Chen, L., Zhou, Y., Wang, K., Chemnick, L.G., Ryder, O.A., Wang, W., Zhang, G., Qiu, Q., 2017. Draft genome of the Milu (Elaphurus Davidianus). GigaScience 7. doi:10.1093/gigascience/gix130

Zhang, W., Zhang, Z., Shen, F., Hou, R., Lv, X., Yue, B., 2006. Highly conserved D-loop-like nuclear mitochondrial sequences (Numts) in tiger (Panthera tigris). Journal of Genetics 85, 107–116. doi:10.1007/bf02729016

Zhao, P., Xu, Huailiang, Li, D., Xie, M., Xu, Huaming, Wu, J., Wen, A., Ni, Q., Zhang, M., Wang, Q., Zhu, G., He, T., Yao, Y., 2017. The complete mitochondrial genome sequence and phylogenetic analysis of white-lipped deer (cervus albirostris). Conservation Genetics Resources 10, 741–745. doi:10.1007/s12686-017-0919-6

10 Appendices

10.1 Tables

Table 10.1: Details of the 90 fallow deer primers and their success throughout the amplification and polymorphism testing process.

Primer Name	Repeat motif (bp)	Primer Sequence (F)/(R)	Initial Amp.	3% Gel	Sanger Seq. & Frag. Analysis	Poly-morphic	HWE & ONT
Fallow_F_1 Fallow_R_1	GACT	CGACTCGATGGACATGGGTT TGGCCGAGAGAATTCCATGG	Yes				
Fallow_F_2 Fallow_R_2	TCAT	CCCAAGCCTCCATCCTTCTC CCACATGCAAACCACACCTC	Yes	Yes			
Fallow_F_4 Fallow_R_4	GATA	TGTCTCCTGCATTGGAAGGC TGTGTGATCTTGGGCATGTCA	Yes	Yes	Yes	Yes	
Fallow_F_5 Fallow_R_5	GACA	GGACACACAGCCACCTAACA CCCTCCCTGGTGTCTGAAAC	Yes	Yes			
Fallow_F_6 Fallow_R_6	GATG	CAGGTCCCCAGTCATTGAACA GACTTGGACACGACTGAGCA	Yes	Yes	Yes		
Fallow_F_7 Fallow_R_7	GATG	CCTCCTGGCAGAGAACACAG TCTCCCAGCCCTTCTTCAGT	Yes				
Fallow_F_8 Fallow_R_8	CATC	TCCATCCATCCATCCTCCCA TGGATGGATGGGAGGATGGA					
Fallow_F_9 Fallow_R_9	TCTG	GGGAGGTGGTGTCAAAGTGA AGACACACACTGAGACATGCA	Yes	Yes			
Fallow_F_10 Fallow_R_10	TGGA	TGATGGTGGAAGGATGGACG TGGCTCAGTGTCTTTCCTGG	Yes	Yes	Yes		
Fallow_F_12 Fallow_R_12	GTTT	CCCCAATCATTCCGGACAGT AGGAATCGAACTCAGACCCC	Yes	Yes	Yes		
Fallow_F_13 Fallow_R_13	CCAT	GCCAGAGACCTACCATTCCA GGACGAATGCTGGGATGGAT	Yes	Yes			

Fallow_F_14 Fallow_R_14	ACTG	TGCGATTCATGGGGTCACAA AGTGGATATTGGCGGTCTGG	Yes	Yes			
Fallow_F_15 Fallow_R_15	TTTC	TCCATGGTGTTGCAAGGAGT GGCAAGAACCTTGGAGTGGA	Yes				
Fallow_F_17 Fallow_R_17	CTGT	GAAGCCCTGGGTCTTGGTTT GCAAAGGCGCAAACAGACAT	Yes	Yes	Yes		
Fallow_F_18 Fallow_R_18	GGAT	TGTGGACAGATGCATAGGTGG ATCCACTCACCCACCTACCA					
Fallow_F_21 Fallow_R_21	AGTG	TGTTCCAAGTCGCTCTCATAGT GAAAGGCTACCCACTCCAGT	Yes				
Fallow_F_22 Fallow_R_22	GATG	ATGGACAGATGGATGGGTGG ATGGGTCTGTGCCTTGAGTC	Yes	Yes			
Fallow_F_25 Fallow_R_25	CAGC	GCAGGTGTTTGAGGCAGACA CCTGTAAAGCTGGCTCTCCT	Yes				
Fallow_F_26 Fallow_R_26	AGAT	TGGGAGTGTTCAGTCTAGACCA GGAGAGTTTTCCCTTGGCAAAC	Yes	Yes	Yes		
Fallow_F_29 Fallow_R_29	GCAC	GGGCTACAGTTCAAGGGTCA TCCTGAAAACCCTTCCCACG	Yes	Yes			
Fallow_F_30 Fallow_R_30	TCAC	CCTGGCAGGCTACAATCCAT CCAAGATTCAACTTTCTGCCTCA		Yes			
Fallow_F_32 Fallow_R_32	ATAC	AGTAGTGGAGGGTGCTTCTCT GGATGGTGGTGGGAGGAAAG	Yes				
Fallow_F_33 Fallow_R_33	ATTC	CTGTGTGCTCTCCCTTTGCT GCTGGAAATGGAATCAGGCC	Yes	Yes			
Fallow_F_34 Fallow_R_34	GAAA	CGTGTACACCCAAGGCTGAT ACTGGAGTGGGTTGACATGC	Yes				
Fallow_F_35 Fallow_R_35	GAAG	GCAAGAAAGTGAGCTCAGCG ATGCCCACACCATGTTGATT	Yes	Yes			
Fallow_F_37 Fallow_R_37	GGAT	CTGGCTTTATGGGTGCAGGA CTGGTGCTGTGGTCTATGCA	Yes	Yes	Yes		

Fallow_F_38 Fallow_R_38	AGTG	CGAGGGTTGTTGGCAGATGA AGTCCTTGGGGTTGCAAAGA	Yes	Yes			
Fallow_F_39 Fallow_R_39	ATGG	TGGATTGGTGGATCGATGGA TCCATCCATCTATCCACGCA	Yes				
Fallow_F_41 Fallow_R_41	ATGA	CTGAACTTCCAGGGCTTAGCA TGCTCAGTTGGCCAAAGACA	Yes	Yes			
Fallow_F_42 Fallow_R_42	GATA	TCCACTCCAGATTCCCTCCA TCACTCATGAGCCTCTGGAG					
Fallow_F_43 Fallow_R_43	TCAT	CTTCTCTGCCTGCGACATGA CAGGAAGGGCAGTCACAGAG	Yes				
Fallow_F_45 Fallow_R_45	CTGC	CCAGTGCCAAGCCTCAGTTT GATGTGGACACTGAGCGGAA	Yes	Yes			
Fallow_F_46 Fallow_R_46	CATC	GAGTCGGACACAACTGAGCA ATTCTTCCCAGTACCCAGCA	Yes				
Fallow_F_48 Fallow_R_48	TCTT	CCCACTCCAGTACTCTTGCC CGTGGCCCTGGAGTGATTTA	Yes				
Fallow_F_49 Fallow_R_49	TATC	GGGGCCATATGTTGTCGAGG TTGAGAACAAGTTGGGAGGCT	Yes				
Fallow_F_50 Fallow_R_50	ATCC	CCACCCATCCATCCATCCAC AGGTACTCTTCGGTTTTATTGGGT	Yes				
Fallow_F_51 Fallow_R_51	ATTC	GTGTGTACATGTGAGCCAGC CACAGGTGGCATTGATCCCT	Yes				
Fallow_F_53 Fallow_R_53	TATC	ATGAATGCTCTCCTGCTGCC TAAGCTCATGGACGGCAACA	Yes	Yes	Yes	Yes	Yes
Fallow_F_54 Fallow_R_54	ATGG	ATGCCTCAAATGTGCCCTGA GCTCCAGGTTTTGTCTCCCA	Yes				
Fallow_F_56 Fallow_R_56	CCAT	CCGAAGTAGCCCACAGAGTG GTGAGTGGATGGATGGGTGG	Yes				
Fallow_F_57 Fallow_R_57	TATG	CTGTATTCTCAGGGCTGTTGGA TCCCAACACAAGGATCAGACC	Yes	Yes			

Fallow_F_58 Fallow_R_58	AGAT	TCATGTGACATGCCAGGTTG AGTTCCTCTTCACCTTCTGCC	Yes	Yes			
Fallow_F_60 Fallow_R_60	TATG	AACCCATGTCTCCTGCACTG GGGCAAGAAGAAAGGAAAGAGC	Yes				
Fallow_F_61 Fallow_R_61	AAAG	ACTCCAACCACTAATGCCAA GGTTTCCAGCATTGCAAGCA	Yes	Yes			
Fallow_F_62 Fallow_R_62	TCAT	TCTCCGAGGTCTGCTGTAGG TCATGCGTGAATGTGCTACCT	Yes	Yes			
Fallow_F_64 Fallow_R_64	TCTA	GTTCAGAAAGCACAGACTGGC TTGGGCATGGACTCTGTGTA	Yes	Yes	Yes		
Fallow_F_65 Fallow_R_65	AAAG	ACATAGTGTCAGATGAAGCCCA TCCGAACCCAGGGATTGAATC	Yes	Yes			
Fallow_F_66 Fallow_R_66	TTTA	TGGCAGTCTTTGAGGAGCTG GGGACTTGCTCGCTGTTTCA	Yes				
Fallow_F_68 Fallow_R_68	AAAT	GGTTCGATCCCTGGTCCATG CACTTCCTCTTCTGCTGCTCA	Yes				
Fallow_F_69 Fallow_R_69	CTTT	GGGACTGCCTTTCTTTGGGAT GCCTCCAACTAATACAAACAAATGG	Yes				
Fallow_F_70 Fallow_R_70	GATA	CATAAGCTCTTCCCTGGGTCTC AAGTCTGCCTGTCTGTCTGC	Yes	Yes	Yes	Yes	Yes
Fallow_F_73 Fallow_R_73	TCAC	CCCACTCCAGTATCCTTGCC ATGGCACCTGAGTTTCTGCA	Yes				
Fallow_F_74 Fallow_R_74	ATAG	GCAGGCTACAATCCATGGGT CGACCCAGGGATGGAAAGTG	Yes				
Fallow_F_76 Fallow_R_76	GGAT	GGGACAAGGCTTTCACGTTC TCCTTCCATCCGTCCATCCT	Yes	Yes			
Fallow_F_77 Fallow_R_77	AAAT	TGTGAAGTGATTGGCCTCCA AGAAAGAACAAAGCTAGTGGAGG					
Fallow_F_78 Fallow_R_78	TTTG	CCCATGTCCAATGCCCAGAA CGGGACTTCTCTGACAGTGC	Yes	Yes			

Fallow_F_81 Fallow_R_81	ATAG	TGGTACGAACTGAGAGGCAC TGTGACTGAGTGACTGAACTGA	Yes				
Fallow_F_82 Fallow_R_82	AATT	TCTTACGTCTCCTGCACTGG TGTAAAATTGCCCTCCCTCCC	Yes				
Fallow_F_85 Fallow_R_85	AGAT	TCTGTACAAAGGGGTGGCAG CCTGACCCCTGCAGTGAATT	Yes	Yes			
Fallow_F_86 Fallow_R_86	TATT	CCGTTGGTAGCTTGATAGGGA ACTGGTCAGGGTCCTTAGCT	Yes				
Fallow_F_87 Fallow_R_87	CTTT	ATTGTGCTCCAGGGGCTTAG CATGGGGTTGCAAAGAGTCG	Yes				
Fallow_F_89 Fallow_R_89	TAGA	ATCCTCTGCCATCCCCTTCT ATCTGTTTGTCTACCCACCACA	Yes	Yes	Yes	Yes	Yes
Fallow_F_90 Fallow_R_90	AAGA	ATGCAGGAGTTTGGGTTTGC TCCCATGGACAGAGAAGCCT	Yes				
Fallow_F_92 Fallow_R_92	CATC	TGTGATAGCGCTGGACTGAG TTATCAGTGCCCAGGACAGC		Yes			
Fallow_F_93 Fallow_R_93	TATG	CTGGTGGTGGATGCAAGCA CACCAGGAAACCCTCTCACC	Yes	Yes			
Fallow_F_94 Fallow_R_94	ATCA	CAGCCAGCCATCAATCAATCC GCACCACTCGTTGTTAGGGA	Yes	Yes	Yes	Yes	
Fallow_F_97 Fallow_R_97	TCAG	AGGTGGGGAACATTAGGGGT GTTGGTGATGGACAGGGAGG	Yes				
Fallow_F_98 Fallow_R_98	ATCC	ACTGGTAAATGGGAAGTGAGGA CACGCTGAAACATTACTACCCA	Yes				
Fallow_F_101 Fallow_R_101	ACTT	TCCCATAGACAGAGAGGCCT AGGAGGGAAGACATGTGGGA	Yes	Yes	Yes	Yes	
Fallow_F_102 Fallow_R_102	GTGA	TCCATGGACAGAGGAGCCTA TGGGGTCTCAAAGAGTCGGA	Yes				
Fallow_F_105 Fallow_R_105	TTAT	TGTCTCTTTCCACAGGTTAGGG CGGGATGGGGAATACGTGTA	Yes				

Fallow_F_106 Fallow_R_106	TACC	GAGTTTGTCTGTTGGCCTGT GCCTTGAACCATGCTTGCAT	Yes	Yes			
Fallow_F_109 Fallow_R_109	ACAT	AGTAGGCATAGTTGGTCCCCA GGCTAGTACCAGTTCAGGGC	Yes	Yes	Yes		
Fallow_F_110 Fallow_R_110	TTTG	CACAGGGGAAGGTAGCAAGT ACAAGGGTGGAGGGAGATGG	Yes	Yes			
Fallow_F_111 Fallow_R_111	TAGA	GCCATTCTCTTTTCCAGGGGA ATCCGTCTCTGCCTCCTCTT	Yes				
Fallow_F_113 Fallow_R_113	CACT	CTTGCTCCTCCACATCCCAG GGCATTCTGGAGAGGTGAGC	Yes	Yes			
Fallow_F_114 Fallow_R_114	ATTT	TTACCATCTGAGCCACACTGC TCGGGATGGGGAATACATGT	Yes				
Fallow_F_117 Fallow_R_117	TACA	ATTCCCTTCACAGTCAGCCC CGGCAGGTTACAGTCCATGA	Yes				
Fallow_F_118 Fallow_R_118	GACC	AGAGAGACAAGGTGGACAGACT GTGAGGTTGGAGGTGAGACG	Yes	Yes	Yes	Yes	Yes
Fallow_F_120 Fallow_R_120	ATTC	AGAAAGGGCTGGGAAGACAC CAGTTCTCCCGATGCCCTTG	Yes				
Fallow_F_121 Fallow_R_121	TATC	AGGGGCCATCAGACCTTCTA TTCTTCATGTGCCTCTGTGTGA	Yes	Yes			
Fallow_F_122 Fallow_R_122	TTTG	AGAGTACTGGAGTGGGGTGC GGCAGCAGGAGAACATGGAA	Yes				
Fallow_F_123 Fallow_R_123	ATAG	GGAGACCTCGACTGGATTCC CCACATTGCAGGCAGACTCT	Yes				
Fallow_F_124 Fallow_R_124	TATC	CCCTTACTGTTCTGTGCCACT TCTGGACAGGCAGGTGGATT	Yes	Yes	Yes	Yes	Yes
Fallow_F_125 Fallow_R_125	ATCC	GGTGAGTCCTGCTGTCCTTG ACTCCTGTGATCGTACATGGA	Yes				
Fallow_F_126 Fallow_R_126	AAAC	GGCAAAGACCCTGAGTTCACT TAGCGCATCTCCACTCTTCC	Yes	Yes			

Fallow_F_127 Fallow_R_127	ATAG	GGGGCCTTGCATGTCATGTA ACTGCAGGAGAAGGTGACGA	Yes	Yes			
Fallow_F_129 Fallow_R_129	CTAT	GGCTGTAGAGATGTTGGAGTCA TCCTGGGAGATGTATGGTCAGT	Yes	Yes	Yes	Yes	Yes
Fallow_F_130 Fallow_R_130	TGGA	AGCCAGCAGTGTTAACATCCA GCATGTGCGATGTGTTTCA	Yes	Yes			
Fallow_F_131 Fallow_R_131	TTCC	CCCACTCCAGTACTCTTGCC CCGTGAGCAACTGAGTCTGT	Yes				

Table 10.2: Complete fragment analysis results for 190 samples analysed across 6 loci. Red indicates samples which could not be analysed.

Sample	Fallow53		Fallow70		Fallow89		Fallow118		Fallow124		Fallow129	
	Peak 1	Peak 2	Peak 1	Peak 2	Peak 1	Peak 2	Peak 1	Peak 2	Peak 1	Peak 2	Peak 1	Peak 2
Dun 5												
Dun 6	152.38		168.04		100.18		177.52	197.56	182.84	186.87	210.79	219.10
Dun 7			168.1		112.18		177.43		182.86		210.81	
Dun 8	152.33		168.01		100.18		177.41	197.49	182.78		210.79	
Dun 9	152.41		168.02		100.18	112.11	197.54		182.80		210.84	219.12
Dun 10	152.33		168.06		100.18		197.5		182.80		210.78	
Dun 11	152.30		167.92		100.24	112.19	177.46	197.54	182.83		210.79	
Dun 12	152.40		168.02		100.18	112.09	177.41		181.83		210.78	
Dun 13	152.33		167.94		100.19		177.43	197.52	181.84		210.76	219.03
Dun 14	152.33		167.96		100.12	100.06	197.5		182.81		210.75	
Dun 15	152.34		168.02		112.11		177.46	197.56	181.84		209.73	
Dun 16	152.32		167.96		100.18	112.10	197.51		181.82		210.74	219.04
Dun 17	143.58		168		100.19		197.44		182.81		210.78	
Dun 18	152.32		168.04		100.19		177.38	197.51	181.85		209.68	
Dun 19	152.34		168.09		100.12	112.09	197.49		182.77	186.79	210.75	218.98
Dun 20	152.31		167.96		100.19		197.52		182.82	186.81	210.72	219.01
Dun 21	152.26		167.9		99.94		177.29		181.87		210.78	
Dun 22	152.31		167.94		100.18		177.35		182.81		210.78	219.04
Dun 23	152.34		167.99		100.19	112.09	177.47	197.52	181.88		210.80	
Dun 24	152.30		167.93		100.18		177.43	197.47	181.86		209.73	

Cro 21	139.32		156.08		100.19	112.16	177.42		182.84	186.86	210.81	219.09
Cro 22	139.32		156.02		112.10		177.43		186.82		210.74	
Cro 23									186.82			
Cro 24	139.31		156.05		100.12	112.11	177.41		182.79	186.85	210.78	
Cro 25	139.31		156.01		112.14		177.48		182.83	186.86	210.84	
Cro 26	139.31		156.04									
Cro 27	139.32		156.05		100.18		177.44		186.90		210.82	
Cro 28	139.31		156.05		112.14		177.46		182.90	186.90	210.81	
Cro 29	139.31	152.30	168.12		112.11		177.48		181.90		219.11	
Cro 30	139.31		155.99		99.10		177.44		182.87		210.79	
Cro 31	139.25		156.02		100.18		177.49					
Cro 32	139.31				112.12		177.48		186.86		210.78	
Cro 33	139.31		155.93		100.25	112.25	177.41		182.88	186.94	210.80	
Cro 34	139.31	152.33	155.97		100.32		177.44		186.06		210.89	219.14
Cro 35	139.38		156.03		100.19		177.52		182.88	186.94	210.83	219.13
Cro 36	139.31		156.07		100.26	112.22	177.43		183.03	187.03	210.83	
Cro 37	139.31		155.93		112.24		177.42		182.96		210.82	
Cro 38	139.31	152.32	156.07		100.26		177.41		182.99		210.84	
Cro 39	139.25		156.03		100.26	112.25	177.42				210.84	
Cro 40	139.31		155.96		100.32		177.39		181.92	185.97	210.80	
Kno 1	139.31	152.39	168.12		112.19		197.57		183.01		210.84	219.13
Kno 2	148.11		168.11				197.5		183.03		210.90	219.17
Kno 3	143.74	152.37	156.08		112.18		197.54		183.00	187.04	210.82	

Kno 4	139.31	143.73	168.11		112.20		177.49		183.03		219.16	
Kno 5	143.75	152.34	168.14		108.21	112.21	177.45		186.98		210.88	219.18
Kno 6	152.37		168.16		100.26	108.22	197.57		187.03		210.84	219.19
Kno 7	152.35		168.15		112.19		173.43	197.54	187.00		210.84	219.15
Kno 8	148.07		168.12		112.19				183.00	187.06	210.85	219.14
Kno 9												
Kno 10												
Kno 11												
Kno 12	143.73		156.09	168.09	96.15	108.24	197.5		183.00	187.04	210.81	219.14
Kno 13	143.72		168.17		108.28	112.27	177.45		183.00	187.06	210.85	219.10
Kno 14			168.12		108.22	112.21	177.43	197.51	182.98	186.97	210.82	
Kno 15	139.31		168.13		96.16		197.49		182.99	187.05	219.15	
Kno 16	143.68	152.32	168.08		108.19	112.24	197.51		187.03		210.81	219.11
Kno 17	143.75		168.12		108.19	112.20	177.46	197.57	187.01		210.76	
Kno 18	143.73		168.12		108.27		177.43		182.94		219.01	
Kno 19	143.75	152.36	168.12		108.23		177.54	197.55	183.03	187.07	210.76	
Kno 20	143.70		168.11				177.46	191.51	187.05		219.03	
Ldn 31	143.72		168.06	172.13	112.22		177.43	197.55	181.96		210.78	219.06
Ldn 32												
Ldn 33			168.1		112.21		197.55		183.00		218.98	
Ldn 34			168.11		112.19							
Ldn 35	143.75	148.11	168.1	172.17	112.27		197.52		182.03		210.72	219.00
Ldn 36	139.31		168.12	172.19	112.21		197.58					

Ldn 37	143.71		168.12		112.20		197.49		186.04		210.74	
Ldn 38	143.70		168.12		112.19		177.51		182.99	187.00	210.68	218.98
Ldn 39	143.75	152.33	168.12		100.32	112.24	177.49	197.57	182.98		219.01	
Ldn 40	143.67		168.07	172.15	100.25	112.17	177.47	197.53	182.96		210.68	218.98
Ldn 41	152.34		168.09		100.32				183.01			
Ldn 42	143.76		168.1	172.1	112.23		177.47		182.94			
Ldn 43	143.73	152.38	168.03		112.23		177.53	197.59	182.99		210.74	218.98
Ldn 44	143.75				112.16		197.52		183.01		210.73	218.95
Ldn 45	139.38		168.03	172.09	112.18		177.42		181.96		210.75	218.98
Ldn 46	139.31		168.03		112.26		177.44		183.01		210.70	
Ldn 47	139.31	152.35	168.07	172.12	100.32	112.26	177.45	197.57	181.96		217.93	
Ldn 48	143.70	152.31	168.06	172.1	112.19		177.43	197.53	182.95		210.71	218.97
Ldn 49	143.72	152.29	168.09	172.08	100.25	112.16	181.45	197.56	181.93	185.99	209.72	217.98
Ldn 50	143.70		156.03	168.09	100.25	112.14	197.53		181.95		210.68	
Dev 20			168		112.16		173.5	197.54	181.98	185.98	209.61	
Dev 21	143.73	156.45	167.99		99.87		197.52		181.91		209.71	217.97
Dev 22	143.71	156.44	168.08		100.26	112.19	197.5		181.90	185.94	209.67	
Dev 23			167.9		100.25	112.18	197.53		181.92		209.70	217.96
Dev 24	156.46		167.87		100.25	112.16	197.52		181.88		209.55	
Dev 25	156.35		168.04		112.17		197.54		181.86		209.69	
Dev 26	143.75	156.46	168.09		112.11		197.56		181.93	185.96	209.64	
Dev 27	152.25		167.97		112.12		197.54		181.86		209.68	
Dev 28	156.33		155.98	168.08	111.98		197.53		181.91		209.52	

Dev 29	156.44		167.96		100.19	112.16	197.51		181.91		209.65	
Dev 30	139.31		168.04		100.19	112.13	197.56		181.89		209.65	217.92
Dev 31	143.70	156.41	167.96		111.98		197.52		181.88		209.68	217.97
Dev 32	143.72	152.33	167.98		99.87		197.5		181.84		209.60	
Dev 33	156.43		168.04		100.19	112.14	197.53		181.91		209.68	217.98
Dev 34	143.66		168.02		100.19		197.55		181.89	185.93	209.68	
Dev 35	139.37	156.37	167.89		99.93		177.47		181.88		209.68	217.97
Dev 36	139.37	152.38	167.9		99.87		197.58		181.93		209.58	
Dev 37	139.25		156.07	168.06	100.19		197.52		181.88		209.76	218.00
Dev 38	152.35		167.99		112.14		173.49		181.94		209.66	
Dev 39	156.43		167.85		100.19	112.19	197.52		181.86		209.61	
Ham 1	152.38		167.95		112.11		197.51		181.87	185.89	210.72	
Ham 2	152.29		167.94		112.14		197.53		181.88	185.95	209.65	217.93
Ham 3	152.33		167.98		112.07		197.49		181.93	185.95	209.67	
Ham 4	152.35		168.04	172.01	100.19	112.13	197.53		185.92	189.99	209.72	
Ham 5	139.31	152.32	167.89		112.13		197.57		181.90		209.53	
Ham 6	139.31		167.96		100.19		177.43	197.55	181.89		209.71	
Ham 7	139.37		156.01	168.05			177.47		181.85		209.72	217.99
Ham 8	152.29		167.89		91.95	112.06	197.51		181.90	185.88	209.49	
Ham 9	139.31	152.27	167.87		91.97		177.47	197.49	181.86		209.47	
Ham 10	152.29		167.97		112.08		197.51		181.87		209.61	
Ham 11	152.29		167.9		112.11		197.47		181.85	185.88	209.68	
Ham 12	139.37	152.35	168.03		112.09		177.45	197.51	181.81		209.68	

Ham 13	139.31	152.32	167.87		99.81		197.55		181.84		209.45	
Ham 14	151.34		167.89		111.95		177.45	197.54	181.86	185.87	209.57	
Ham 15	151.30		168.01		112.11		197.52		185.92	189.94	209.66	
Ham 16	138.36	151.32	167.89		112.06		177.47	197.55	181.79		209.53	
Ham 17	151.34		167.92		112.11		197.53		185.86	189.94	209.59	
Ham 18	151.38		168.04		112.07		197.55		185.89		209.71	
Ham 19	138.29	151.34	167.88		111.95		177.46	197.52	181.84	185.87	209.74	217.96
Suf 1	142.67	147.04	167.95		112.12		177.43	197.49	181.83	189.91	210.70	
Suf 2	143.73		167.99		112.07		177.45	197.55	181.86		209.66	
Suf 3	135.04	152.32	168.01		112.13		197.53		181.84		209.72	
Suf 4	143.68	148.09	167.94		111.96		197.52		181.80		209.67	
Suf 5	148.07		167.88		112.04		177.45	197.49	181.81		209.69	
Suf 6	143.69	148.05	167.85		111.91		177.37	197.47	181.83		209.41	
Suf 7	143.71	152.34	167.95		100.13	112.15	177.39		181.79		209.69	
Suf 8	143.73	148.10	167.88		112.01		177.44	197.52	181.86		209.50	
Suf 9	152.35		168		100.19	112.08	177.43	197.55	181.75		209.68	
Suf 10	143.73	152.32	167.91		99.81		177.35		181.80		209.61	
Suf 11			168		111.96		197.49		181.79		209.67	
Suf 12			167.99		100.19	112.19	177.48	197.53	181.88		209.68	
Suf 13			167.98		112.15		197.56		181.76		209.71	
Suf 14			167.96		100.19	112.14	197.54		181.85	189.91	209.69	
Suf 15			168		100.12	112.09	197.49		181.80	189.86	209.69	
Suf 16			167.93		100.12	112.11	197.53		181.80	189.86	209.64	

Suf 17			168.05		99.94		197.5	201.63	181.86	189.93	209.76	218.02
Suf 18			167.93		112.04		197.56		181.82		209.57	
Suf 19			168.03		100.12	112.16	177.43	197.49	181.84		209.75	
Suf 20	143.72	152.30	168.02		112.08		197.54		181.83		209.79	
Ply 1	139.37		168.07		100.18		197.51		185.85		209.73	217.98
Ply 2			168.92		100.18	112.10						
Ply 3			169.39									
Ply 4			169.82									
Ply 5	151.32		168.09		112.09		177.51	201.62	181.85	185.86	209.77	
Ply 6			168.06	172.07	112.13				181.89	185.91	209.80	218.08
Ply 7												
Ply 8												
Ply 9												
Ply 10					100.18							
Ply 11			168.04		112.10		177.51	197.58	181.84	185.92	209.81	
Ply 12			168.09	172.08	100.18				185.91		209.83	
Ply 13												
Ply 14					100.18	112.15			185.88		210.82	
Ply 15	151.33		168.01		100.18	112.10	177.41	197.49	185.93		209.80	
Ply 16												
Ply 17					100.18				185.89			
Ply 18			168.05		100.18	112.15	197.49	201.58	181.89		209.78	
Ply 19			168.38									

Ply 20			168.86									
Nfo 1			168.06		112.15		177.51	197.52	182.85		219.12	
Nfo 2	152.36		156.06	168.06			197.56		182.81	190.89	210.77	219.02
Nfo 3	143.74	148.08	156.03	168.11	112.11		177.45	197.49	182.87		210.77	
Nfo 4	139.31		156.01	168.09	112.17		177.41	197.47	182.85	190.91	210.69	
Nfo 5	139.30		168.01		112.11		177.45	197.56	182.87		210.79	219.06
Nfo 6												
Nfo 7	130.78		168.1		100.18				182.84			
Nfo 8	152.36		168.02		100.18	112.15	197.53		182.80	186.84	210.70	
Nfo 9	152.33		168.01		100.24		193.49	197.5	182.82	186.90	210.78	
Nfo 10	139.31		167.95		91.91		177.45	197.56	182.86		210.77	
Nfo 13			156.04		100.18	112.18	177.47	197.56	186.87		218.96	
Nfo 14	139.31		167.97		112.09		197.48		182.87		210.66	
Nfo 15	152.28		167.98				197.51		182.86		210.68	218.98
Nfo 16	139.31		168.02		112.09		197.55		182.86		210.72	218.93
Nfo 17	139.31		167.99		100.24	112.14	197.51		182.85		210.73	
Nfo 18	139.31		167.92		112.05		197.54		182.84	186.91	210.69	
Nfo 19	139.30		167.97		112.07		197.52		182.80		210.65	
Nfo 20	139.31		156.06	168.06	91.97	112.13	197.57		182.91	186.96	210.70	218.97
Som 1	139.37		168.1		100.18	112.15	197.56		182.87	186.88	210.74	218.97
Som 2			156.03		100.24	108.17	177.5	197.55	182.81		210.65	
Gal 1	152.34		156	168.03	112.17		197.56		182.85	186.91	210.68	
Gal 2	152.31		156.03	168.07	112.17		177.43	197.55	182.83		210.65	

Gal 3	152.32		168.03		112.18		177.45	197.52	182.82	186.87	210.66	
Gal 4	152.38		167.97		112.07		177.43		182.83	186.86	210.68	
Gal 5			167.96		112.15		177.44		182.80	186.85	210.72	
Gal 6	152.36		155.97	168.06	112.06		177.36		182.92	186.95	210.70	
Gal 7	152.35		155.98	168.02	112.12		177.42		182.88		210.71	
Gal 8	152.34		167.95		112.19		177.49		182.84		210.67	
Gal 9	152.34		155.85		112.11		177.49	197.59	182.88		210.68	
Rir 1	152.31		167.87		100.18	112.11	177.41		182.84		210.61	
Rir 2			168.03		100.17	112.14	177.43		182.90		210.61	218.90
Number of Alleles Per Locus	6		5		6		8		5		5	

10.2 Manufacturer's Protocols

10.2.1 Gentra® Puregene Tissue Kit

Protocol: DNA Purification from Tissue Using the Gentra Puregene Tissue Kit

This protocol is for purification of genomic DNA from 5–10 mg or 50–100 mg fresh or frozen solid tissue using the Gentra Puregene Tissue Kit.

Important point before starting

- In some steps of the procedure, one of two choices can be made. Choose ■ if processing 5–10 mg tissue; choose ▲ if processing 50–100 mg tissue.

Things to do before starting

- Preheat water baths to 37°C for use in step 3, 55°C for use in step 2b, and 65°C for use in steps 2a and 15 of the procedure.

Procedure

1. **Dissect tissue sample quickly and freeze in liquid nitrogen.**

   Grind ■ 5–10 mg or ▲ 50–100 mg frozen or fresh tissue in liquid nitrogen with a mortar and pestle. Work quickly and keep tissue on ice at all times, including when tissue is being weighed.

2. **Dispense ■ 300 µl or ▲ 3 ml Cell Lysis Solution into a ■ 1.5 ml or ▲ 15 ml grinder tube on ice, and add the ground tissue from the previous step. Complete cell lysis by following step 2a or 2b below:**

2a. **Heat at 65°C for 15 min to 1 h.**

2b. **If maximum yield is required, add ■ 1.5 µl or ▲ 15 µl Puregene Proteinase K, mix by inverting 25 times, and incubate at 55°C for 3 h or until tissue has completely lysed. Invert tube periodically during the incubation.**

   The sample can be incubated at 55°C overnight for maximum yields.

3. **Add ■ 1.5 µl or ▲ 15 µl RNase A Solution, and mix the sample by inverting 25 times. Incubate at 37°C for 15–60 min.**

4. **Incubate for ■ 1 min or ▲ 3 min on ice to quickly cool the sample.**

5. **Add ■ 100 µl or ▲ 1 ml Protein Precipitation Solution, and vortex vigorously for 20 s at high speed.**

6. **Centrifuge for ■ 3 min at 13,000–16,000 x *g* or ▲ 10 min at 2000 x *g*.**

   The precipitated proteins should form a tight pellet. If the protein pellet is not tight, incubate on ice for 5 min and repeat the centrifugation.

7. **Pipet ■ 300 µl or ▲ 3 ml isopropanol into a clean ■ 1.5 ml microcentrifuge tube or ▲ 15 ml centrifuge tube and add the supernatant from the previous step by pouring carefully.**

   Be sure the protein pellet is not dislodged during pouring.

   Note: If the DNA yield is expected to be low (<1 µg) add ■ 0.5 µl Glycogen Solution (cat. no. 158930).

8. **Mix by inverting gently 50 times.**
9. **Centrifuge for ■ 1 min at 13,000–16,000 x *g* or ▲ 3 min at 2000 x *g*.**
10. **Carefully discard the supernatant, and drain the tube by inverting on a clean piece of absorbent paper, taking care that the pellet remains in the tube.**
11. **Add ■ 300 µl or ▲ 3 ml of 70% ethanol and invert several times to wash the DNA pellet.**
12. **Centrifuge for ■ 1 min at 13,000–16,000 x *g* or ▲ 1 min at 2000 x *g*.**
13. **Carefully discard the supernatant. Drain the tube on a clean piece of absorbent paper, taking care that the pellet remains in the tube. Allow to air dry for ■ 5 min or ▲ 5–10 min.**

    The pellet might be loose and easily dislodged. Avoid over-drying the DNA pellet, as the DNA will be difficult to dissolve.

14. **Add ■ 100 µl or ▲ 400 µl DNA Hydration Solution and vortex for 5 s at medium speed to mix.**
15. **Incubate at 65°C for 1 h to dissolve the DNA.**
16. **Incubate at room temperature (15–25°C) overnight with gentle shaking. Ensure tube cap is tightly closed to avoid leakage. Samples can then be centrifuged briefly and transferred to a storage tube.**

10.2.2 DNA Clean & Concentrator-5 PCR purification kit

Buffer Preparation

- ✓ *Before starting:* Add 24 ml 100% ethanol (26 ml 95% ethanol) to the 6 ml **DNA Wash Buffer** concentrate. Add 96 ml 100% ethanol (104 ml 95% ethanol) to the 24 ml **DNA Wash Buffer** concentrate.
- ✓ DNA Wash Buffer included with D4001S and D4001T is supplied ready-to-use and does not require the addition of ethanol prior to use.

Protocol

All centrifugation steps should be performed between 10,000 - 16,000 x g.

1. In a 1.5 ml microcentrifuge tube, add 2-7 volumes of **DNA Binding Buffer** to each volume of DNA sample (see table below). Mix briefly by vortexing.

Application	DNA Binding Buffer : Sample	Example
Plasmid, genomic DNA (>2 kb)	2 : 1	200 µl : 100 µl
PCR product, DNA fragment	5 : 1	500 µl : 100 µl
ssDNA[1] (*e.g.* cDNA, M13 phage)	7 : 1	700 µl : 100 µl

For efficient recovery of genomic or large DNA (> 20 kb to > 200 kb), use the **Genomic DNA Clean & Concentrator™ (Cat. Nos. D4010, D4011)**.

2. Transfer mixture to a provided **Zymo-Spin™ Column**[2] in a **Collection Tube**.
3. Centrifuge for 30 seconds. Discard the flow-through.
4. Add 200 µl **DNA Wash Buffer** to the column. Centrifuge for 30 seconds. Repeat the wash step.
5. Add ≥ 6 µl **DNA Elution Buffer**[3] or water[4] directly to the column matrix and incubate at room temperature for one minute. Transfer the column to a 1.5 ml microcentrifuge tube and centrifuge for 30 seconds to elute the DNA.

   Ultra-pure DNA is now ready for use.

For **Assistance**, please contact Zymo Research Technical Support at 1-888-882-9682 or e-mail tech@zymoresearch.com.

Notes:

[1] For ssDNA purification, see **Appendix A** on page 5.

[2] The sample capacity of the column is 800 µl. Therefore, it may be necessary to load and spin a column multiple times if a sample has a volume larger than 800 µl.

[3] **DNA Elution Buffer**: 10 mM Tris-HCl, pH 8.5, 0.1 mM EDTA

[4] Elution of DNA from the column is dependent on pH and temperature. If water is used, make sure the pH is >6.0. Waiting 1 minute prior to elution may improve the yield of larger (> 6 kb) DNA. For even larger DNA (> 10 kb), the total yield may be improved by eluting the DNA with 60-70 °C **DNA Elution Buffer**.

10.2.3 BigDye™ Terminator v3.1 Cycle Sequencing Kit

Perform cycle sequencing

Prepare the reactions

Set up the sequencing reactions

IMPORTANT! Protect dye terminators from light. Cover the reaction mix and sequencing plates with aluminum foil before use.

1. Completely thaw the contents of the BigDye™ Terminator v3.1 Cycle Sequencing Kit and your primers and store on ice.
2. Vortex the tubes for 2 to 3 seconds, then centrifuge briefly (2 to 3 seconds) with a benchtop microcentrifuge to collect contents at the bottom of the tubes.
3. Add components as indicated:

   IMPORTANT! Change pipette tips after each transfer.

   IMPORTANT! For control reactions, use 4 µL of the control primers for 20 µL and 10 µL reactions. Control primer concentration = 0.8pmol/µL.

Component	Standard reaction (20 µL)			Standard reaction (10 µL)		
	Quantity per reaction	Example Forward	Example Reverse	Quantity per reaction	Example Forward	Example Reverse
BigDye™ Terminator 3.1 Ready Reaction Mix	8 µL	8 µL	8 µL	4 µL	4 µL	4 µL
Forward primer (3.2 µM)	3.2 pmol	2 µL	—	3.2 pmol	1 µL	—
Reverse primer (3.2 µM)		—	2 µL		—	1 µL
Deionized water (RNase/DNase-free)	Varies based on template and primer volume	8 µL	8 µL	Varies based on template and primer volume	4 µL	4 µL
Template	See "Template quantity" on page 12	2 µL[1],[2]	2 µL[1],[2]	See "Template quantity" on page 12	1 µL[1],[2]	1 µL[1],[2]
Total volume	**20 µL**	**20 µL**	**20 µL**	**10 µL**	**10 µL**	**10 µL**

[1] e.g. 150-300ng/µL of dsDNA

[2] Concentration of template may affect volume, if template volume differs, adjust the volume of water in the reaction mix.

Note: Store on ice and protected from light.

4. Seal the plate with MicroAmp™ Clear Adhesive Film.

5. Vortex the plate for 2 to 3 seconds, then centrifuge briefly in a swinging bucket centrifuge to collect contents to the bottom of the wells (5 to 10 seconds) at 1,000 x g.

   Note: Bubbles may be present within the wells, but do not adversely affect the reaction.

Using BigDye™ Terminator v1.1 & v3.1 5X Sequencing Buffer to dilute sequencing reactions

Some cycle sequence reactions may be optimized using diluted BigDye™ Terminator Ready Reaction Mix. The BigDye™ Terminator Ready Reaction Mix is provided at a 2.5X concentration and can be diluted using BigDye™ Terminator v1.1 & v3.1 5X Sequencing Buffer to a final end reaction concentration of 1X.

Calculate the volume of BigDye™ Terminator v1.1 & v3.1 5X Sequencing Buffer to use:

0.5 * ((total reaction volume)/2.5) - volume of BigDye™ Terminator Ready Reaction Mix)

Note: Dilution of the BigDye™ Terminator v1.1 & v3.1 5X Sequencing Buffer without optimization, may cause deterioration of sequencing quality. We can not guarantee the performance of BigDye™ chemistry when it is diluted.

Component	Diluted reaction (0.5X)		
	Quantity per reaction	Example Forward	Example Reverse
BigDye™ Terminator 3.1 Ready Reaction Mix	4 µL	4 µL	4 µL
BigDye™ Terminator v1.1 & v3.1 5X Sequencing Buffer	2 µL	2 µL	2 µL
Forward primer (3.2 µM)	3.2 pmol	2 µL[1]	—
Reverse primer (3.2 µM)		—	2 µL[1]
Deionized water (RNase/DNase-free)	Varies based on template and primer volume	10 µL	10 µL
Template	See "Template quantity" on page 12	2 µL[2],[3]	2 µL[2],[3]
Total volume	**20 µL**	**20 µL**	**20 µL**

[1] The control primer is provided at 0.8pmol/µL. Use 8 µL to obtain a total primer quantity of 3.2 pmol per 20 µL reaction. If primer volume differs, adjust the volume of water in the reaction mix.

[2] e.g. 150-300ng/µL of dsDNA

[3] Concentration of template may affect volume, if template volume differs please adjust the volume of water in the reaction mix.

Perform cycle sequencing

Run the sequencing reactions

1. Place the tubes or plate(s) in a thermal cycler and set the correct volume:
   - 20 µL for microcentrifuge tubes or 96-well reaction plates
   - 10 µL for 384-well reaction plates
2. Perform cycle sequencing:

Parameter	Stage/step				
	Incubate	25 cycles			Hold
		Denature	Anneal	Extend	
Ramp rate	—	1°C/second			
Temperature	96°C	96°C	50°C	60°C	4°C
Time (mm:ss)	01:00	00:10	00:05	04:00[1]	Hold until ready to purify.

[1] Shorter extension times can be used for short templates.

3. Briefly centrifuge the reactions and proceed to Chapter 4, "Purify the sequencing reactions".