UNIVERSITY OF LIVERPOOL

The Application of a New Mass Spectrometry Technique Using Non-Invasive Biological samples for Conservation and Ecology Studies

Thesis submitted in accordance with the requirements of the University of Liverpool for the degree of Doctor in Philosophy by

Natalie Koch

April 2023

Table of Contents

Acknowledgements 5
Abstract 6
Covid-19 Statement 7
Abbreviations 8
Chapter 1: Introduction 10
1.1 Population Monitoring Methods for Conservation and Ecology Studies 11
1.1.1 Observational Monitoring 11
1.1.2 Tagging and Live-Trapping Methods 12
1.2 Non-Invasive Techniques to Investigate Mammalian Population 14
1.2.1 Camera Traps 15
1.2.2 Acoustic Monitoring 16
1.3 DNA Analysis of Non-Invasive Biological Samples 16
1.3.1 Hair 18
1.3.2 Faeces 18
1.3.3 Other types of non-invasive samples. 19
1.3.4 Disadvantages of DNA analysis 20
1.4 Mass Spectrometry 21
1.4.1 Ion Sources 22
1.4.2 Mass Analysers 23
1.5 Mass Spectrometry in Biology 25
1.6 Rapid Evaporative Ionisation Mass Spectrometry 26
1.8 Aims 30
Chapter 2: A REIMS Analysis of Faecal Pellets to Distinguish Sex, Maturity, and Strain of Lab Mice .. 32
2.1 Introduction 33
2.2 Method and Materials 36
2.2.1 Collection of faecal pellets from lab animals 36
2.2.3 Establishing a preliminary method for burning faecal pellets and a model to classify sex, maturity, and strain 37
2.2.4 Random Forest Analysis 44
2.2.5 Changing the REIMS method to increase classification accuracy (Coagulate Method) 49
2.2.6 Repeating study with an increase in the number of samples used and the amount of pellet hydration (Coagulate/Hydration Method). 53
2.3 Results 54
2.3.1 Optimising the REIMS Method 54
2.3.2 Results from the Coagulate Method 60
2.3.2 Results from the Coagulate/Hydration Method 64

2.3.2.1 Classification of Sex ... 64
2.3.4. Classification of Age ... 80
2.3.4 Classification of Age and Sex ... 87
2.3.5 Classification of Strain ... 91
2.4 Discussion ... 94
2.4.1 Optimising the REIMS Method ... 94
2.4.2 Using REIMS to Distinguish Intra-species Variation ... 99
2.4.3 Conclusion ... 102
Chapter 3: An Investigation into the Use of REIMS as a Non-Invasive Method of Pregnancy Detection in Okapi (*Okapia johnstoni*) ... 104
3.1 Introduction ... 105
3.2 Method ... 110
3.2.1 Faecal Sample Collection from Okapis in Chester Zoo ... 110
3.2.2 Using REIMS to Burn and Analyse Okapi Faecal Pellets. ... 111
3.3 Results ... 113
3.3.1 Using REIMS to distinguish pregnant and non-pregnant samples. ... 113
3.2.2 Changing Model Parameters to Increase Classification Accuracy ... 118
3.2.3. Early Versus Late Pregnancy ... 121
3.2.4 The Effect of Progesterone Concentration ... 124
3.2.5 The Effect of Storage Time on the Faecal Pellets ... 126
3.2.6 Comparison of Homogenised and Whole Faecal Pellets ... 134
3.4 Discussion ... 141
Chapter 4: REIMS Analysis of Stored Vole Pellets ... 146
The effect of storage time on the classification of RIEMS signatures using faecal samples from Bank and Field Voles ... 146
4.1 Introduction ... 147
4.2 Method ... 149
4.2.1 Wild Vole Trapping and Breeding ... 149
4.2.2. Sample Collection ... 151
4.2.3 Using REIMS to Burn and Analyse Faecal Pellets of Bank and Field Voles ... 152
4.3 Results ... 154
4.3.4 Discrimination of Vole Pellets by Storage Collection Year ... 154
4.3.5 Using Older Samples to Predict Species of Newer Samples. ... 163
4.3.1 Discrimination of Vole Species of Faecal Pellets Stored For up to Four Years ... 158
4.3.2 Discrimination of Sex of Vole Faecal Pellets Stored up to Four Years ... 163
4.3 Discussion ... 174
Chapter 5: Application of REIMS to Species Profiling in Field Studies ... 177

5.1 Introduction 178
5.2 Method 181
5.3 Results 186
5.3.1 Determining The Species of Faecal Pellets Collected from Inactive Traps 186
5.3.2 Species Distribution of Rodents Across Four Different Field Sites 189
5.4 Discussion 193
5.4.1 Comparison of the Three Different Methods Used to Determine Species Distribution 193
5.4.2 The Affect of the Population Boom 197
5.4.3 Conclusion 199
5.5 Ethical Statement 200
Chapter 6: Discussion 201
6.1 Key Findings 202
6.2 Further Work 205
6.3 Conclusion 209
Chapter 7: Publications 211
Publication contributions completing this PhD 211
7.1 Rapid identification of species, sex and maturity by mass spectrometric analysis of animal faeces 212
7.2 The application of rapid evaporative ionization mass spectrometry in the analysis of Drosophila species—a potential new tool in entomology 213
7.3 Social status and ejaculate composition in the house mouse. 214
References 215

Acknowledgements

I would like to thank my supervisors, Rob Beynon and Jane Hurst, whose support has been invaluable from helping me obtain funding for the project right through to completion. I will forever be grateful for their advice and guidance over the last four years they have helped me immensely to develop as a person and a scientific researcher. I feel very fortunate to have had the opportunity to work with such excellent supervisors.

I am also highly grateful for all the support and funding I have received from Waters both here in the UK and in Hungary, especially Mike Morris, Emrys Jones and Julia Balog, who were instrumental in making my CASE placement both enjoyable and informative.

I am thankful to everyone from the Centre of Proteome Research; this PhD journey would have been much harder without their continuous encouragement and advice. I am particularly grateful to Philip Brownridge and Joscelyn Harris, who helped train me on instrumentation and provided invaluable advice and assistance over the last four years.

I appreciate all the help and encouragement from my colleagues in the Mammalian Behaviour and Evolution group; they have made these demanding four years very enjoyable. I am particularly thankful to Amanda Davidson for all the training and support she provided in such an enthusiastic manner and to John Waters and Simon Gibson for their assistance and guidance with the animals.

I want to thank the staff at Chester Zoo, particularly Becky Mogey, for their support, enthusiasm and provision of samples for this project.

I am thankful to the undergraduate students who helped with this project, Emily Munro, Amelia Perring and Elizabeth Goodfellow, especially for their continuous enthusiasm after the impact of COVID.

I am exceptionally grateful to the BBSRC for funding this project and the events they organised to help develop my skills as a researcher.

I am incredibly thankful for my parent's support and encouragement over the years; their support allowed me to accept valuable opportunities across the globe so to develop as a person allowing me to manage the challenges of this PhD. I am also thankful to my husband, Simon whose support made the completion of the project possible. Finally, I am eternally grateful to my late Papa Ivan Dunne, who taught me the value of education and inspired me to pursue a career in science and obtain a PhD.

// Abstract

Effective conservation strategies are required to increase biodiversity. More information about a species increases the chances of choosing policies that enhance survival. Many historical population monitoring methods, such as live trapping, were invasive or difficult to implement for particular species or locations. This has led to the development of molecular analysis of non-invasive samples, including faeces. Faecal samples are easy to collect and store; they are a potentially rich source of information and do not require direct observation of the animal. Faecal samples can be used to obtain genetic information, but techniques are labour-intensive and time-consuming, and the abundance of hormones in faecal samples degrades with time.

The potential of a new ambient mass spectrometry technique to analyse faecal samples was investigated in this study. Rapid Evaporative Ionisation Mass spectrometry (REIMS) was developed for medicine to distinguish between cancerous and healthy tissue. It was used in food security to determine if the origin of a food source is as advertised. We have shown that REIMS can discriminate different species of rodents. This study aimed to determine the scope of REIMS to differentiate faecal samples of laboratory animals by analysing the sex, maturity and strain of different lab mice. The discrimination ability of REIMS was also explored in captive zoo animals to determine whether REIMS could be used to detect pregnancy. The practicality of REIMS in field studies was tested, and the ability to use REIMS as a method for population monitoring was determined by establishing the species distribution of rodents in multiple field sites. The power of REIMS to distinguish between more subtle differences; sex, maturity, strain and pregnancy; was limited compared to species. The composition of faecal samples did change with storage time, but REIMS could still distinguish between species of samples that had been in the freezer for over two years. REIMS established the species distribution of three rodent species across four field sites. Therefore, REIMS can be used as an additional non-invasive method to aid conservation and ecology studies.

Covid-19 Statement

The recent pandemic has significantly influenced this project, including the exclusion of two projects, delays in sample collection, complications in the analysis of zoo-derived samples and rescheduling of fieldwork. The rescheduling of fieldwork subsequently delayed a project to determine if REIMS could analyse the social status of male mice by comparing dominance and subordinance. This data would have been compared to the analysis of urine using ESI-TOF-MS mass spectrometry to explore the signature of major urinary proteins (MUPs). Before the lockdown in March 2019, I had started investigating the ability to use liquid-chromatography mass spectrometry for faecal analysis. Before using mass spectrometry, I had to develop a method of running faecal samples on acrylamide gels, and I was still developing the technique when the covid lockdown began. I did not continue with the project when the lockdown ended as it would have required close supervision contrary to covid policies at that time. The time needed to collect faecal samples increased significantly after the implementation of social distancing as only one person was allowed in the animal rooms at any time; this had a knock-on effect that pushed back all projects. The analysis of faecal samples in zoo animals was meant to involve multiple animal species, including identifying the sex of bird faeces. Undergraduate students were going to assist with the preliminary studies of analysis of faecal samples from zoo animals. Chester Zoo staff helping with the project were furloughed, and therefore only samples that had already been collected could be used in the analysis. Undergraduate students were not allowed in the labs, so their projects had to be changed to computer-based; this reduced the zoo chapter to just okapi faeces analysis.

Abbreviations

3Rs	Replacement, Reduction and Refinement
Laboratory Mouse strains	
BALB/cOLaHsd	BALB/c
C.C3-H2^{k}/LilMcdJ	BALB.k
C57BL/6JOLaHsd	C57BL/6
CE	Capillary electrophoresis
Coag	Coagulate
DA	Dalton
DART	Direct Analysis in Real Time
DC	Direct current
DNA	Deoxyribonucleic Acid
EI	Electron Ionisation
ESI	Electrospray Ionisation
FAB	Fast Atom Bombardment
GC	Gas Chromatography
GPS	Global Positioning System
Hsd:ICR (ICR(CD-1))	ICR(CD-1)
IUCN	International Union for Conservation of Nature
Lab	Laboratory
LC	Liquid Chromatography
LC-MS	Liquid Chromatography Mass Spectrometry
LDA	Linear Discriminant Analysis
LeuENK	Leu-enkephalin
MALDI	Matrix-assisted laser desorption/ionisation
MBE	Mammalian Behaviour and Evolution Group, Leahurst Campus, University of Liverpool
MS	Mass Spectrometry
MS/MS	Tandem Mass Spectrometry
MUPs	Major Urinary Proteins
m/z	Mass to charge ratio
NGS	Next Generation Sequencing
OMB	Offline Model Builder
PA	Prediction accuracy

PCA	Principal Component Analysis
PCR	Polymerase Chain Reaction
PIT	Passive integrated transponder
RF	Radiofrequency
Rf	Random Forest accuracy
REIMS	Rapid Evaporative Ionisation Mass Spectrometry
UAS	Unmanned aircraft system

Chapter 1: Introduction

1.1 Population Monitoring Methods for Conservation and Ecology Studies

Various methods can be used to obtain information about multiple species living in different environments. The data collected can aid optimal conservation management (Eggert et al., 2003); be used for pest control (Campbell et al., 2002); estimate population numbers (Bellemain et al., 2005); detect endangered animals (Ernest et al., 2000) and track particular individuals (Walton et al., 2018). Efficient methods for population monitoring are increasingly important due to the rapid decline of biodiversity. The IUCN lists 25% of all its assessed mammals as threatened and 14% of its birds. “Threatened” also includes those ranked as vulnerable, endangered or critically endangered (IUCN 2017). The biodiversity decline is so rapid that the earth is considered to be in its sixth mass extinction, with an average of two vertebrate species going extinct each year for the last 100 years. It would take 10,000 years for the same number of vertebrates to become extinct under normal background extinction rates (Ceballos et al., 2017). The causes of the population decline of vertebrates can be linked to the rapid expansion of the human population which has led to deforestation, habitat loss, climate change, and the introduction of invasive species (Ceballos et al. 2017).

There are multiple methods used to obtain information about a species’ behaviour, distribution and condition, including capture-mark-recapture methods (Morin et al., 2016, Bellemain et al., 2005), camera trapping (Rowcliffe et al., 2014), acoustic detectors (Lucas et al., 2015), tagging (Laplanche et al., 2015) and observation (Tena et al., 2017). Ideally, monitoring should be conducted without affecting the animal’s behaviour and be cost-effective and easy to implement. Every method used for monitoring has its advantages and disadvantages; disadvantages include being invasive, costly or impractical for certain species. Genetic and hormone analysis can provide non-invasive alternatives to traditional trapping and observational methods. However, these methods can be very time-consuming, require large amounts of sample preparation, and the methods used vary between species. They also may work for some species in specific locations and not others. Mass spectrometry is a potential type of analysis of biological samples that could address these issues.

1.1.1 Observational Monitoring

Historically monitoring species distribution was accomplished by having observers count animals; this could be achieved by walking transects or from a fixed position record all

observed animals. A more recent development, unmanned aircraft systems (UAS or drones), can be combined with Global Positioning System (GPS) technology and automatic detection to determine what species are present in an area and far they are distributed. Drones could detect koalas in the canopy using thermal signatures; the accuracy of detection by drones at lower heights (<60 m) was as accurate as conducting ground surveys (Gonzalez et al., 2016). The drone's height is critical; when they are too low, they disturb the animals, but increasing the height increases the rate of counting errors. Guanacos (a relative of llamas) were found to react to drones at heights greater than could be used for monitoring purposes (Schroeder et al., 2020). Other factors affecting the use of drones are location, amount of vegetation, the choice of camera and the time of day. Laws may restrict the use of drones, including flying beyond the operator's sight and flying at night, near people or infrastructure (Gonzalez and Johnson, 2017).

Counting could be achieved indirectly by observing and counting for signs of animals such as nests, faeces, tracks, partially eaten food and disturbance of plants. A study performed in South Africa compared an observational count to dung counts to estimate the species diversity of large mammals. The observational method required more sampling effort (more walking) than counting dung; dung was observed 45 times for every direct observation. The dung count results had a higher herbivore species richness compared to observational counts when surveying areas less than 25 km^2. Dung counts were more reliable for rare animals than direct observation (Cromsigt et al., 2009). One advantage of large mammals is the ease of identification of their faeces which are potentially still identifiable up to two months after defecation (if unaffected by other animals and climate). This is less true for smaller mammals whose faeces can be indistinguishable from each other and more challenging to find. The consequence of which is that dung counts cannot be used for all species (Figure 1.1). Finally, observational counts are labour intensive and underestimate rare or elusive species; they can be impractical under challenging terrains; dense rainforest and mountain areas are particularly difficult for observational methods (Witmer, 2005).

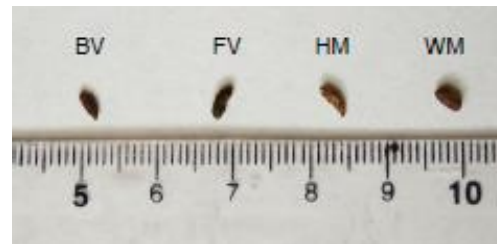
1.1.2 Tagging and Live-Trapping Methods

Mark-recapture methods are used to estimate population sizes without counting all individuals. A small number of wild animals are captured, marked, and released. Capture is repeated, and the number of marked animals is recorded. In small populations, animals with

African Elephant

Rhinoceros

Giraffe



Bank Vole, Field Vole, House Mouse, Wood Mouse

Figure 1.1. The faeces produced by a) an Africa elephant © Africa Geographic, b) a rhinoceros ©Shutterstock, c) a giraffe ©Shutterstock and d) four species of rodents © Nicola Davidson. The variation observed with the rodent faeces occur across species and vary from pellet to pellet. The large African herbivores produce faeces with distinguishable features that can be used for identification. The faeces of rodents are indistinguishable between the four species.

marks are more likely to be caught again; the opposite is true with large populations. There are multiple ways of marking an animal, and this will vary by species. For small mammals marking methods include ear-clippings, fur-clippings, toe-clipping and passive integrated transponder tagging (PIT-tagging). These marking methods are invasive, and retention of the marks can be problematic; small mammals may lose toes and physical tags in the wild (Jung et al., 2019). Tags can also be used to monitor and track animals through GPS and accelerometers, but tagging can harm the animal and may change their behaviour (Ropert-Coudert and Wilson, 2005). Tags have damaged the fins of sharks (Hammerschlag et al., 2014), lowered breeding opportunities and chick production of tagged penguins (Gauthier-Clerc et al., 2004) and lowered the pup growth rate of tagged Antarctic fur seal mothers

(Hammerschlag et al., 2014). There is the added risk of anaesthesia for larger mammals as the animal must be sedated to be tagged (Fahlman et al., 2005). Tagging can be expensive and labour-intensive, and electronic tags are expensive, particularly when considering the number of tags that fail to transmit data for the entire length of the study (Hammerschlag et al., 2014, Ropert-Coudert and Wilson, 2005).

Live trapping can be an effective method of monitoring small mammals. Live trapping of harvest mice was compared to nest counts to establish if harvest mice were present in different sites in Nottingham. Harvest mice nests were only observed in two of the four study sites, but harvest mice were caught in live traps in all sites. The observational method of nest counting underestimated the harvest mice population (Kettel et al., 2016). For population estimates, it is assumed that each individual has the same chance of being captured when using live traps. Some animals, however, may be considered 'trap-happy'. These animals may have been trapped before and will purposely go into traps for food and security. The probability of small mammals entering traps may increase if there is an increased predation risk leading to an overestimate of population numbers (Brehm and Mortelliti, 2018). Other studies suggest that live trapping is biased towards age, social status, or sex, as small mammals will avoid traps unless they have a strong need for the resources (bedding and food) found in the traps (Stryjek et al., 2019). Live trapping may be stressful for the animals; stress hormone (cortisol) concentrations of red squirrels increased with prolonged trapping. Stress levels varied with the type of trap used, but the traps that caused the least stress had the lowest capture rate (Bosson et al., 2012).

1.2 Non-Invasive Techniques to Investigate Mammalian Population

The impact of animal research and their stress and welfare is more of a priority than it once was. The Animals (Scientific Procedures) Act, implemented in the UK in 1986, regulates the use of protected animals (vertebrates and cephalopods) in research. It states that animals must be cared for in accordance with the best standards of modern animal husbandry. This was updated in 2006 to state reasonable steps should be taken to ensure the animal's welfare and needs (Animals (Scientific Procedures) Act 1986, 2012). Invasive methods have also been shown to impact animal welfare by increasing stress and influencing behaviour negatively (Gouveia and Hurst, 2019, Wolfensohn et al., 2018). Researchers must consider method

choices carefully before starting any project leading to an increasing interest in the analyses of non-invasive samples such as faeces and urine.

1.2.1 Camera Traps

Camera traps are a non-invasive population monitoring method used to monitor rare, elusive and nocturnal animals that usually evade direct observation (Trolliet et al., 2014, Wearn and Glover-Kapfer, 2019). For example, snow leopards were observed for the first time in Gaurishankar Conservation Area, which may be an important bridge between East and West Nepal. Such evidence can be used to prevent habitat destruction and increase the conservation of this endangered species (Koju et al., 2020). Camera traps can also be used to record animal behaviour, reducing the effect of human presence on behaviour and allowing multiple observers to review captured footage, reducing observer bias (Caravaggi et al., 2017). Camera traps can monitor individuals through natural markings, replacing traditional mark-capture methods (Macaulay et al., 2019). Camera traps were used to estimate the home range of brocket deer in Brazil and compared to home range estimates from GPS data. Deer were captured and anaesthetised using a dart gun. The captured deer were marked with an ear tag and given an individually marked radio-transmitter collar with a very-high-frequency transmitter and GPS. The collars recorded geographic point locations every 13 hours for one year. The sampling areas were divided into 20, 25-ha plots and camera traps were placed in each plot over one year, the location of the camera within the plot was changed every 30 days. Natural markings such as cuts and scars were used to identify deer captured on camera that did not have a radio collar. The GPS locations were used to estimate the home range of the deer, this was compared to the camera trap footage. The camera traps, however greatly underestimated the home ranges when compared to GPS data. The home ranges estimates of the camera traps were 80% of the estimates using GPS data. The results also differed for habitat use, GPS suggested that grasslands were the most utilised, but the camera traps suggested grasslands were the least used. The cameras in open areas such as grasslands are more exposed to environmental influences that may interrupt the camera’s operation, thus reducing capture events. Four individuals were identified from 21 camera trap records; this was only 5.3% of the total records of non-collared deer (Grotta-Netto et al., 2021). The results from this study suggest that camera traps are unsuitable for tracking individual deer through natural markings or establishing a home range. Natural markings can be a non-invasive

method of identifying individuals, but it requires the animals to have distinctive markings that remain stable with time. Yet cuts on ears can change with time, and small mammals can be almost identical when observed on cameras with limited resolution.

1.2.2 Acoustic Monitoring

Acoustic monitoring was developed to monitor cetaceans but has since been used for terrestrial studies (Enari et al., 2019). The advantage of acoustics is that sound can be collected over a relatively long distance, several sound-producing species can be monitored at once, and sounds can be identified using recognition algorithms (Hill et al., 2018). An investigation into acoustic monitoring of three call types of sika deer and five call types of macaques found that sounds were recorded faster than images caught using camera traps. However, recognition algorithms used to identify the calls produced many false positives. Some of the calls produced by the deer were only produced by dominant males in the presence of females; therefore, these calls could not be used to estimate population numbers (Enari et al., 2019).

1.3 DNA Analysis of Non-Invasive Biological Samples

Five non-invasive methods were compared for their ability to survey carnivores in North America. These were camera traps, track plates (collect footprints on paper using soot), scent stations, snow-tracking (track surveying after a fresh snowfall) and scat surveys. The carnivores surveyed included bears, coyotes, dogs, raccoons, opossums, skunks, weasels, martens, and foxes. The optimal method varied for the different species; camera traps worked best for most species except coyotes and smaller mammals. Scat surveys showed the highest abundance of coyotes, and the track plates showed the highest abundance of the smaller carnivores (weasels and martin). Track plates cannot be used for larger animals; snow tracking is biased towards animals active during winter and unsuitable for all environments (Gompper et al., 2006). Traditional methods may not be the best option for all studies involving the monitoring of animals; DNA analysis may offer an alternative to population monitoring.

Technological advancements have allowed individuals and species to be identified using genetic techniques to analyse biological samples. This has many advantages, including tracking and monitoring an individual animal and through genotyping, can inform efforts to

maintain genetic diversity. A genetic bottleneck is when a population declines significantly, which affects gene diversity. Inbreeding increases, causing a reduction in population fitness, and the ability of the species to adapt to change decreases (Lande, 1988). Genetics methods can be used to monitor populations at risk of a genetic bottleneck, and direct resources to those most at risk.

DNA can be extracted and analysed from various non-invasive samples. Identifying genetic diversity at the genome level will allow genetic threats to populations to be identified and the effect of humans on genetic variation (Hunter et al., 2018). Genetic analysis of biological samples is commonly achieved either by selective amplification of specific regions of the genome by the polymerase chain reaction (PCR) or by exhaustive genomic sequencing using massively parallel next-generation sequencing (NGS) approaches. PCR amplifies the DNA in a sample, primers are used to target the locus for amplification, and different primers are required depending on the genes and species being investigated. The amplified DNA is then analysed using gel electrophoresis or sequenced. NGS describes various sequencing technologies that can directly determine the nucleic acid sequence of DNA and RNA. The first step of NGS is to create a sequence library from the DNA sample; the isolated DNA sample is fragmented and then ligated with specific adaptor sequences on each end. The DNA fragments are attached to a solid surface (flow cell) and are amplified, usually by PCR. Different sequencing instruments read the DNA using fluorescently labelled nucleotides; each nucleotide will have a different label, and the fluorescent labels indicate the order of nucleotides in the DNA fragment. Millions or billions of strands are read and combined using assembly algorithms to produce the whole genome sequence. The challenge in using NGS is the amount of data that needs to be handled and analysed, which requires proficiency in bioinformatics (Tan et al., 2019). The technique can help build a species' life history and determine genetic defects and disease resistance. This information can help direct resources and influence decisions to obtain optimum conservation management. For example, a genome study found the golden jackal of Africa was more closely related to the grey wolf than the Eurasian golden jackal. This data led to the proposal that the African jackal should be considered a separate species from the Eurasian golden jackal, assigned its own conservation status, and considered separately when making conservation decisions (Koepfli et al., 2015).

1.3.1 Hair

Hair traps can be placed on the ground or attached to trees to collect hair samples from mammals, the traps will vary depending on the investigated species. Barbed wire and especially designed mats baited with scents and placed on the ground or tress have collected hair samples from bears (Berezowska-Cnota et al., 2017). The morphology of the hair may identify the species of the hair sample, or DNA analysis of the hair could be used to determine the species of origin (McKelvey et al., 2006), sex (Taberlet et al., 1993) and even individual identification (Weaver et al., 2005). For example, there have been increased efforts recently to reintroduce beavers to Europe. It is vital to monitor these reintroductions to determine the effect on habitat and how the beaver population will change with time. Beavers can be challenging to count using traditional methods as they are nocturnal and live in dens. A study by the University of Wrocław assessed the efficiency of passive hair traps. The hair traps were set up to monitor European beavers over 27 days, 12 hair samples were collected for a trapping rate of 0.7 per day. Using hair samples to monitor beavers is much more efficient than live-trapping or direct observation (Sobkowiak et al., 2021). Hair traps were used in the UK to determine the population and average group size of the European badger; it found that the population had nearly doubled since the 1980s. Hair traps (n=72) were randomly placed near a badger set; this was less than desired due to a lack of permission from landowners (Abu-Rabie et al., 2021). Hair traps are not always successful at collecting samples from all species. Hair traps were placed in forests in Malaysia with scent baits to encourage carnivores to interact with them and leave behind hair samples. Thirty traps were placed over one year but only two animals over the year rubbed the hair traps enough to leave a sample (Hedges et al., 2015). Even though hair samples can provide useful information beyond species identification hair is not a suitable sample choice for all mammal species and cannot be used for non-mammals.

1.3.2 Faeces

Faecal samples for biological monitoring have several benefits. It requires no handling or contact with the animal, making it easier for the researcher to collect. It is less stressful for the animal as sampling does not interfere with natural behaviours. For example, the population of brown bears (*Ursus arctos*) was estimated in Sweden using DNA from faeces to establish individual identification of samples. Volunteers collected 1904 faecal samples over

two years from brown bears. DNA was extracted from the faeces using a specifically designed commercial kit, Qiamp DNA Stool Kit. The DNA was initially amplified using PCR with one microsatellite. The faecal samples that showed a DNA band on an agarose gel were reanalysed using six microsatellite primer pairs. Microsatellites are small repeating DNA sequences subject to higher mutation rates and can be used to measure relatedness between species and individuals. Different primers were used for sex identification which required several pre-amplification steps. After amplification, PCR was used for genotyping individuals and each sample with four replicates. The methods used could analyse the DNA of 75% of the collected samples. This study compared the results from the DNA analysis to other methods, including radio collars. It determined that identifying individuals from the DNA analysis of faeces was reliable. The DNA analysis of all samples took 13 months to complete (20 hours of laboratory work a week) and did not include preliminary studies to establish suitable primers (Bellemain et al., 2005). Although this method can identify individuals, it would not be suitable if time was a limiting factor.

The species of carnivores historically have been identified through faeces using features such as size, shape, scent and dietary content. However, relying on gross appearance and features alone can be inadequate for some species. DNA barcoding can be used to identify a species from faecal samples; barcoding uses a standardised method that can be used for all mammals worldwide, giving it an advantage over other DNA analysis techniques. A short DNA sequence unique to a species can be compared to a reference database to identify the species. This technique can be used to detect the species of animals eaten by predators. This technique is still subject to the usual PCR errors; the benefit is it only relies on small sections of DNA, so degradation of the sample is less of an issue (Zeale et al., 2011).

1.3.3 Other types of non-invasive samples.

Many mammals produce scent marks for olfactory communication, including marking territories; DNA from the secretions produced from scent marking can be used for analysis (Malherbe et al., 2009). Up to four scent mark samples were collected from three captive tigers by swabbing the marked area. DNA was extracted from the swab using the QIAGEN DNeasy Blood and Tissue Kit protocol. Five microsatellite loci were used to identify individuals of all but one sample (Caragiulo et al., 2015). DNA was successfully extracted from the solid

scent marks produced from the anal pouch of Hyaenas (Malherbe et al., 2009) and the scent secretions produced by the scent glands of giant pandas (Ding et al., 1998).

DNA has also been extracted from wolf urine; wolves urinate more often than defecate, so urine samples are easier to collect. The urinary DNA was successfully amplified and analysed using PCR for 80% of the collected samples. Urine with DNA concentrations over 32.6 pg/µL could be used to determine the sex and individual identification of the sample. Urine could only be collected from males as males urinate more to mark territories, and it was only possible to collect urine samples when there was snow. (Hausknecht et al., 2007). Urine was only observed when a fresh sheet of snow was collected along with the urine; the snow-urine mixture was analysed. Without the snow, the urine would not have been observed in the ground before it could be collected. Although this method could be used with wolves' urine, it could not be used for other species that inhabit regions with no snow or produce small volumes of urine.

1.3.4 Disadvantages of DNA analysis

A problem with genetic analysis of non-invasive samples is the DNA extracted from these samples can be of low quality and quantity. DNA will degrade in samples that are left exposed, and the degradation rate will be higher in tropical countries (Caragiulo et al., 2015, Smith and Wang, 2014). DNA degradation can make many samples unsuitable for analysis as primers find it harder to bind to the DNA. This can lead to smaller sample sizes, which is particularly problematic for rare and elusive species with small population sizes. Lower-quality DNA increases the chance of genotyping errors, which occur at a higher rate in faecal samples due to PCR inhibitors. PCR inhibitors are any substance that affects the PCR; they may be derived from the sample or introduced during the PCR process (Schrader et al., 2012). Samples that have been stored for longer increase the chance of primers targeting the wrong loci. Allelic dropout is also typical for non-invasive samples when only part of the targeted gene is amplified (Shestak et al., 2021). An investigation of the effect of sample size and genotyping error on estimates of genetic variation found that sample sizes needed at least ten individuals (Smith and Wang, 2014). A study that collected genetic information from hair and faecal samples from bears in Italy had to exclude 36% of hair samples caught in hair traps, 85% of the hair samples found opportunistically, and 73% of the faecal samples as the DNA was not of a high enough quality (De Barba et al., 2010). The DNA analyses from several sample types

(tissue, blood, hair and faeces) of carnivores could differentiate different sub-species of foxes but could not differentiate between species within the canine family (De Barba et al., 2014).

In summary, genetic analysis of biological samples is a valuable tool to compare the genetics of species and determine relatedness. Not all methods are reliable for all applications, and no universal method is currently available (Arif et al., 2011). Genetic analysis is very labour-intensive and time-consuming and may not be appropriate for observing specific phenotypes. An alternative to genetic analysis could be mass spectrometry.

1.4 Mass Spectrometry

Mass spectrometry is an analytic technique that measures the mass-to-charge ratio of ions. For mass spectrometry analysis to occur, the ions must be in the gas phase, as virtually all mass spectrometers operate in a vacuum. When the ions enter the mass spectrometer, they are resolved according to their mass-to-charge ratio and their intensity is detected. When abundance is plotted against the mass-to-charge ratio, the result is a mass spectrum. The highest peak is referred to as the 'base peak' and is often assigned a relative abundance of 100%. In the gas phase, the energies can be manipulated such that the ion fragments into smaller ions. The mass spectrum can be unique as each ion can fragment differently, allowing the mass spectrum to be used to identify different molecules or elements.

All mass spectrometers contain a sample inlet, an ionisation source, a mass analyser, and a detector. The inlet introduces the ions into the mass spectrometer and may be combined with the ionisation source. An ionisation source is where the ionisation of the sample occurs as particles must carry an electric charge to be analysed. Once the ions are formed, they pass through the mass analyser; the motion of ions through the analyser will depend on many parameters, including their mass-to-charge ratio. The ions will therefore be ejected from the analyser at different time points to be measured. Ions will then encounter the detector, producing a signal (usually electrical) that can be measured. Ion analyses must occur under a vacuum so sample ions reach the detector without colliding with gas molecules, except for ion mobility mass spectrometry. (Figure 1.2). In ion mobility, ions move through a region with inert gas, and various ions will behave differently after colliding with the background gas affecting the ion's transit time. For more complicated samples, chromatography techniques can resolve analysts before being introduced to the mass

spectrometer. Gas chromatography (GC-MS), liquid chromatography (LC-MS) and capillary electrophoresis (CE-MS) can all be coupled with a mass spectrometer (de Hoffmann and Stroobant, 2007).

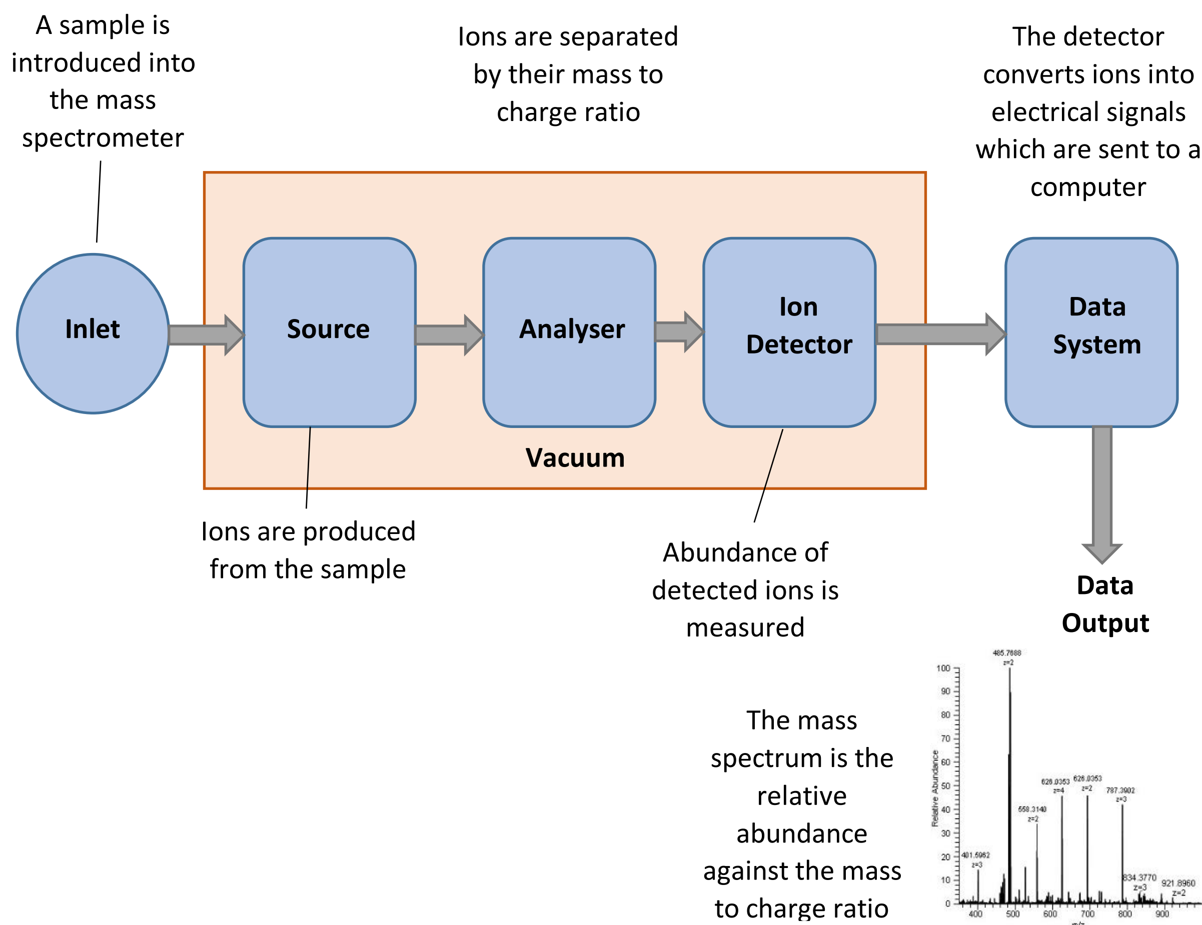


Figure 1.2. The elements are found in all mass spectrometers, although the inlet and source can be combined. The analyses and detector must be in a vacuum to prevent the sample ions colliding with gas molecules. The data system can also be used to change settings of the mass spectrometer such as voltages

1.4.1 Ion Sources

There are multiple ion sources; the first developed was Electron Ionisation (EI), originally known as Electron Impact ionisation. High-energy electrons produce ions from a sample in the gas phase. Electrons are emitted by a cathode and are accelerated through an ionisation chamber towards an anode. The sample enters the ionisation chamber and is bombarded by electrons which cause the molecules within the sample to fragment, producing ions (Nier,

1947). EI is a hard ionisation technique that produces significant fragmentation; soft ionisation uses less excess energy, so less fragmentation occurs.

Fast Atom Bombardment (FAB) is a soft ionisation technique that can analyse non-volatile analytes. The analyte is dissolved into a matrix solution (usually glycerol) and placed on a target plate. A beam of neutral atoms is directed toward the analyte, energy is transferred to the analyte causing desorption and ionisation, and the ions are accelerated into the mass spectrometer (de Hoffmann and Stroobant, 2007).

Electrospray Ionisation (ESI) is particualty valuable for fragile volatile bio-molecules. The sample is dissolved in a solvent and ejected through a narrow metal or glass capillary when using an EI source. A high voltage is applied to the tip; the emerging sample is broken down into an aerosol of highly charged droplets, the 'electrospray'. Applying a high temperature and a gas (usually nitrogen) directs the electrospray towards the mass spectrometer down a pressure and potential gradient. The charged droplets decrease in size as the solvent evaporates and increases surface charge density. This continues until a critical point is reached and the ions at the droplet's surfaces are ejected into the gaseous phase. (Ho et al., 2003). The ions are directed into the mass spectrometer through a skimmer (ion funnel) which focus the ions then an electric field accelerates the ions towards the mass analyser (Kelly et al., 2010).

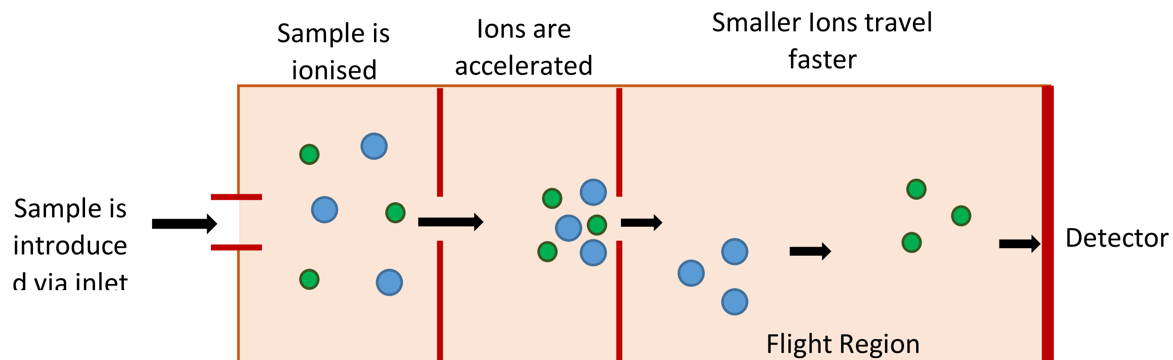
Matrix-Assisted Laser Desorption Ionisation can ionise large, thermally liable, non-volatile compounds. A UV laser is directed to a sample mixed into a highly absorbing crystalline matrix. The matrix absorbs the UV light and converts it to heat energy; the outer surface of the matrix heats rapidly and vaporises with the sample, during which the sample molecules are ionised (de Hoffmann and Stroobant, 2007). MALDI is a similar technique to FAB, but MALDI can ionise analytes with much higher molecular masses

1.4.2 Mass Analysers

Time-of-Flight analysers separate ions by their mass-to-charge ratio (m/z). An electric field accelerates ions through a field-free region (flight tube) of a mass spectrometer. Ions of the same charge will have the same kinetic energy, but velocity depends on mass to charge ratio. Ions with smaller masses move faster through the mass spectrometer. Reflectors can be added to the flight tube to increase the flight path; the ions spend more time in flight, so

more separation and higher mass resolution can occur (Figure 1.3) (Mamyrin, 2001, Chernushevich et al., 2001).

a) **TOF – Mass Spectrometer**



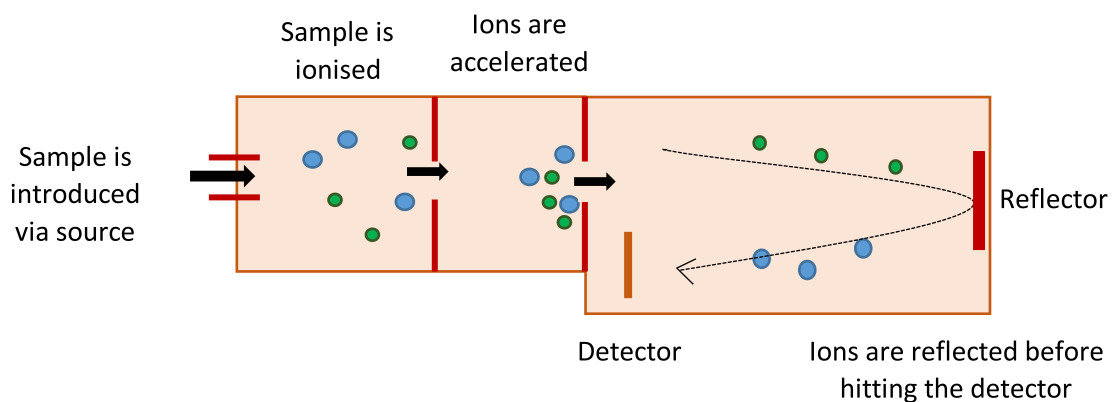


Figure 1.3. a) A TOF mass spectrometer with a linear flight paths, the samples ions are accelerated to enter the flight region with the same amount of energy. Small ions will travel through the flight region faster than large ions. b) A TOF mass spectrometer with reflector, this increases the flight paths and ions have more time to separate by size.

Using oscillating DC and RF electric fields, quadrupole analysers separate ions by mass-to-charge ratios. A quadrupole consists of four rods, and direct and alternate currents are applied so opposite rods have the same voltage and perpendicular rods have opposite polarities. An oscillating electric field is produced between the rods; how an ion moves through a quadrupole depends on the m/z. The voltages of the rods can be changed so ions of a particular m/z ratio will move through the whole quadrupole while other ions will be

deflected and will not hit the detector. Quadrupoles can analyse a single m/z value and be combined with Time-of-flight mass spectrometer analysers (Chernushevich et al., 2001, de Hoffmann and Stroobant, 2007).

Ion traps are regions that can hold ions for an extended time, there are various types of ion traps, and they work similarly to quadrupoles. Ions become trapped in a voltage field by applying an alternate current to a ring electrode and a direct current to end caps above and below the ring. The ions will remain in the space between the rods until the voltage changes; depending on the applied charge, ions of a particular mass-to-charge ratio will be pushed out through holes in the end caps and hit the detector (March, 1997). These techniques can be combined into one instrument known as tandem mass spectrometry (MS/MS). MS/MS is any method with more than one mass analysis stage; it usually involves the analyses of the productions of a particular precursor ion (de Hoffmann and Stroobant, 2007).

1.5 Mass Spectrometry in Biology

Since its invention in the early 1900s, mass spectrometry was first used to measure the mass of the atom and was mainly only used by industrial chemists until the 1950s. Over time, however, its importance in researching organic material grew, especially when it was discovered that a mass spectrometer could be used to determine the structure of molecules. The identification of biological samples by mass spectrometry began in earnest in the early 1970s. Before this, analyses were only possible for volatile substances as electron and chemical ionisation required the molecules to already be in the gas phase. Ions from liquid and solids states could be analysed by developing desorption ionisation methods, including fast atom bombardment (FAB) (de Hoffmann and Stroobant, 2007). These developments allowed biomolecular analysis by mass spectrometry and detecting compounds in much smaller quantities of tissue samples. A study of gout and its treatment was one of the first to use mass spectrometry on biological tissue. It detected the presence of five molecules (hypoxanthine, xanthine, uric acid, allopurinol and oxipurinol) in skeletal muscle tissue (Snedden and Parker, 1971). The subsequent development of mass spectrometry in biological research came in the 1980s. Advancements including Ion traps and MALDI, allowed for the analyses of macromolecules, including proteins (Griffiths, 2008).

The biological samples from animals already used in mass spectrometry analysis vary greatly, including tissue from mice (Hsu et al., 2015), milk from ungulates (Yang et al., 2014), hair from carnivores (Hollemeyer et al., 2007) and deer antler velvet (Arzmi et al., 2021). MALDI-TOF-MS was used to identify the presence of peptides in hair samples. The identified peptides of type 1 and 11 keratin proteins were used to create a flowchart for identification at the genus level of the hair sample. Hair samples were collected from museum skin samples of mammals from North America. This method could separate goats from sheep and a lynx from a puma. However, a lynx couldn't be separated from a closer relative, the bobcat, and wolves and coyotes could not be differentiated (Solazzo, 2017).

LC-MS is widely used for the detection of hormones in animals (Weltering et al., 2012) from urine, including detecting steroids (doping) in horse racing (Wong et al., 2012) and illegal growth hormones in livestock (Regal et al., 2009). The advantage of using LC-MS over other methods such as assays, is its ability to simultaneously measure many hormones and metabolites. The disadvantages are the overall capital cost (instrument cost) and the required sample preparation steps (Murtagh et al., 2013). LC-MS can be used with faecal samples; several types of hormones, including cortisol, testosterone and progesterone, were measured from the faecal samples of capuchin monkeys (Weltring et al., 2012). Cortisol was also shown via LC-MS to increase in cows that experienced stress due to transport and novel housing conditions (Mostl et al., 2002). Mass spectrometry may offer a new technique to analyse faecal and urine samples; this would have seemed impossible years ago as these samples would have been considered "too dirty" to enter a mass spectrometer. The same sentiment was made about macromolecules in the 1950s (Griffiths, 2008), and once again, new ionisation techniques have extended the capacity of analysis of non-invasive samples.

1.6 Rapid Evaporative Ionisation Mass Spectrometry

Zoltan Takáts and his team developed a new type of technology that uses mass spectrometry with surgical instruments. The new method, termed "Rapid Evaporative Ionisation Mass Spectrometry (REIMS)", was able to evaporate biological tissue quickly (Schafer et al., 2009). The technique was based on electrosurgical cutting, whereby an electric current was applied to a tissue sample using a surgical electrode. The burning caused the sample's water content

to evaporate and ionise; as a result, the tissue was broken down and a ‘smoke-like’ aerosol was produced. The aerosol was transported to a mass spectrometer using a Venturi pump to be analysed. In the first REIMS study, porcine liver and kidney samples were used, and the mass spectra produced ions in both polarities between m/z 600 to 900. Further, tandem mass spectrometry work showed the predominant ions were glycerophospholipids. The spectra changed between tissue types, and principal component analysis could be used for tissue

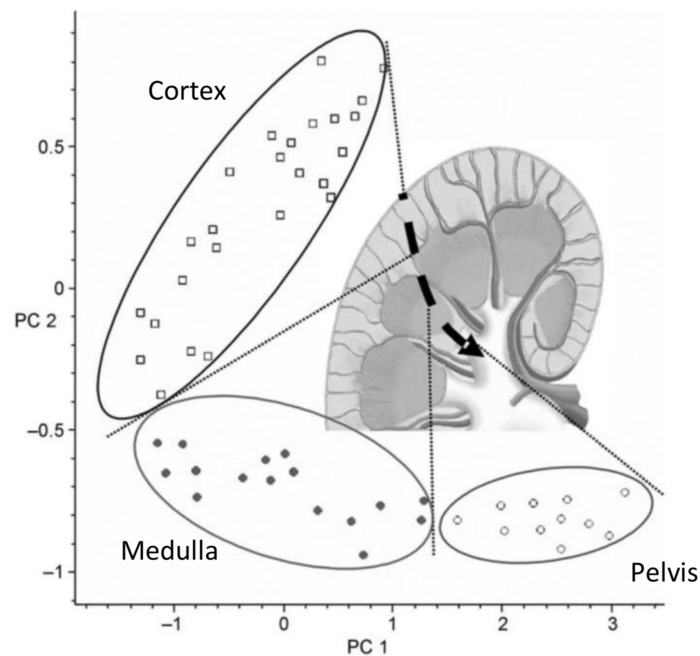


Figure 1.4. The PCA plot from Schafer et al., 2009 shows the first two principal components could separate tissue samples from different parts of a kidney that had been analysed using REIMS. The arrow indicates the direction of the transect by the electrosurgical electrode

identification (Figure 1.4) (Schafer et al., 2009).

Another early REIMS study found the spectra produced from burning various tissue samples from canine organs, including lung, liver, pancreas and muscle tissue, could be separated using PCA. Similar results were found for samples from cancerous tissue from the same organs. The differences between the spectra were due to differences in the abundance of ions from lipid components rather than unique molecules specific to one type of tissue. The identity of the ions was confirmed using MS/MS experiments and included sulfatides, phosphatidic acids, eicosanoids, lysophospholipids and cardiolipins. The study was repeated to identify tissue origin from rats, including kidney, liver, lung, spleen and testicle tissue. The

rats were fed on three different diets with increasing levels of fatty acids since phospholipids are thought to be responsible for the differences in spectra. The change in diet, however did not affect the ability to classify the different tissue types. Some separation was observed between the three diets in myocardial tissue, but this did not affect the ability to identify the tissue. Spectral changes were also observed with the rats' age but did not affect tissue origin classification (Balog et al., 2010). REIMS was initially envisioned to be used with cancer *in situ*; the advantage of REIMS is that the speed and lack of sample preparation steps meant tissue samples could be identified in real-time during surgery using standard surgical tools (Edward et al., 2017).

REIMS was able to identify the species origin of meat which demonstrated the potential of REIMS to be used for species identification. REIMS could distinguish beef from horse meat; a leave-one-out cross-validation of the data produced 100% classification accuracy of horse and beef. A leave-20%-out cross-validation produced a 97.5% classification for types of beef (Balog et al., 2016). REIMS can identify the species of five types of white fish (cod, coley, haddock, pollock and whiting). A monopolar electrosurgical knife was used to cut the tissue samples (n=478) 8-12 times, and each cut lasted 3-5 s. The knife was connected to the REIMS ionisation source and a Xevo G2-XS quadrupole time-of-flight (QTof) mass spectrometer. The raw data was uploaded to a prototype recognition software from Waters (OMB) and were lockmass corrected using LeuENk and normalised. LDA analysis was performed on the top 25 PCA components of 80% of the samples that were correctly identified, a leave-20%-out cross-validation of the same data produced a classification accuracy of 98.99%. The training model was able to correctly identify 98 of 99 test samples, one cod sample was not identified. The catch method of haddock, line-caught (n=35) or trawl caught (n=65) could be identified, and a leave-20%-out cross-validation of the PCA-LDA models resulted in a 95% correct classification of the catch method. It is unknown what caused the differences between the methods; the two methods caught fish from different depths, so that the fish may have had different diets, or the two methods caused different stress markers (Black et al., 2017).

REIMS has been used to identify the species and sex of five *Drosophila* species. A total of 800 adult *Drosophila* were killed by freezing and stored for up to 6 days at -20°C. Samples were then defrosted before being burned with a monopolar diathermy electrode. A TOF-MS analysed the aerosol produced from burning an entire specimen with the REIMS ionisation source. The mass spectra were binned and analysed by PCA-LDA, emphasising the resolution of the five species. The separation correlated with phylogeny, and the most related species (*D. melanogaster and D. simulans*) had the most similar linear discriminant values (Figure 1.5). The samples could also be separated by sex even when all species were combined in the same PCA-LDA model (Wagner et al., 2020)

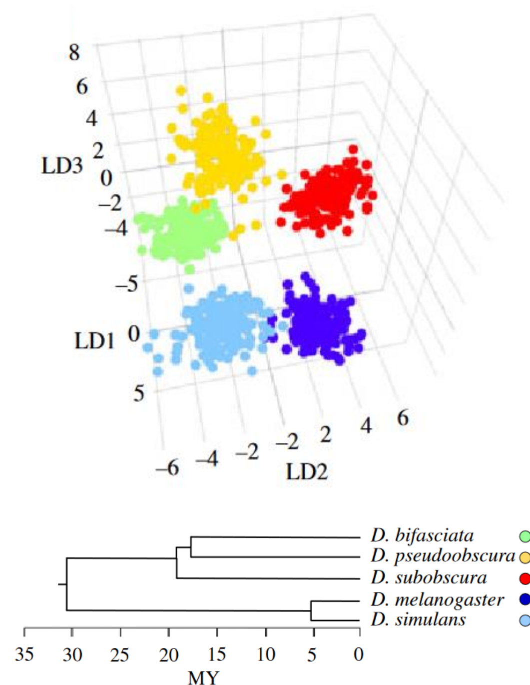


Figure 1.5. Adapter from (Wagner et al., 2020), the species discrimination of five Drosophila species (800 samples in total) by REIMS. PCA-LDA separation show that the data points correlate with the phylogenetic relatedness of the five species.

These studies showed that to a high degree of accuracy, REIMS could be used for species identification when using tissue samples. However, a non-invasive sample such as faeces would be required for ecological purposes. REIMS can be used to identify the species of insects and mammals, but these methods have used either tissue samples or whole organisms

and are, therefore, invasive. There are many advantages of using faecal samples to investigate the phenotypic features of an individual.

1.8 Aims

Traditional methods for conservation and ecology studies are not always suitable for all species and conditions. Many genetic analyses method are suitable for obtaining the information needed from animals to help with their conservation. However, these methods can be time-consuming and expensive, limiting their usefulness. REIMS analyses of faeces may provide an additional method to help with conservation and ecology studies.

This study aimed to determine how valuable REIMS analysis of faecal samples could be for conservation and ecological studies. Deciding resource distribution and conservation management is simpler if more information about target species is known. REIMS could also offer a more accessible and non-invasive method to obtain the required information. A previous investigation showed that it was possible to identify the species of rodent a faecal pellet originated from using REIMS.

The first aim of this study addressed in chapter 2 was to investigate differences within a species, including sex, age and strain of lab mice. This would test the capability of REIMS to detect features that are arguably more similar than species. The study would aim to obtain the highest classification possible for the three factors by changing various parameters of the REIMS method. By determing what parameters produced the highest classification accuracy would help establish a step-by-step approach for all REIMS analyses of faecal samples.

Since breeding programmes play a significant role in conservation, chapter 3 aimed to determine if REIMS could distinguish hormonal differences such as pregnancy. The aim was achieved by investigating the faecal pellets of two captive okapis. Random forests were used to classify faecal samples collected when the animal was pregnant or not. This project would also help determine if faeces from ruminants could be analysed with REIMS and if it could distinguish between individuals. The results from this project suggested that long-term storage of the faecal pellets may have impacted classification accuracy.

Many conservation projects take part over many years, so it is vital to know the effect storing samples in a freezer could have. The REIMS instrument may drift over time, so keeping

pellets and burning them all in one sitting would be more beneficial than when they are collected. Th study in chapter four aimed to determine how much storage influences REIMS analysis. Faecal pellets were collected from bank and field voles over four years, kept in a freezer, and then analysed with REIMS simultaneously. Random forests were used to classify the samples by species and sex to determine if classification accuracy decreased the longer samples were stored.

For REIMS analysis to benefit conservationists and ecologists, it must be able to contribute to fieldwork studies. The final project (chapter 5) aimed to determine the species distribution of rodents from various field sites using REIMS. The results were compared to two other methods, camera traps and live trapping. If the results using REIMS were comparable to the other methods, it would confirm the suitability of the REIMS method for conservation and ecology projects.

Chapter 2: A REIMS Analysis of Faecal Pellets to Distinguish Sex, Maturity, and Strain of Lab Mice

2.1 Introduction

REIMS was first used with faecal samples to classify the species of different rodents. Faecal pellets were collected from five species of lab rodents, including bank voles, field voles, wood mice, house mice and a randomly segregating cross of Wistar × Brown Norway laboratory rats. Faecal signals produced spectra that were ion rich in the 600 to 900 m/z range that rarely differed between individuals of the same species. The pattern of ions rather than ion identification was used to analyse the faecal pellets. REIMS could differentiate between the five species using random forest models with an accuracy of 83%. Rats had the highest identification accuracy at 95%. Samples were also stored under different conditions, including being left for up to 28 days at ambient temperature, at -21°C, 4°C and 21°C. Changing the storage of the faecal pellets did not affect the ability to classify species. House mice were placed on four different diets over four weeks; there was a slight decline in house mouse classification but not enough to suggest that diet had an effect. Faecal samples were also collected from wild animals, except house mice which were exceptionally difficult to catch. Lab house mice on varied diets were used with the other four wild-caught species. Identification accuracy of at least 91% was obtained for each wild animal species; the highest of 97% was for rats and field voles, while bank voles had the lowest accuracy. This random forest model could predict 25 unknown mice samples and 11 bank voles with an accuracy of 94%. This preliminary study shows the potential of REIMS to differentiate between faecal pellets (Davidson et al., 2019).

For most animal species, the ratio of males to females in a population is 1:1; sex ratios can be influenced by environmental factors, genetic disorders, behaviour differences or human influence (hunters may choose one sex over the other). In the wild, if the difference in sex ratio becomes too extreme, it could have detrimental effects; it can slow down population growth or increase the rate of adverse genetic effects (Wedekind, 2012). The Kakapo became extinct in certain regions due to the sex ratio becoming weighted towards males because of high levels of predation of females (Tella, 2001). Conservationists may require the sex ratio of a population when making decisions about what animals to release back into the wild. Many animals do not display sexual di-morphism at any stage of development or only show characteristic differences after maturation making it difficult to determine sex on observation alone. Due to deforestation, the maned sloth (*Bradypus*

torquatus) is listed as vulnerable to extinction (IUCN, 2017) and is difficult to sex. Some differences exist, such as females having larger bodies and smaller, lighter-coloured manes. These differences are only observed in reproductively active individuals; sloths reach sexual maturity after two to three years (Lara-Ruiz and Chiarello, 2005). Sloths can be challenging to find in the wild as they are arboreal, but they come down to the ground once a week to defecate (Pauli et al., 2014). Being able to determine if these faecal samples belong to males or females could be beneficial to conservationists.

A species generation length is one of the parameters that can be used by the IUCN red list to determine the risk of a species becoming extinct. Generation length is the average age of parents of the current brood; to be calculated, it requires the age of a female's first reproduction (Luba et al., 2020). A study that examined population trends from two large databases aimed to determine if certain risks could predispose mammals to population decline. Biological data were collected from the Living Planet database, including body mass, litter size, home range and age at first birth. Information about a species population decline was collected from the IUCN mammal species database. A multivariate model was created to find traits more likely to cause a population decline. It found that animals with a higher age when giving birth and under the threat of habitat loss were significantly declining (Collen et al., 2011). It was found that changes to the age structure due to human influence on groups of Asian elephants had affected the population dynamics. Data was collected from Myanmar elephants between 1970 and 2014. Early in the study, the population was weighted towards younger individuals, then became balanced by the year 2000. Elephants could become reproductive from age 12, reaching a peak between 18 and 24 and then declining from 55 years old. The population increased until 1980, decreased until 2000 and increased again. This pattern correlated with age structure; the population declined when the age balance was more varied, and there was a lower number of older individuals than young juveniles. The study also found that the decline was higher than expected due to environmental factors, which may have caused the decline in birth rates (Jackson et al., 2020). Knowing the age dynamics would be helpful for conservationists as they can redirect sources to the populations with more significant differences.

Conservationists would benefit from obtaining more than the species from faecal samples. This study aimed to determine if REIMS could be used to classify differences within

a species, such as age and sex and between different strains. Using REIMS to distinguish between different strains of lab mice would suggest its potential to separate sub-species. Faecal samples were collected from four strains of lab mice to investigate if REIMS could identify if a faecal pellet was produced from a male or female, a juvenile or adult, or what strain of lab mouse. The classification was achieved using random forests, a learning model that can be used for classification or regression; the model consists of multiple decision trees that vote on a class type. The model must be trained using data where the classification is known; this model can be used to predict samples of unknown classification. When creating a decision tree, the random forests use bootstrapped data and only a subset of all variables. The bootstrapped method is a resampling technique, samples are selected randomly from the dataset, and the same sample could be selected more than once in the bootstrapped dataset. The random forest model prevents overfitting by using only bootstrapped data and a random subset of variables. Over-fitting is when a model uses random differences to establish a relationship between variables; a pattern may be found in a specific dataset but would not be in a very similar dataset.

Principal component and linear discriminant analysis were also utilised. These dimension-reducing techniques allow for better visualisation of the data; they were not used for classification due to the higher probability of over-fitting. Principal component analysis is an unsupervised learning algorithm it creates a new set of variables to separate the data while disregarding class labels. These new components are known as principal components, principal component one finds the maximum variance within the data. The algorithm may produce as many components as samples, but not all components are informative; the top ten components may be responsible for 90% of the variance. These informative components are then used to build the linear discriminant algorithm, which includes class labels. An LDA will create new components to maximise the separation between classes. PCA and LDA are more sensitive to overfitting, and the chance of overfitting increases with a larger dataset. Such as in these studies that potentially had 174 samples and 14000 variables. Therefore, these dimension-reducing techniques were used along with random forests.

The second aim of this investigation was to determine if the random forest classification accuracy could be increased by changing different parameters of the method and data analysis to determine the best protocol to burn faecal pellets using REIMS. The test

to determine the classification of sex, age and strain was repeated several times but changing different settings. Parameters used in the original proof of concept of species separation were altered to determine what variables give the highest classification accuracy for sex, age and strain of lab mice. When it was established what parameters affected classification accuracy, a method for REIMS analysis was developed to help obtain the highest classification accuracies in future studies. A small number of samples were constantly classified as the wrong sex or age by the random forest models. The anogenital distance of a small group of individuals was measured to determine if females had become more male-like due to having more male siblings in utero. It was also investigated if siblings born in groups with a higher sex imbalance were more likely to be misclassified. It would be helpful to know the longevity of a random forest model. Therefore, a random forest model was built to predict the sex of samples collected and analysed with REIMS a year after the samples were used to build the model.

2.2 Method and Materials

2.2.1 Collection of faecal pellets from lab animals

Faecal pellets were collected from laboratory strains of house mice originally from Envigo UK.; three inbred strains C57BL/6JOLaHsd (C57BL/6), BALB/cOLaHsd (BALB/c), C.C3-H2^{k}/LilMcdJ (BALB.k) and one random-bred laboratory strain Hsd:ICR (ICR(CD-1)). All strains of mice were housed in same-sex sibling groups of two to four in 48 x 28 x 13 cm cages (MB1, North Kent Plastics, UK), except older males, which were housed individually in 48 x 15 x 13 cm cages (M3, North Kent Plastics, UK). All animals had access to food, 5FL2 EURodent Diet (IPS Product Supplies Limited, London, UK) and water *ad libitum.* All mice cages contained Corn Cob Absorb 10/14 substrate (IPS Product Supplies Limited, London, UK) to line the base. Cardboard tubes and paper wool nest material for enrichment were provided to all animals. During sample collection, mice were moved to individual clean cages for a maximum of two hours or once the individual had produced at least five pellets. Pellets produced by an individual were removed from the cage using metal tweezers, placed in a 1.5 ml Eppendorf tube, and stored at -18°C. After the two hours, mice were returned to their home cage regardless of defaecation.

2.2.3 Establishing a preliminary method for burning faecal pellets and a model to classify sex, maturity, and strain

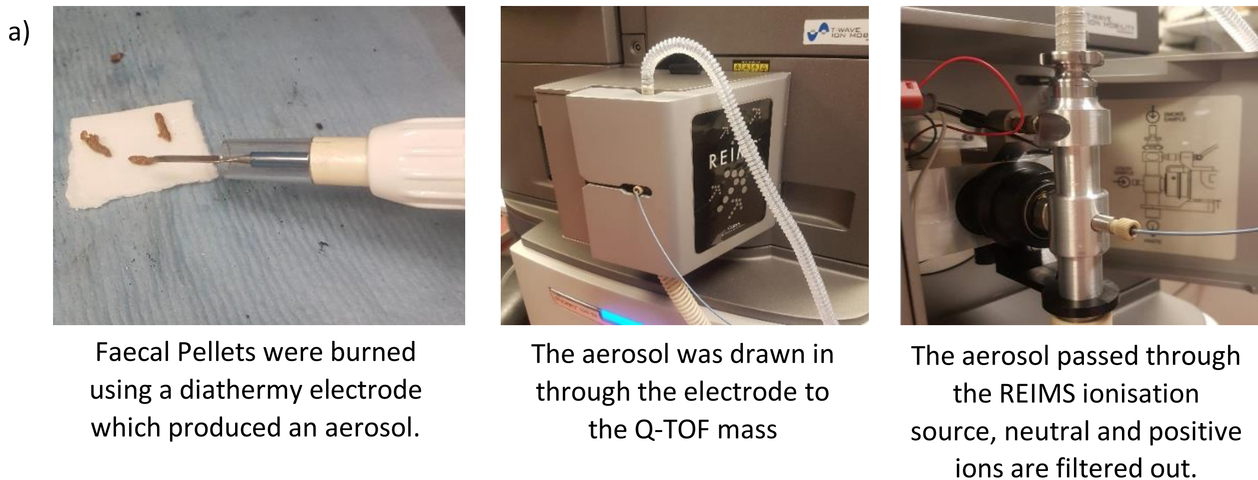
A total of 51 individual lab mice of four strains were chosen randomly (Table 2.1). Sampling took place in a Ductless Fume box (Air Science Liverpool, UK), and three faecal pellets from each individual were analysed. The pellets were placed onto glass microfiber paper (GFP, GE Healthcare Whatman) and up to 100 µl of MilliQ water was pipetted over each pellet to help conduction. A monopolar electrosurgical pencil powered by a VIO 50 C electrosurgical generator was used in cut mode at 35 volts to heat the samples for 2 to 5 seconds, producing an aerosol. The aerosol was suctioned through the pencil by a Venturi gas jet pump powered by nitrogen into a 3 m evacuation tube attached to the end of the pencil. The aerosol was drawn into the mass spectrometer (Synapt G2-Si, Waters, Wilmslow, UK) using the instrument's vacuum through the REIMS ionisation source, which contained a specifically

Table 2.1. The number of individuals of each sex, age and strain that were used to build the random forest models. Three faecal pellets were burned for each individual using REIMS, the spectra produced were binned and normalised in LiveID, before being uploaded to RStudio.

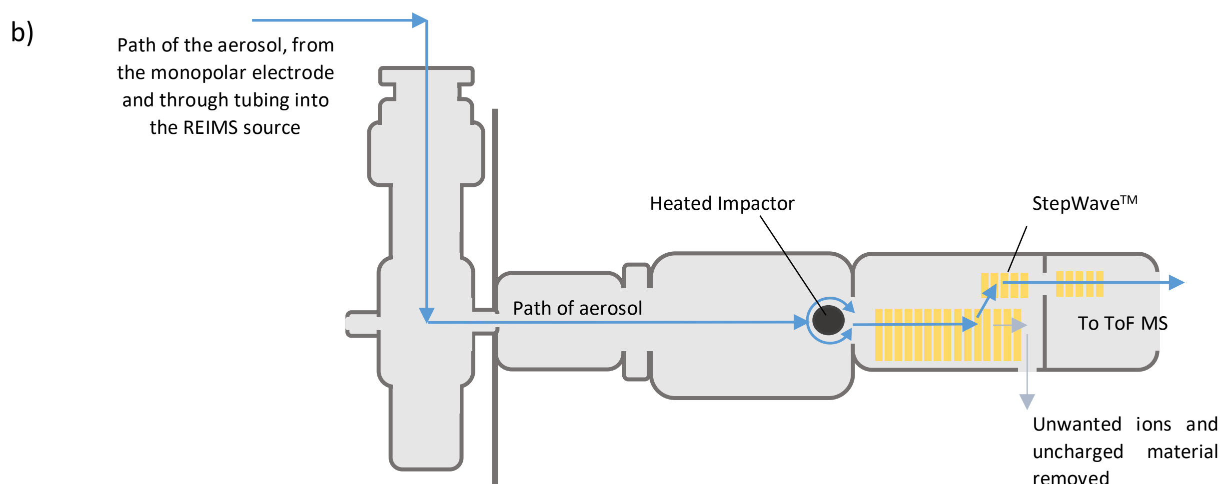
Sex	Age	Strain	Number of Samples
Female	Adult (>36 Days)	C57BL/6	6
		BALB/c	3
		BALB.k	0
		ICR(CD-1)	8
	Juvenile (<36 Days)	C57BL/6	0
		BALB/c	6
		BALB.k	0
		ICR(CD-1)	0
Male	Adult (>52 Days)	C57BL/6	5
		BALB/c	8
		BALB.k	4
		ICR(CD-1)	6
	Juvenile (<52 Days)	C57BL/6	1
		BALB/c	4
		BALB.k	0
		ICR(CD-1)	0
		Total	51

designed whistle that filters out larger particles within the aerosol (Figure 2.1). The order the samples were burned was randomised, and it was unknown at the time of burning which individual the samples had come from. Leu-enkephalin (LeuENK) (1.72pmol/µL dissolved in propan-2-ol) (Fisher Scientific) was continuously ejected at 50 µL/min into the inlet capillary of the REIMS ionisation source so LeuENK was constantly flowing through the instrument. Leu-Enk provided a lock mass (544.26 m/z) for an accurate mass measurement during analyses. The sample cone and the heater bias were set to 60 V. The spectra were recorded in full-scan resolution, negative ion mode, at a scan rate of 1 scan per second. Full scan mode specifies that the instrument will detect ions from the full mass range of 50 to 1200 m/z.

a)



b)



c)

Figure 2.1. a) The REIMS set up attached to a Q-TOF mass spectrometer, the monopolar electrosurgical pencil was used to burn three faecal pellets from a lab mouse. The aerosol created from the burning was suctioned through the evacuation tube into the REIMS source via the Venturi where larger particles were filtered out. b) Particles entered the mass spectrometer and hit the impactor where further ionisation takes place, they continue through the StepWave™ which filters neutral particles and positive ions. A video demonstrating the burning of faecal pellets of lab mice.

An acquisition file is created to contain all the data produced from burning one sample, which may contain burn signals from multiple faecal pellets of the same individual. The mass spectrum is the ion abundance against the mass-to-charge ratio of all the burn signals within one acquisition file (Figure 2.2).

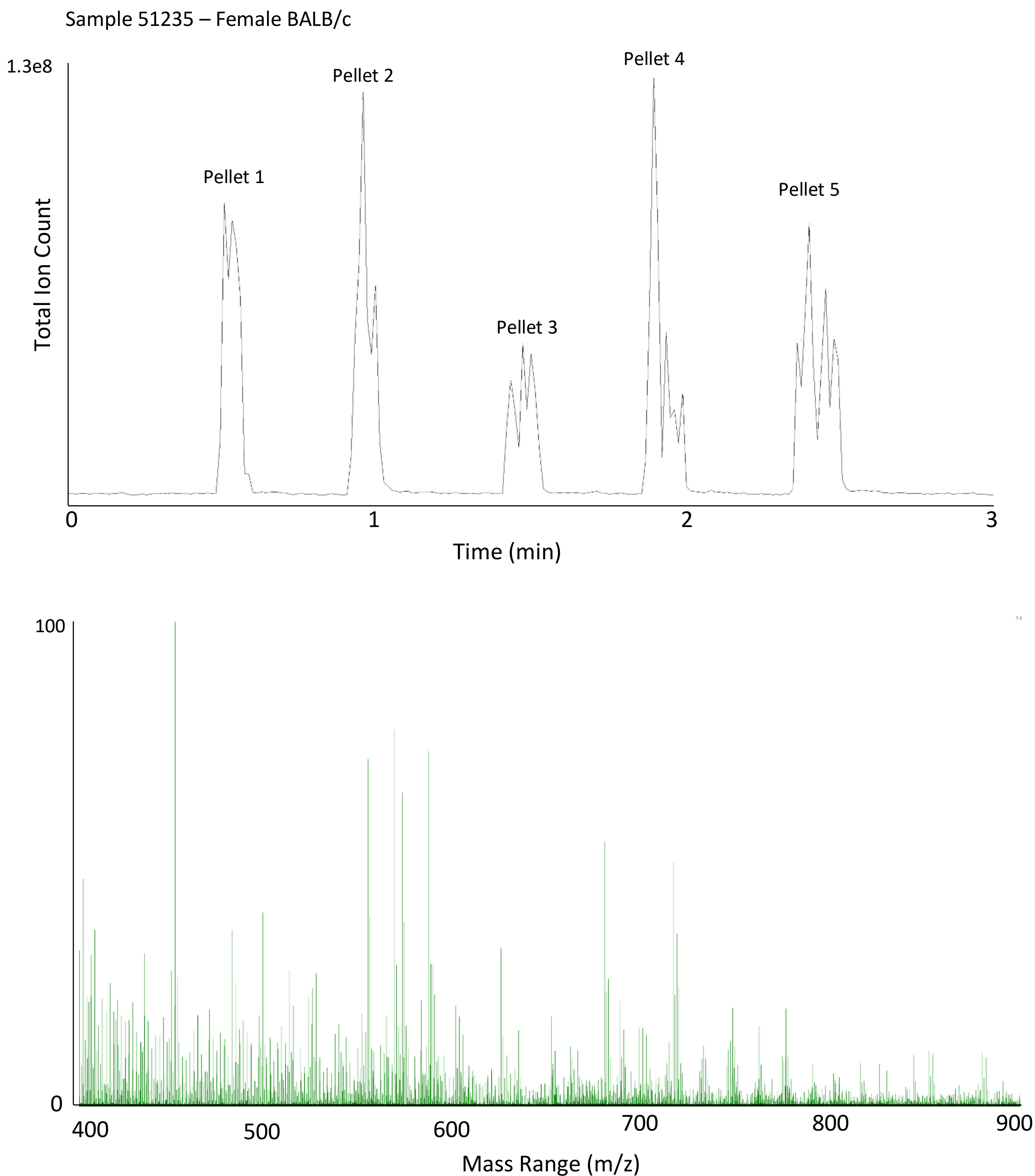
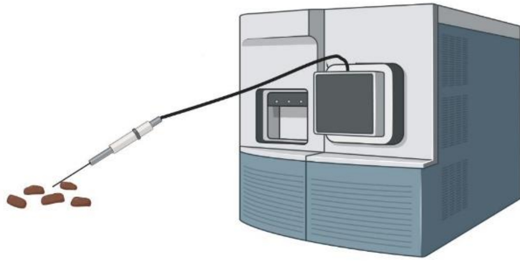


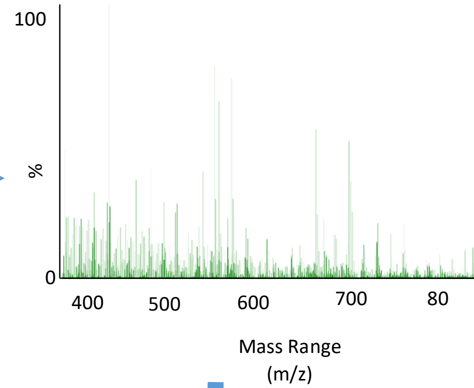
Figure 2.2 a) An example of a chromatogram (sample 51235) from MassLynx software used with REIMS and Q-TOF mass spectrometer which shows the number of negative ions that pass through the mass spectrometer when a pellet is burned. b) An example from the same individual of the mass spectrum from MassLynx of one faecal pellet with background subtraction. This data is stored in an acquisition file that is uploaded to LiveID to be processed before analysis in R. Each sample has one acquisition file which many contain the burn signals of several faecal pellets.

The mass spectra (acquisition files) were imported into LiveID version 1.1.872.736 (Waters). Within LiveID, the mass range was reduced to 400 to 1100 m/z; the data were binned to 0.1 m/z; lock-mass corrected using the LeuENK peak, and LiveID automatically set the intensity threshold. The intensity threshold is the limit that separates the burn signal from the background signal. The binned data were then exported as a .csv file and uploaded to R version 3.4.2 for further analysis (Figure 2.3).

A faecal sample was burned and analysed with REIMS

A mass spectral fingerprint was produced for each sample

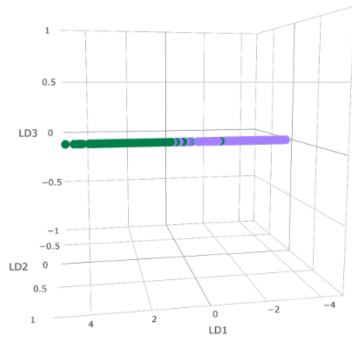




The spectra were uploaded to LiveID

In LiveID the data is binned and normalised

The spectra could be analysed by PCA-LDA in LiveID



The relative intensities of the binned data were exported

Sample	Sex	Age	Strain	400.05	400.15
50631	Female	Adult	ICR	0.000122	0.000168
50631	Female	Adult	ICR	5.29E-05	0.000112
50631	Female	Adult	ICR	2.73E-05	0.000169
50631	Female	Adult	ICR	0.00012	0.000187
50632	Female	Adult	ICR	5.22E-05	0.000186
50632	Female	Adult	ICR	7.41E-05	0.000247
50632	Female	Adult	ICR	0.000112	0.000363

The data was uploaded to R for further analysis

Figure 2.3 The processing steps required between burning the faecal pellets and analysing the data in R. Some of the settings within LiveID such as the intensity threshold for establishing burn signal from background noise are built into the software and cannot be changed. The binning resolution and the mass range are features within LiveID that can be changed.

2.2.4 Random Forest Analysis

A detailed explanation of random forests is in Box 1. When building a random forest model, only two parameters can be changed, including the number of individual decision trees to be used (ntree) and the number of variables to be selected for each tree's subset data (mtry). Random Forest models were created using the 'randomForest' package (Liaw and Weiner, 2002), 70% of the spectra were used to build the training model, and the model was then used to predict the classification of the remaining 30% of the data. For each model built, the optimal 'mtry' was selected using the tuneRF function (Liaw and Weiner, 2002). The plot function determined the optimum number of trees (ntrees) to use, which plots the class error against the number of trees. Increasing the number of trees reduces the error, but running the model takes longer. The optimal number of trees to use in the model is the minimum number required, which will significantly lower the class error. Each factor had two results; the random forest accuracy (rf value) was the percentage of correctly assigned individuals from the random forest model using the training data. This model was then used to assign a class to several unknown samples (the test data); the percentage of these individuals correctly assigned was the prediction accuracy (pa value).

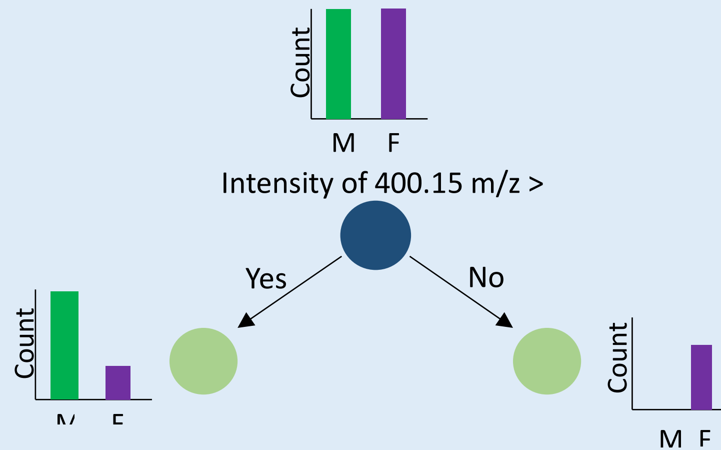
A random forest model was created to classify the samples as male or female (sex), as juvenile or adult (age) or as C57BL/6, BALB/c, BALB.K or ICR(CD-1) (strain). A random forest model was built for each factor using the training data, and these models were used to predict the classification of samples in the test data. Classification accuracy is produced based on the number of individuals the model correctly assigned with the suitable class. Since random forest models use a random subset of values to create each model, the accuracy can change slightly for each run. The random forest model was re-created ten times for the same data set, and the average was calculated. The data set was split into training and test data before each model so both sets had different samples each run. Multiple R packages were used for generating statistics and creating figures; a description of all packages used is summarised in Table 2.2. Waters offers two pieces of software available for analysis of the REIMS data, LiveID™ version 1.1.872.736 and Offline Model Builder version 1.1.28.0 (OMB). OMB offers the same processing abilities as LiveID plus additional options but is less user-friendly than LiveID. Both software provides the option to create a PCA-LDA model and to carry out a "leave-20%-out" cross-validation. The data was also randomised; each sample was assigned

as either ‘Random One’ or ‘Random Two’ to compare how much greater the classification accuracy for a particular trait is compared to a randomly assigned classification. Samples were randomised using an online random number generator (Haahr, 1998).

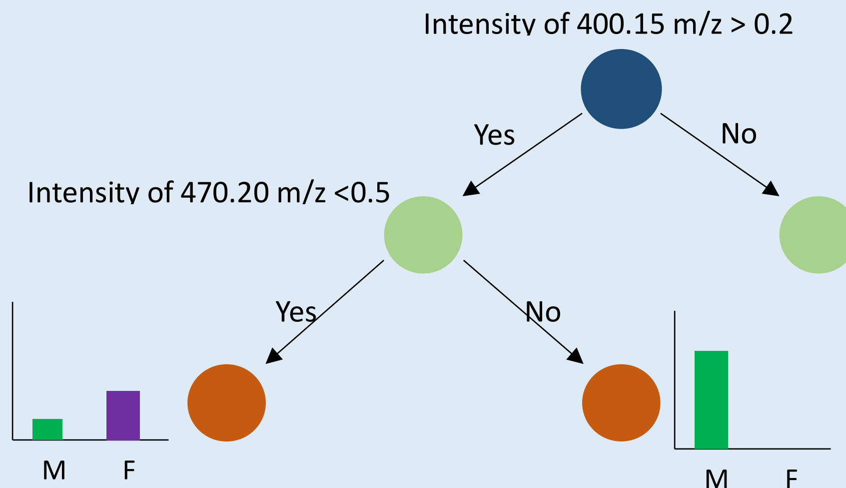
Box 1 – Random Forests

A decision tree is created from a bootstrapped dataset and only a random subset of the variables (mass bins) are used.

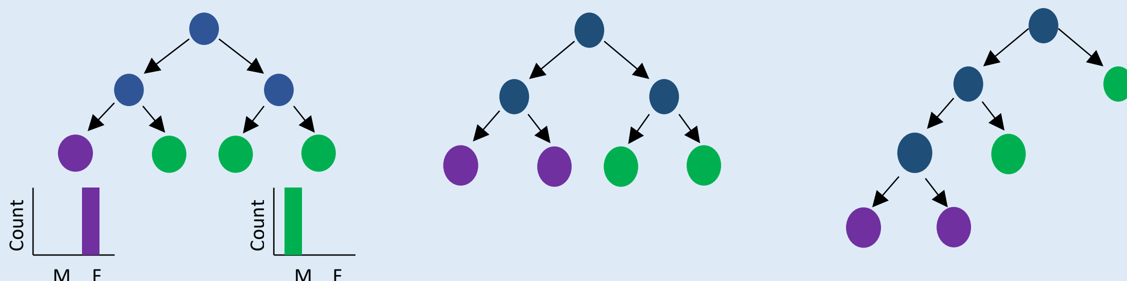
The first node of the decision tree is the mass bin that best splits the data into the selected classes, for this example the classes are male (M) and female (F).



It is unlikely that the first node completely separates all samples as males and females. A second node is required to split the data further it is established using the same two steps, independent from the first node.



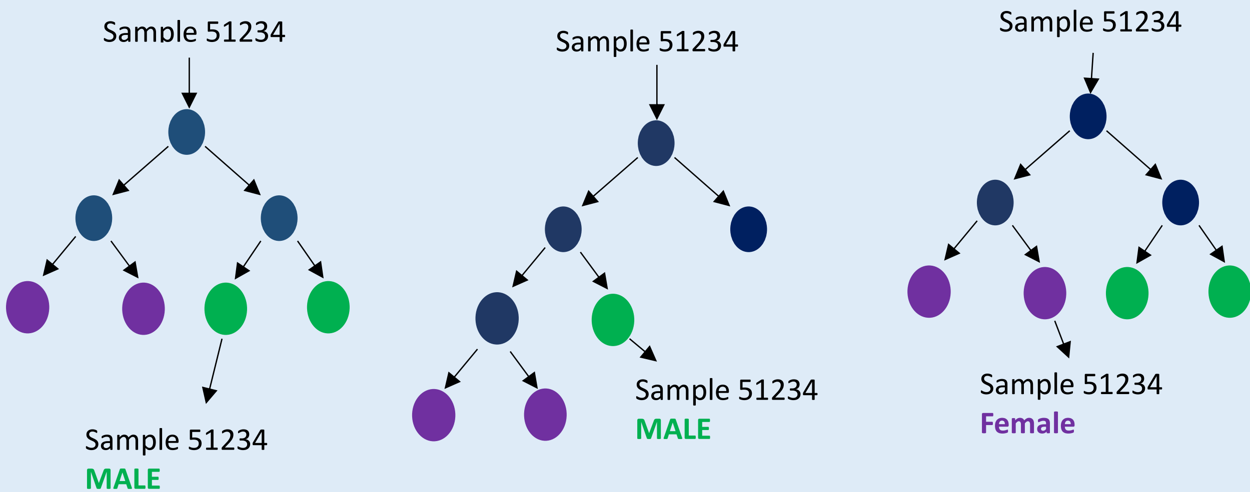
These steps continue, building one node at a time until all samples are divided, and the tree is complete. The model repeats this for as many trees as chosen by the user (ntrees).



Since the model is built using bootstrapped data there are samples that were not involved in building some trees these are the unbagged data.

Box 1 Continued – Random Forests

Each of the unbagged samples are run through the decision trees that they were not involved in building.



Each tree votes on what classification the sample is, the sample is classified as the class that had the most votes.

The out-of-bag error is the percentage of samples that were assigned the wrong classification, the random forest accuracy is 100 – out-of-bag error.

Sample	No of Trees voted Female	No of Trees voted Male	Mode Class	Actual Class	Correctly assigned
Sample 51234	20	80	Male	Male	Yes
Sample 51235	100	30	Female	Male	No
Sample 51236	39	110	Male	Male	Yes
Sample 51237	41	95	Male	Male	Yes
Sample 51238	112	10	Female	Female	Yes
			Out-of-bag error		20%

Random Forest accuracy (Rf) = 100 – out-of-bag error

The random forest model is created with 70% of the total number of samples (unless otherwise stated) this is known as the training set.

The remaining 30% of the samples are the test set and are used to confirm how accurate the model is at predicting unknown samples.

- Each sample of the test set is run through all the decisions trees within the training model
- Each tree votes on a classification for the samples
- The samples are assigned with the modal classification.
- The predicated classification is compared to the actual classification.

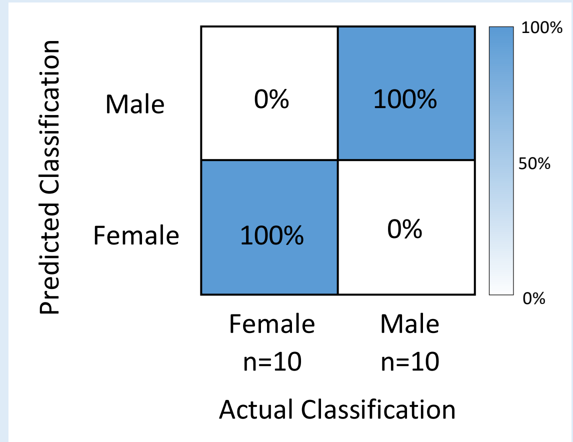
The percentage of samples that were assigned with the correct class is the prediction accuracy.

Box 1 Continued – Random Forests

For each model there is an overall classification accuracy but there is also an accuracy for each classification

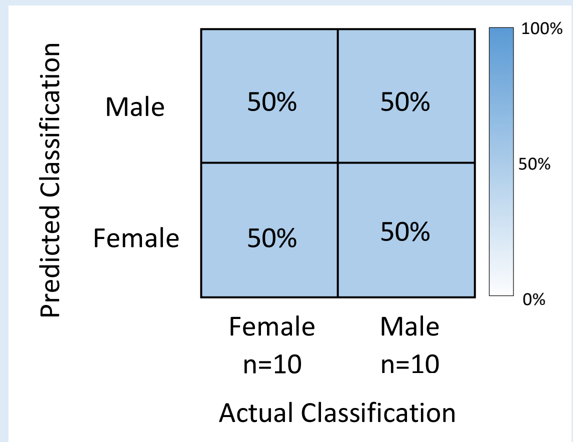
Using a random forest model of 10 males and 10 females as an example.

Actual Classification		Female	Male	Accuracy
	Female	10	0	100%
	Male	0	10	100%



10 of the 20 samples are correctly identified therefore the overall accuracy is 50%

Actual Classification		Female	Male	Accuracy
	Female	5	5	50%
	Male	5	5	50%



15 of the 20 samples are correctly identified therefore the overall accuracy is 75%

Actual Classification		Female	Male	Accuracy
	Female	9	1	90%
	Male	4	6	60%

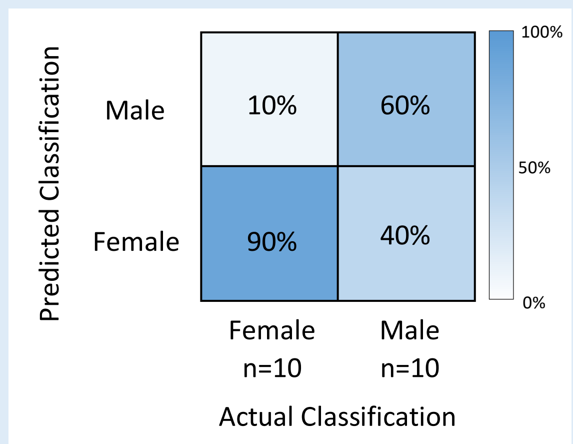


Table 2.2 All packages used within RStudio

Name of Package	Function	Reference
ggplot2	Used to create graphics including barplots, boxplots, histograms and piecharts	(Wickham, 2016)
dplyr	Functions for data manipulation including filtering datasets and grouping data to calculate means	(Wickham et al., 2021)
RandomForest	Required to build a forest of trees using random inputs	(Liaw and Weiner, 2002)
RandomForestExplainer	A set of function to help explain which variables are most important in a random forest.	(Paluszynska and Biecek, 2017)
ROSE	Functions to deal with binary classification with imbalanced classes.	(Lunardon et al., 2013)
rstatix	Contains functions for performing basic statistical tests including ANOVA and Kruskal-Wallis	(Kassambara, 2021)

2.2.5 Changing the REIMS method to increase classification accuracy (Coagulate Method)

The burning of samples was inconsistent, with some pellets producing better burn signals than others. Changes to the protocol were made to decrease the number of burns per pellet. Nine new methods labelled A through to I were created; changes included the voltage of the pencil and whether it was used with cut or coagulate mode (Table 2.3). Cut mode produces a constant radio frequency; this waveform produces heat rapidly. Coagulate mode produces a pulsed radiofrequency and produces less heat than cut mode. The amount of water added to the pellets was increased to help with conductivity. To increase the difference between the background signal and the burn event, a wait time of 30 seconds was introduced before burning the first pellet and after burning the last pellet. Between burning pellets of the same individual, the burn signal decreased to background intensity before the next pellet was burned. Each of the nine different methods was used with pellets from seven individuals (Figure 2.4). Each pellet should have produced one burn event when uploaded to LiveID; when the pellets did not burn easily, they produced multiple burn events. The method that produced the least number of burns was the most successful at burning faecal pellets. Method H (from now on referred to as the coagulate method) produced the least number of

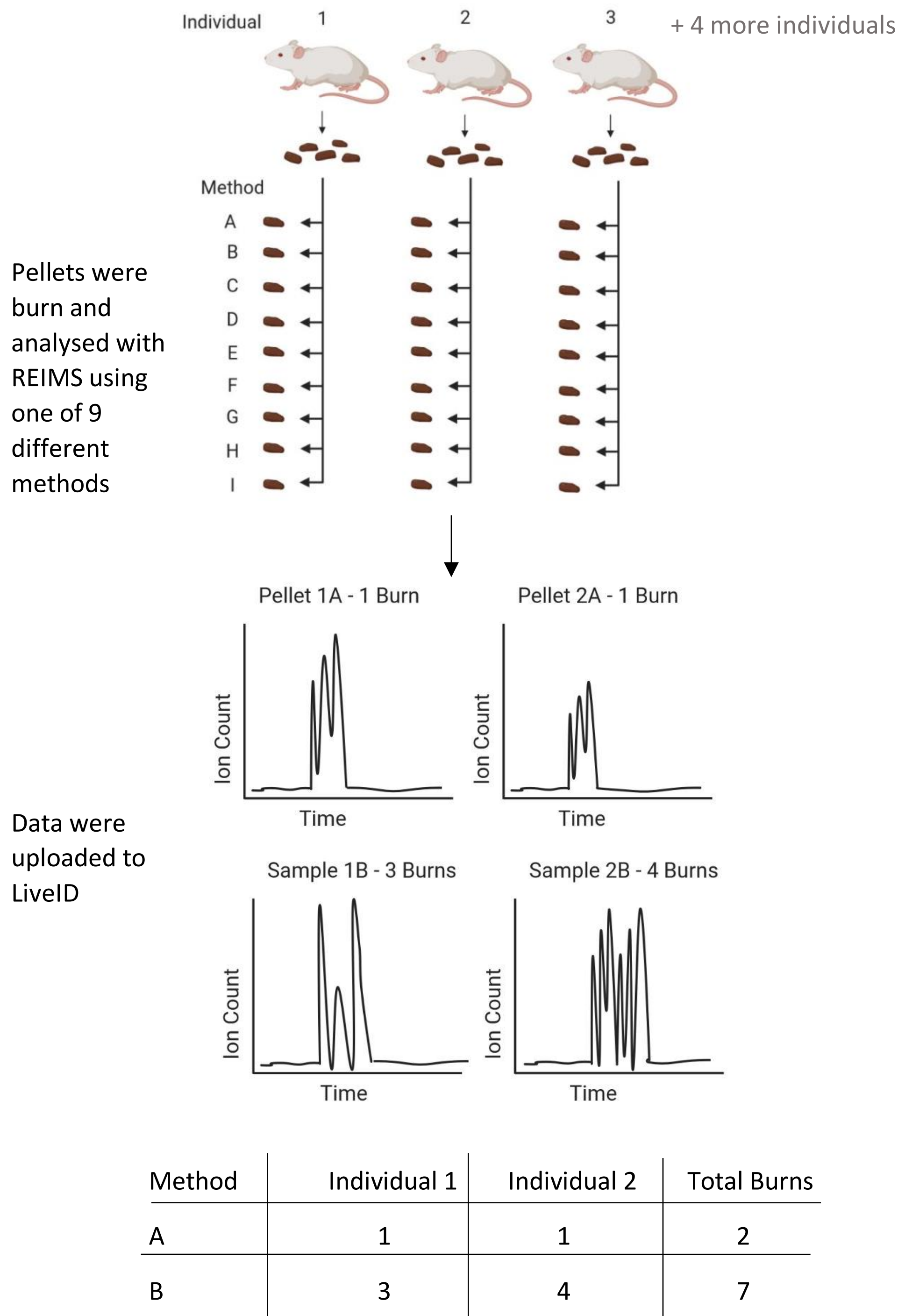
burns per pellet; this method was repeated using 41 individuals to determine the best LiveID parameters to use, including binning resolution and mass range. The coagulate method used coagulate mode at 40 V and 200 µl of water added to each pellet.

Table 2.3 The different changes made to the REIMS method to help improve burning of faecal pellets. The method was changed nine times, changes included the cutting mode of the electrode, the voltage of the electrode and how much water was pipetted on to a pellet before burning. Method H was established as the best method

Method Attempt	Cut or coagulate mode	Voltage (V)	Vol of water added (µl)	Other
A	Cut	40	100	
B	Cut	15	100	
C	Coagulate	15	100	
D	Coagulate	40	100	
E	Cut	25	>200	Water was added until pellet was saturated
F	Cut	25	100	
G	Cut	30	>200	Waited 30 s before/after burning
H	Coagulate	40	>200	Waited 30 s before/after burning
I	Cut	15	>200	Waited 30 s before/after burning

The coagulate method was repeated using 154 new individuals (Table 2.4). The faecal samples were collected using the same protocol as the preliminary method but burned using the settings from the coagulate method. The spectra were then uploaded to LiveID, the mass range was reduced to 400-1100 m/z, and the data binned to 0.05 Da. The data were exported from LiveID to R, where the spectra were averaged per individual. The mean accuracy of ten random forests was recorded for sex, age, and strain. The random forests were repeated for sex, age, and strain with the same spectra but with reduced mass ranges, 50-400 m/z and 400 to 700 m/z and 600-1200 m/z. The ranges were selected based on the general pattern of the mass spectra; relative abundancies were lower for the low and high mass ranges than the centre. The spectra were also assigned as one of two random classifications to confirm that

the ability of the random forest to classify correctly was due to the spectra differences between males or females, adults and juveniles or the various strains.



Method	Individual 1	Individual 2	Total Burns
A	1	1	2
B	3	4	7

Figure 2.4. The workflow to establish the method (A-I) that produces the least number of burns per pellet. Nine pellets were collected from seven individuals (3 shown) and each pellet was burned using a different method. The samples were uploaded to LiveID which assigned several burns depending on when the burn signal passed the threshold intensity. The method with the least number of total burns would be the best for burning faecal pellets.

The coagulate method was repeated three more times with the same 154 individuals but using either one pellet per individual, three pellets per individual or five pellets per individual. The samples were burned using the same settings as the coagulate method, and the spectra were uploaded to LiveID. The mass range was reduced to 400-1100 m/z, and the data binned to 0.05 Da. The spectra were exported to R to create new random forest models. The models were built as before using a mean of ten forests for sex, age, and strain.

Table 2.4. The number of individuals used with the coagulate method, at least three pellets were collected from each individual.

Sex	Age	Strain	Number of Individuals
Female	Adult (>36 Days)	C57BL/6	15
		BALB/c	32
		BALB.k	12
		ICR(CD-1)	3
	Juvenile (<36 Days)	C57BL/6	3
		BALB/c	11
		BALB.k	4
		ICR(CD-1)	3
Male	Adult (>52 Days)	C57BL/6	16
		BALB/c	22
		BALB.k	12
		ICR(CD-1)	12
	Juvenile (<52 Days)	C57BL/6	2
		BALB/c	4
		BALB.k	2
		ICR(CD-1)	1
		Total	154

2.2.6 Repeating study with an increase in the number of samples used and the amount of pellet hydration (Coagulate/Hydration Method).

New faecal pellets were collected from the same 154 lab mice as before if still available and were burned along with pellets from new individuals; pellets were collected from 176 individuals altogether (Table 2.5). The method used in the repeat (Coagulate/hydrated method) consisted of burning 3 to 5 pellets per individual, dependent on the number of pellets available, using 35 V in coagulate mode. The voltage was lowered to 35 V as using 40 V produced many sparks, and 35 V burned the pellets just as well as 40 V did. Enough water was added to the pellet until maximum hydration was reached to help with conductivity. The volume of water added to a pellet varied greatly between 50 and 200 µl due to the varying degrees of dehydration of pellets. The mass spectra were uploaded into Waters LiveID Software; the spectra were normalised, scaled and lock-mass corrected to Leu-enkephalin at m/z 544.26. The spectra were discretised ('binned') to a 0.05 Da bin width, and the m/z range was reduced to 400 to 1000 m/z. The data were exported from LiveID into an Excel file to be

Table 2.5. The final classification models were created using 176 individuals with the coagulate/hydrated method. Five pellets were taken from each individual when possible but at least three pellets were collected.

Sex	Age	Strain	Number of Individuals
Female	Adult (>36 Days)	C57BL/6	13
		BALB/c	36
		BALB.k	12
		ICR(CD-1)	11
	Juvenile (<36 Days)	C57BL/6	2
		BALB/c	6
		BALB.k	2
		ICR(CD-1)	3
Male	Adult (>52 Days)	C57BL/6	14
		BALB/c	18
		BALB.k	12
		ICR(CD-1)	16
	Juvenile (<52 Days)	C57BL/6	9
		BALB/c	8
		BALB.k	6
		ICR(CD-1)	8
		Total	176

used in R for further analysis. The burns produced from burning multiple pellets per individual were averaged within R, so there was one relative intensity value per mass bin per individual to prevent pseudo-replication. Ten random forests were run using all the data to test the accuracy of predicting sex, age, and strain then the data was divided 70:30 to be used as training and test data.

For some models, one of the classes would have many more samples than the other. The under-sampling function from the ROSE package was created to balance datasets randomly (Lunardon et al., 2013). For example, there were many more adult samples than juveniles so the under-sampling function was applied before creating random forest models. The under-sampling function would randomly select adult samples so there was a similar number of adults to juveniles. Over-sampling was also tried, increasing the number of rare samples in a bootstrapped data set. Over-sampling increased the random forest accuracies but did not increase prediction accuracy, possibly due to over-fitting. Models built with oversampling could not predict unknown samples. Therefore, under-sampling was chosen for imbalanced data sets.

2.3 Results

2.3.1 Optimising the REIMS Method

The spectra of different classifications appear almost identical, and differences between spectra cannot be determined through observation (Figure 2.5). The leave-20%-out cross-validations were higher for LiveID (76%) than OMB (Figure 2.10). Each individual was assigned as either ‘Random One’ or ‘Random Two’ using an online random number generator. Both software packages produced lower cross-validations results of the random data than the correctly classified data. OMB gave a cross-validation accuracy of 42%, which would be expected for randomised data; LiveID had a higher than excepted result of 62%. The PCA-LDA plots for the random data showed as much separation as the males and females (Figure 2.6).

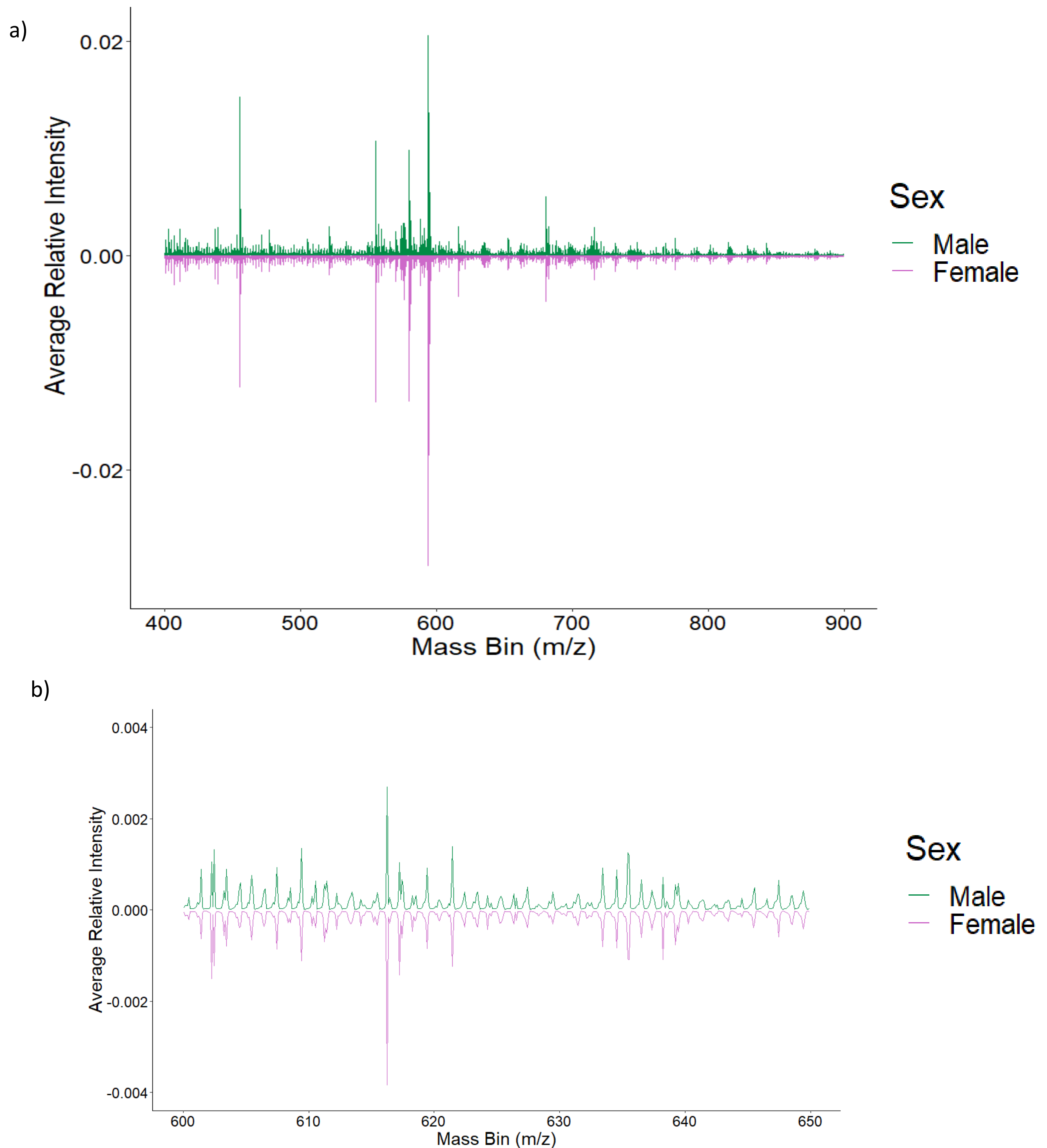
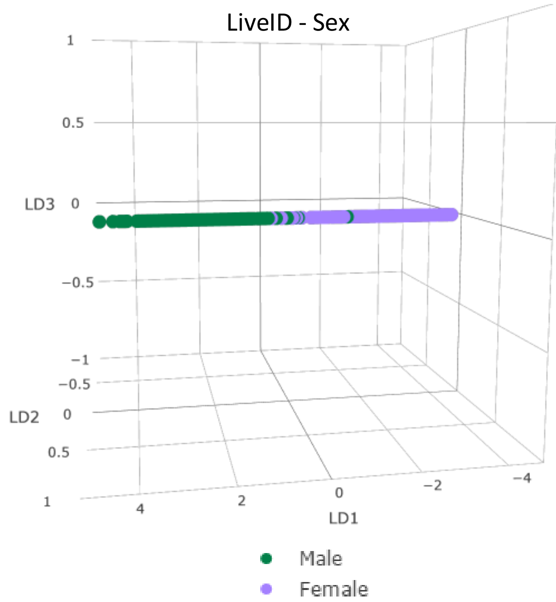


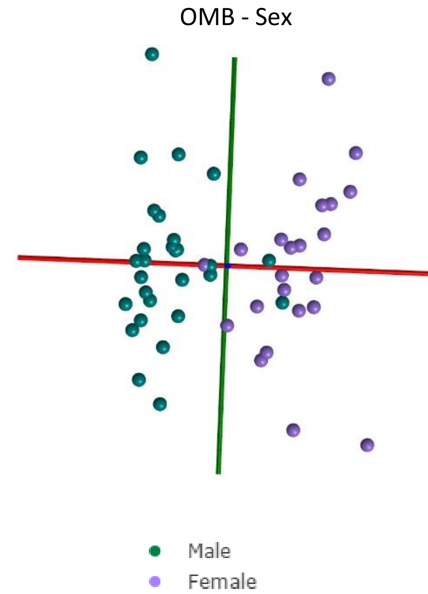
Figure 2.5 a) The average relative intensity of 51 male and female of four lab strains of mice for each mass bin from 400 to 900 m/z b) The average relative intensity for 600 to 650 m/z. With the expectation of the three highest peaks which show a very small difference between males and females there was no observable differences between males and females. The highest peak for both sexes was bin 594.25, the female intensity was 0.008 more than males. LeuENK peak was removed from the spectra.

a)

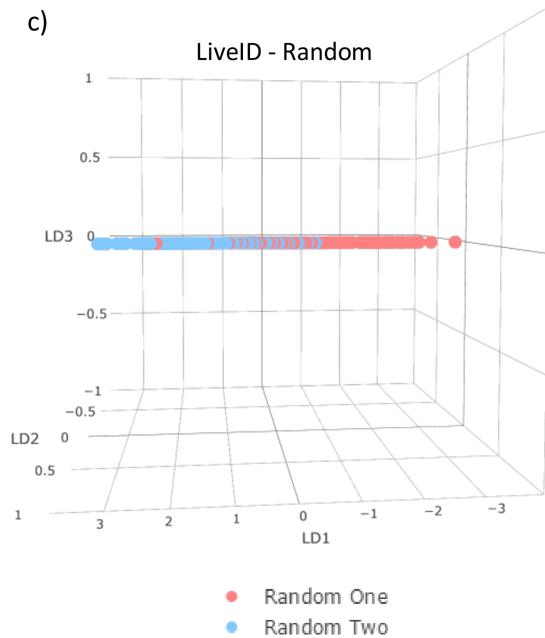
Software and data used	Leave 20% out cross validation
LiveID - Sex	76%
LiveID - Random	62%
OMB - Sex	70%
OMB – Random	42%

b)





c)



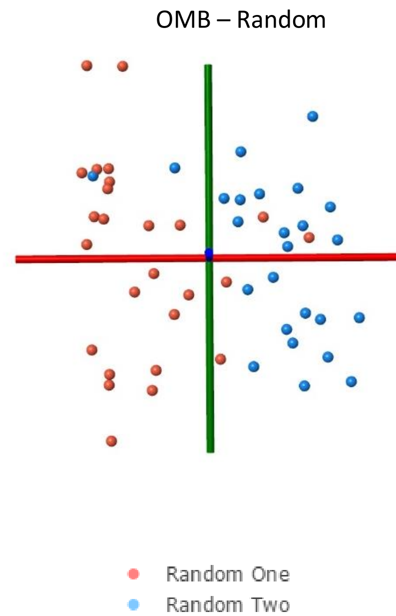


Figure 2.6 a) The prediction accuracy of leave a 20% out cross validations either using the 51 individuals with their correct sex classification or with a randomly assigned classification, either 'random one' or 'random two'. The validations were carried out in both software offered by waters LiveID and OMB b) The PCA-LDA plots from LiveID and OMB for the samples with the correct classification. c) The PCA-LDA plots from LiveID and OMB for the samples with a random classification.

Ten random forest models were created for each factor (sex, age, and strain), and the mean accuracy was calculated. (Figure 2.7). Each of the three factors had random forest accuracies (rf value) above 70%, considerably higher than the rf value of the randomised data of 38%. The pa values for age and strain were above 80%, while sex was 68%. It was still much higher than the randomised data of 43%; however, the standard errors for all values were relatively high (between 6 and 14%).

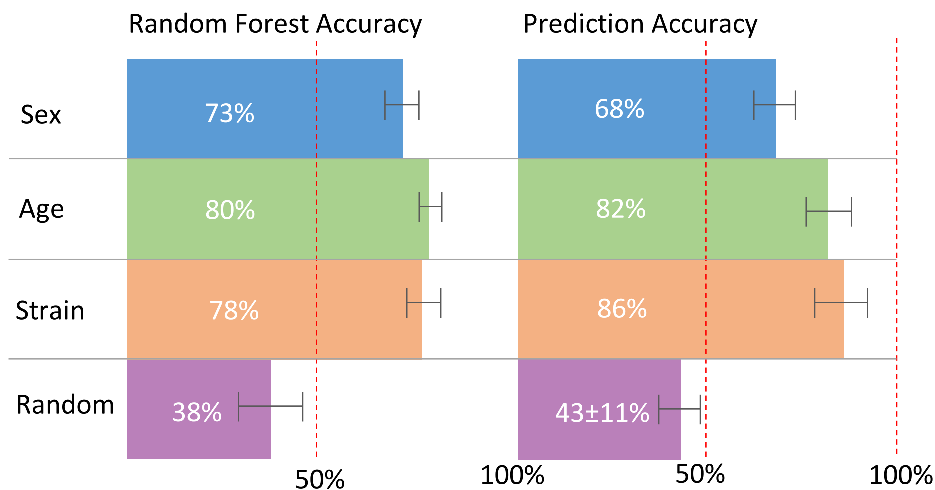


Figure 2.7. The random forest accuracy produced from the training data and the prediction accuracy produced by using the model to predict unknown samples. The accuracies are an average of ten random forests. If the random forests models were unable to find a pattern between classes, it would produce an accuracy of 50% for age and sex and an accuracy of 25% for strain.

The experiment was repeated nine times with a small subset of individuals changing one step of the method each time to reduce the number of burns per pellet (Table 2.6). The spectra produced from different pellets of the same individual that were burned using either coagulate or cutting in this study showed only a tiny difference in intensity with the intensity of Method H (Coagulate Method) higher than Method B (Figure 2.8). The most significant change to the number of burns produced was to add as much water as was needed to saturate each pellet rather than adding a designated volume of water. The amount of water required would change depending on the size and consistency of the pellet, with the best burning achieved when the water was able to saturate the whole pellet.

Table 2.6. The numbers of burns created from burning seven pellets using nine different burning techniques. The burning of faecal pellets was repeated nine times each with one change from the preliminary method . Seven faecal pellets from various individuals were burned for each of the different methods. Method produced the least number of burns per pellet.

Method	Total number of burns	Mean Number of Burns per pellet
Preliminary method	45	6.4
A (Cut, 40 V, 100 µl)	41	5.9
B (Cut, 15 V, 100 µl)	26	3.7
C (Coag, 15 V, 100 µl)	31	4.4
D (Coag, 40 V, 100 µl)	23	3.3
E (Cut, 25 V, >200 µl)	24	3.4
F (Cut, 25 V, 100 µl)	38	5.4
G (Cut, 30 V, >200 µl)	10	1.4
H (Coag, 40 V, >200 µl)	8	1.1
I (Cut 15 V, >200 µl)	12	1.7

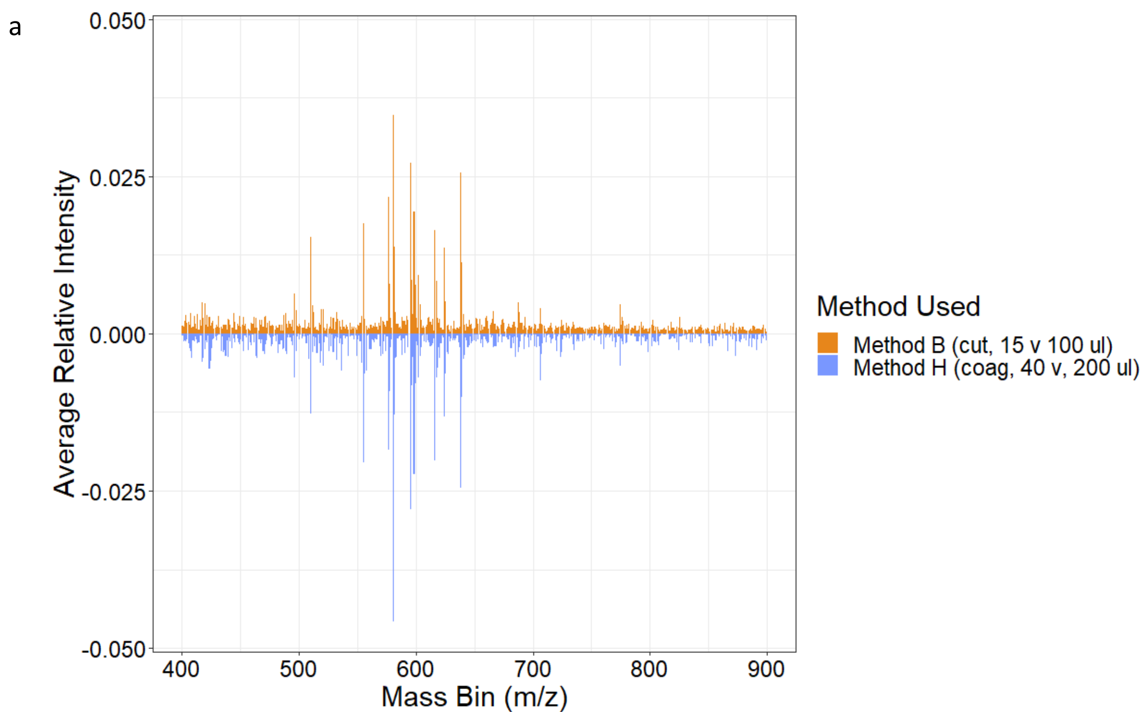


Figure 2.8. One pellet was burned with using a diathermy electrode on cut mode at 15 V, 100 µl of water was added to the pellet before burning (Method B). One pellet was burned set on coagulate mode and 40 V with 200 µl of water added (Method H /Coagulate Method).The mass spectra of two pellets from the same individual burned using two different methods. The spectrum for each pellet shows the same overall pattern but the intensity is higher for the pellet burned using method H.

The accuracy was also improved when the file acquisition began 30 seconds before burning and was allowed to continue for 30 seconds after burning, allowing for a clear distinction between the background signal and the signal created by the aerosol produced from burning. The highest random forest accuracies were obtained with a bin size of 0.05 Da. The random forest accuracy for sex increased from 73% to 75% when the bin size was decreased to 0.05 Da. The random forests for the sex data were repeated with the sum of all the burns instead of the mean or using just one randomly selected burn. The mean of all the burns from each individual produced the highest random forest accuracy. The random forest accuracy increased slightly when the mass range was reduced to 400 – 700 m/z (Figure 2.9)

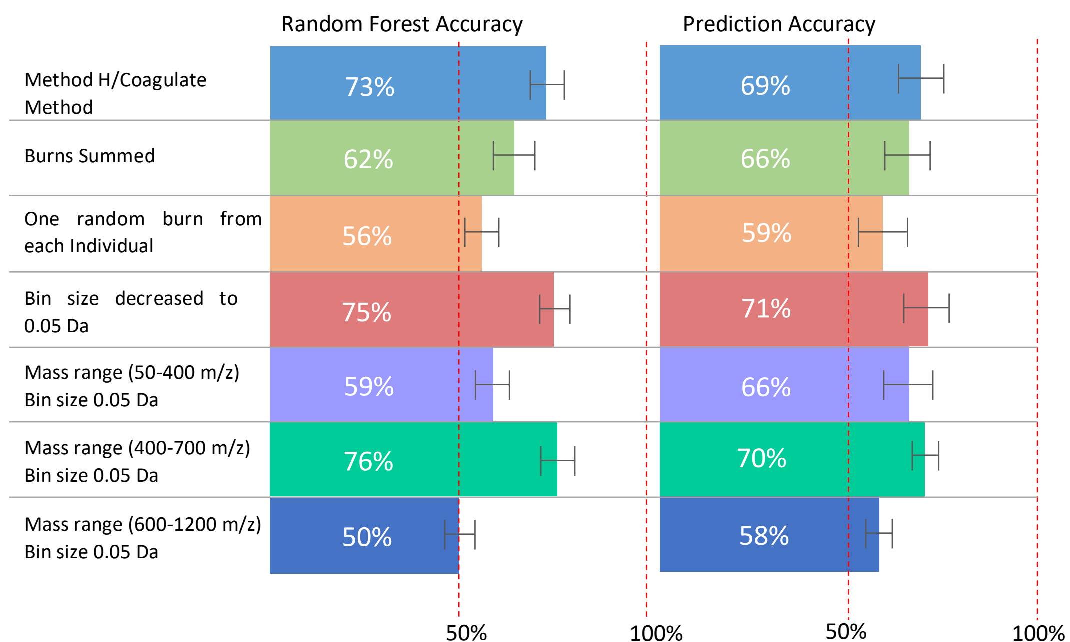


Figure 2.9. Results from random forests models for male and females. Each model used the same data, but one parameter was changed. The spectra were summed instead of the mean taken. A random burn was selected from each individual rather than using the average of all burns. The bin size decrease to 0.05 Da and the mass range was changed. For each parameter an average of ten random forests was obtained.

2.3.2 Results from the Coagulate Method

Random forests produced higher accuracies with higher sample numbers; therefore, collecting samples from more individuals increased the overall accuracy and reduced the standard deviation. Using the coagulate method and more individuals produced an increase in the random forest accuracy of both sex and age but only produced an increase in prediction accuracy for sex (Figure 2.10). The standard deviation decreased in all cases apart from strain. With the initial method using 51 individuals, BALB.K samples were consistently misclassified, and in the newer model, 63% were misclassified and assigned as BALB/C. The random forest accuracy increased by 12% when BALB.k individuals were removed from the dataset, and the prediction accuracy increased by 15%. There was also an increase in accuracy when the BALB/c and BALB.k individuals were combined into one "strain" designated as BALB (Figure 2.11). The classification accuracy for BALB/c and BALB was 96%, but the accuracies for the other two strains (ICR and C57BL/6) were much lower, especially when BALB/c and BALB.k were combined. Even though combining the two BALB strains gave a slightly higher random forest accuracy, BALB.k individuals were removed from the data set for subsequent strain models. BALB.k may be too closely related to BALB/c to be considered a separate strain (they only differ by an H2k mutation) but combining them caused too much of an imbalance in the dataset.

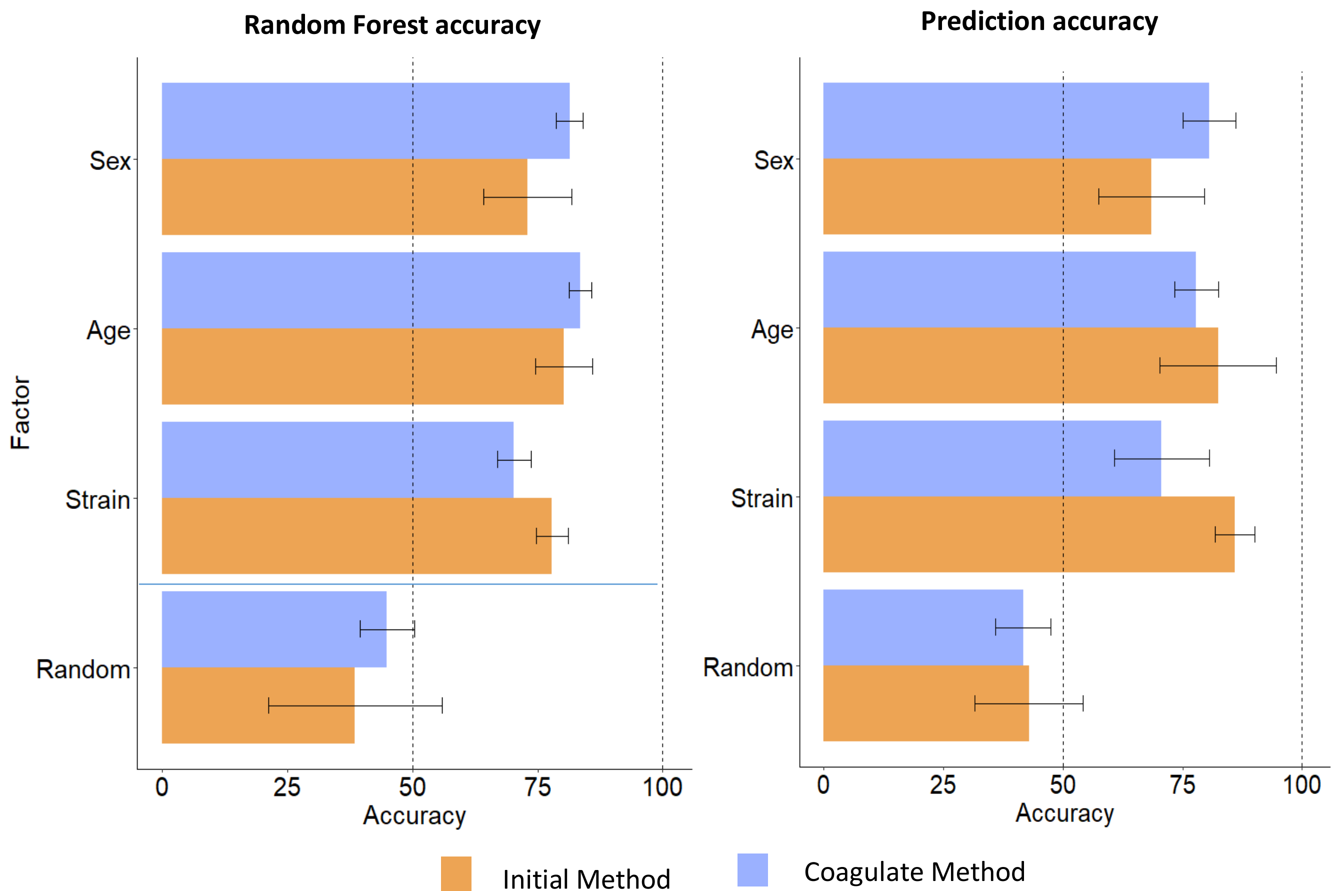


Figure 2.10. The random forest accuracy (%) and prediction accuracy (%) using the preliminary method and the coagulate method. Ten random forests were used to classify males and females (sex), adults and juveniles (age) and between different strains of lab mice (strain). For the coagulate method pellets were burned using the coagulate mode on 40 V and with 200 µl water added to each pellet. The random forest model had a bin size of 0.05 a mass range of 400 to 1100 m/z and a total of 151 individuals were used. The initial method used cut mode, 35 V, 100 µl of water added and a bin size of 0.1.

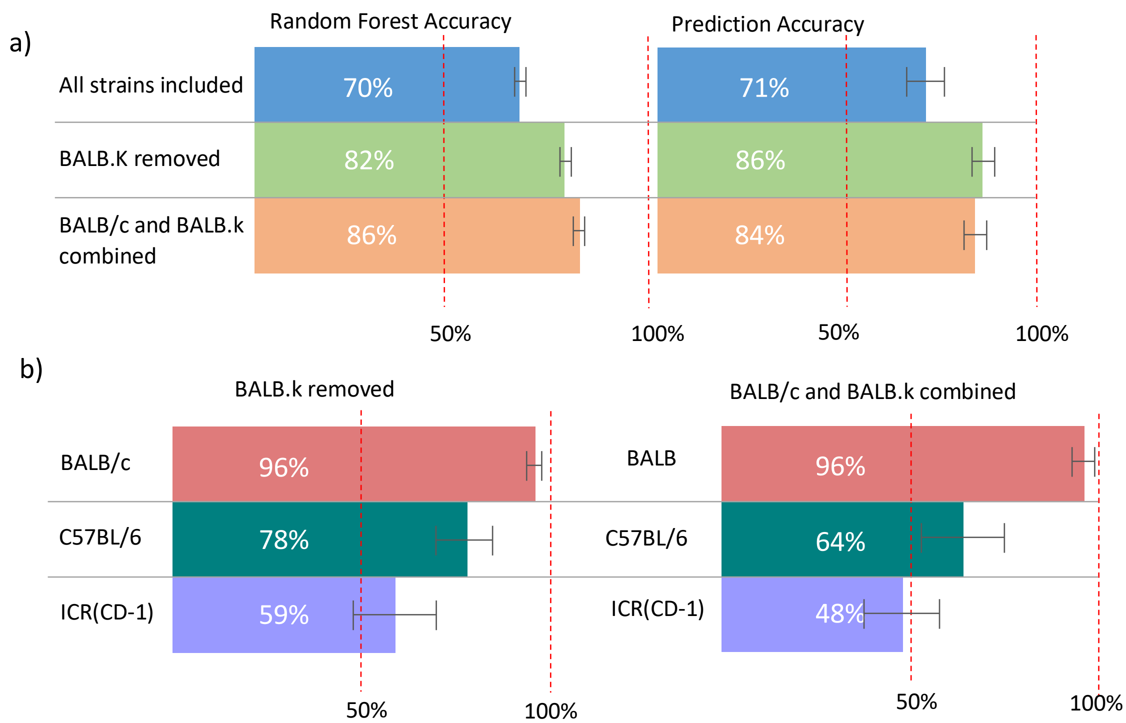


Figure 2.11. a) The random forest accuracy (%) and prediction accuracy (%) using the optimised method for when all four strains are used in the model, when all BALB.k individuals are removed from the data and when BALB/c and BALB.k individuals are given the same classification (BALB). b) The individual classification accuracies for each strain when BALB.k is removed from the data set and when combined with BALB/c.

There was a problem with an imbalance in the data; there were more adults (n=121) than juveniles (n=30) and nearly twice as many BALB/c individuals than any other strain. This caused the overall random forest and prediction accuracies to be relatively high, even though most juveniles were classified as adults (Figure 2.12). When BALB/c and BALB.k individuals were combined, the unbalance between strains increased even more. Although the overall prediction accuracy increased because 96% of the BALB individuals were correctly assigned, only 64% of C57BL/6 and 48% of ICR individuals were assigned correctly (Figure 2.11). This may be why the prediction accuracies decreased for strain and age when more individuals were added, increasing the group imbalance.

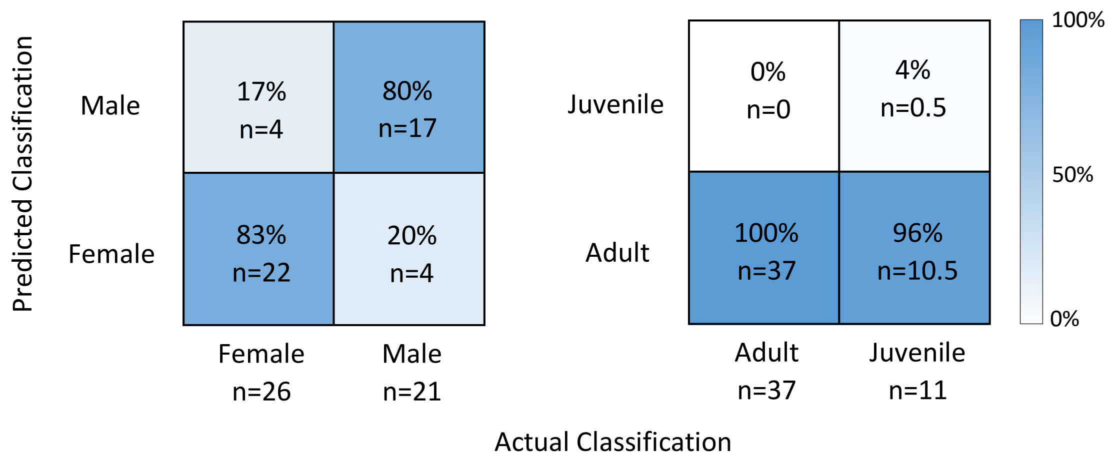


Figure 2.12. The prediction accuracies of random forest model for sex and age. The average prediction accuracies for sex and age were 80 and 78%. The model for sex was as accurate at predicting females as it was for males. The model for age however predicted everything as an adult but the lack of juveniles meant the overall accuracy was still high.

The coagulation method was also used to compare the optimum number of pellets per individual. The method was used to burn five pellets each from 128 individuals (only individuals with five pellets were used), but the analysis was carried out three times using the burns from one pellet, then the average burns of three and five pellets. The pellets included in the analysis with one or three pellets were chosen randomly, and the random forest and prediction accuracy for sex were calculated. The random forest accuracy increased with the number of pellets, and the standard deviation decreased slightly (Figure 2.13). There was little difference between the accuracy for three pellets and five pellets confirming that individuals that only produced three pellets would still be acceptable to include in further studies.

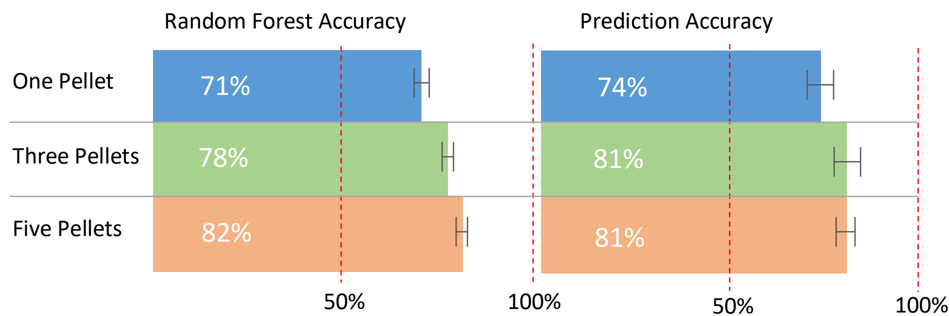


Figure 2.13. The random forest and prediction accuracy using the same model properties (bin 0.05 and mass range 400-1100 m/z) using 128 individuals but burning either one, three or five faecal pellets. All pellets were burned using coagulate mode on 40 V. All accuracies are an average of ten random forests.

2.3.2 Results from the Coagulate/Hydration Method

2.3.2.1 Classification of Sex

The random forest model used to predict sex gave a prediction accuracy of 78% (Figure 2.14). The model was run another ten times but using all samples (samples were not split 70:30 into training and test) to establish what mass bins were the most responsible for the differences between males and females. When only individuals over 55 days old were included, the model accuracy increased by 8%, while the prediction accuracy increased to 81%. The top five most discriminant mass bins were also established; four out of five of the top mass bins were the same when both juveniles and adults were included in the model. The remaining mass bins were within the top 10 for both groups. The other four mass bins were ranked the same for both models (Figure 2.15). Out of the 44 juveniles misclassified, eleven of them were assigned the wrong sex in every model. There were only thirteen juvenile females, seven of whom were consistently misclassified. The random forest accuracy for sex with juveniles was only 76% for males and 30% for females. Only one of the top mass bins was the same for juveniles compared to adults (Bin 469.275) (Figure 2.16). Comparing adults and juveniles for the top bins that separate sex the most confirms a significant difference between adult and juvenile females but not males.

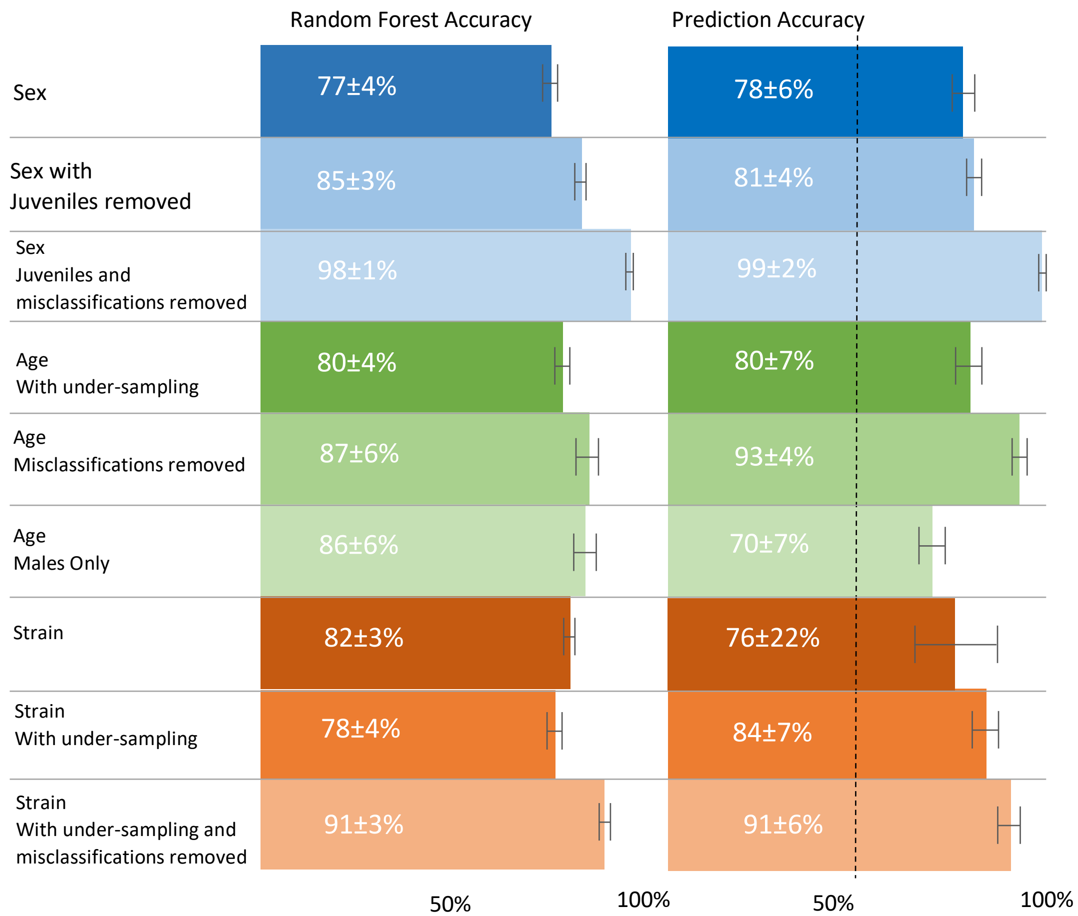


Figure 2.14. Random Forest and prediction accuracy for the 176 sample. Under-sampling indicates that the under-sampling function from the ROSE package was used before the random forests were performed to ensure each class had the same number of samples. A list of samples was created for each of the three groups (sex, age, and strain) that were always misclassified in the random forest models to be removed from subsequent models.

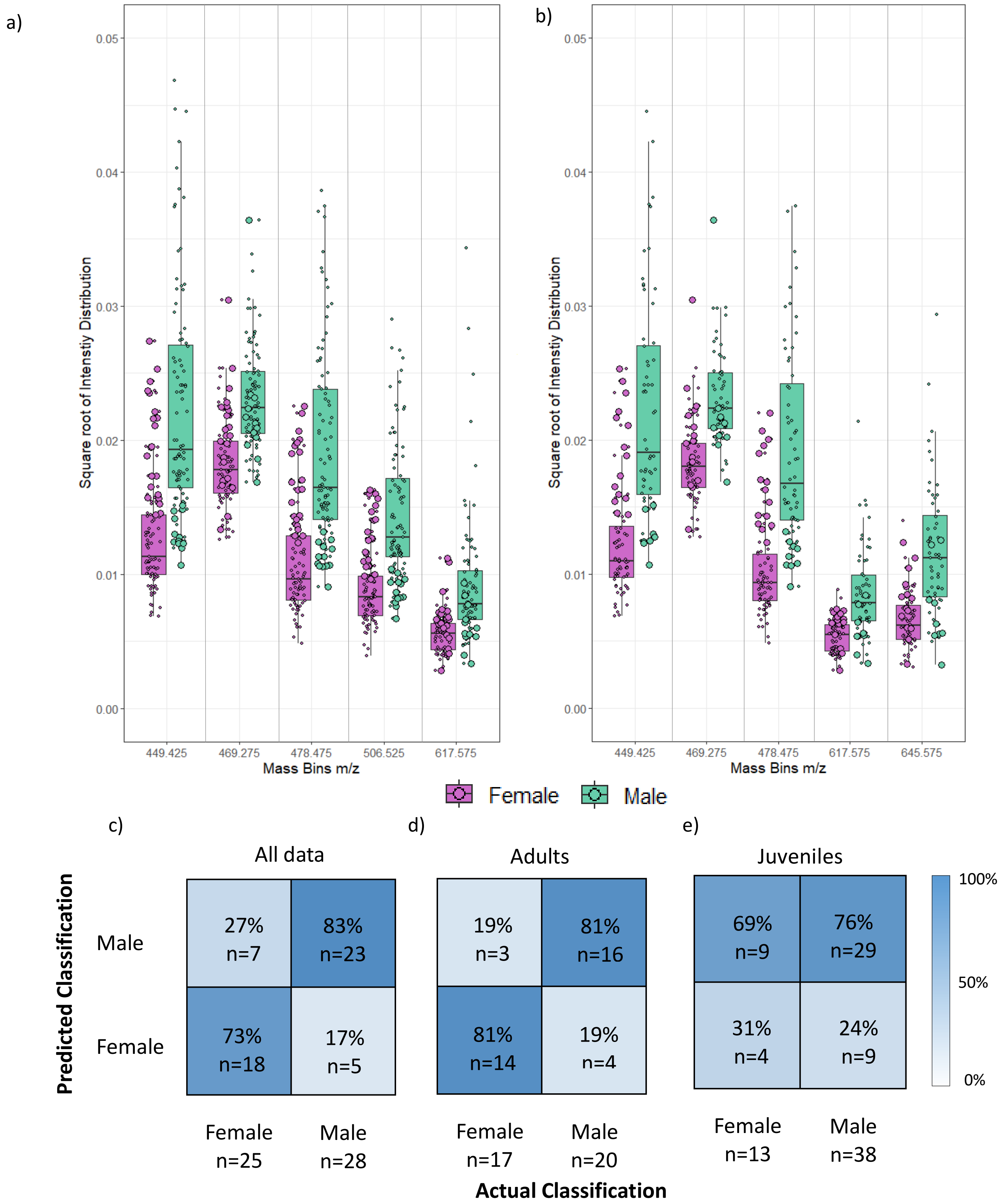


Figure 2.15. a) The intensities of the top five mass bins most responsible for the differences between males and females from an average of ten random forest runs when all samples were included. A mass bin value represents the centre of the bin size, mass bin 449.425 is all the intensities between 449.400 to 449.450 for a bin size of 0.05.The increased size in points indicates those samples that were wrongly identified in all ten models. b) The intensities of the top five mass bins when juveniles (under 55 days) have been removed from the data set. c) The average prediction accuracy of ten random forest models when all data was included d) when juveniles were removed. e) The average of ten random forest accuracies of juvenile samples only, there were not enough juvenile samples to split the data into training and test set.

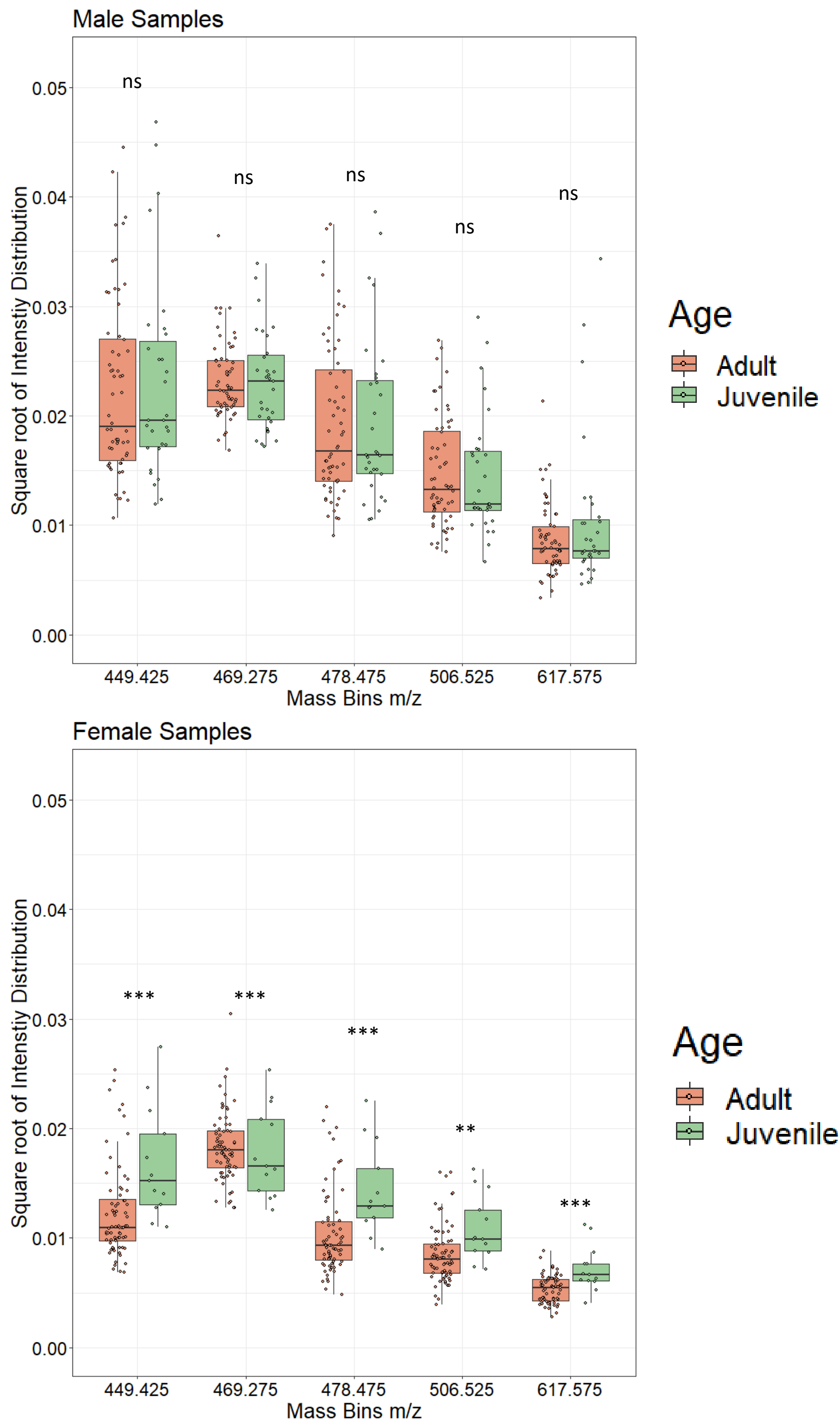


Figure 2.16. The relative intensities of adults and juveniles for the top five most discriminant mass bins to separate sex. There was no significant difference between intensities of male adults and juveniles. There was a significant difference between females, the spectra of females changed with age. This increased the probability of female juveniles being classified as males.

It was established that the same samples among the adults were consistently misclassified each time they were run. Two female BALB/c sisters showed similar burn events, with a maximum ion count of 1x10 (Figure 2.17). When observing the boxplots of the top five mass bins, the misclassified male individuals had lower intensities than the lower interquartile range. In contrast, the misclassified females had intensities higher than the higher interquartile range (Figure 2.15). Compared to ten mass bins at random, the intensities of misclassified individuals were more varied (Figure 2.18). It is not that these individuals have a lower-than-average intensity across the whole spectra for males and higher than average for females; they differ only in those mass bins responsible for sex variation. If these individuals have spectra more similar to the opposite sex, then it would be expected that removing them from the model would dramatically increase the classification accuracy. Removing these samples from the model will affect classification accuracy differently depending on why they were misclassified. Classification accuracy would increase if the samples were misclassified because the most discriminant mass bins' intensity values differed from that class's average. Removing the samples would not affect classification accuracy if it were another unknown reason for the misclassification. Removing these individuals and juveniles increased the accuracy dramatically to 98% for the model and 99 % for prediction accuracy (Figure 2.18). Greater separation was observed when an LDA was performed on the top 20 PCA components when specific samples were removed compared to an LDA on the top 40 PCA components with all samples included (Figure 2.19). Again, four of the most discriminant mass bins were the same for the model using all samples compared to the model with samples removed (Figure 2.20).

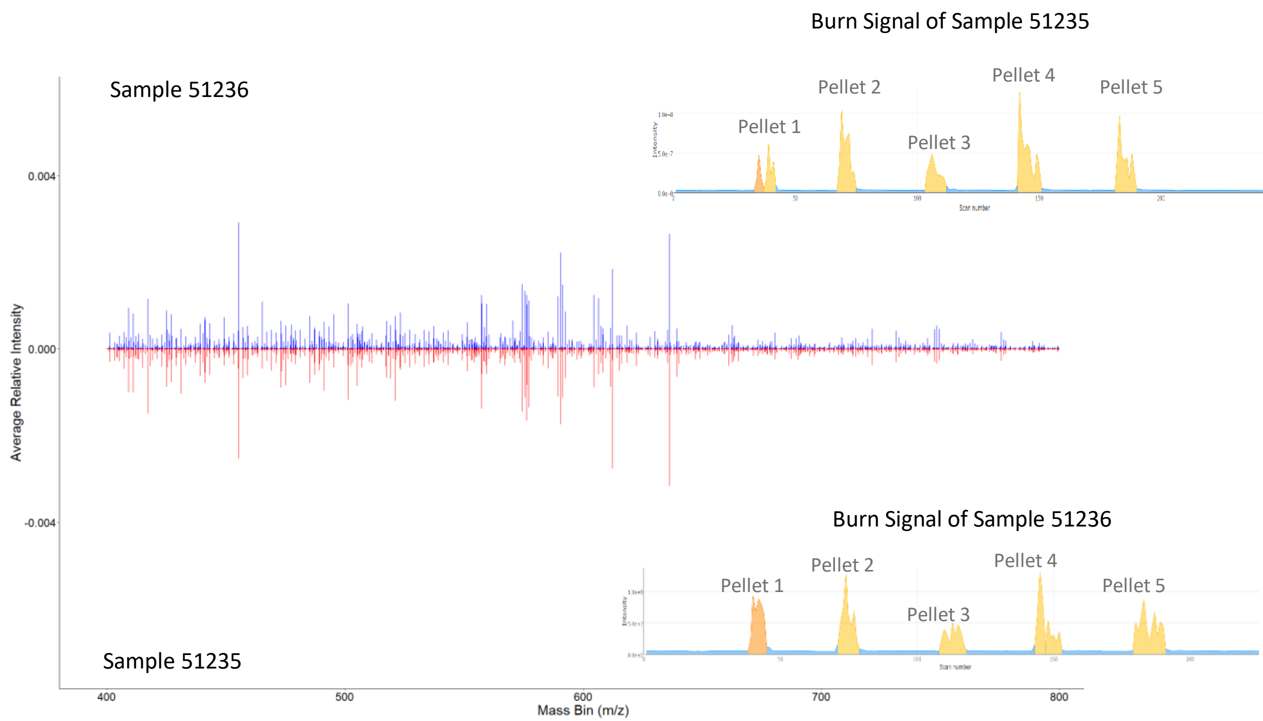
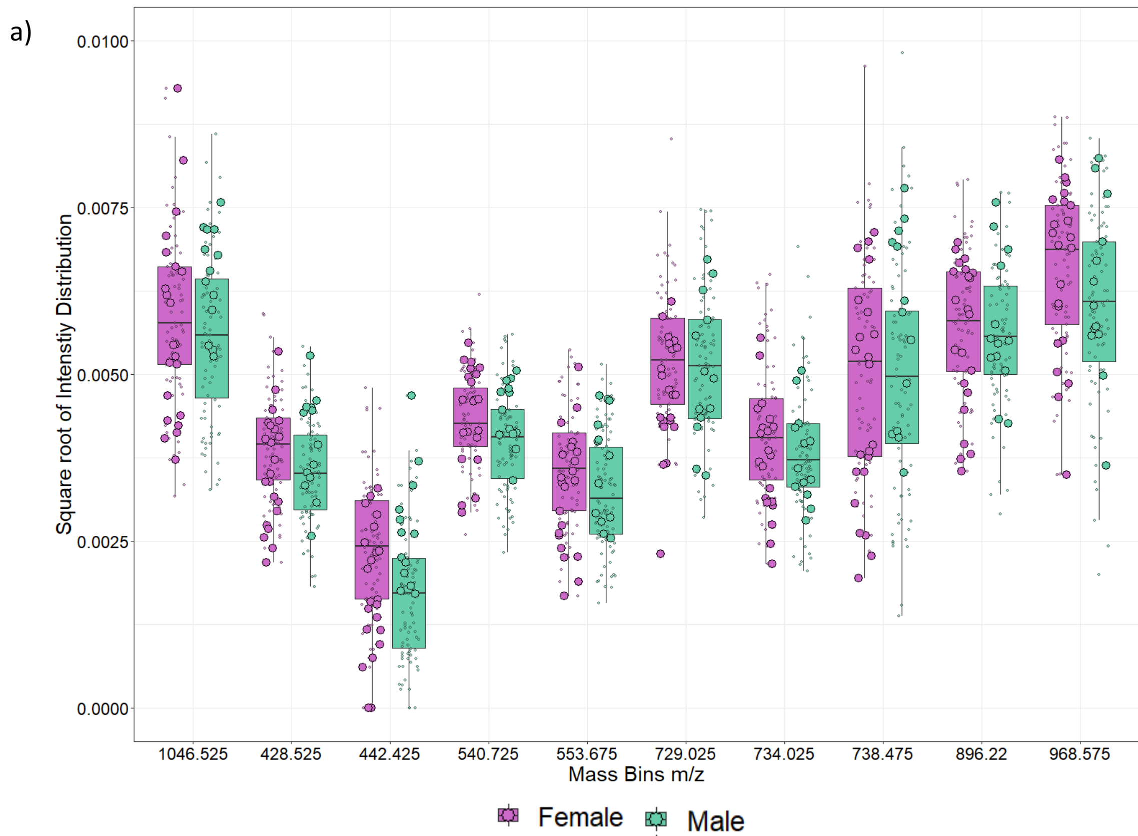


Figure 2.17. The mass spectra and burn events for two different samples. Samples 51236 and 51235 were both BALB/C sisters but 51235 was always correctly identified as female whereas 51235 was always classified incorrectly as a male. LeuENK peaks have been removed as they are of a much higher intensity the other peaks cannot be observed.



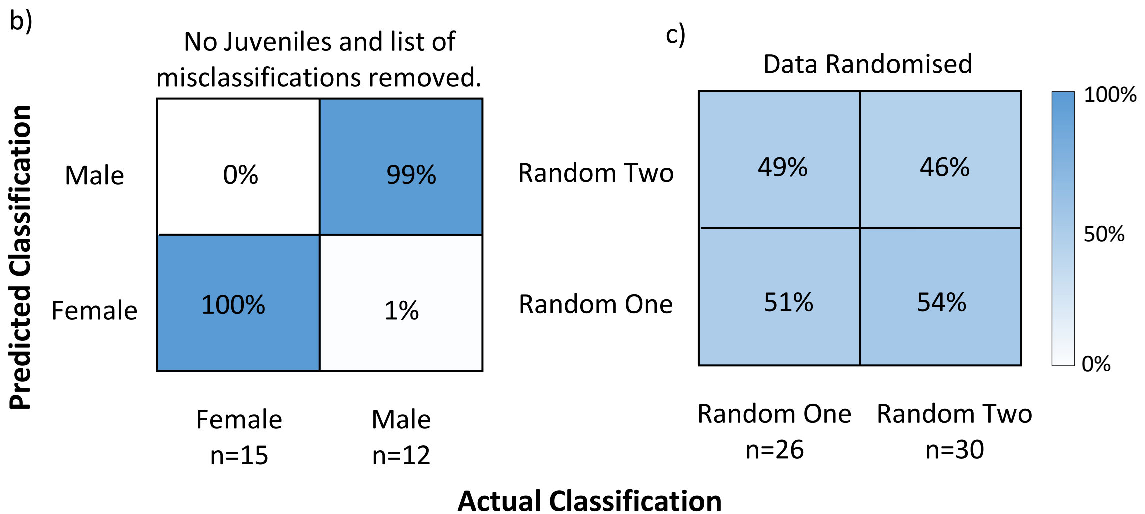


Figure 2.18. a) The intensities of males and females for ten mass bins selected at random. The larger points are the individuals that were classified as the wrong sex in all ten random forest runs. The intensities of the misclassified individuals are much more varied for the random bins compared to the top five mass bins that are most responsible for the difference between male and females. b) The average prediction accuracy of ten random forest models when juveniles and the samples misclassified in all ten models using all the samples were removed from the model. c) The average prediction accuracy of ten random forest models when all samples were randomly assigned as either 'random one' or 'random two'.

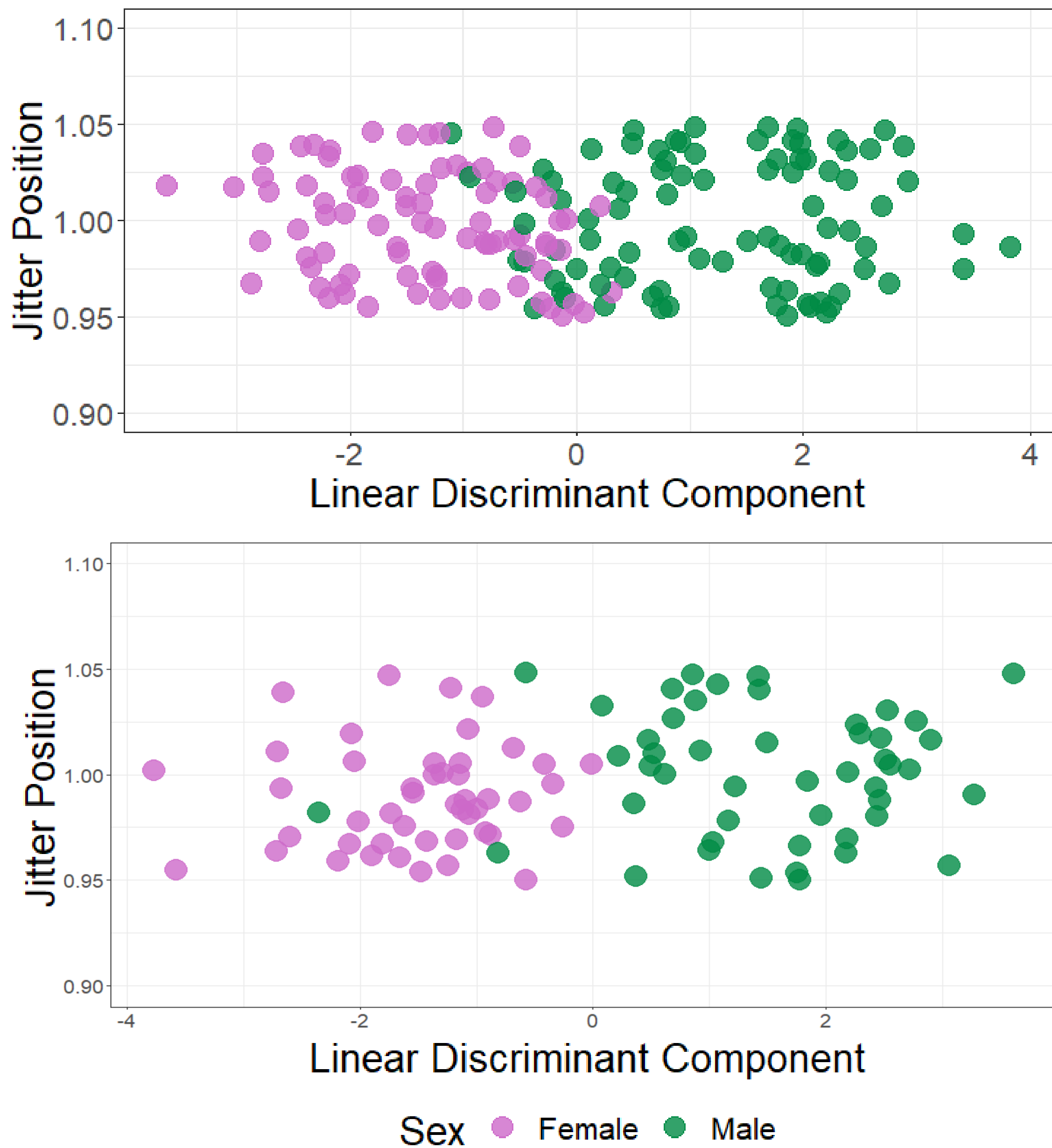


Figure 2.19. a) A PCA-LDA plot of sex using the top 40 PCA components with all samples included in the model. b) A PCA-LDA plot of sex using the top 20 PCA components with juveniles and specific adult samples removed from the model. The adult samples removed were those that the random forest models constantly misclassified. The jitter position is a randomly assigned value to disperse the results vertically so individual sample points are visible

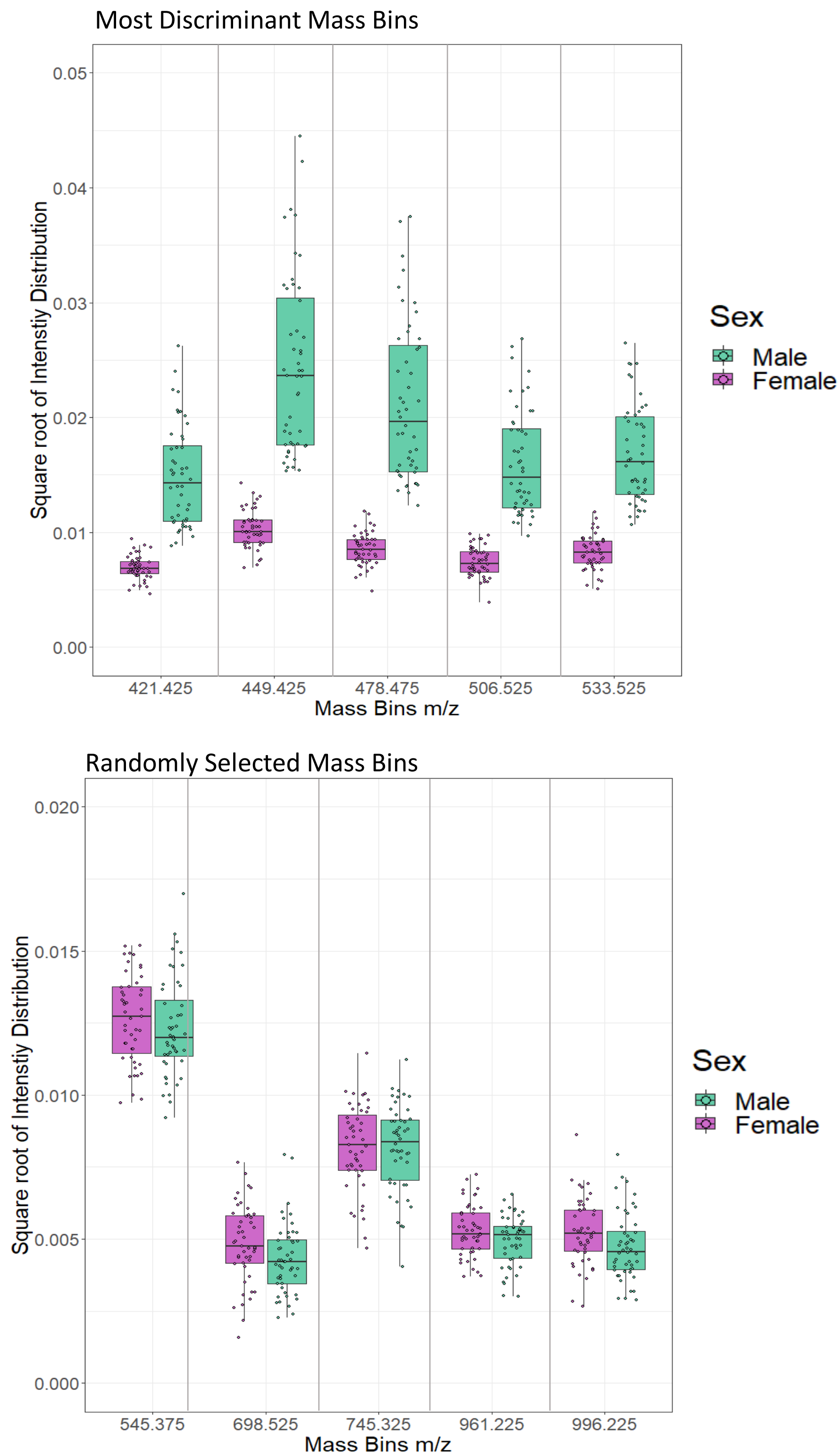


Figure 2.20. The intensities of the top five mass bins most responsible for the differences between males and females from an average of ten random forest runs when juveniles and the samples misclassified in all ten models using all the samples were removed. The intensities of five randomly selected mass bins, selected using an online random number generator

The below-average intensities for males and above-average for female misclassifications suggest that females that were misclassified had a mass spectrum more like that of males and vice versa. New samples were collected from 22 individuals, and their anogenital distance was measured (Group I). All the males that were measured were correctly assigned as male; their distance varied between 0.6 and 1.2 cm. The females that were measured were from four different sibling groups. The females were all correctly assigned as females except for three individuals from the same sibling group. The anogenital distance for females ranged from 0.5 to 0.8 cm with an average of 0.63, and the three misclassified females have distances of 0.53, 0.66 and 0.71. The second group of individuals were collected (Group II), and the male anogenital distance ranged from 1.07 to 1.98 cm with an average of 1.5. Overall, the model correctly identified eight out of ten males, and the two misclassified males had a distance of 1.07 and 1.29 cm. The anogenital distance for the females ranged from 0.46 and 0.98 cm, with an average of 0.69 cm. Only four out of 11 females were correctly identified as female, three of them and an anogenital distance below the average of 0.69 and one with a distance of 0.75 cm. A t-test confirmed no significant difference between the anogenital distance of females correctly identified and misclassified (p=0.24). However, this would need to be repeated with more individuals measured to confirm an association between anogenital distance and classification.

The original 176 individuals came from 36 sibling groups; 19 had more males than females, 15 had more females than males, and two groups had even numbers. Samples were not collected from every individual from a sibling group. Although eight sibling groups were in the dataset without female samples, only one had no females born into the group. Out of the 14 groups with more males than females, five had females that were not correctly identified, and the number of females varied between one and four in a sibling group. Six groups had all the females identified correctly, and the other groups had 33%, 50% and 75% of the females correctly identified. Overall, this gave an average prediction accuracy of 54% for females in sibling groups with more males. The prediction accuracy of females in groups with more females than males was 73%. A t-test gave a p-value of 0.25; therefore, there is no significant difference between the ability of REIMS to identify a female from a sibling group with more females or more males. The p-value decreased to 0.07 when using only the sibling groups with more than twice the number of individuals as the opposite sex (Figure 2.21).

Three groups of siblings had no males, two of these groups had a 100% prediction accuracy, and the third had two females, but only one was correctly identified. There was less difference between sibling groups when looking at the number of males correctly classified (p-value = 0.60). The average of males predicted correctly in groups with more females was 72% and 80% for groups with more males.

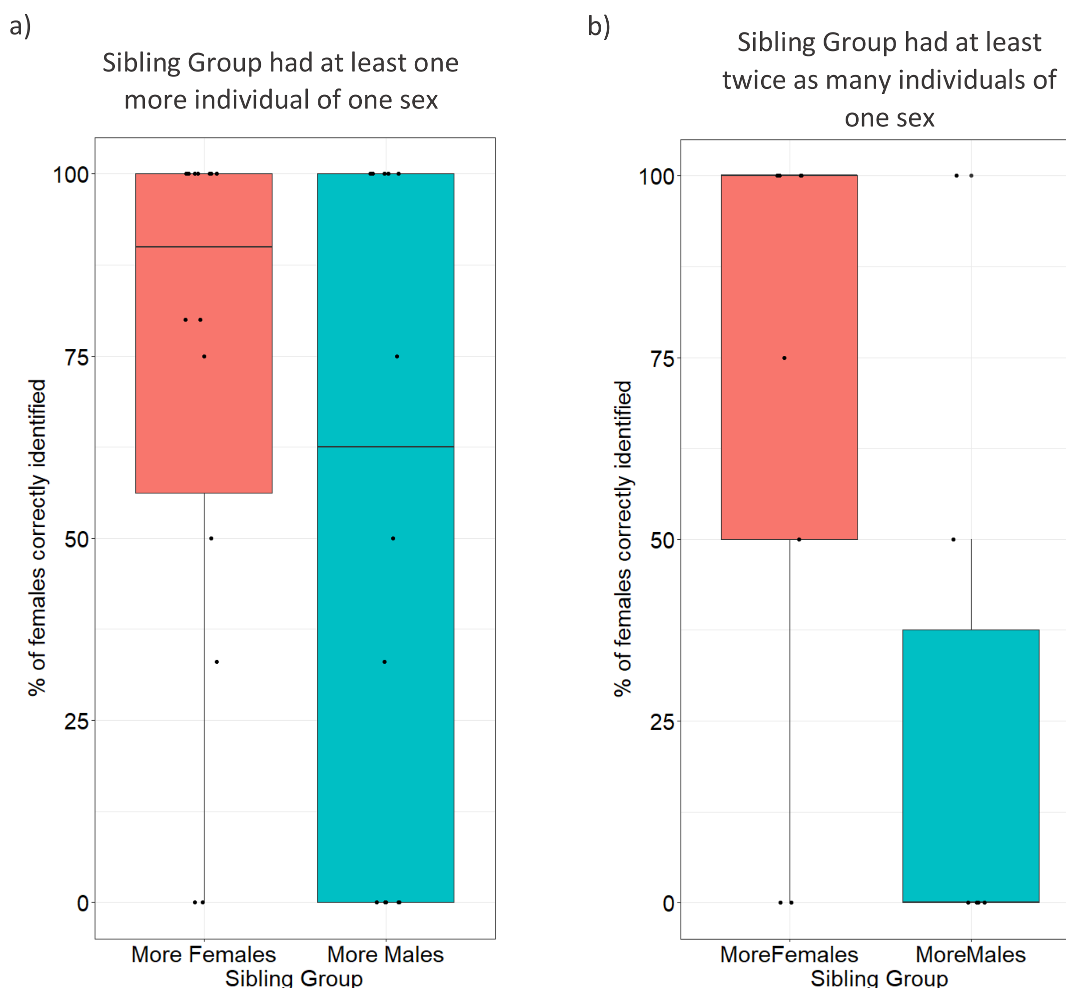


Figure 2.21. a) The percentage of females from the total number of females within a sibling group that were correctly identified by random forests. Separated based on if the sibling group had more females than males (More Females) or if there were more males than females (More Males). b) The percentage of females from the total number of females within a sibling group that were correctly identified by random forests. Separated based on if the sibling group had at least twice as many females than males (More Females) or if there were at least twice as many males than females (More Males).

Some of the most discriminant mass bins were only one Dalton apart. Comparing the intensities of the top eight mass bins to each other shows that the one Dalton apart mass bins

had a stronger correlation. This suggests that these bins may be a result of carbon isotopes. Phospholipids and fatty acids consist of hydrocarbon chains. The presence of a ^{13}C would increase the atomic mass by one. Therefore, one of each of the pairs of bins that were one Dalton apart was removed from the list of the top five mass bins (Figure 2.22). Two pairs of mass bins had a difference of 28 Daltons, which could be due to isomeric compounds or fragments differing by two CH_2 entities, each with a molecular mass of 14. Isotopes were removed from the lists of the most discriminant mass bins, as they show the same intensity distribution as each other. Including the next informative mass bin that was not an isotope would be more valuable.

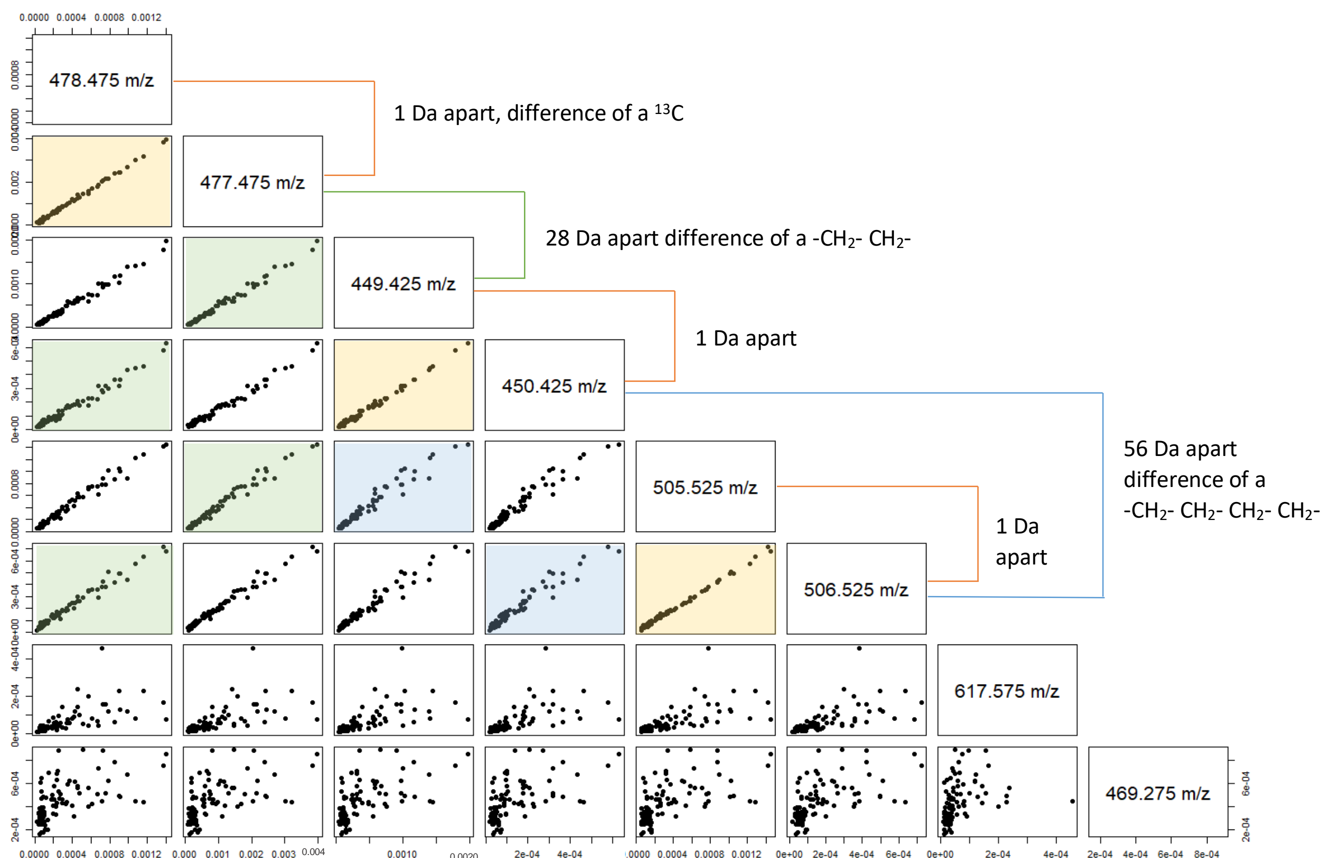


Figure 2.22. The intensities of each of the top eight mass bins most responsible for the differences between sexes against the other seven. The bins were the top eight most discriminant from an average of ten random forests using all 176 samples. The highlighted plots are the pairs of mass bins that are either one, 28 or 56 Dalton apart from each other. These bins show a stronger correlation compared to the other two bins (617.575 and 469.275) suggesting they may be isotopes (1 da apart) or have a fatty acid chain that either has two (28 Da apart) or four (56 Da apart) more hydrocarbons.

The model can separate adult mice based on sex when the samples are collected and burned simultaneously. New samples were collected and burned over a year after the previous sample collection. The model was created using all the adult samples from before, with the established list of samples that were consistently misclassified removed as the training data for 98 individuals in the training set. Removing the misclassified samples could help the training data build a random forest model with the best chance of finding the differences between classifications. The model was used to predict the test data, which consisted of 22 new individuals: 10 females and 12 Males (Group I). An average of five random forests gave an average prediction accuracy of 82%. Only five random forests were used as the random forest accuracy did not change between runs as the same data was used to build the model each time. The model consistently predicted seven females and 11 males correctly

each time. This was repeated with the second group (Group II) of 21 individuals, 11 females and ten males, collected two weeks after Group I. The prediction accuracy was only 47%, as the model classified all the individuals as males. Group III, which contained 11 males and 11 females, collected and burned a month after group I also gave an accuracy of just 50% as all individuals were classified as male. All samples were combined to create a new model minus one sample from the new groups. This model was used to predict the left-out sample and was repeated, and a new model was created until each new sample was tested. Of the 65 new samples, 28 were correctly identified, giving an accuracy of 65%. This is lower than with the first model but higher than when using only older samples in the model. The intensity distributions for groups II and III for the top three mass bins were much higher than the original model and group I (Figure 2.23). The average spectra also had a higher overall ion count across the spectra for groups II and III than the original 176 samples. (Figure 2.24). The increase in intensity was contributed to an increase in instrument performance due to a new StepWave^{TM}.

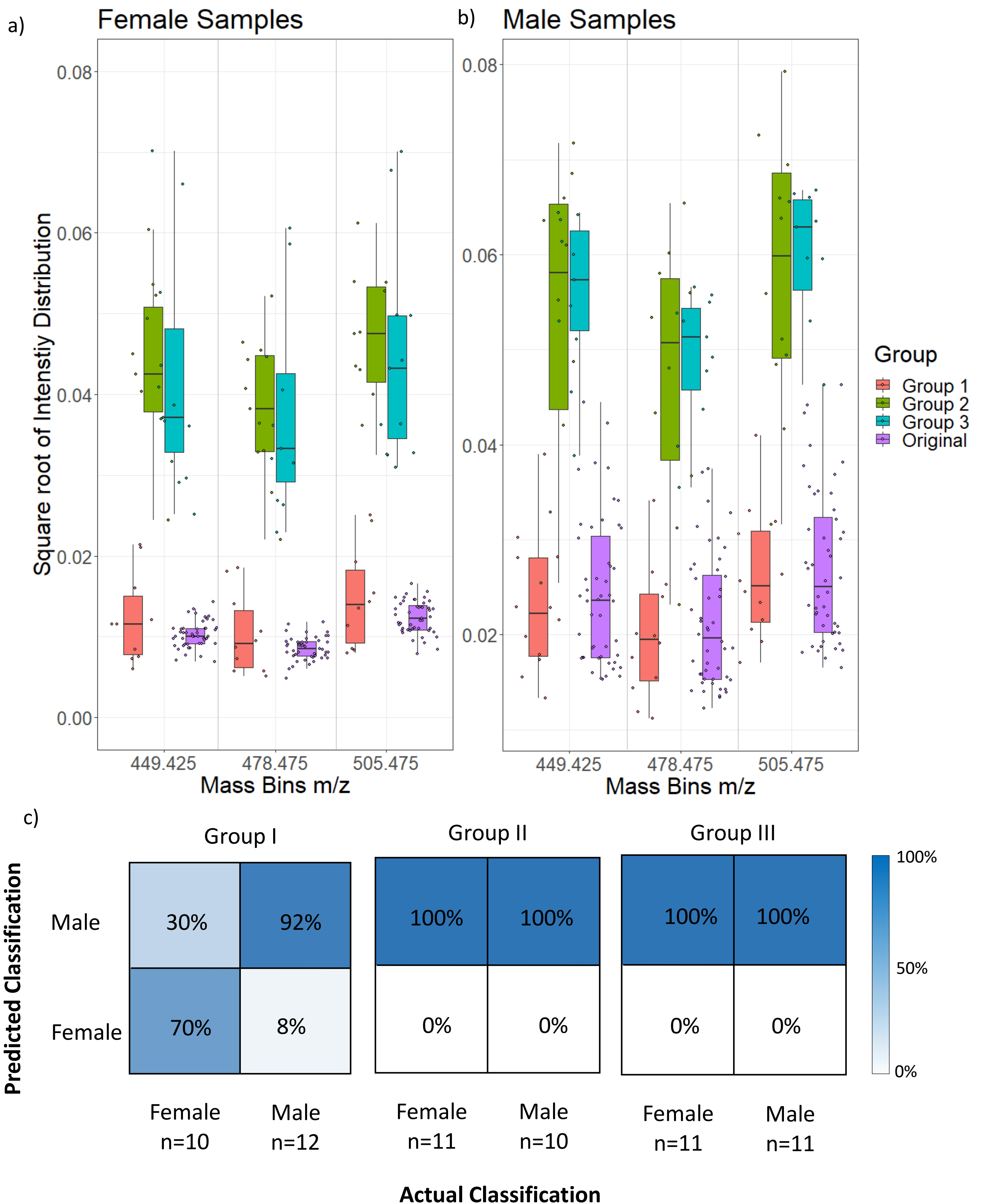
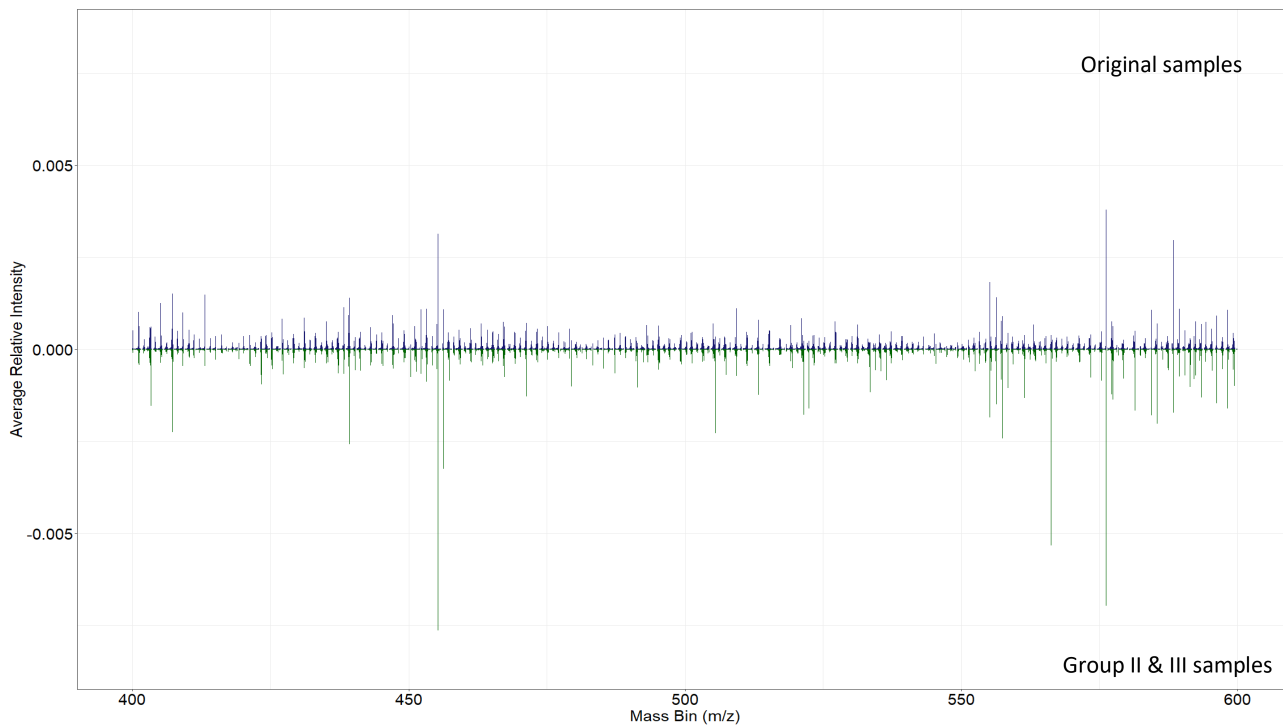


Figure 2.23 a) The relative intensities of female samples for the top three mass bins that separate sex for each of the four collection groups. Original represents the samples used to train the model from the 176 samples with juveniles and misclassifications removed. Groups I, II, and III were collected a year later, two weeks apart. b) The male samples from the same collection groups. c) The predictions accuracy for each of the three collection groups, the original group was used as the training data for groups I to III.



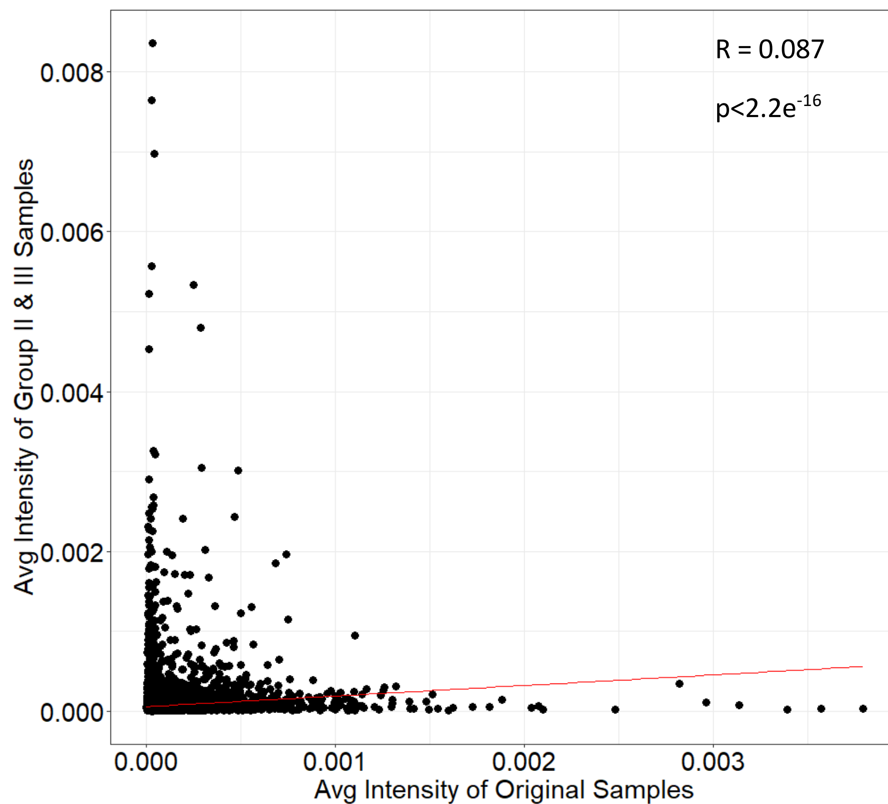


Figure 2.2 The average intensities of all samples from the original data set (n=176) and for the samples from groups 2 and 3 (n=43). There were observable differences between the spectra and no correlation between the intensity of samples from the different groups. The samples from groups 2 and 3 were analysed with REIMS a year after the original samples.

2.3.4. Classification of Age

The random forest and prediction accuracy gave an average result of 80% from ten models when using the under-sampling function (Figure 2.14). Under-sampling was performed before the ten models, so a different group of adults was present in the dataset for each model. Unlike in previous models, the training and test data were split 80:20 to increase the number of juvenile samples in the training data. The data was split randomly, so the number of juveniles in the training data would change slightly each time, potentially leaving just four juvenile samples in the test data. Two models had only one juvenile misclassified, six of the ten runs had two juvenile samples misclassified, and the other two had three juvenile misclassified. Since the number of juveniles varied between 4 and 11, the percentage of juveniles misclassified varied between 12% and 50%, averaging 30%. The same juveniles were consistently misclassified, so the overall accuracy of predicting juveniles may be higher with increased juvenile numbers (Figure 2.25).

The random forests were repeated ten times, but with all the data, so there was no split into training and test data. The top mass bins and a list of misclassified samples in each model were also established for this data set. Out of 176 individuals, 24 were consistently misclassified; they were assigned the wrong age classification in at least 9 out of ten models. Nine samples were misclassified in at least two of the model runs. One sample was misclassified in just one run of the model. This could have occurred by chance due to the random nature of random forests, so this individual was classified as always correctly identified. Unlike with sex, the samples that were misclassified and the individuals that were assigned the wrong age had varied intensities for the top mass bins, and they were not always the highest or lowest values (Figure 2.26).

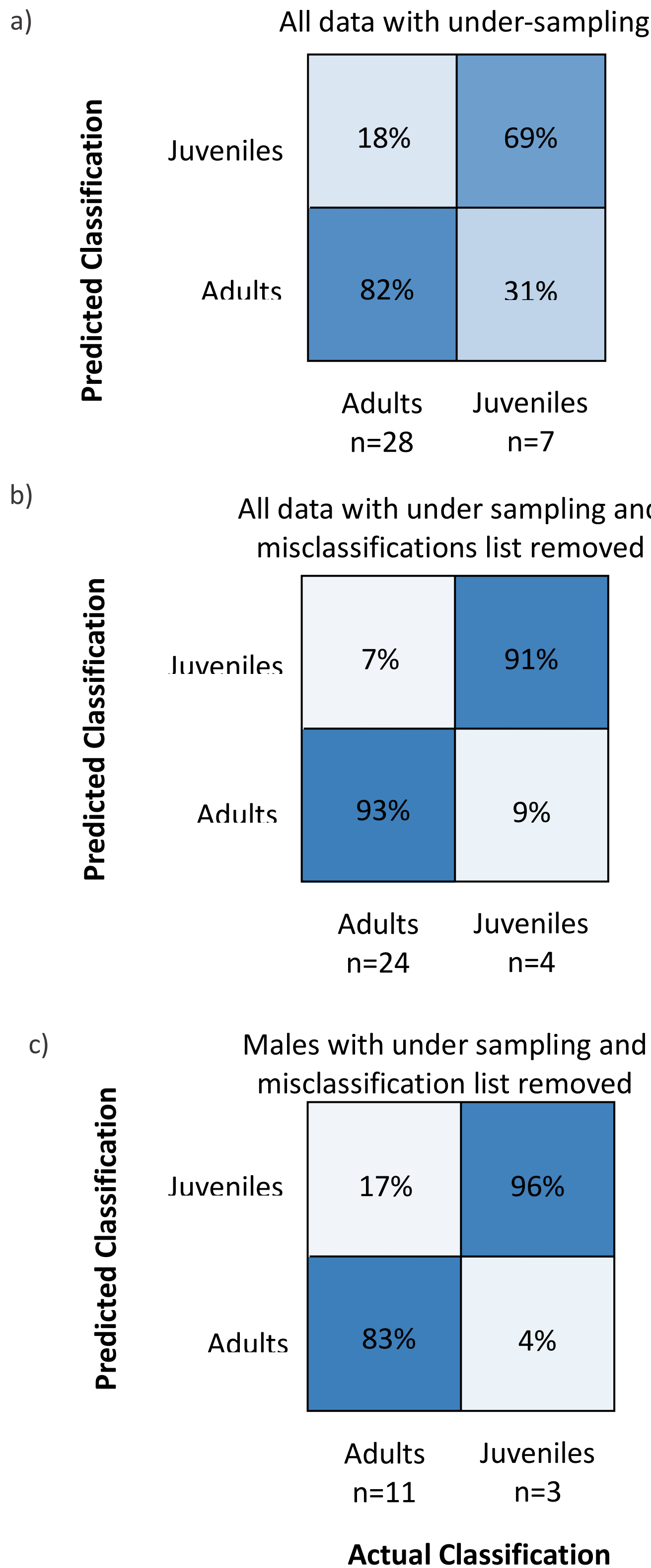


Figure 2.25. a) The average prediction accuracy of adults and juveniles from ten random forest models using the under-sampling function. b) The average prediction accuracy of adults and juveniles from ten random forest models using the under-sampling function and with the samples that were misclassified in all ten of the previous random forest models removed. c) The average prediction accuracy of male adults and juveniles from ten random forest models using the under-sampling function and with the samples that were misclassified in all ten of the previous random forest models removed.

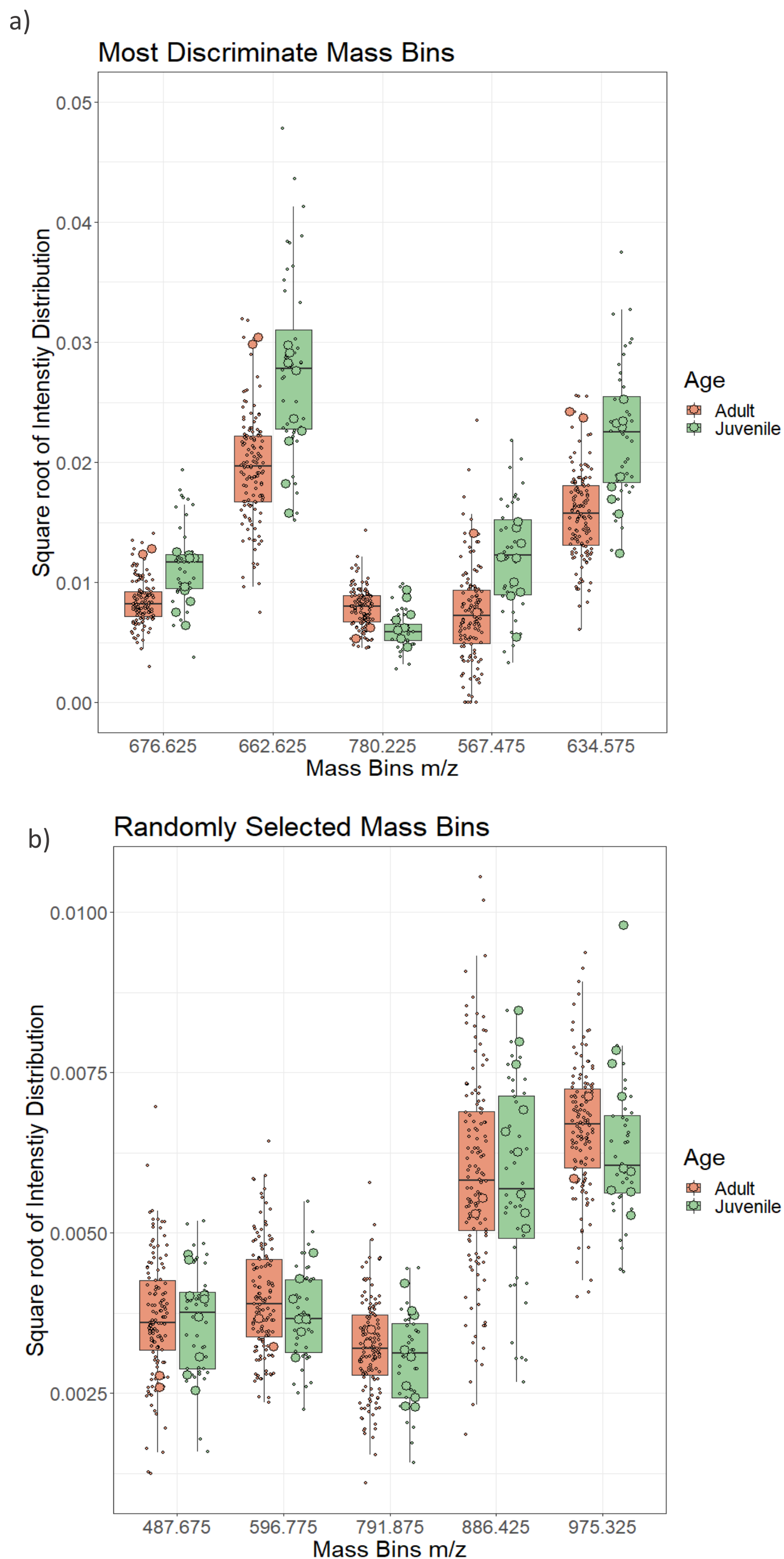


Figure 2.26. a) The average relative intensities of the five most discriminant mass bins of adults and juveniles. The larger points are the samples that were misclassified as the wrong age. With sex the misclassified samples were either above or below the interquartile range whereas the age misclassified samples have more varied intensities. b) The average relative intensities of randomly selected mass bins of adults and juveniles

The individuals that were misclassified were those closest to the age boundary (Figure 2.27). Females were classified as adults when they reached 44 days, but maturity was probably reached before this point. However, no female samples were collected from individuals between 28 and 44 days old; all-female juveniles were between 25 and 28 days old. Samples were only collected from juveniles after weaning at 21 days old. An extra four days were added to allow the individuals to habituate to their post-weaning surroundings to

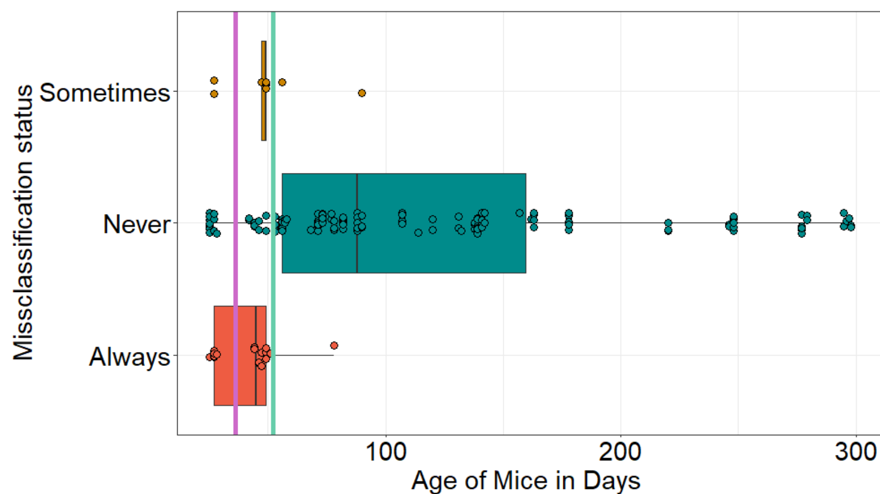


Figure 2.27. The age of all 176 mice in days and whether they were misclassified in all ten random forest models (always), between 2 and 8 random forest models (sometimes) or if they were never misclassified (never). The purple line indicates the age boundary for females and the green for males.

reduce potential stress for the young juveniles. Males mature later than females, and adults were assigned to individuals 52 days old and above. The male juveniles ranged in age from 25 to 51 days old. Age is a continuous variable, and the spectrum changes as the animal ages into adulthood (Figure 2.28). For the top five most discriminant mass bins, the juvenile intensity decreases as they age, and for females, the most significant drop was observed after 25 days.

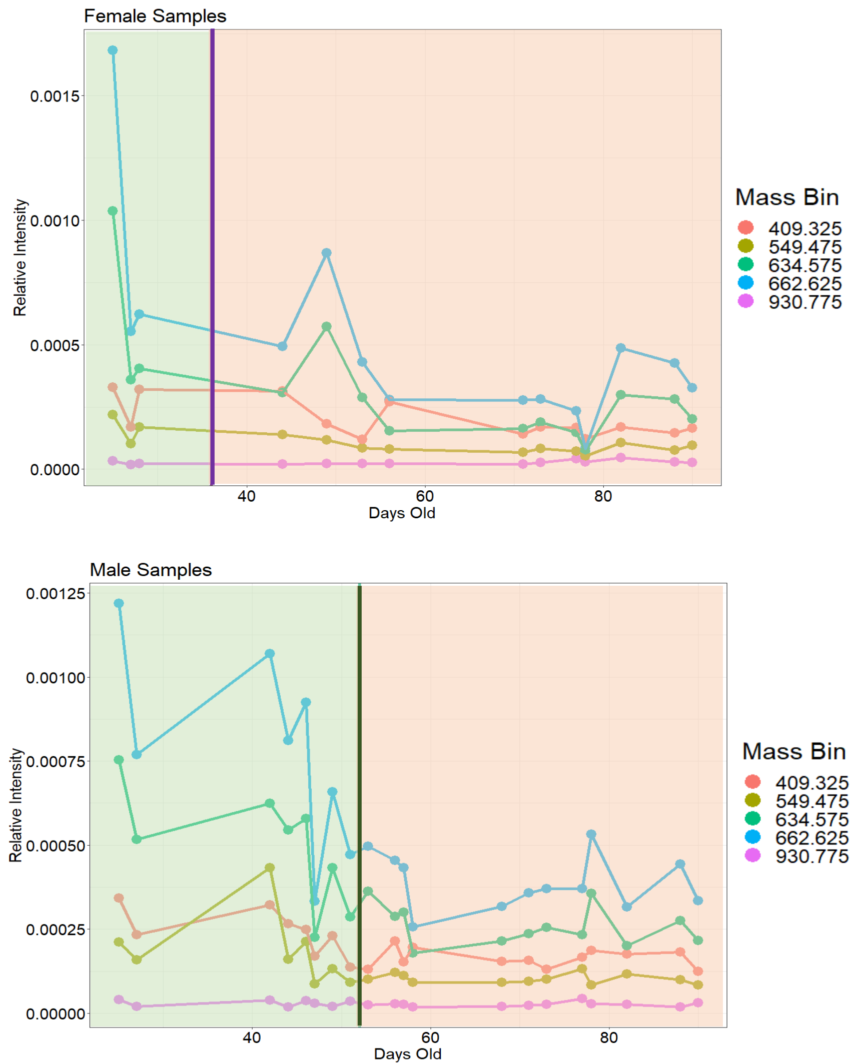


Figure 2.28 The intensities of females under 100 days old for the top five mass bins responsible for the differences between adults and juveniles. The age boundary for females is 36 days old. The intensities of males under 100 days old for the top five mass bins responsible for the differences between adults and juveniles. The age boundary for males is 52 days old.

The males also showed a decline in the top mass bin intensity after 25 days, but it decreased less than the females. Males showed another drop before and after the 52-day mark, suggesting that maturing into adults affects the spectra. Ten random forests were repeated with under-sampling but with the consistently misclassified samples removed. The average random forest accuracy was 87%, while the prediction accuracy was 93%. Although there was an increase compared to the previous accuracy of 80%, the increase was less compared to sex. This model was repeated using only male samples, the random forest accuracy was 86%,

and the prediction accuracy was 85%. Removing the females did not affect the random forest accuracy, and the prediction accuracy decreased slightly. The classification of adults performed better when both males and females were included, and having a lower sample number could have more of an effect on accuracy than using both sexes (Figure 2.25). The PCA-LDA separation is also greater between adults and juveniles when the individuals closest to the age boundary were removed Figure 2.29)

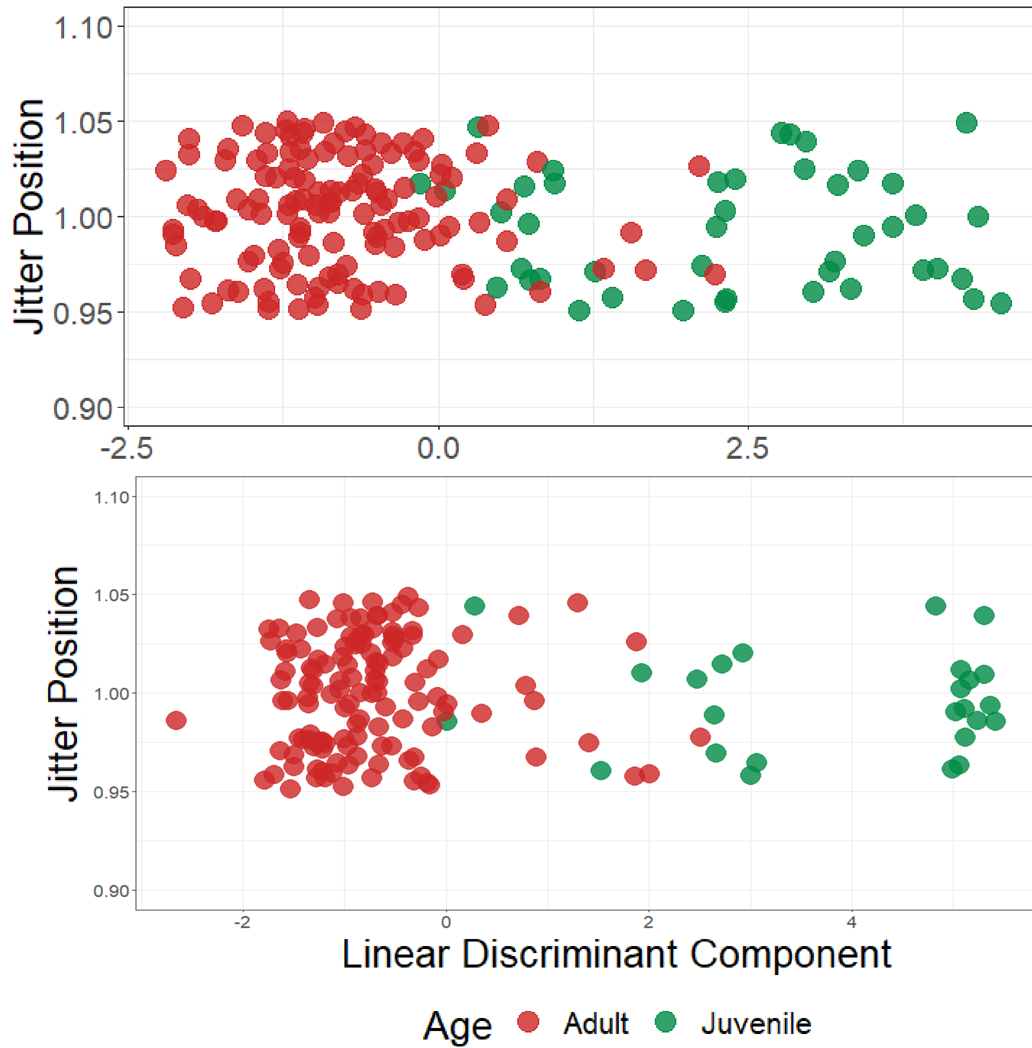


Figure 2.29. a) A PCA-LDA plot of age using the top 70 PCA components with all samples included in the model. b) A PCA-LDA plot of age using the top 20 PCA components with the samples closest to the age boundary removed. The jitter position is a randomly assigned value to disperse the results vertically so individual sample points are visible.

2.3.4 Classification of Age and Sex

The average random forest accuracy when age and sex were included was 71%, and the prediction accuracy was 64%. This was lower than when only sex or age was included in the models. The random forest correctly identified adults more than juveniles, with adult females as the group most correctly identified (77%), but none of the juvenile females was correctly identified (Figure 2.30). There was a significant difference between the top five mass bins separating the four groups, and the adult females had a lower intensity distribution than the other groups. Female juveniles had intensities more similar to both male groups than females. The top mass bins that separated age and sex were very similar to those that separated males and females. An LDA on the top 40 components also showed a separation between all adults and juveniles, between adult females and males but not between juvenile males and females (Figure 2.31). When the male and female juveniles were grouped, the random forest accuracy for adult females, adult males, and juveniles remained 71%; the prediction accuracy increased to 76%. Female adults still had the highest classification accuracy at 81%, and juveniles had the lowest at 65% (Figure 2.32). The top five most discriminant mass bins remained the same as when juveniles were separated by sex.

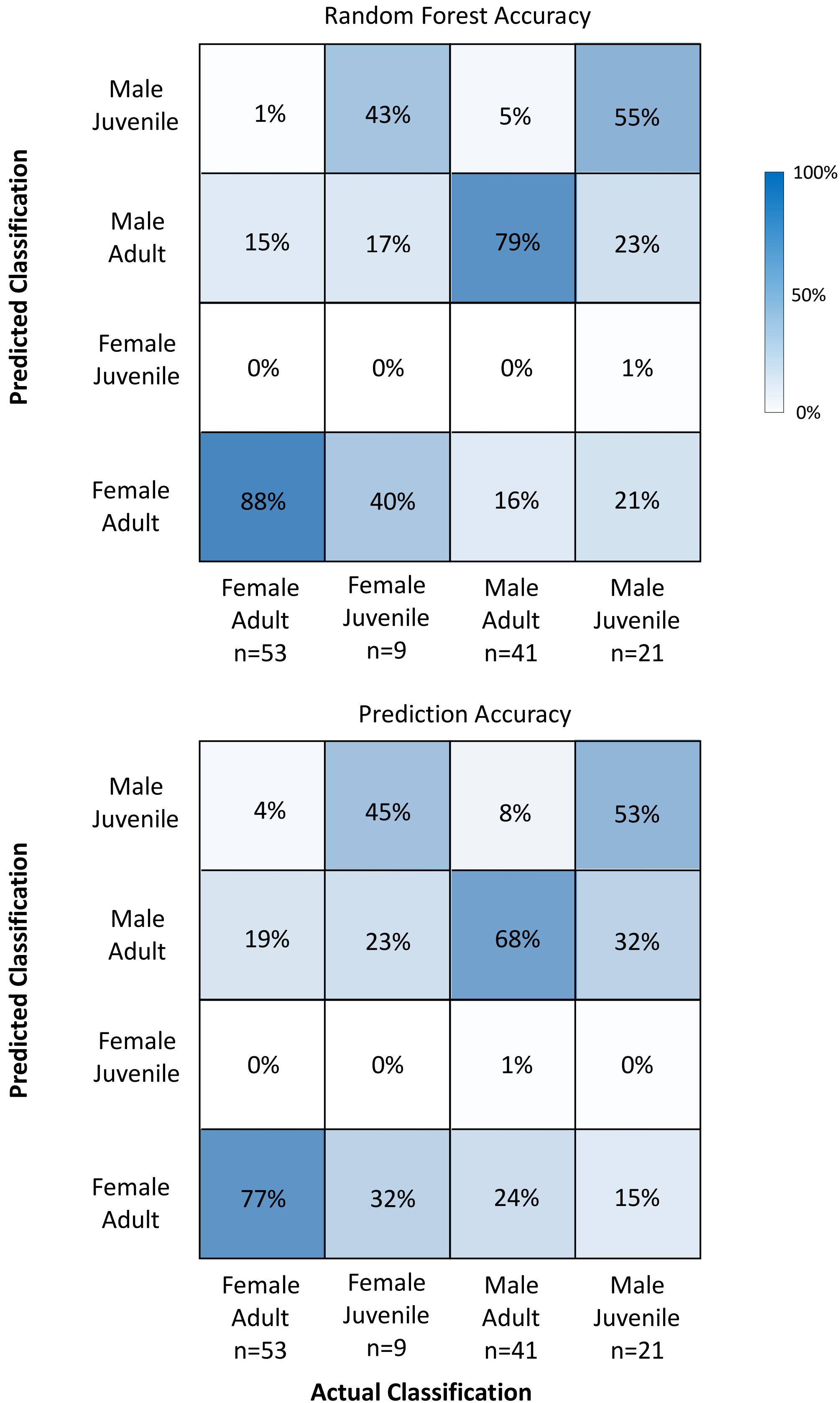


Figure 2.30. The random forest and prediction accuracies from an average of ten random forests. The random forests were built with a different subset of data each time with about 70% of the samples. This was used to predict the classification of the remaining 30% as adult male, adult female, juvenile male, or juvenile female.

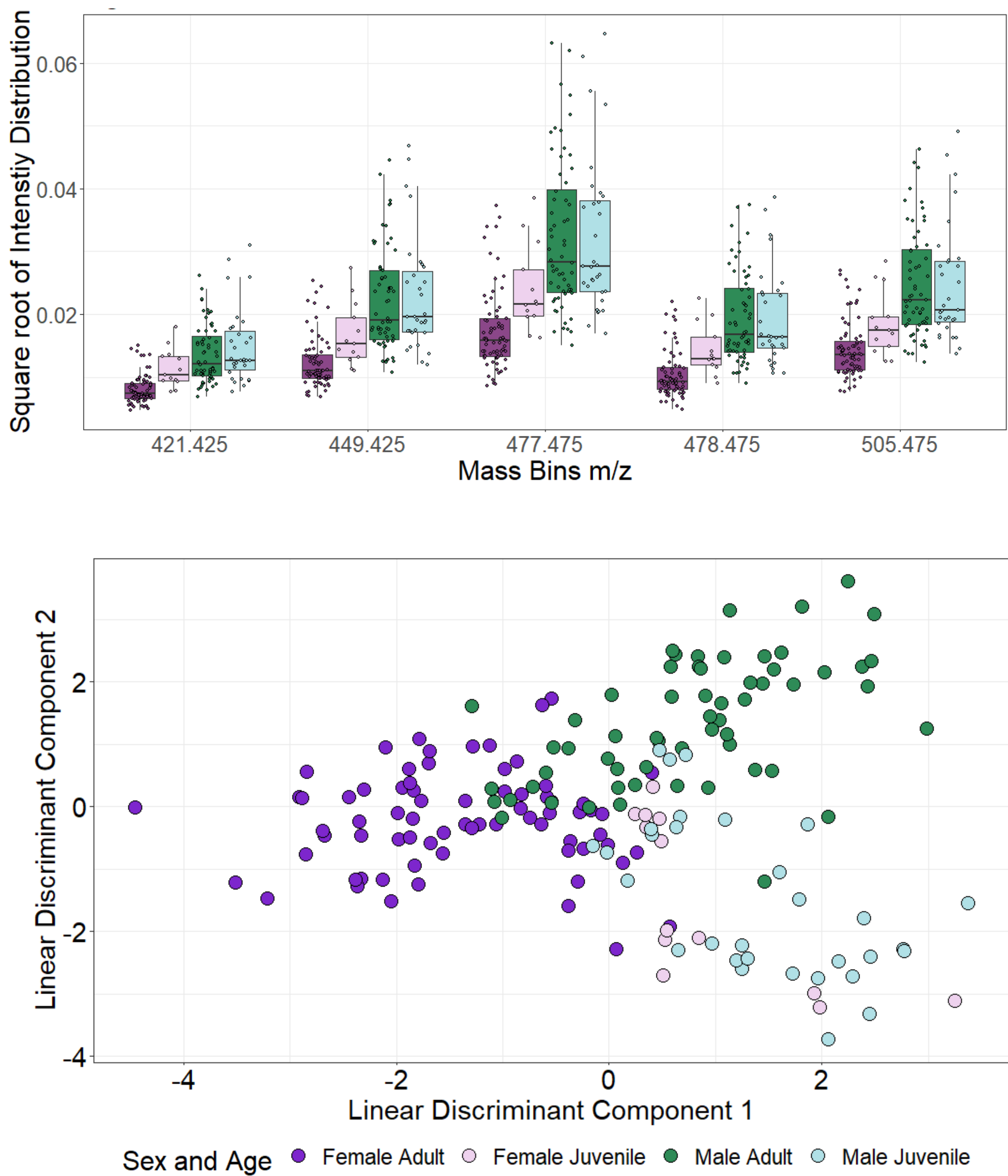


Figure 2.31. The intensities of the top five mass bins responsible for the differences between adult females, adult males, juvenile females, and juvenile males. b) A PCA-LDA plot of strain using the top 40 pca components.

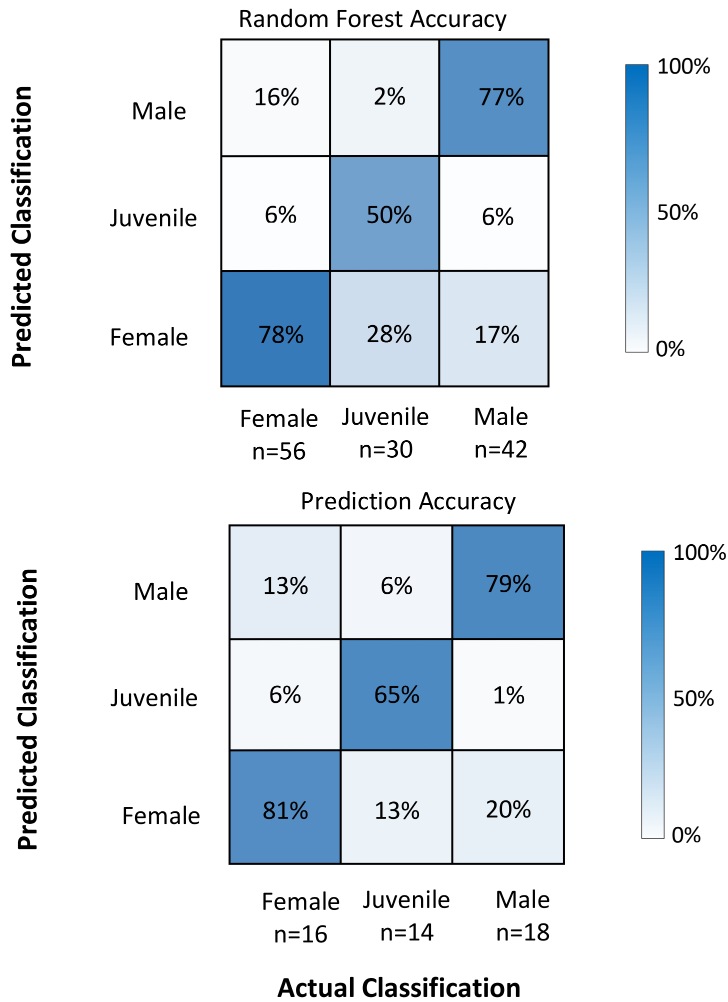


Figure 2.32. The random forest and prediction accuracies from an average of ten random forests. The random forests were built with a different subset of data each time with about 70% of the samples. This was used to predict the classification of the remaining 30% as adult male, adult female or juvenile.

2.3.5 Classification of Strain

With the BALB.k samples removed, a random forest accuracy of 82% and a prediction accuracy of 76% were achieved but, accuracy was inconsistent between each strain. BALB/c had the highest accuracy of 95%, whereas the other two strains were 68% and 71%, which was a range of 27%. Most of those samples that were misclassified were assigned as BALB/c. There were nearly twice as many BALB/c individuals as the other two strains. It was thought that this unbalancing in the group numbers might be causing a decrease in accuracy, like with age. Using the under-sampling function only increased the overall random forest accuracy by 2% to 78%, and the prediction accuracy increased by 8% to 84% (Figure 2.14). The individual accuracies were more even this time, with BALB/c decreasing to 81%, C57BL/6 increasing to 72%, and ICR to 80%. Like with age and sex, it was the same individuals being misclassified each time, and when these samples were removed from the data set, the random forest accuracy and the prediction increased to 91%. There was also less difference between the individual accuracies even though under-sampling was not used. BALB/c was the highest, with 97%, followed by 90% for ICR and 79% for C57BL/6, giving a range of 18% (Figure 2.33). When using the under-sampling function with strain, only two of the top five mass bins were the same as when using all the data. There was a significant difference between the intensities of the top five mass bins, and separation was observed on an LDA of the top 80 PCA components. (Figure 2.34).

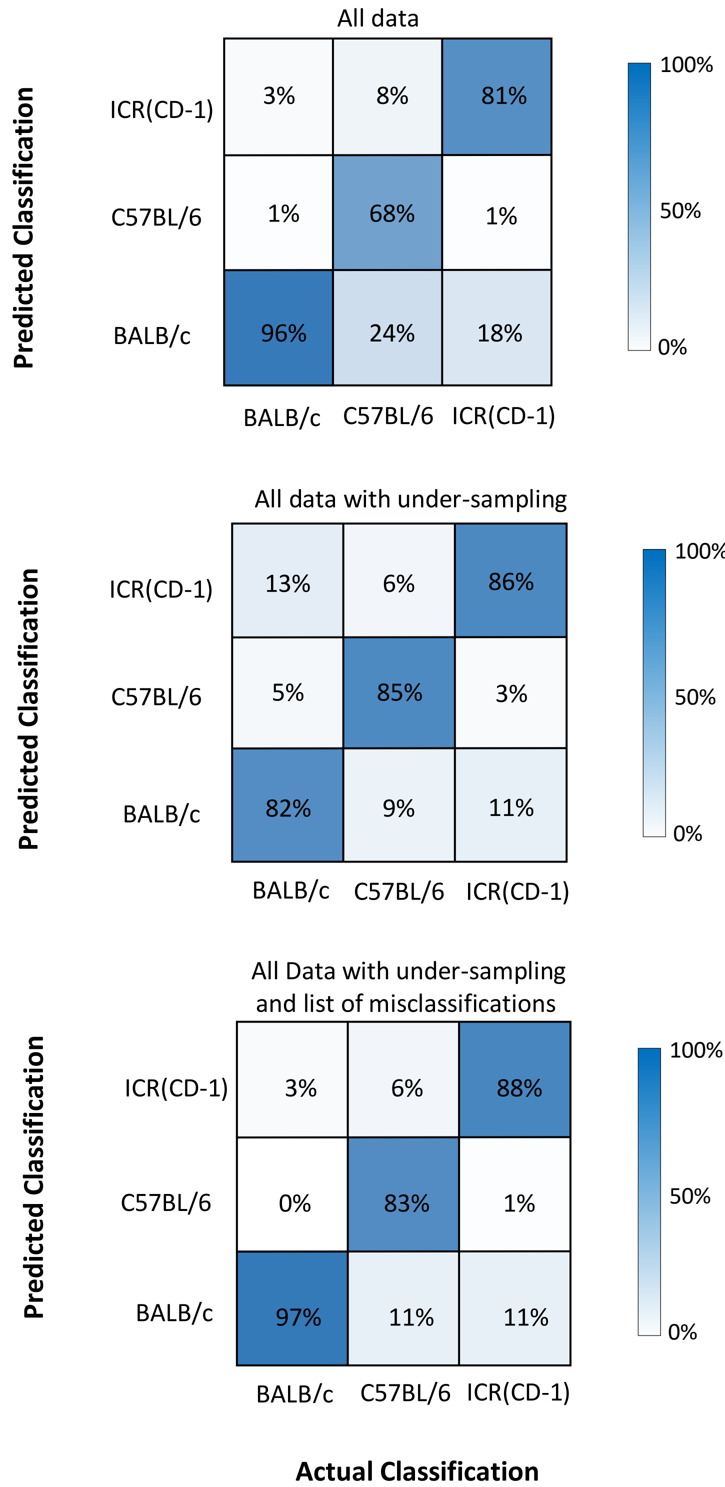


Figure 2.33. The prediction accuracies from an average of ten random forests for when all the data is used, when the under-sampling function was used with all the data and when the under-sampling function was used with the misclassifications removed from the data set.

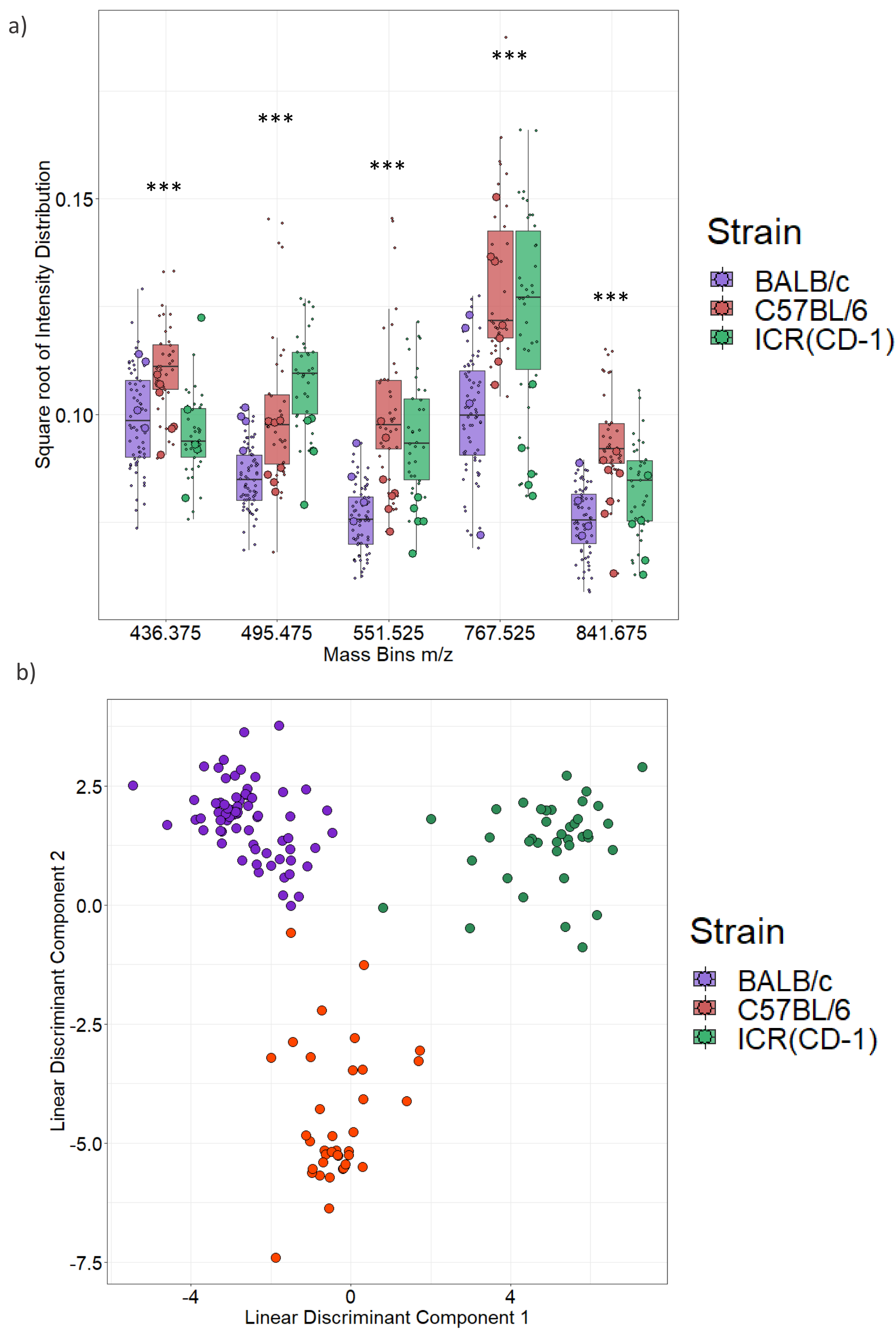


Figure 2.34. The intensities of the top five mass bins responsible for the differences between each strain.(Kruskal-Wallis p<6.9e-15 for all mass bins) b) A PCA-LDA plot of strain using the top 80 pca components.

2.4 Discussion

2.4.1 Optimising the REIMS Method

REIMS can produce a spectrum containing phenotypic information from faecal pellets. The ability of random forests to distinguish between spectra can be improved by changing the burning methodology and processing/model parameters. Adding water increased the accuracy; water helps with conductivity, and the optimum volume of water added depended on the sample. Some samples were drier than others, requiring more water, but smaller samples would disintegrate if too much water were added. Therefore, the amount of water to add to a sample should be done at the user's discretion. For mouse pellets, 200 µl evenly distributed over the three pellets was efficient at obtaining conductivity without causing disintegration of pellets. Not all pellets produced by an individual are the same size and consistency. The faecal matter consistency may depend on the time since the individual last ate and the amount of water and food consumed. Therefore, if possible multiple pellets should be used for each individual, increasing sample processing time. The same analysis (random forests) was carried out using the same individuals but either using one, three or five pellets. The prediction accuracy also increased the more individuals included in the analysis, but only after individuals were removed to ensure that each class contained the same number of individuals. The prediction accuracy increased the more pellets were used, but there was not much of an increase between three and five pellets, so three pellets are adequate to use if five pellets are unavailable. Five was chosen as the maximum because this is the highest number of available pellets for all individuals. Obtaining more pellets from individuals in one sitting would require placing the mice in the collection cages for more than two hours. Keeping the mice from water and food for this long would have been detrimental to their welfare.

The spectra were uploaded to Waters' software LiveID, and the software can be used to normalise the data and build PCA-LDA models. Various parameters can be changed while building the model, including the standard deviation, bin size and mass range. Using the waters software has limitations as they only allow specific parameters to be controlled by the user. Unseen algorithms may occur within the software when performing PCA and LDA tests, leading to over-fitting. Therefore, the spectra produced using REIMS were analysed using random forests. Random forests cannot separate random data, and R allows the user more

control over changing parameters. A test set can confirm if over-fitting occurs; if there is overfitting, the classification accuracy of the training set will be significantly higher than the test set. An Excel file containing each mass bin and burn intensity can be extracted from LiveID and uploaded to other software, such as R, for further analysis. The bin size can be changed from 0.01 to 1.0; decreasing the bin size will increase the total number of bins using the same mass range. Having more bins may expose minor differences that may otherwise be hidden by more significant, consistent peaks but requires a much larger data file, leading to longer processing times. With the mice, samples decreasing the bin size increased the accuracy and a decrease in the mass range would help with processing time. With the mouse pellets, the accuracy increased when only the middle of the mass range was used (400 to 700 m/z). When carrying out a REIMS study, some time should be spent determining the best bin size and mass range, which may change depending on the questions being asked. A bin size of 0.1 was adequate for species separation (Davidson et al., 2019), but classification accuracy improved with a smaller bin size within species. This was probably due to the difference between species being more significant than those within a species.

Random forests were used rather than relying on just PCA-LDA for analysis due to the potential for PCA-LDAs to cause overfitting. This would explain the low cross-validation results for the randomised data, even though the LDA showed high separation. The water's software may perform additional filtering steps that can not be prevented to force a separation between data. Performing random forests in R gives the user more control over the analysis, and this method prevents overfitting, as random forests cannot separate random data. Therefore, once the model was built within LiveID, the binned data was extracted as a .csv file which was uploaded to R for analysis. Each individual must have one list of mass bin intensities for the random forests, but LiveID gives a list for each burn. The intensities were either averaged, summed, or a random burn was selected; the highest prediction accuracy was obtained when the data were averaged.

A problem that may have been contributing to the high variation between models was the varying consistency of burning pellets. Each pellet should produce one burn; therefore, each individual should have had three burns. Many individuals produced more burns than pellets, with some having up to seventeen burn events. Some pellets burn better than others because they're bigger or have a better consistency. Occasionally, pellets would have to be flattened,

and then water added to obtain a burn, some pellets would not burn regardless of the water added. It was unknown what caused a difference in pellets consistency, as there were problematic pellets to burn across all sex, strains, and ages. The consistency was based on an individual rather than any other factor. It could depend on the individual's size or the length of time since the animal last ate; since the animals were fed *ad litum,* it cannot be known when they last ate. Burning multiple pellets and using an average ensures the spectra represent a pattern for an individual animal and not just a particular faecal pellet. Therefore, when analysing faecal pellets, multiple pellets should be burned per individual when possible.

The efficiency of burning may have been affected by the mode choice of the diathermy electrode between cut or coagulate. A study using the monopolar electrode and REIMS to distinguish between cancerous and non-cancerous breast tissue compared coagulate and cutting modes. A leave-one-out cross-validation was used for analysis; cut mode produced an accuracy of 95.8%, while coagulate mode produced a similar accuracy of 94.7%. In cut mode, the high-intensity spectra were in the phospholipid range (600-850 m/z) and the triglyceride range (850-1000 m/z), while coagulate mode had higher intensities in the triglyceride range, the mass range for analysis was set to 600-1000 m/z. The study concluded using cut and coagulate spectra within their analysis, as both modes would be used during surgery (Edward et al., 2017). Using the coagulate mode in this study produced a slight decrease in burn number, possibly due to the coagulate button being more comfortable than the cut mode button. With coagulate mode, it was easier to maintain continuous contact with the pellet while burning (Figure 2.12).

Changing analysis parameters such as bin size and mass ranges could increase random forest accuracies. The optimum parameters may depend on the type of faecal pellets to be burned and the proposed research question. The amount of smoke produced from the pellets could be observed changing while adjusting the electrode voltage; when the voltage was too low, the electrode was insufficient at burning the pellets, and so little to no smoke was produced. Sparks were produced when the voltage was too high, and the pellet disintegrated more quickly, causing less smoke and leading to more burns per pellet. For mouse pellets, the optimum voltage was 30 V when using cut mode and 35 V with coagulate mode. Changing between cut and coagulate mode and changing voltage only significantly reduced the number of burns produced.

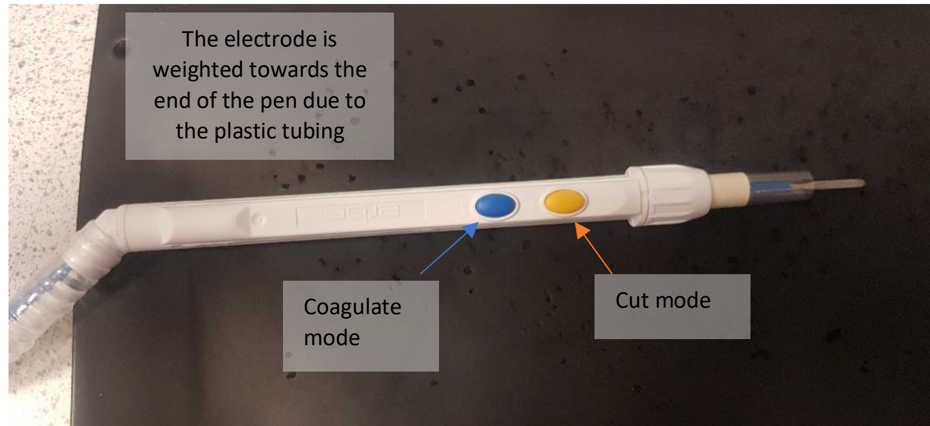


Figure 2.12. Due to weight distribution, the pen is easier to hold towards the end with the tubing, for those with smaller fingers this means the coagulate button is closer and easier to hold for longer periods of time. The position of the buttons would not affect every user and may only make a difference when burning for significant periods of time.

Specific parameters will also increase processing time; depending on the number of variables, this could be quite a difference overall. Therefore, each time a REIMS project is approached, time should be taken to establish the best method to burn the pellets and analyse the spectra. This could be performed by just using a small subset of pellets. Box 2 summarises each step that should be taken to help optimise the REIMS method, and this approach could be taken for other types of biological samples.

Box 2

REIMS workflow for sampling faecal samples

Does the Faeces produce a burn signal?

Step 1: Establish Best Faecal Composition to Burn

- Should the whole pellet be used – Is the faecal composition uniform?
- Does water need to be added before burning and how much?
- Does the pellet need to be prepared first e.g., mixed with water or alcohol?
- Should the faeces be dried first?
- Do the faeces need to be filtered, for herbivores vegetation may need to be removed?

Step 2: Establish whether to use cut or coagulate mode on the diathermy electrode

- Consider if using a different type of electrode is better (tweezers or bipolar probe)

Step 3: Establish what voltage to use for the diathermy electrode

Step 4: Determine how much faeces to use per sample (number of pellets), and how many burns per faecal sample.

Do the spectra produced by REIMS contain discriminate information?

Step 5: Upload the spectra to waters software, either LiveID or OMB

Step 6: Establish what model parameters to use

- Mass range – start with a wide range (e.g., 400-1100)
- Bin Size – start with a large bin size (0.5)

Step 7: Run a PCA-LDA, to determine if the samples separate

Step 8: Export spectra from the Waters software for further analysis in R using random forests

Step 9: Establish classification accuracy

- Create ten random forests and establish the average classification accuracy and top discriminate mass bins

Step 10: Improve classification accuracy

- Change the mass range
- Decrease the bin size
- Repeat model with top mass bins only
- Filter samples e.g., remove juveniles from models discriminating sex

2.4.2 Using REIMS to Distinguish Intra-species Variation

The final overall prediction accuracy for assigning the correct sex to an individual was 99%. Although more steps were required to obtain this high accuracy than species identification, it shows REIMS can identify differences within the same species. When observing the boxplots of the top five most discriminant mass bins, the misclassified male individuals had lower intensities than the lower interquartile range. In contrast, the misclassified females had intensities higher than the higher interquartile range. When compared to five mass bins at random, the intensities of misclassified individuals were more varied, so it is not that these particular individuals have a lower-than-average intensity across the whole spectra for males and higher than average for females. They differ only in those mass bins responsible for sex variation. It was thought that these differences might be caused by bi-directional blood flow in the uterus. When foetuses of rodents are in the uterus, blood and steroids pass between siblings; therefore, a female lying between two males is exposed to higher levels of testosterone than a female beside one or no males. The anogenital distance of females surrounded by more males in the uterus is higher than other females (Vom Saal and Dhar, 1992). The individuals with anogenital distance measured did not show a correlation between length and classification accuracy. Measurements were only taken from four sibling groups so that further research may show a stronger correlation. Even if the bi-directional flow of steroids affected the masculinisation of females, other parameters must also have an effect. It could be that bi-directional flow only contributes to increased hormone levels to the point of changing the spectrum when a female is surrounded on either side by two males. The sex ratio cannot be used to predict what order individuals lie in the uterus.

Females from sibling groups with twice as many males as females did have a lower classification accuracy than other females, although the results were not statistically significant. It could be that the bi-directional flow is making some females have a more male-like spectrum, but this is not the only factor affecting prediction accuracy. A potential influence may be the time since an individual last ate. When the pellet was produced, a difference may be observed between the first pellet produced during the two-hour collection period and the last pellet produced. Further work using anogenital distances would need to be carried out to confirm the effect bi-directional blood flow has on an individual's ability to be classified.

REIMS can classify the sex of faecal pellets when pellets are collected and burned at once. When the ability to classify new samples using a model built with older samples was tested, the model could only correctly identify females in one of the three groups. All three new groups were collected a year after the testing model was built, and there was a two-week gap between groups I and II. Between these two weeks, the StepWave^{TM} was cleaned, which could have caused a change to the average spectra produced from the faecal pellets of lab mice. The change in spectra was a higher overall intensity suggesting instrument performance increased. Individuals from group I was able to be classified so REIMS could be used to burn samples for a significant amount of time after creating the original model. The REIMS signal, however, is susceptible to change and should be checked when burning new samples. If the signal changes, adding new known samples to the original model could help identify new samples.

REIMS can be used to classify lab mice as adults and juveniles. The highest overall accuracy (93%) for age was not as high as sex, but this was expected as age is a continuous variable. The prediction accuracy may improve if a model could be built with more individuals of a younger age (less than 30 days old). The changes between juveniles and adults happen much earlier in females than males. Females showed the most significant difference between 25 and 28 days old, whereas males changed the most between 40 and 60 days old. Mice are considered sexually mature after 35 days but are not mature adults until three months old (Flurkey et al., 2007). The relative intensities for the mice changed more before 60 days old than after, suggesting the spectra are responding to sexual development as opposed to other types of development. Based on the spectra produced by REIMS, 35 days for sexual maturity could be considered an average between both sexes. The spectra for females changed before 35 days (day 28), and for males, it was after day 45. It seems that more mass bins might be responsible for the differences between adults and juveniles, and the top five changed more when changing model parameters compared to sex and strain. This could be because what mass bins show the most significant change is dependent on the average age in days of the individuals classified as juveniles. The classification accuracy was higher for male juveniles than for females. Females showed more adult-like spectra much earlier than males. The classification of juveniles may improve if more juveniles below 30 days of age were added to the model. Therefore, when approaching an age study using REIMS, it may be beneficial to

base the age boundary on changes in the spectra rather than on an age chosen from the literature and to select each sex separately.

REIMS could not be used to classify sex and age within one model. The differences between sexes are more prominent than age, and the random forests only separated adults by sex. There were more adults than juveniles, and an imbalanced data set significantly reduced classification accuracy. Separating adult males and females from juveniles may be possible if the model included more juvenile samples. Juvenile males and females may be too similar to separate; if changes in intensities due to sex occurs after sexual maturity, then increasing the sample number will have no effect.

Lab strains can show differences in appearance and behaviour (Sultana et al., 2019, Jacome et al., 2011). BALB.k differs from the BALB/c strain by an H2k mutation, so the strains could be too similar for the random forests to differentiate between. (The Jackson Laboratory, 2013). BALB/c and BALB.k could be combined into one class, but this caused an imbalanced data set, so BALB.k were removed from the models regarding strain. BALB/C and C57BL/6 are inbred strains and should be more like each other than the outbred strain ICR. In most models, C57BL/6 individuals did have the lowest classification accuracy but were only classified as BALB/c slightly more than as ICR(CD-1). BALB/c individuals were mostly correctly identified but were more likely to be classified as ICR than C57BL/6. The strains are, therefore, different enough for the REIMS to separate. As with BALB.k, BALB/c and C57BL/6 differ at the H2 site of the major histocompatibility class 1 gene, BALB/c is H2d and C57BL/6 is H2b (The Jackson Laboratory, 2013). Removing the misclassifications from the strain data did not dramatically improve the accuracy of all three strains compared with sex and age. The overall accuracy improved because the classification accuracy of C57BL/6 improved, BALB/c already had quite a high classification accuracy, and ICR(CD-1) only increased slightly. This could be because some C57BL/6 strains are more like BALB/c, making it hard to distinguish. Or, unlike age and sex, there is not one factor causing some individuals to be misclassified. It is promising that REIMS can show separation between strains as it suggests that if it can separate animals from the same species with different immune responses, it should successfully separate sub-species in wild animals.

The classification accuracy of various models increased dramatically when specific samples were removed from the data. The acceptability of removing samples is dependent

on the purpose of the model. When a random forest model is used to determine if there is a difference between samples, all samples should be included in the model. If a random forest model is being used to predict the classification of another data set, it may be practical to remove samples to obtain the highest random forest accuracy. A higher random forest classification (training set accuracy) increases the chance of test samples being correctly identified. If a high random forest accuracy were obtained because removing samples led to over-fitting, then the test set's prediction accuracy would not increase. If overfitting occurred, the most important variables separating the training set would differ from the test set. If the increase in random forest accuracy due to removing samples were due to an increased probability of important variables being in the decision trees rather than overfitting, then the test set's prediction accuracy would increase. An increase in both random forest and prediction accuracy was observed in this study when samples were removed. The benefit of REIMS analysis for ecologists and conservationists would be identifying unknown wild samples using a previously built model with known samples. The random forest models must always be created with known samples so it would be possible to refine the list of known samples according to a list of parameters. The prediction model should be built to give the highest chance of correctly identifying unknown samples. If this were to be carried out with wild samples, it would be impossible to know the exact date of birth to exclude all individuals below a specific date. Random forest models must be built with known samples; some live trapping would be needed to establish a model to predict unknown samples. The random forest model could be built using samples from caught individuals above a certain weight or size to ensure they were adults. This model could predict faecal samples collected using a non-invasive method. It would be more valuable if a model built using laboratory samples could predict samples from the wild; then, no live trapping would be required.

2.4.3 Conclusion

REIMS can be used to distinguish intra-species variation of faecal pellets but requires more steps to analyse the results compared to species identification. The highest accuracies were received when only one variable was present; with sex and strain, the accuracy improved when only using adults. This would need to be considered when carrying out a REIMS project with multiple questions so each class being investigated has enough individuals, and each

group should have a relatively equal number of individuals. More steps were required in the analysis stage to build a model capable of identification of these more subtle changes than when analysing species. The REIMS method of analysing faeces is still much more universal than other methods, such as PCR techniques. A few steps can be changed to help overall accuracy at the instrument stage. These steps are minimal, and most biologists with no prior mass spectrometry knowledge should be able to conduct REIMS research. Even with taking the time to optimise the REIMS protocol, REIMS is still a much faster and more efficient method for phenotypic analysis of faeces.

Chapter 3: An Investigation into the Use of REIMS as a Non-Invasive Method of Pregnancy Detection in Okapi (*Okapia johnstoni*)

3.1 Introduction

Since the beginning of the 2000s, the World Association of Zoos and Aquariums (WAZA) has aimed to become a world-leading conservation organisation. Their first review of their conservation projects in 2009 suggested that zoos contribute to conserving threatened species. However, the lack of funds individual zoos receive can be a limiting factor. The review suggests that more collaboration among zoos could help increase their potential as a conservation force (Gusset and Dick, 2010). Conservation has not always been a priority in zoos, and there were criticisms that zoos mainly used conservation for publicity to encourage more visitors to the zoo. Most of the research in zoos is still husbandry-based, but the research that could potentially help *ex-situ* and *in-situ* conservation efforts is increasing. Between 1998 and 2018, the European Association of Zoos and Aquaria contributed to 3345 peer-reviewed articles, a threefold increase in the second decade (Welden et al., 2019). The change in attitude towards conservation in zoos may result from a response to climate change, habitat loss, an increase in disease and invasive species and other human-caused environmental changes leading to a dramatic loss in biodiversity. Zoos can help with research projects that would be more difficult to do in the field, especially for rare species and research that needs to be carried out in a controlled environment (Minteer and Collins, 2013). One of the most substantial contributions zoos can make to conservation is breeding and reintroducing individuals into the wild, which will help declining populations.

Zoos help manage breeding programs, including extensive record-keeping to ensure genetic variability (Mallinson, 1995). Zoos rely on these breeding programs to prevent population numbers from declining. Zoos alone cannot be responsible for wild populations. Animals bred in captivity for many generations will not have the genetic diversity required to sustain a wild population. A decline in genetic diversity can lead to a shorter life span, poorer health and reduced reproduction (Minteer and Collins, 2013). It has been suggested that integration of zoo breeding programs with field based conservation projects will help ensure population sustainability. This would require zoo breeding programs to join larger conservation management projects and have captive animals exchanged for wild animals. If a programme of this magnitude was to occur, it would require management of genetic information, improved reintroduction success rates, the effect of environmental variations,

the effect of different husbandry techniques and control of disease risk. This all creates pressure on research by zoos both within the zoo and in the field (Lacy, 2013).

Chester Zoo is one of the leading zoos for zoo-based conservation and has been involved with multiple *ex-situ* conservation projects (Chester Zoo, 2021). As an example, they have worked with other European zoos to help with the Cikananga Conservation Breeding Centre (CCBC). CCBC release captive birds of Indonesia's most endangered species, including the Javan green magpie, into protected areas. The zoos sent experts to help with the husbandry needs of the birds, as healthy animals with good welfare are more likely to breed successfully (Owen et al., 2014). Chester Zoo has been helping increase the population of the Monsteny Brook Newt in Spain by breeding newts in the zoo for reintroduction to Monsteny National Park. Research carried out in Uganda helped establish the habitat requirements of the giant pangolin using camera traps and GPS biotags (Chester Zoo, 2021).

The rapid decline of biodiversity has increased the need to prioritise conservation research, but the welfare of individual captive animals cannot be dismissed. Animal welfare should be considered in all captive animal research. Efforts should be made to increase the use of non-invasive techniques, such as obtaining hormones from urine or faeces rather than blood samples. Many zoos rely on faecal and urine samples to inform their research programmes as it easier and non-invasive to collect. Chester Zoo has been monitoring the steroid concentration of faeces of their five female Asian elephants (*Elephas maximus*) since 2007. The females had synchronised oestrous cycles except for a 14-to-20-week period where three of the females became acyclic. Irregular cyclicity has been associated with lower fertility. The faecal progestogen and glucocorticoid metabolites were measured using enzyme immunoassays. Due to the extensive records kept by the zoo, all management changes that occurred were noted, and the zoo could check if they were having a negative impact on the elephants. The lack of change in faecal progestogen concentration was not due to the change in their daily routine. Although the zoo could not show a direct cause of the temporary pause in cyclicity, it did show the importance of regular monitoring of individuals (Edwards et al., 2016).

One species Chester Zoo is helping conserve is the okapi (*Okapia johnstoni).* Okapis, relatives of the giraffe, are classified as endangered; their numbers have declined by 50% in the last 24 years. They have a small geographic range (14000 km^2) and are only found in the forests of the Democratic Republic of the Congo (IUCN, 2017) (Figure 3.1). The main threat to okapi is habitat loss, as legal and illegal deforestation occurs for residential and commercial

Figure 3.1. Okapi in Chester Zoo 2020 (© Natalie Koch), Chester Zoo had two female okapis during the study collection period. The geographic region inhabited by wild okapi (Africa Geographic Stories, 2018)

development and the logging industry. Okapis are protected under Congolese and International law, but hunting and armed military groups in certain areas contribute to their decline. The okapi population is distributed throughout the Congo forests, and an increase in deforestation increases the risk they will be physically isolated from each other, causing long-term detrimental effects on genetic diversity (Nixon and Lusenge, 2008). The okapis are the only extant species of their genus, and without intervention to help conserve their numbers, may become extinct in the wild.

The okapi is a challenging animal to study in the wild as they live in such dense forests that observation and tracking the animal is difficult (the Zoological Society of London only became aware of their existence in 1901). The Mbuti who lived in the same forests as the okapi considered them a forest spirit due to their elusiveness and would not hunt them (Okapi Conservation Project, 2020). Chester Zoo and the Uganda Wildlife Authority have been conducting surveys for Okapi in Semuliki National Park. It has helped support the Okapi Conservation Project in the Ituri Forest of DRC (Roffe, 2021). With only 76 okapis in European zoos, it is vital to have efficient breeding programmes to support a reintroduction project.

Chester Zoo has been monitoring the progesterone metabolite cyclicity of their female okapi. Regular cycling suggests the animal is in good reproductive health, and a significant increase in concentration suggests the animal is pregnant. The zoo's endocrinology team performs hormone assays on faecal samples to measure progesterone metabolite concentration. When an individual becomes pregnant, their progesterone metabolite cyclicity increases (Kusuda et al., 2007); eventually, it will increase to levels significantly higher than the baseline for confirmation that the individual is pregnant (Figure 3.2). This increase, however, may only occur at 300 days. Usually, pregnancy confirmation can be made five to nine months into gestation, and the okapi gestation is 15 months (Schwarzenberger et al., 1996). Progesterone levels can change quickly, so there must be a substantial backlog of progesterone levels to detect significant increases. The zoo must constantly monitor the progesterone metabolite cyclicity of their okapi. It would therefore benefit from a pregnancy test that would only require one faecal sample that can determine pregnancy much earlier than using progesterone measurements. The zoo can adapt the husbandry needs of an animal they know is pregnant to increase the chance of a healthy pregnancy and offspring (Schwarzenberger et al., 1999).

This study aimed to determine if the REIMS signature obtained from okapi faeces could be used to predict pregnancy. It was hoped that the spectra from archived samples from okapi at Chester Zoo when they had and had not been pregnant using REIMS could be used to build a predictive random forest model. This random forest would then have been used to classify new samples as pregnant or not pregnant. The classification accuracies for the initial models were lower than expected. Therefore, the second aim was to increase the classification accuracy by grouping the data by pregnancy stage and reducing the number of

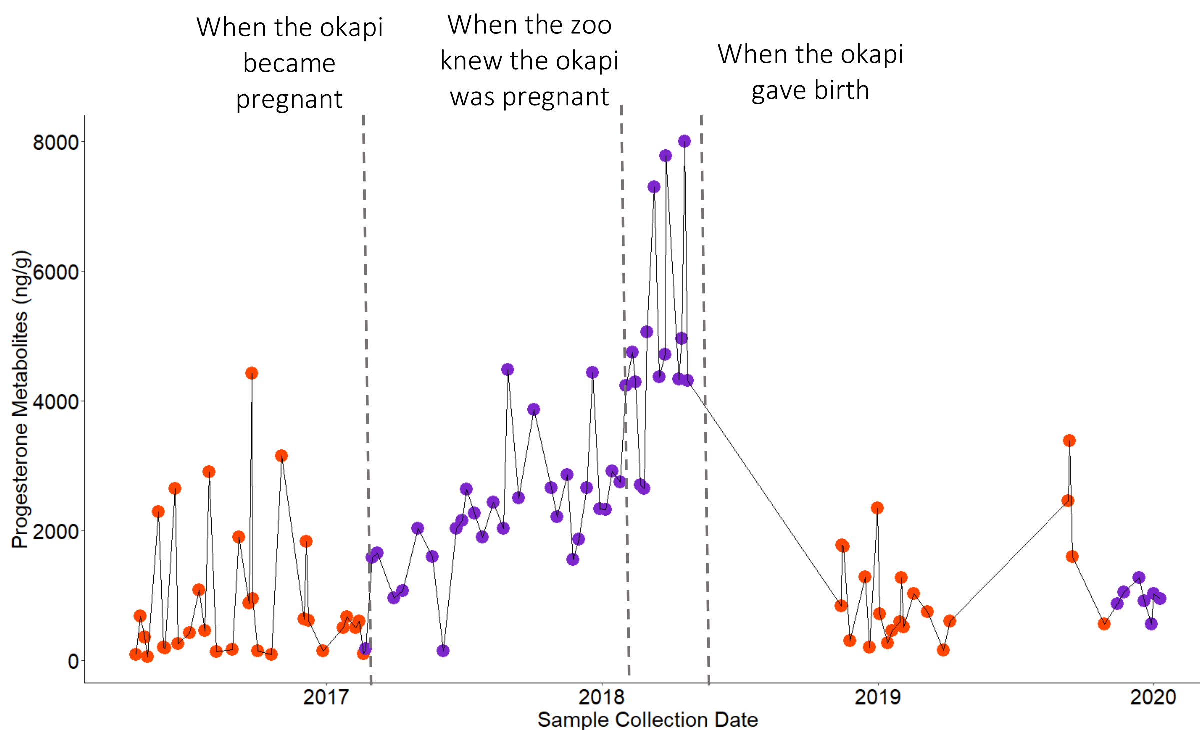


Figure 3.2. The progesterone metabolites of faecal pellets collected from an okapi at Chester Zoo. The zoo does not know the okapi is pregnant until there is a dramatic increase in progesterone. Once the okapi gives birth the zoo can work backwards to establish the date of conception

variables (mass bins). Changing parameters did not increase classification accuracy significantly. The results suggested that either sample preparation or storage length was affecting the ability of REIMS to classify the faecal pellets. New samples were collected from the two okapis currently at Chester Zoo, but since one of the okapis had not been pregnant, individual identification was tested instead of pregnancy. Random forest models built using the new samples could distinguish which okapi produced the faecal pellets but could not for the older samples. This confirmed that either sample preparation or storage time has a negative impact on the utility of the REIMS spectrum.

3.2 Method

3.2.1 Faecal Sample Collection from Okapis in Chester Zoo

Chester Zoo provided archived faecal samples from Stuma and K'tusha; both females were born in the zoo, Stuma in 2005 and K'tusha in 2013. They were pregnant twice between the first and last sample collection; K'tusha was only two and a half months pregnant with her second calf when the collection ended. There were 208 faecal samples provided, 52 samples from each individual when they were pregnant and 52 when they were not (Table 3.1). Samples were collected as part of Chester Zoo's routine procedures to monitor health and

Table 3.1. The number of faecal samples collected by Chester Zoo from two female okapis. Both individuals were pregnant twice during the collection period. Each of the non-pregnant samples were assigned as low, increasing, peak or decreasing depending on their position in the follicular cycle (see Figure 2.1)

Individual	Pregnancy Status	Follicular cycle status	Total Number
K'tusha	Pregnant		52
	Not Pregnant	Low	13
		Increasing	13
		Peak	13
		Decreasing	13
Stuma	Pregnant		52
	Not Pregnant	Low	13
		Increasing	13
		Peak	13
		Decreasing	13
		Total	208

maintain their breeding programme and were collected up to eight hours after defecation. Samples could not be collected consistently; intervals between collected samples varied between one week and a month. The sample collection rate did increase when the animal was known to be pregnant. Samples from Stuma were collected from July 2011 to August 2017 and from K'tusha from April 2016 to January 2020. The in-house endocrinology team at Chester Zoo had already measured the progesterone metabolites. The measurements were tracked over time to monitor the follicular cycle. Each sample from when an individual was not pregnant could be classified as increasing, decreasing, peak or low, depending on where it was collected during the cycle (Figure 3.3). When okapis defecate, they produce pellets

approximately 2 cm long and 1cm wide. For each sample, multiple pellets were homogenised, placed into a 15 ml conical sample tube, and stored at – 20°C.

Faecal pellets were collected from the current okapi at Chester Zoo, K'tusha (n=43) and Ada (n=42) in November 2019. Ada was a female born in London Zoo in 2017 and moved to Chester in 2019. Samples were collected as before, but faecal pellets were not amalgamated. Several whole pellets produced simultaneously were stored at -20°C in small plastic sample bags.

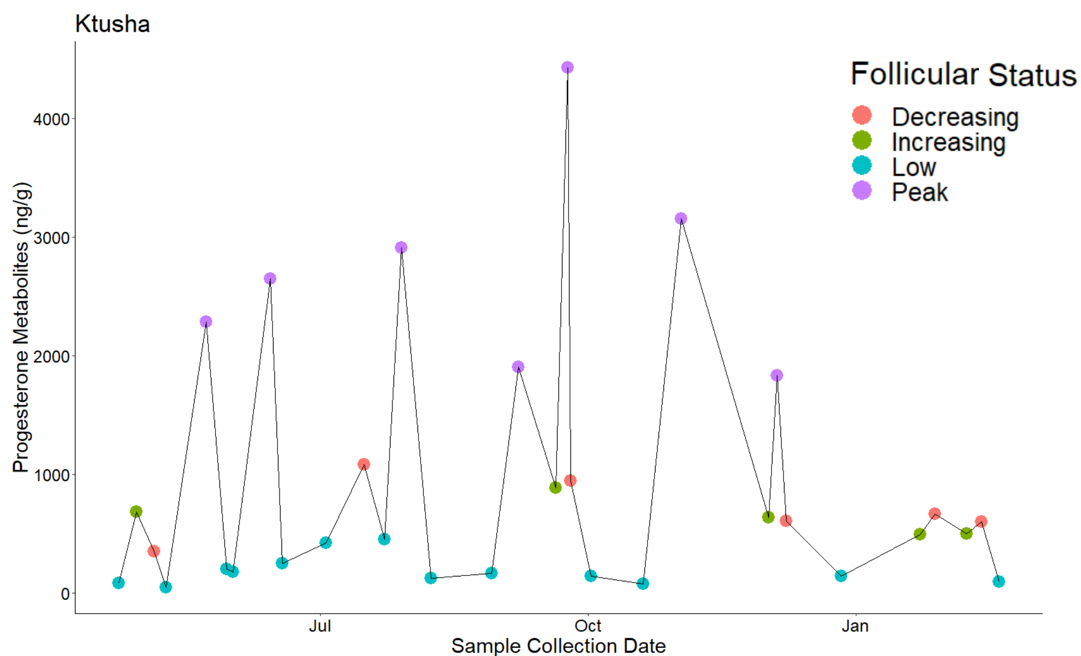


Figure 3.3. The progesterone metabolites of faecal pellets collected from K'tusha for the first year of collection. All samples are from before K'tusha became pregnant. The progesterone metabolites are cyclic, each sample was assigned as increasing, peak, low or decreasing depending on when it was collected during the follicular cycle.

3.2.2 Using REIMS to Burn and Analyse Okapi Faecal Pellets.

The diathermy electrode was set to 30 v, and water was added using a 1.5 ml dropper until the sample became saturated. Unlike rodent samples, burning did not completely disintegrate the sample. To ensure REIMS sampling was consistent, each sample was burned five times for 10s, and at least 30 seconds of acquisition time was added before and after burning a sample. Whole samples were burned using the same setting as the homogenised samples. Water was added to the whole pellet until complete hydration, and the whole pellet was burned within one acquisition file. Leu-enkephalin was continuously ejected at 50 µL/min

into the inlet capillary of the REIMS ionisation source to provide a lock mass. The sample cone and the heater bias were set to 60 V. The spectra were recorded in full-scan resolution, negative ion mode, at a scan rate of 1 scan per second from 50-1200 m/z. Samples were burned blind and in a random order determined by a random number generator (Haahr, 1998). A sample file may contain more than one burn, but these were averaged when uploaded to R. Before being uploaded to R, the spectra were lock-mass corrected, binned to 0.1, and the mass range was reduced to 400 to 1100 m/z within LiveID.

3.3 Results

3.3.1 Using REIMS to distinguish pregnant and non-pregnant samples.

A preliminary study was conducted to determine if REIMS could discriminate between faecal samples produced from a pregnant or non-pregnant individual. A total of 208 faecal pellets from two okapis that had been homogenised were analysed with REIMS. The spectra produced from burning were uploaded to LiveID, where the spectra were normalised and binned to 0.1 Da. The averaged spectra of pregnant samples had a similar pattern to those of non-pregnant samples. Comparing the average spectra of just one individual showed little difference between pregnant and non-pregnant samples. Minor differences between pregnant and non-pregnant samples can be observed when the intensities are compared for a small mass range (445 to 450 Da) (Figure 3.3). The average intensities across this range of mass bins were significantly different ($p<0.004$). This suggests that there may be differences between faecal pellets due to pregnancy, but these differences are too small to be observed on the spectra. This was similar to the previous study; the spectra for males and females were similar but could be distinguished using random forests.

The spectra were analysed using random forests using a randomly selected training set (70% of all spectra) to predict the pregnancy status of the individuals in the test set (30% of all spectra). This was repeated ten times with both individuals in the model and one individual at a time to give an averaged random forest and prediction accuracy (Figure 3.4). The random forest accuracies were slightly above 50%, suggesting that pregnant and not pregnant samples could not be distinguished. There was, however, a significant difference between the top five most discriminant mass bins when samples of both individuals were included in the model (Figure 3.5). The top five most discriminant mass bins for pregnancy varied between each model. Eight mass bins occurred in the average top ten most discriminant mass bins of all three models and were significantly different between pregnant and non-pregnant samples. This suggested some consistent differences between pregnant and non-pregnant samples. Although the most discriminate mass bins varied between the

two individuals, they could still be used to show separation between pregnant and non-pregnant samples of the other individual (Figure 3.6).

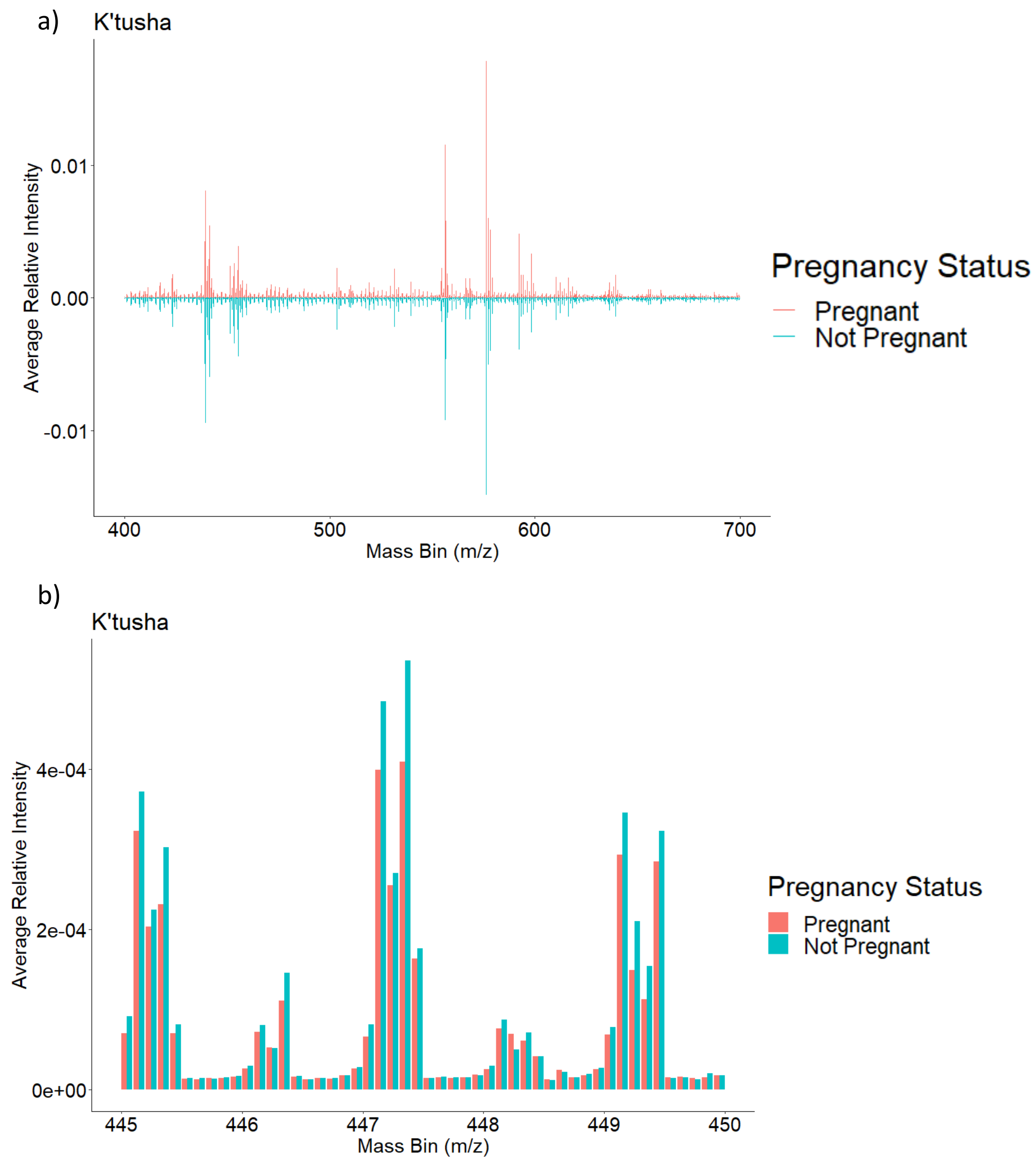
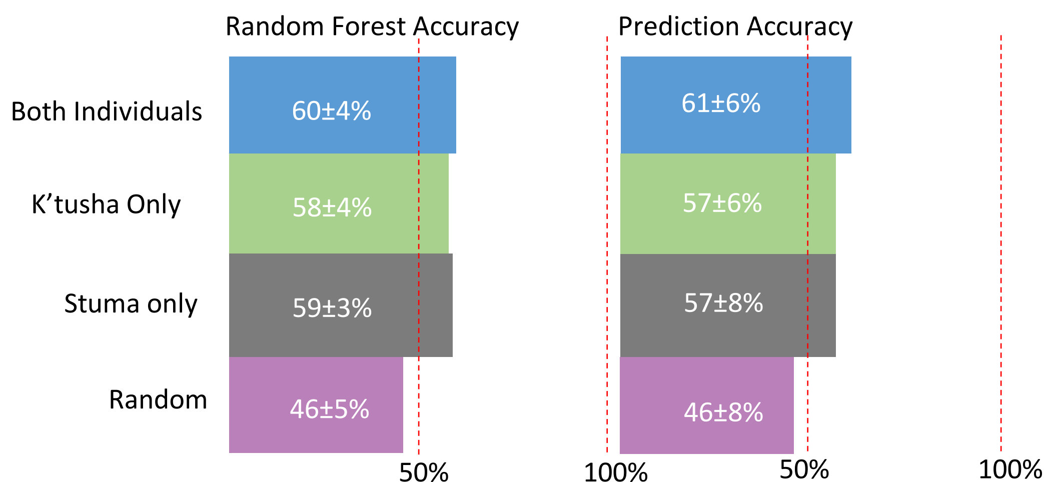


Figure 3.3 a) The averaged spectra of all pregnant (n=52) and non-pregnant samples (n=52) of K'tusha. b) Small differences are observable between mass bins when comparing a small mass range (5 Da), the non-pregnant samples had intensities higher than pregnant samples in this range. The average intensities across this range of mass bins were significantly different ($p<0.004$)



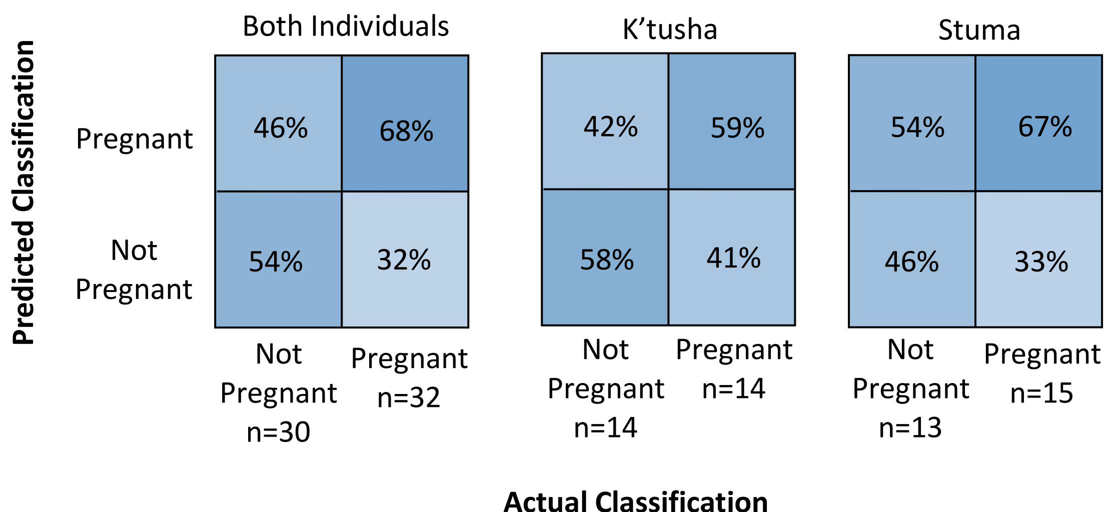


Figure 3.4 The random forest accuracy and prediction accuracy when the mass range was set from 50 to 1200 m/z and the data binned to 0.1 Da for the two individual Okapi (Stuma and K'tusha) combined and on their own. Accuracies were from an average of ten random forest models and the data was split 70:30 into training and test data. The Random class uses the same data as the other models but each sample as either been randomly assigned as 'random one' or 'random two'. If there is no difference between two groups, the random forest accuracy would lie close to 50%. These results suggest there is no difference between pregnant and non-pregnant samples.

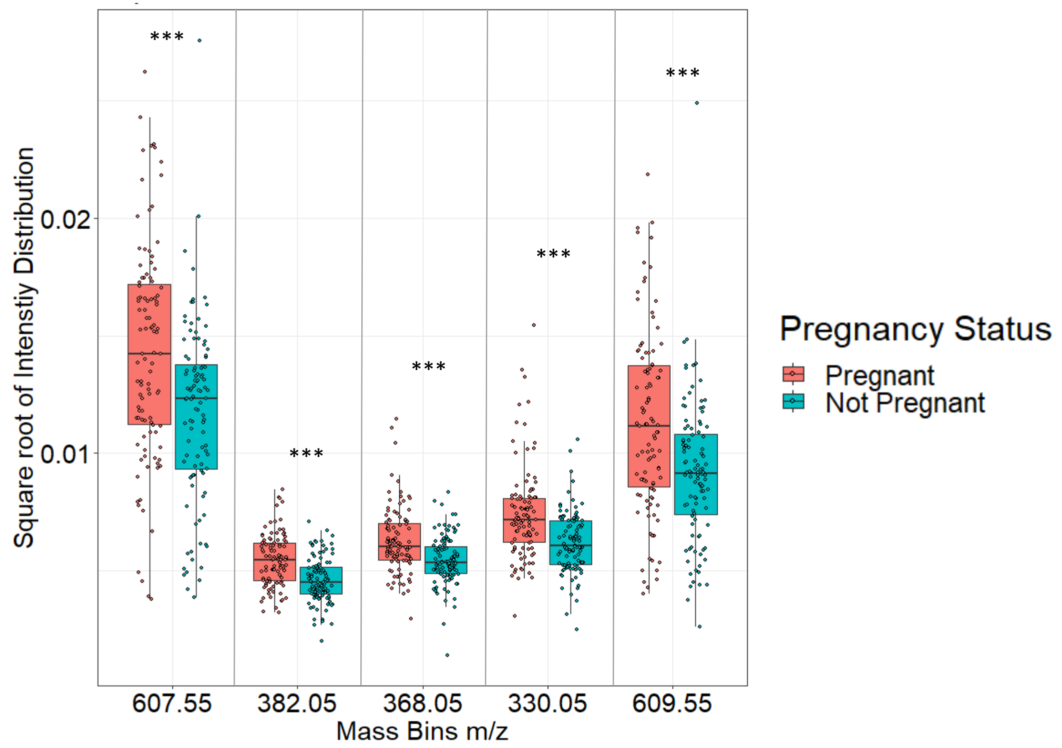
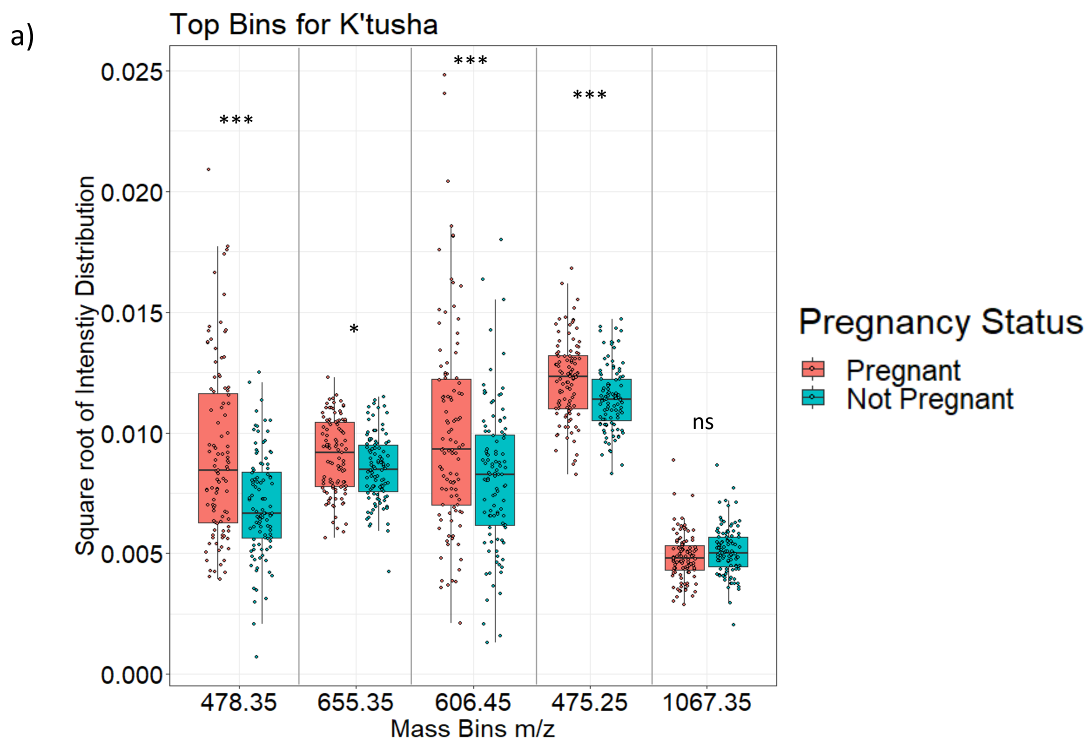


Figure 3.5 The relative intensities of the top five most discriminate mass bins of all pregnant and non-pregnant samples, all five bins show a significant different between pregnant and non-pregnant samples.

a)



b)

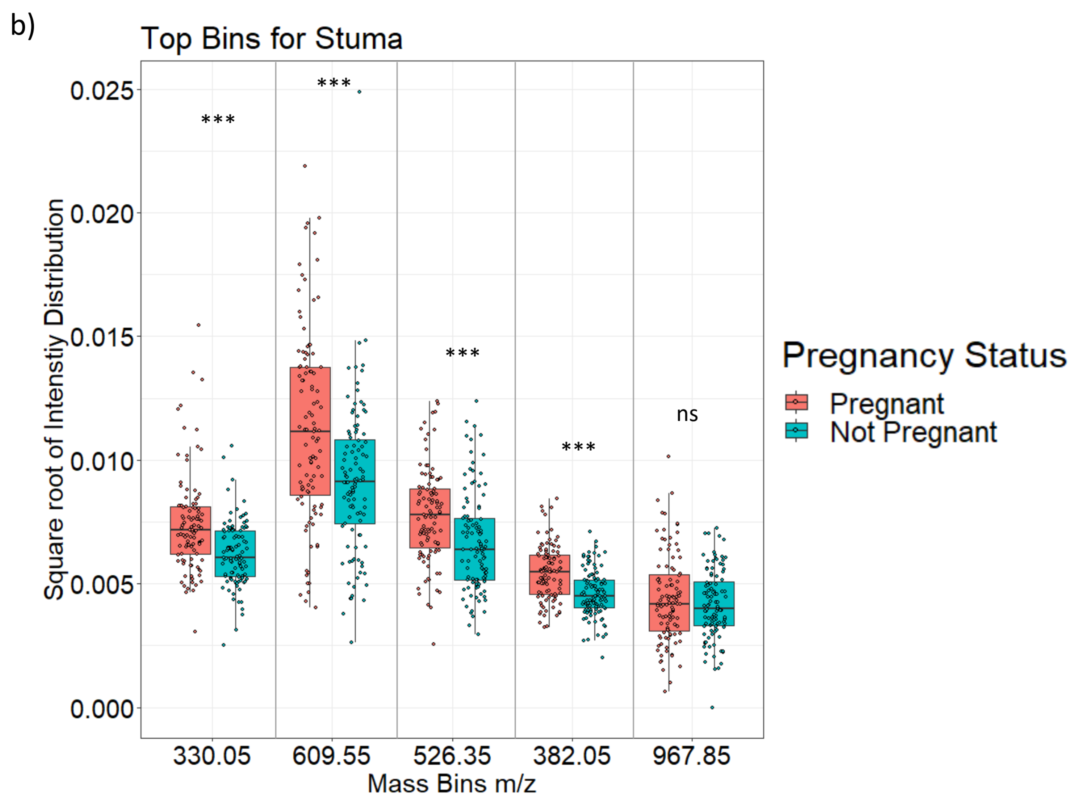


Figure 3.6 a) The relative intensities from all samples (K'tusha and Stuma) of the top five most discriminate mass bins that distinguished pregnant and non-pregnant samples of K'tusha. Four of the mass bins showed a significant different between pregnant and non-pregnant samples. b) The relative intensities from all samples (K'tusha and Stuma) of the top five most discriminate mass bins that distinguished pregnant and non-pregnant samples of Stuma. Four of the mass bins showed a significant different between pregnant and non-pregnant samples.

3.2.2 Changing Model Parameters to Increase Classification Accuracy

Changes were made to the data processing to increase prediction accuracy. The mass range was reduced to 300 to 900 m/z, as none of the top ten most discriminant mass bins were outside this range. The bin size was decreased, and the raw spectra were run through Progenesis Bridge (version 1.0.29) before they were uploaded to LiveID. Progenesis Bridge is part of the MassLynx software, which combines all the burns from one acquisition file into one burn signal. These changes, however, only showed a slight increase in the random forest and prediction accuracy (Figure 3.7). A new model was created to ensure that there was no sample number in the training data that was lowering the expected random forest accuracy.

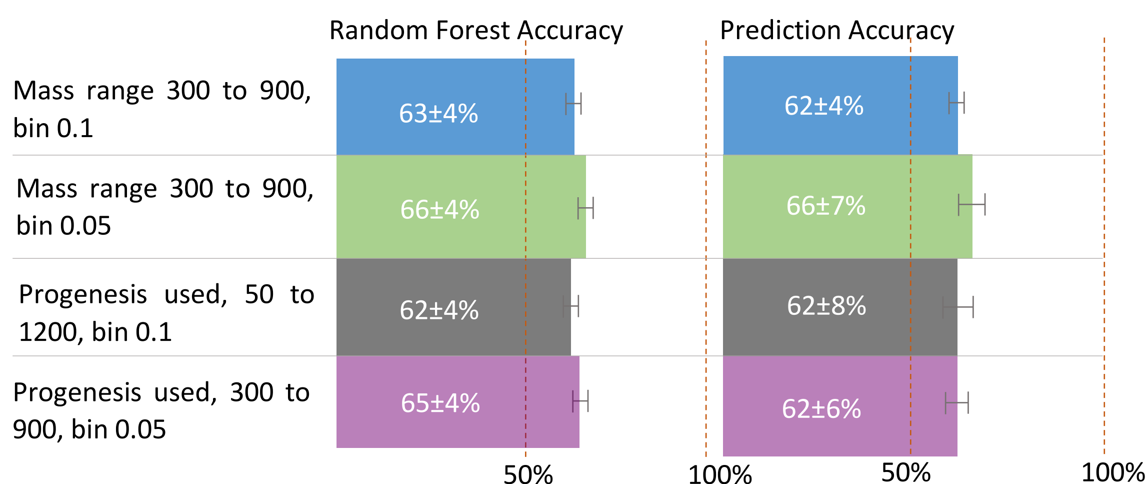


Figure 3.7 Changes were made to the data set including decreasing the mass range and bin size in LiveID and using Progenesis to average the burns before being uploaded to LiveID. The random forest accuracies are an average of ten with both individuals included. The changes produced accuracies higher than the 50% expected by chance but are still too low to suggest significant differences between samples.

A training model was built that contained all samples but one. The model was then used to predict the left-out sample; this was repeated for every sample. The accuracy of pregnant samples increased to 70%, but the classification accuracy of non-pregnant samples remained at 63%, suggesting there had been enough samples in the training model. (Figure 3.8).

If the models could not determine the differences between samples because there were too many un-discriminant variables (too much noise), reducing the number of variables would have increased classification accuracy. A random forest was created to determine the

top 100 most discriminant mass bins, and then a random forest model was created using these 100 mass bins only, repeated ten times. The top 100 were used as random forests require an extensive data set to obtain reliable results. Using only top five or top ten could give a high random forest accuracy but it would not be a suitable predictive model. The prediction accuracy was much lower than the random forest accuracy due to over-fitting when using only five or ten mass bins. The average random forest accuracy increased to 78%, and the prediction accuracy to 76%. These results are higher than the classification accuracy obtained by chance and suggest small changes between pregnant and non-pregnant samples (Figure 3.9).

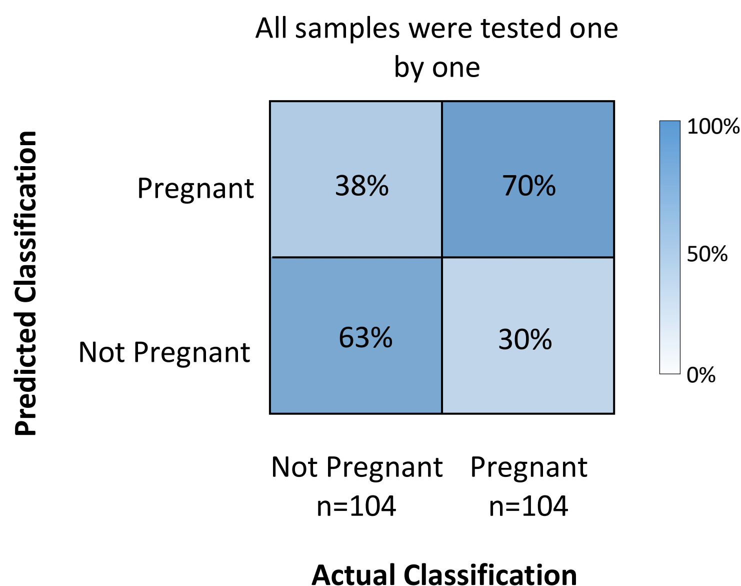


Figure 3.8 A random forest model was created using all the data but one sample. The model was used to predict the classification of the one left out sample, this was repeated for all 208 samples. In LiveID the data was binned to 0.05, mass range was reduced to 300 to 900 m/z and the spectra were lock mass corrected before being uploaded to R for the random forest analysis.

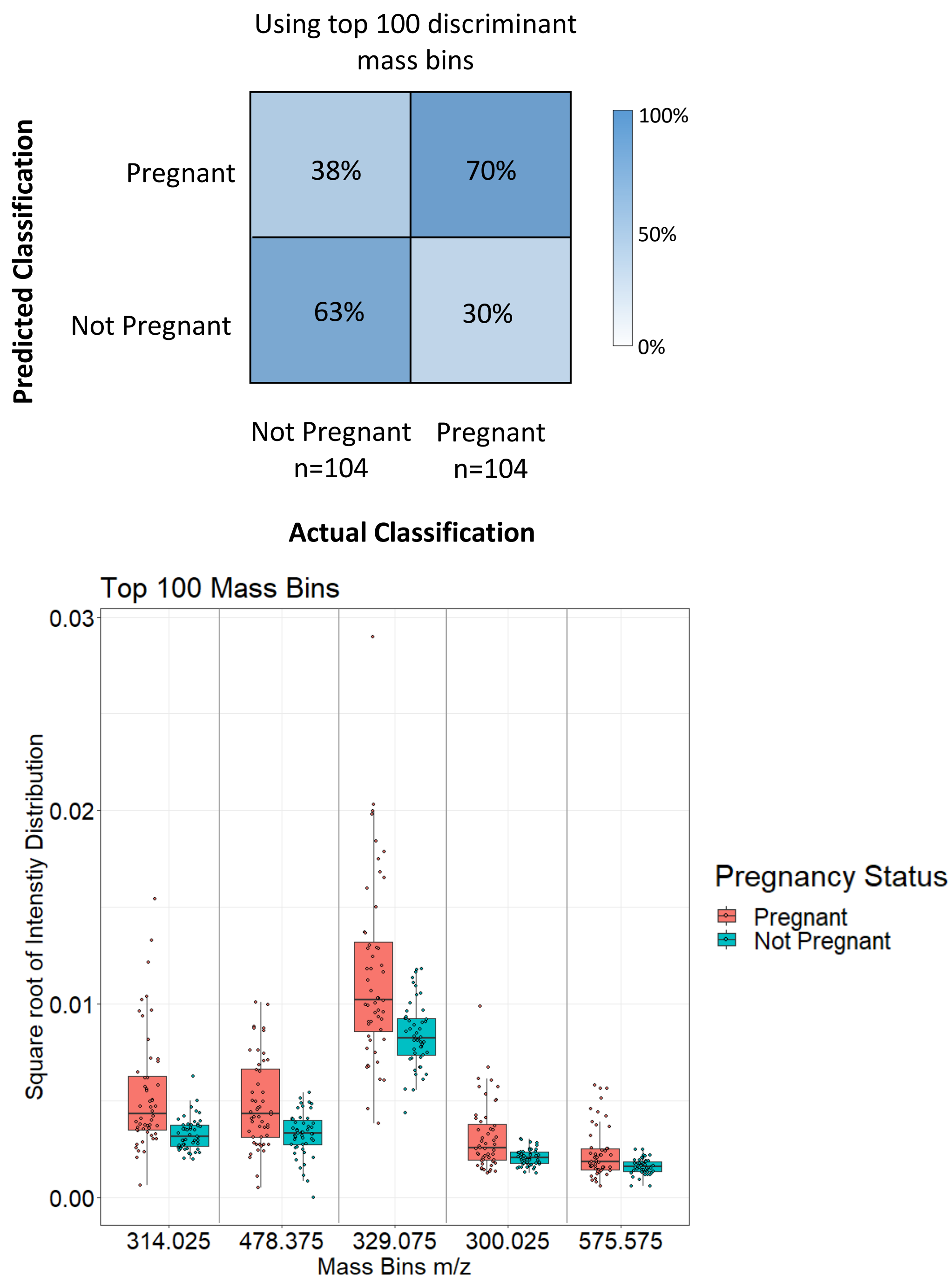


Figure 3.9. Random forest confusion matrix for the average individual classification of all samples using the top 100 discriminate mass bins only. The relative intensities of the top five most discriminate mass bins in the mass spectrum that distinguish pregnant and non-pregnant samples when using top 100 mass bins only.

3.2.3. Early Versus Late Pregnancy

The pregnant samples were split into late and early-pregnancy groups to potentially increase the difference between pregnant and non-pregnant samples. For K'tusha, the first 15 samples collected during her first pregnancy were assigned as early, and the last 15 were assigned as late, from a total of 43 samples. Stuma had fewer samples, so the first ten samples from both pregnancies were assigned as 'early' and the last ten collected as 'late'. K'tusha only had samples from the first half of her second pregnancy collected, so these were omitted. A random forest model was created to distinguish between early and late pregnancy. It produced a classification accuracy of 59±3%, suggesting no difference between the spectra of late and early pregnant samples. Only one of the top five most discriminant bins of pregnant and non-pregnant samples showed a significant difference between early and late samples. However, separation was observed with an LDA on the top five principal components. (Figure 3.10). If late-pregnancy samples varied more from non-pregnant samples, the classification accuracy would be higher when using only late-pregnancy samples compared to early-pregnancy samples. However, the random forest classification accuracy for early pregnant and non-pregnant samples and late pregnancy and non-pregnant samples was 77% when using the top 100 mass bins (Figure 3.11). These results suggest no significant differences between samples collected early or late in the okapi's pregnancy.

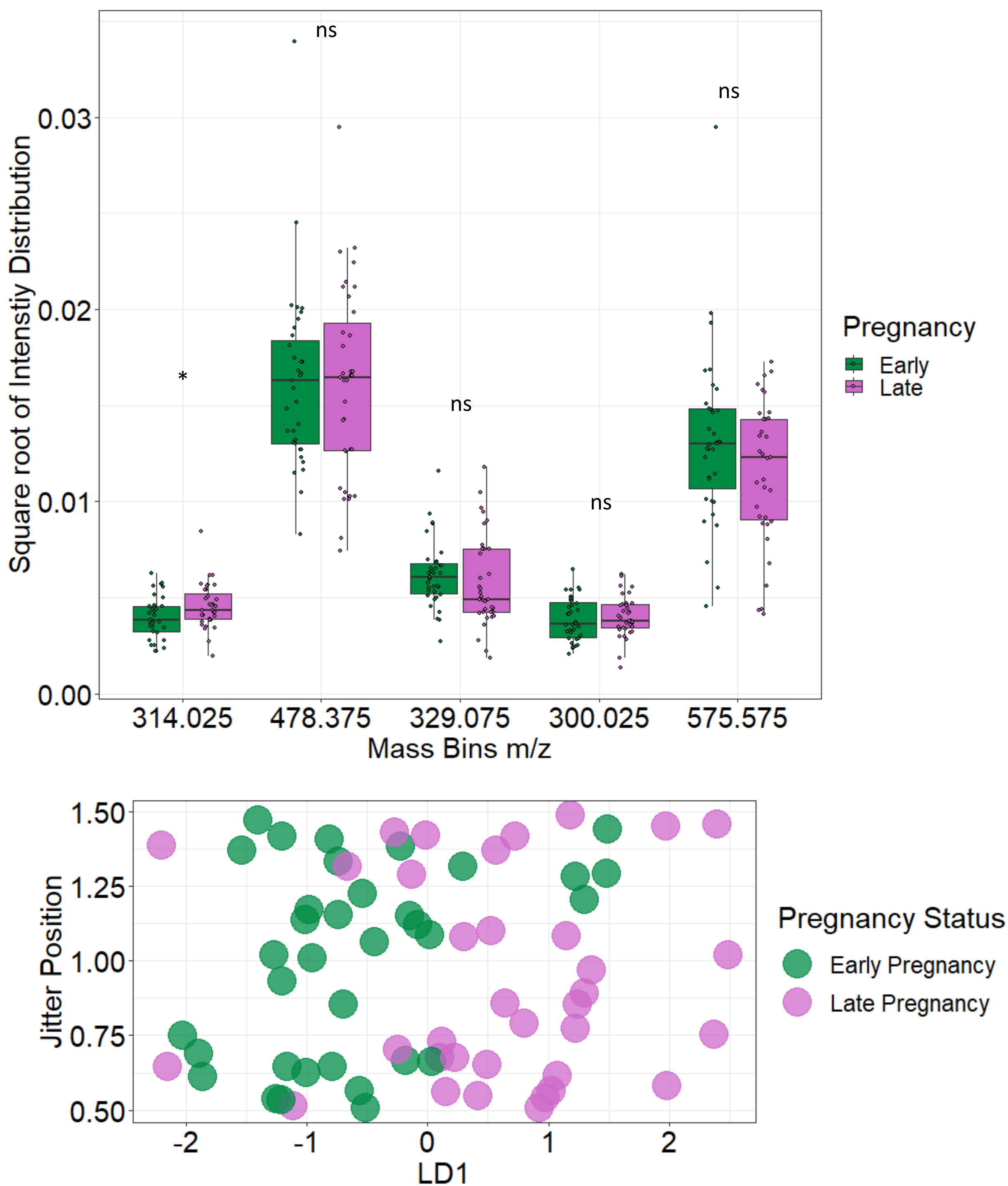


Figure 3.10. a) The relative intensities of the top five most discriminate mass bins of early and late pregnancy. Only one mass bin (314.025) had a significant difference between early and late pregnancy suggesting that the REIMS spectra do not change during pregnancy. b) The first linear discriminant component from the top five pca components which were responsible for 99% of the variation. The jitter position is randomly assigned to each sample to spread out samples vertically for visualisation.

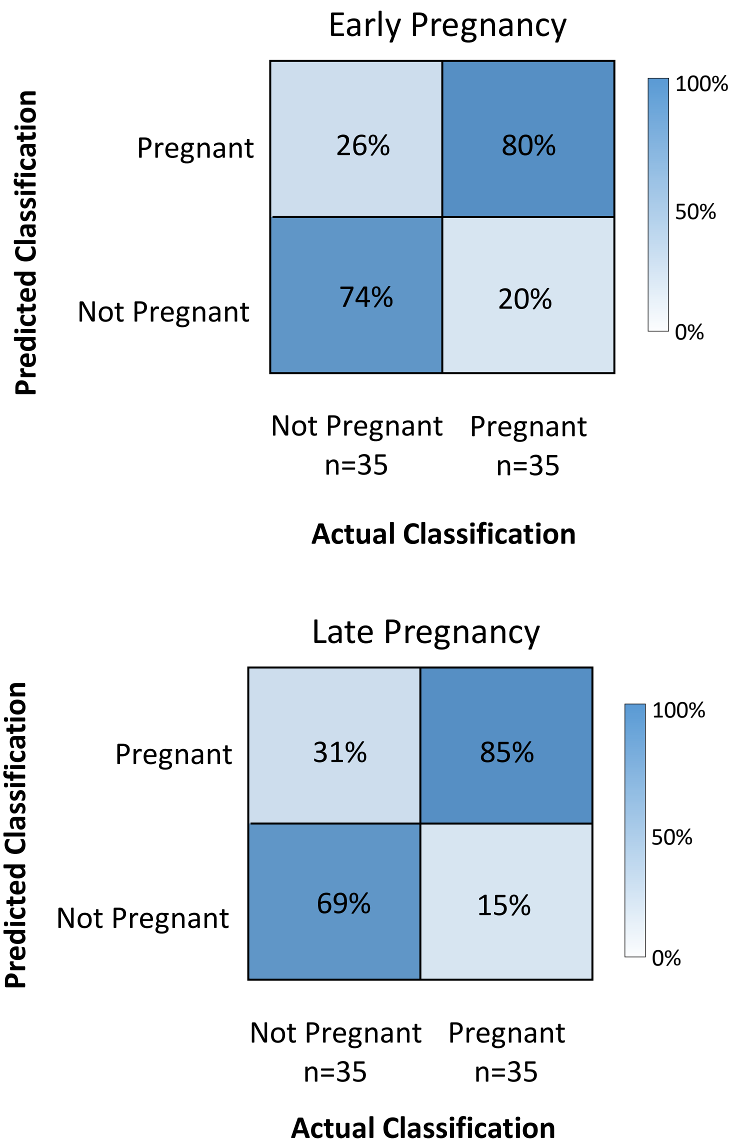


Figure 3.11 The confusion matrices of an average of ten random forests using samples from either the first three month of pregnancy (Early) or the last three months of pregnancy (Late) and randomly selected non-pregnant. The random forest model could predict early and late pregnancy better than non-pregnant samples, classification accuracy was slightly higher for late pregnancy samples than early samples.

3.2.4 The Effect of Progesterone Concentration

The non-pregnant samples may have too much variation among themselves to allow for consistent differences between pregnant and non-pregnant samples. The non-pregnant samples were classified into four groups based on their progestogen concentration, decreasing, increasing, low and peak. Those samples with the highest concentrations were classified as peak and had similar concentrations to samples early in pregnancy (Figure 3.12 a + b). It would be difficult to classify non-pregnant samples if a quarter of the samples had intensities the same as pregnant samples. It was thought that if the samples classified as peak were removed from the data set, the random forest accuracy would increase. Removing the peak samples, however, did not affect the random forest accuracy of non-pregnant samples (Figure 3.12 c). They may still be too much variation among the non-pregnant samples, but this is unlikely to be caused by different progesterone levels.

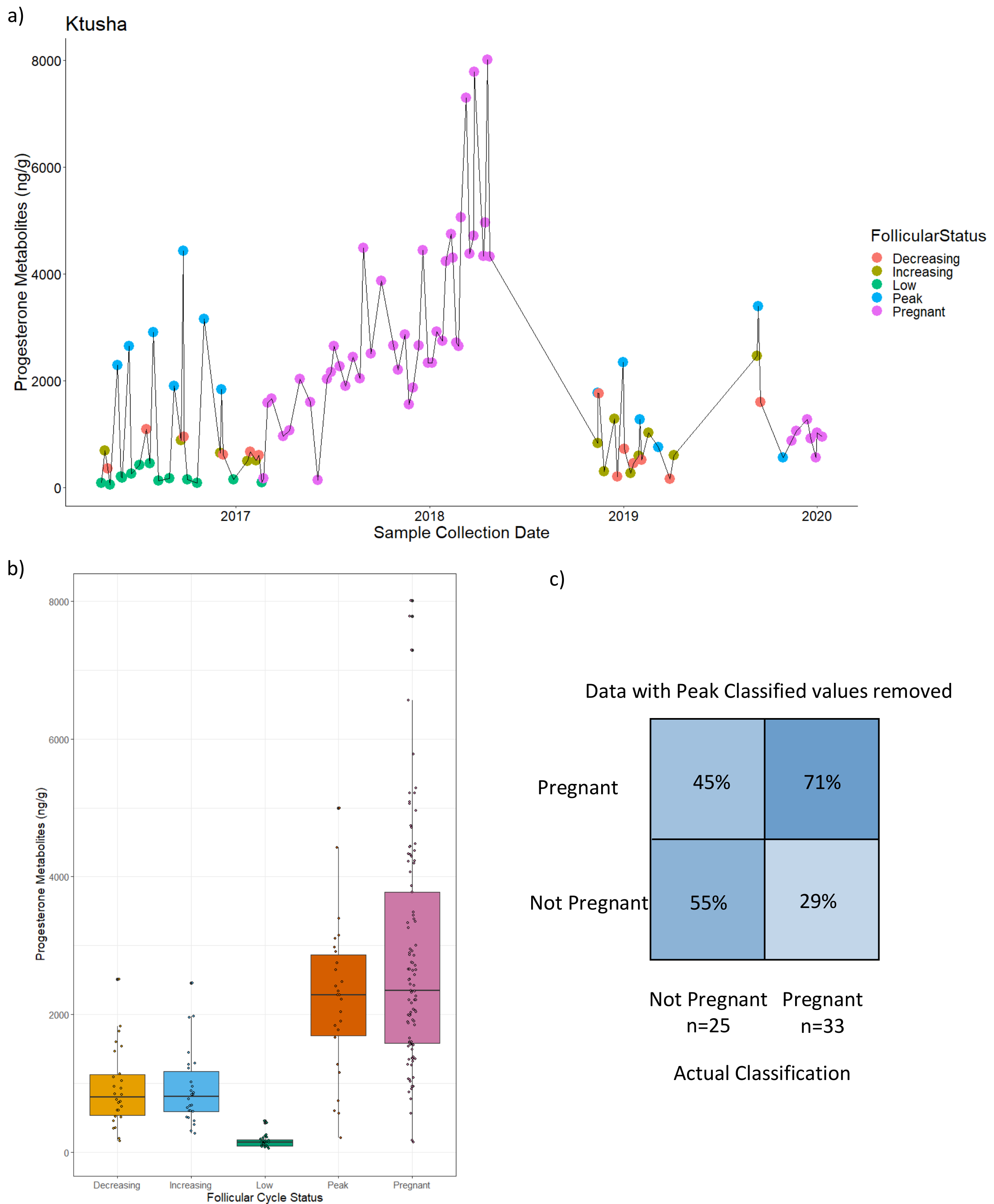


Figure 3.12 a). The progesterone concentration of each of the four follicular cycle groups of the non-pregnant and pregnant samples of both okapi individuals. Samples classified as peak have progesterone concentrations more similar to pregnant samples than other non-pregnant samples. b) The relative intensities for the top five mass bins most responsible for the differences between pregnant and non-pregnant samples. The samples classified as peak have intensities more similar to other non-pregnant samples than with pregnant samples. c) The ability of random forests to classify pregnant and non-pregnant samples did not increase by removing the non-pregnant samples classified as peak

3.2.5 The Effect of Storage Time on the Faecal Pellets

The time between collecting the okapi faecal samples varied greatly, with eleven years between the first and last samples collected. The significant length of time in storage at -20°C might have influenced the ability of REIMS to classify the okapi samples. Random forests could quite accurately classify samples based on their collection date of non-pregnant samples. Random forests could predict K'tusha's samples from 2016-2017 (the oldest 30 samples) with 80% accuracy and samples from 2019-2020 (the newest 30 samples) with 78% accuracy. This increased to 93% when using the oldest 15 and newest 15 samples. The increase in accuracy when using fewer samples suggested that the differences between the oldest and newest samples increased when the time between sample collections increased. Random forests could be used to predict Stuma's samples from 2009-2010 (the oldest 30 samples) with 71% accuracy and samples from 2016-2017 (the newest 30 samples) with 89% accuracy (Figure 3.13)

An LDA-PCA plot (LDA on the top five PCA components) separated both individuals' oldest and newest samples (Figure 3.14). There was a significant difference between old and new samples of the top five most discriminant mass bins (Figure 3.15). These bins differed from the most discriminant mass bins of pregnancy (Figure 3.5). This suggests that although the samples could be separated based on collection date, that does necessarily confirm it was influencing the ability to classify pregnancy.

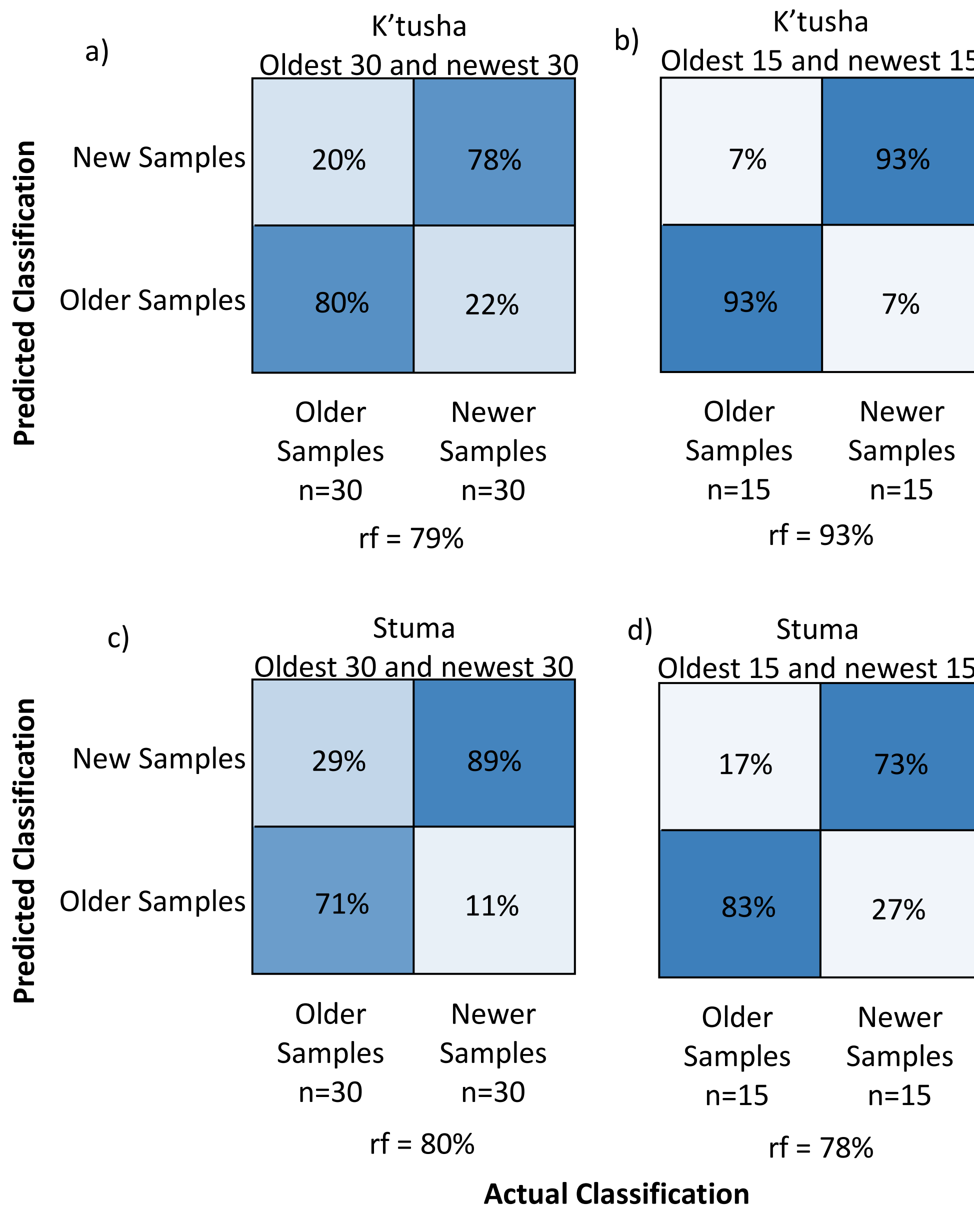


Figure 3.13 The classification accuracy of samples from K'tusha and Stuma based on collection date of the faecal sample. a) The random forest confusion matrix for the 30 oldest and 30 newest samples from K'tusha. b) The random forest confusion matrix for the 15 oldest and 15 newest samples from K'tusha, by reducing the number of samples it increases the time between the two groups. c) The random forest confusion matrix for the 30 oldest and 30 newest samples from Stuma d) The random forest confusion matrix for the 15 oldest and 15 newest samples from Stuma

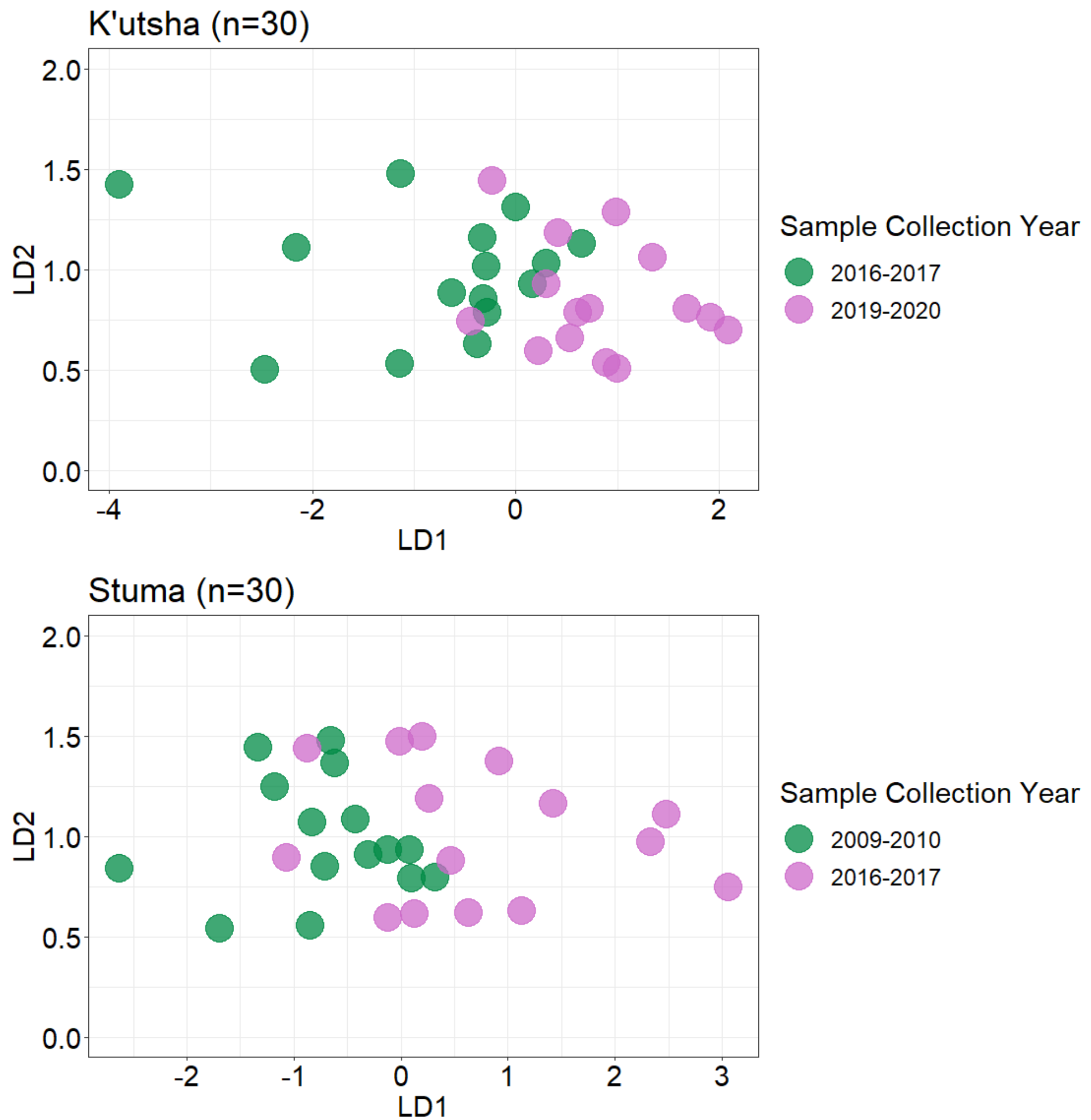


Figure 3.14 Discriminant function analysis of the top 5 principal component for classification of faecal samples on their collection dates for both okapis, K'tusha and Stuma. The separation of samples suggests there is a difference between spectra depending on the length of time the sample was in storage.

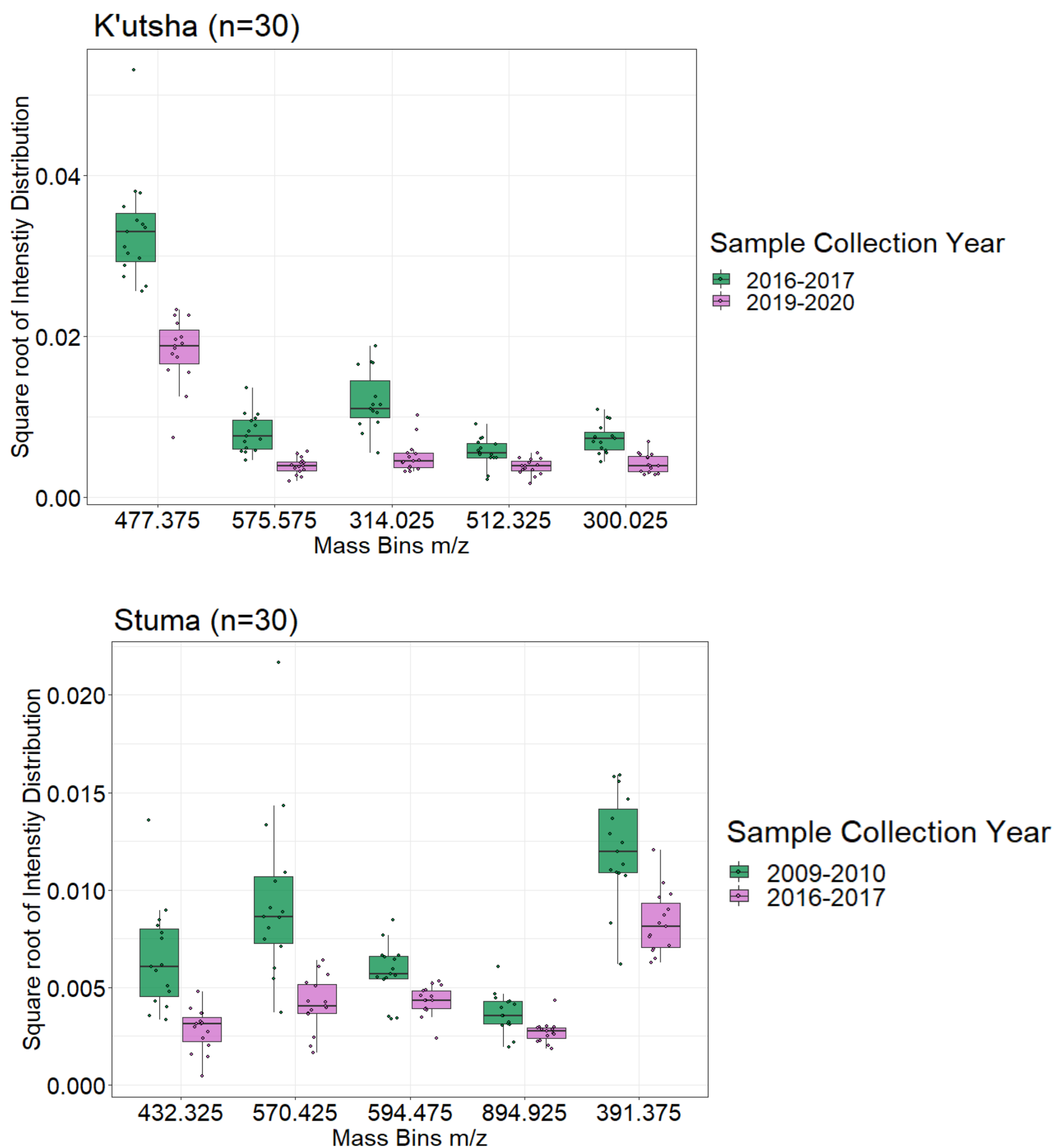


Figure 3.15 The relative intensities of the top five most discriminate mass bins that separate samples based on collection date. The two collection groups for both individuals show a highly significant differences suggesting that the REIMS spectral fingerprint changes due to the length of time a sample is in storage.

It was surmised that if the collection date does influence the REIMS spectra, classification accuracy would be improved if the samples were subdivided into groups based on the sample collection date. The data was split into four sub-groups and running random forests with the data from each group showed increased classification accuracy. Group I contained samples from Stuma from 2010 to 2012, and Group II included samples of both individuals from 2014 to 2017. Group III contained samples from Stuma's first pregnancy and the non-pregnant samples from before this. Group IV contained the samples from before and during K'tusha's first pregnancy (Figure 3.16). Group II had the lowest classification accuracy out of the four groups, possibly due to having the most extensive sample range and samples

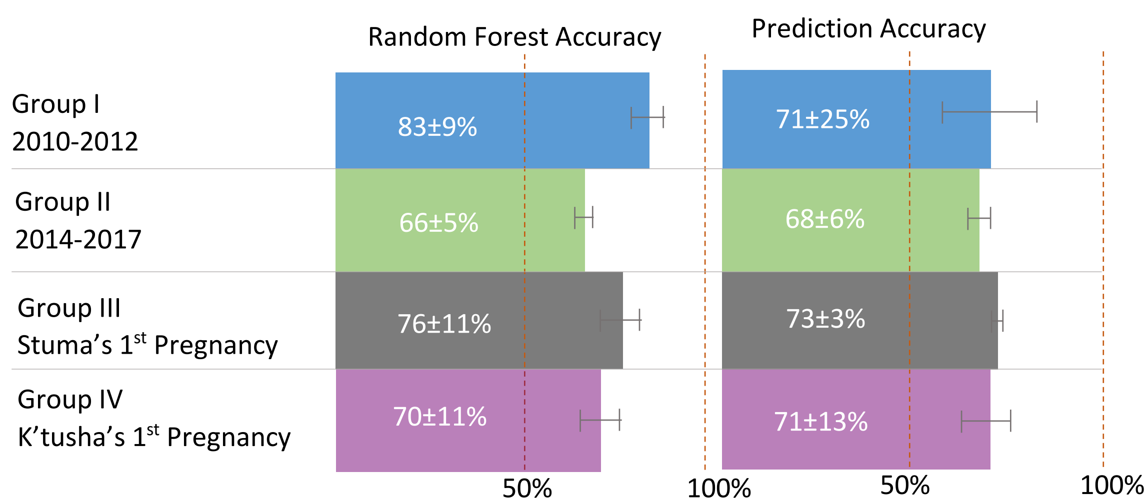


Figure 3.16 The data were split into four groups based on collection data. Group 1 contained all samples between 2010 and 2012. Group 2 contained all the samples from 2014 and 2017. Group 3 contained the samples from Stuma's first pregnancy and the non-pregnant samples before the pregnancy. Group 4 contained the samples from K'tusha's first pregnancy and the non-pregnant samples before the pregnancy. The data were binned to 0.05 da and the mass range reduced to 300 to 900 m/z. The random forest and prediction accuracies were higher than when all samples were included in one model.

from both individuals. The top mass bins varied between the four groups, suggesting that the mass spectra are different, which may have been caused by the length of time in storage. The top five most discriminant mass bins varied for each group except for mass bin 314.025, which was in the top five for groups II and IV. Mass bin 314.025 was also in the top five bins most discriminant of sample collection year for K'tusha and early and late pregnancy. The top 19 most discriminant mass bins between collection groups showed a significant difference

between all pregnant and non-pregnant samples (Figure 3.17). Mass Bin 314.024 (in the most changed significantly for the non-pregnant samples, negatively for K'tusha (Pearson correlation, R = -0.55 p = 2.4e-5) and positively for Stuma (Pearson correlation R = 0.5 p = 0.00014). The mass bin did not show a significant difference for pregnant samples of either individual (Figure 3.18).

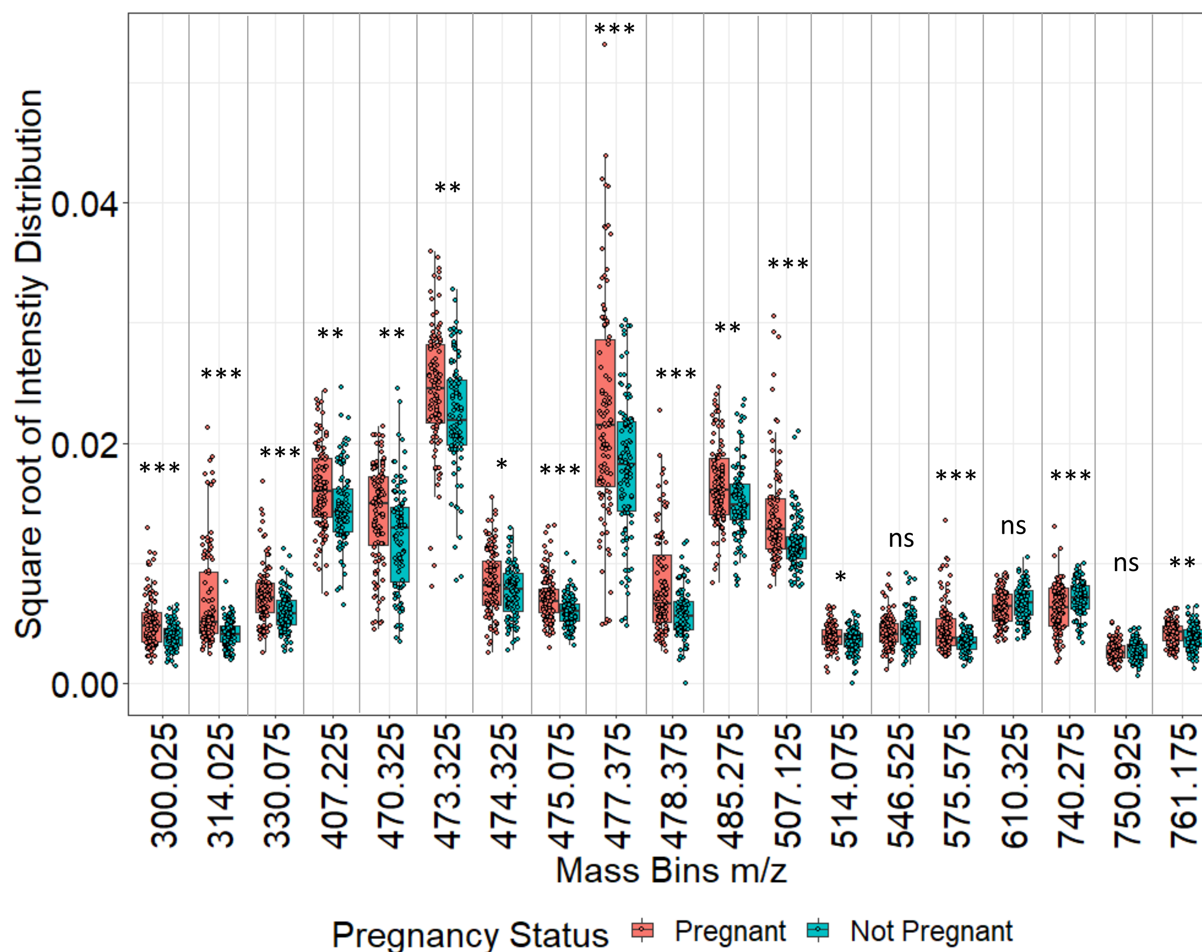


Figure 3.17 The relative intensities of all samples for the top five most discriminate mass bins of pregnancy for each of the four sub-groups based on collection dates. Mass Bin 314.025 was in the top five for two of the sub-groups. Even though the top mass bins changed between the four groups all but three of the mass bins showed a significant difference between pregnant and non-pregnant samples.

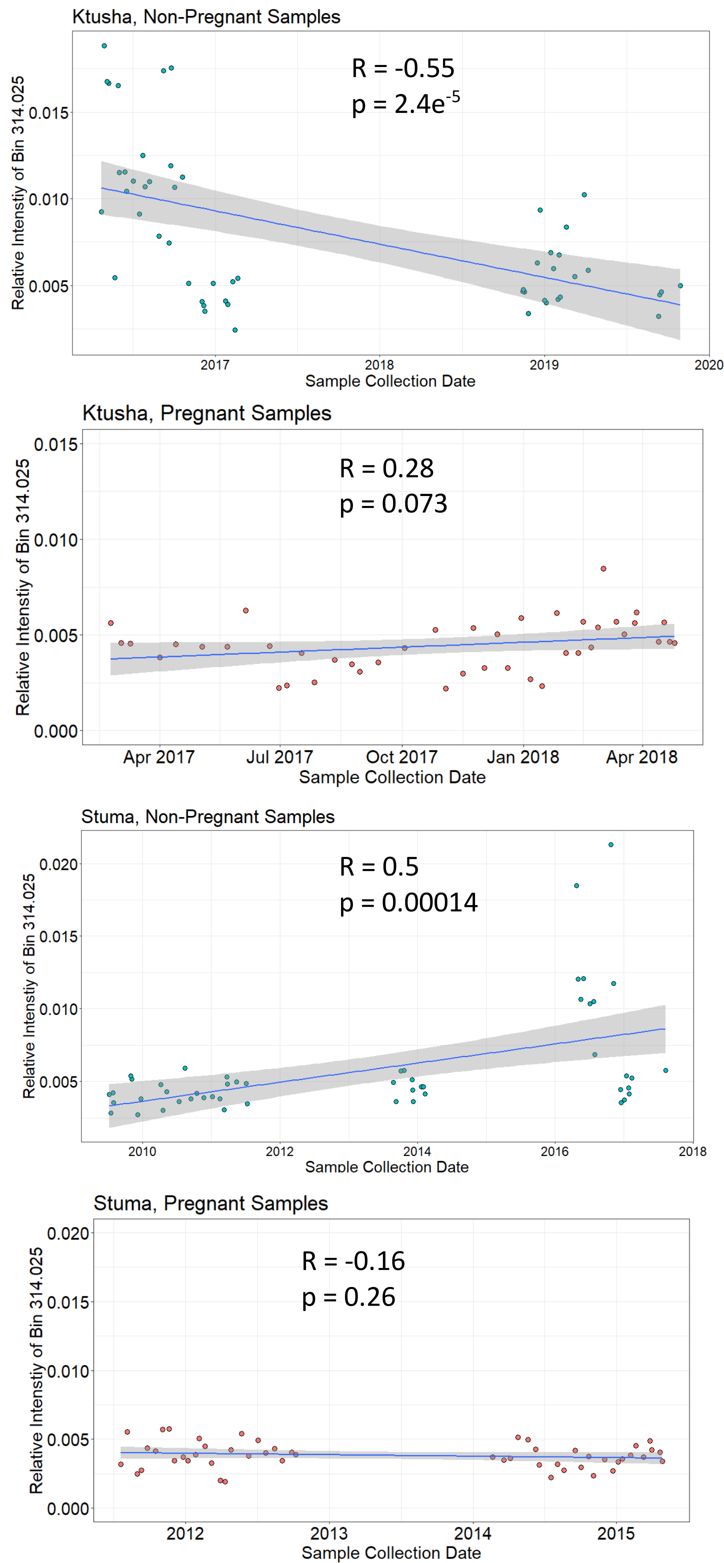


Figure 3.18 Mass Bin 314.025 was in the top five most discriminate mass bins for several models suggesting it changes the most between samples. The intensity of the bin decreases over time for non-pregnant samples of K'tusha but increased for Stuma. The intensities of pregnant samples did not change with time for either individual. Correlation established using Pearson's correlation.

3.2.6 Comparison of Homogenised and Whole Faecal Pellets

The spectra of three randomly selected non-pregnant amalgamated samples from K'tusha were visibly different from each other both in spectral pattern and intensity (Figure 3.19). Three randomly selected non-pregnant whole pellets from K'tusha showed a much more similar pattern to each other compared to the homogenised samples (Figure 3.20). Random forests could distinguish between whole and homogenised non-pregnant pellets of K'tusha with 99% accuracy (Figure 3.21). There was a significant difference between the top five most discriminant mass bins, and PC1 and PC2 could separate all but five whole pellets from homogenised pellets (Figure 3.22). This suggests that the sample preparation of the historical samples held by the zoo has greatly influenced the faecal profile.

A PCA could not separate the non-pregnant homogenised samples by individual but could separate whole pellets (Figure 3.23). Random forest models could distinguish between pellets of Ada and K'tusha with 93% accuracy; prediction accuracy was 94%. The random forest accuracy was much lower for the homogenised non-pregnant Stuma and K'tusha samples, random forest accuracy was 72%, and prediction accuracy was 66% (Figure 3.24). This also suggested that the more recently collected whole pellets burned by REIMS were more information-rich than the historical homogenised samples. Therefore, this study should be repeated using whole faecal pellets to establish if REIMS can determine pregnancy from okapi samples.

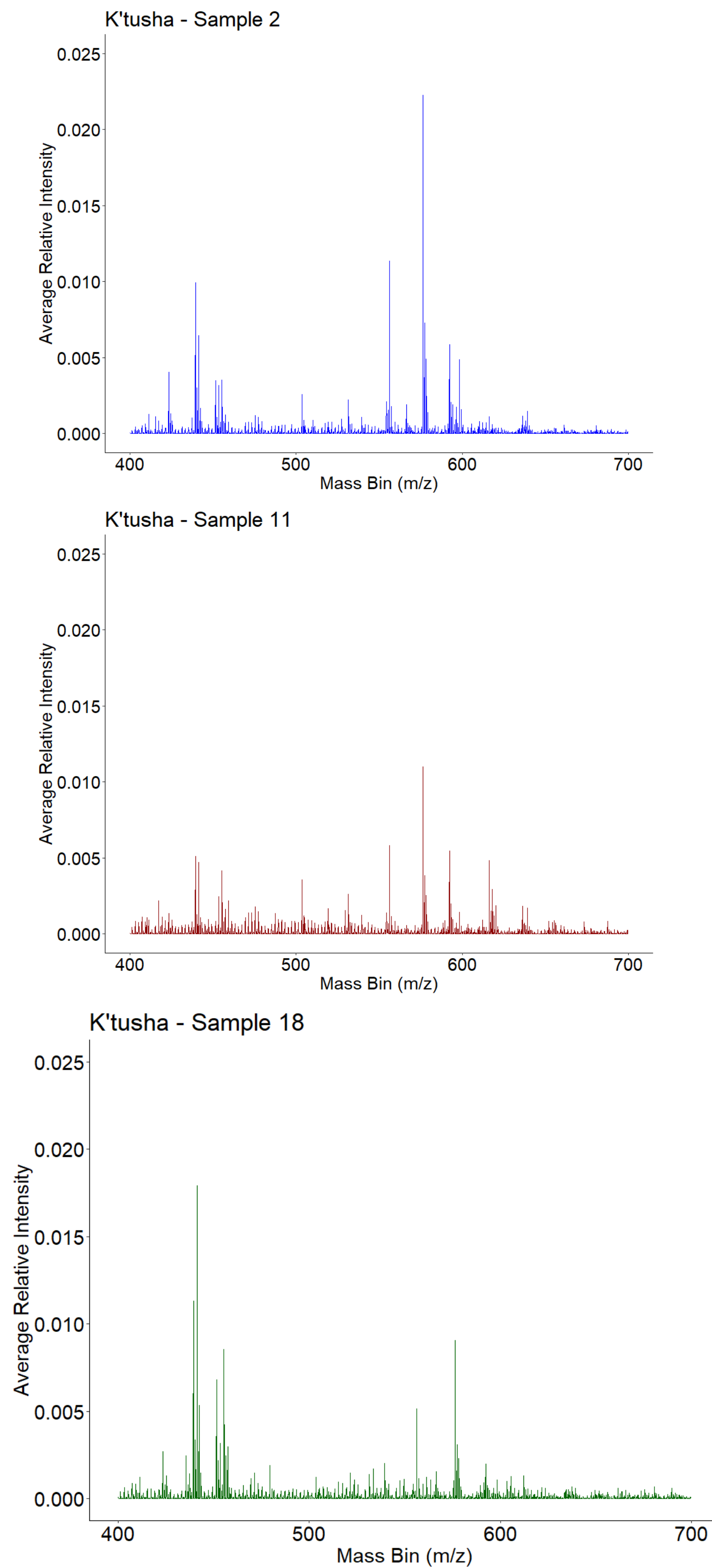


Figure 3.19 The spectra of three randomly selected non-pregnant homogenised samples from K'tusha. The spectra show greater differences between individual pellets than compared to the averaged intensities of pregnant and non-pregnant samples.

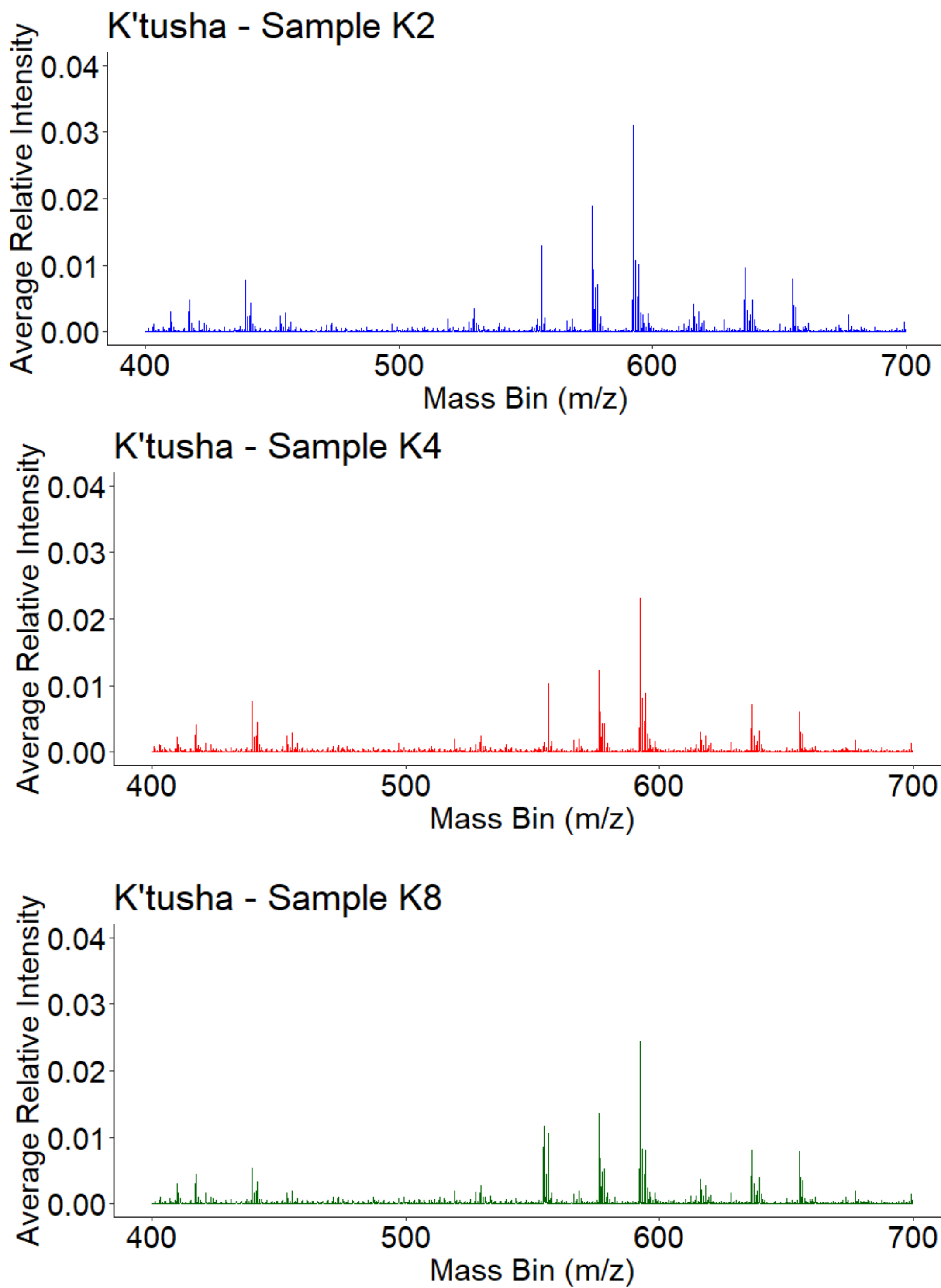


Figure 3.20 The spectra of three randomly selected non-pregnant whole samples from K'tusha. The spectra are more similar to each other compared to the homogenised samples.

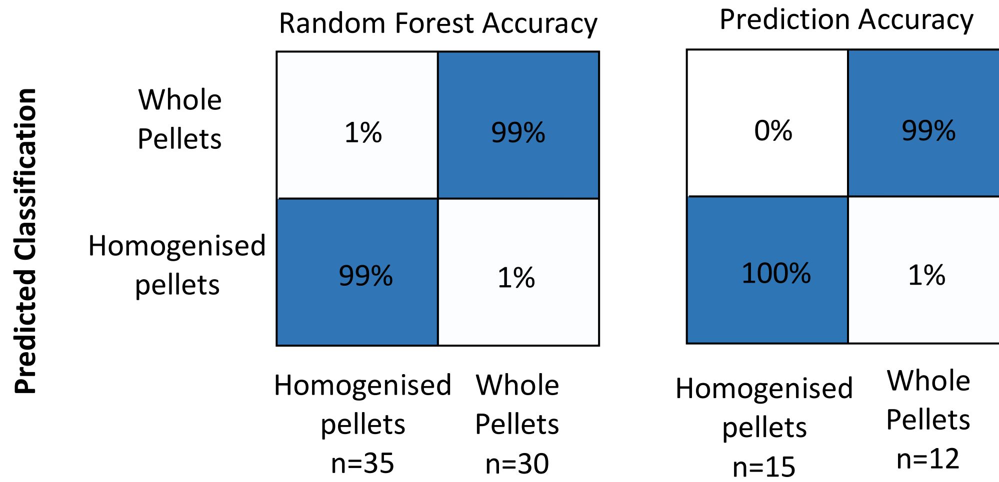


Figure 3.21. The random forest and prediction accuracy of non-pregnant samples from K'tusha. These results suggest there is a significant difference between the spectra produced from whole pellets compared to homogenised samples.

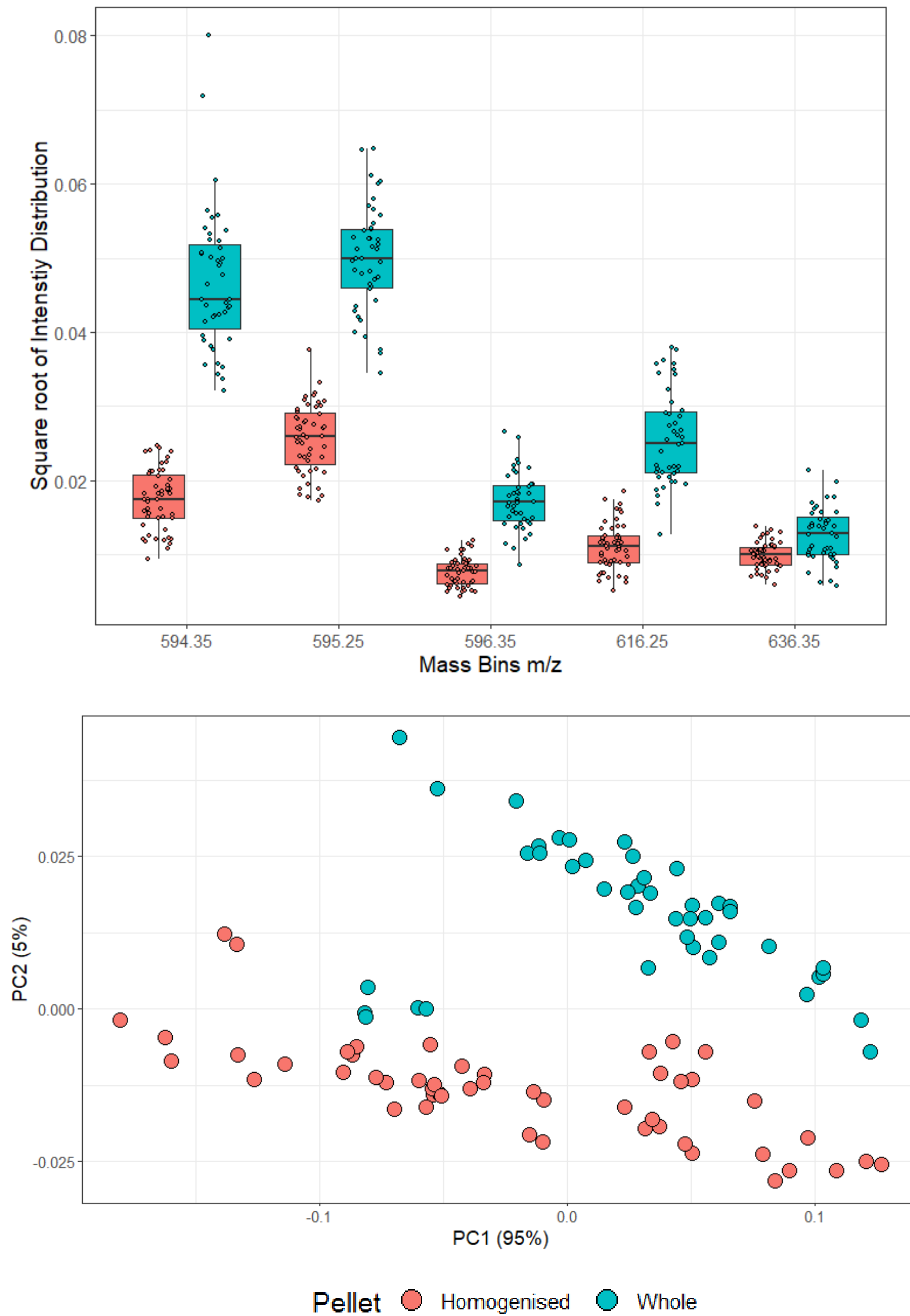
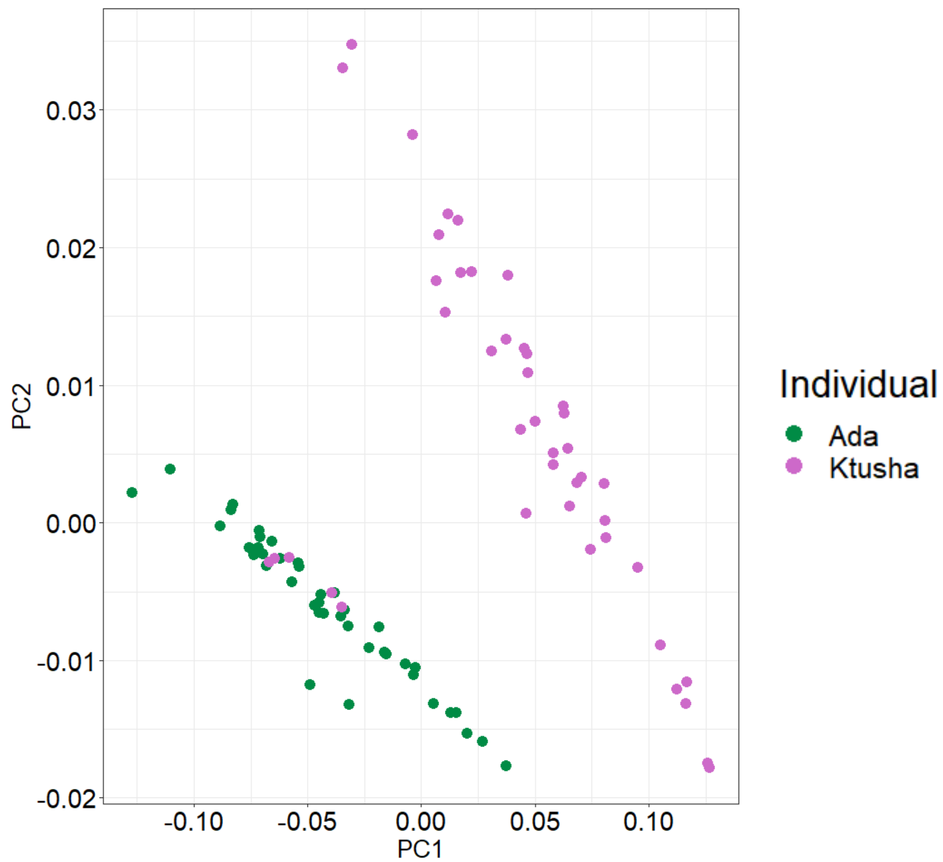


Figure 3.22. The square root relative intensities of the top five mass bins most responsible for the differences between the homogenised and whole pellets. The first two principal components that separate homogenised and whole pellets. Faecal samples are from K'tusha when not pregnant.



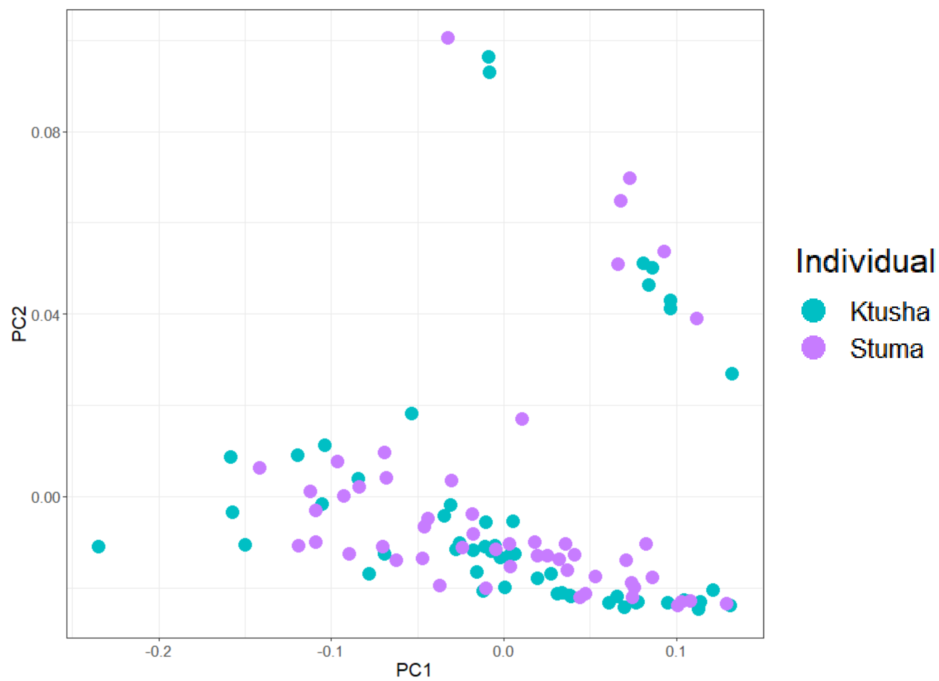
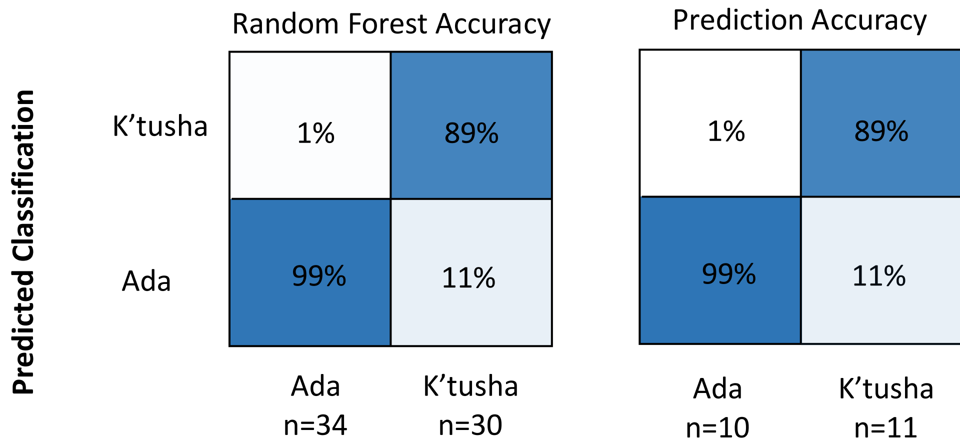


Figure 3.23 a) PC 1 (95%) against PC2 (5%) for whole pellet samples burned from Ada (n=42) and Ktusha (n=43) the current female okapi at Chester Zoo. b) PCA plot for the stored non-pregnant samples from Stuma (n=52) and Ktusha (n=52). Stuma and K'tusha were the only two female okapis until Stuma's death in 2017. Ada and Ktusha show separation whereas Stuma and Ktusha do not. Burning whole pellets rather than amalgamated pellets may provide a more information rich spectra.

Whole Pellets



Homogenised Pellets

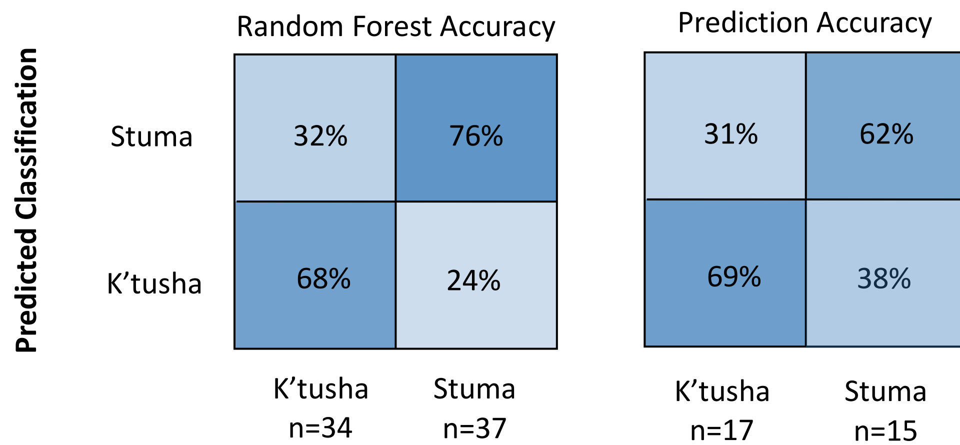


Figure 3.24. Random forests could distinguish between individuals to a higher accuracy for the whole pellets compared to the homogenised samples. Only sample from animals when they were not pregnant were used. The homogenised pellets had also been stored for much longer than the whole pellets.

3.4 Discussion

REIMS cannot be used as a method of testing for pregnancy in okapi using homogenised archived samples. REIMS may still have the potential to be a valuable tool for zoo research. Random forests could not classify pregnant and non-pregnant samples, but specific mass bins did show a significant difference between samples. Reducing the number of mass bins increased prediction accuracy, as did only using samples from a particular time. REIMS could distinguish between individuals of samples burned as whole pellets but could not separate individuals of the combined samples. This suggests that sample quality could significantly affect the spectra obtained from burning.

Okapis are herbivores classified as browsing ruminants (Clauss et al., 2006); browsers feed on leaves and grazers feed on grass. Ruminants are a sub-order of mammals that have four-chambered stomachs. The first chamber is the rumen, where bacteria and protists convert the cellulose of the plant material into simpler compounds to be absorbed (Mason et al., 2011). Browsers retain food particles in their rumen for a shorter time than grazers and, consequently, have a lesser ability to break down fibrous material. In ruminants, the higher digestive ability is associated with smaller faecal particles. The modulus of fineness (MOF) is a measurement used to record the particle size of faeces. A higher MOF means larger particle sizes and, therefore, smaller amounts of digested material. In a study that measured the MOF of 16 species of browsing ruminants, the okapi had the second-highest MOF, with their relative, the giraffe having the highest (Clauss et al., 2002). This suggests that faecal pellets produced by okapi have less digested faecal material and more undigested fibrous material than other mammals. The protocol used by breaking down the pellets and placing them into sample tubes is intended to remove the undigested materials and increase the number of faecal particles in the samples. As a result, what undigested material was present in the samples was much less consistent than with whole pellets. Some samples would have leaves and other twigs, which may have been from a different plant. Breaking the pellets down may have led to higher variation between samples.

Faecal sample preparation has been explored with human faecal samples. The microbiome was compared for differently prepared human samples. Samples were classified as fresh, homogenised (in a blender), frozen, frozen and homogenised (in a blender) and homogenised in a pneumatic mixer. The bacteria diversity was similar for each sample

regardless of the preparation method, but each taxon's abundance varied. The frozen and homogenised samples had more *Faecalibacterium, Streptococcus*, and *Bifidobacterium* and less *Oscillospira,* Bacteroides, and Parabacteroides than fresh samples (Hsieh et al., 2016). Another study that compared fresh, homogenised, and washed with anaerobic dilution fluid (Biopsy wash) found similar results. Bacteria diversity was the same, but the abundance of proteobacteria varied between the biopsy wash samples and the fresh and homogenised samples (Mukhopadhya et al., 2022). It is unknown how much the microbiome influences the REIMS signature. If it does have an effect, then varying abundances of bacteria in the homogenised pellets would make them harder to analyse than the whole pellets.

Mass spectrometry has been used to discriminate vegetation similarly to REIMS but with a different ion-ionisation source. Direct analysis in Real-Time (DART) ionisation is an ion source coupled with a TOF mass spectrometer. A gas (helium or nitrogen) is introduced through the DART source between two electrodes producing ions with electronic excitation. This plasma then travels through more electrodes so only neutral molecules remain. As these molecules leave the source, they pass through the sample, causing ionisation and then continue into the mass spectrometer's inlet (Hoffmann and Stroobant, 2007). DART has been used to identify different species of oak trees (Cody et al., 2012) and legumes (Lancaster and Espinoza, 2012) and between wild and cultivated agarwood (Espinoza et al., 2014) with high accuracy. DART was also used to distinguish between Douglas-fir trees from two different locations, and random forests produced accuracies of 70% and 76% (Finch et al., 2017). Although mass spectrometry can be used to classify vegetation, it was unlikely to have an effect here. When burning the homogenised samples, the digested material disintegrated on burning, producing lots of smoke; the undigested material did not burn or produce smoke. The REIMS signature was more likely influenced by the digested material than the undigested vegetation. The diets of okapi (in a zoo context) should have been similar enough not to have had an effect, as DART could distinguish between species (Cody et al., 2012) but was less able to distinguish differences within species (Finch et al., 2017). Whole pellets could therefore produce, more reliant signals for classification.

As previously stated, okapis have a larger MOF compared to other mammals. REIMS could still be potentially be used to test pregnancy in other species with a lower MOF. Near-infrared reflectance spectroscopy (NIR) is a method used to analyse the faeces of herbivores.

NIRS is a rapid, non-invasive analytical technique that measures the light of a sample in the near-infrared region of the electromagnetic spectrum (700–2500 nm). It has been used to determine the sex and species of red and fallow deer, red deer were successfully identified in 81% of samples, and fallow deer was 77%. The classification of sex was only 60%, although they did include the two species within the sex classification (Tolleson et al., 2005). The classification of species and sex identification was less than that of using REIMS to identify rodent species. It could be due to the different methods (REIMS or NIRS) or because herbivore faeces have more variation and are, therefore, more challenging to classify. NIR can predict the diets of grazers through the analysis of faeces, but only within specific subsets. For example, the data collected from tropical regions could not be used to predict the diets of grazers from temperate regions (Dixon and Coates, 2009).

The homogenised samples were stored in a freezer for a significant period. In this study, the pregnancy status accuracy increased when the data was split into subsets based on the collection date. Plotting the okapi samples through the total collection time showed less variation than occurred month to month, which suggests that the spectra were not changing with the length of storage time. Tracking the intensity of one mass bin showed changes for non-pregnant samples but not pregnant samples, but the changes increased for one individual and not the other. These results also suggest that storage time does not consistently impact the spectra.

The individual's age at the time of collection could have been having an effect. The relative intensities of mice did change with age, but only until they reached maturation. There was little difference between mice 100 days and 300 days old. Samples were only collected from the okapi when they reached breeding age. In this study, the animal's age at collection was directly related to the time the sample was in storage. Therefore, it cannot be determined if the age or the storage time had an effect. NIR studies have been able to use faeces that had been stored for up to two decades, the spectra signature changed with time, but this did not affect the ability to predict the diet content of samples (Landau et al., 2008). The diet of mice did not affect their ability to be classified by REIMS (Davidson et al., 2019). Therefore, a change in spectra does not necessarily affect the ability to classify certain classes.

Changes may have occurred at different time points due to environmental factors. Faecal glucocorticoid metabolites (FCM) have increased in captive polar bears when they

become stressed due to being transported (Hein et al., 2020). They varied in wild dogs from different zoos (Van der Weyde et al., 2016) and showed seasonal patterns in numbats (Curry et al., 2021). Okapis are pregnant for up to 15 months; within this time frame, many factors could have occurred that may have caused a change in the spectra that was more dominant than the change caused by pregnancy. During the sample collection period, there would have been social changes in the group, including the removal of the previous offspring to other zoos, a change in the stud male and the death of Stuma in 2017. It was unknown if any events occurred that could have caused a spectral change. Adding samples from multiple individuals from multiple zoos could help pregnancy classification. If changes occur through environmental factors, the random forests might find it difficult to establish what change in spectra was caused by external factors and what change was caused by pregnancy. If many individuals were included, there is potential for many more changes caused by external factors, but these changes are unlikely to be consistent. If the pregnancy causes an intensity change, then this change should be consistent for all individuals. Including more individuals would make it easier for random forests to establish what mass bins are only responsible for the difference in pregnant and non-pregnant samples.

With confidence, REIMS could identify which individual a whole faecal sample came from. Individual identification would benefit conservationists as they could track animals released into the wild. Known samples could be collected while the animal is in captivity, and these samples could be used to create the random forest model. If the animal is released, the model could be used to predict if samples collected from the wild belonged to the animal. This could reduce the reliance on tracking devices or observational sightings to monitor released animals. None of the three females were related to each other; therefore, if REIMS could separate Ada and K'tusha but not Stuma and K'tusha, it would suggest the different sample preparation methods may be affecting the results. The ability to identify individuals also showed that whole pellets that have not been in long-term storage have information-rich spectra. Random forests could also determine whether a sample from the same non-pregnant individual had been homogenised. This suggests that homogenising the sample is having an impact on the REIMS signature. However, the homogenised samples were also older and had been in storage for significantly longer. The results confirm a difference between the types of faecal samples, but more work is needed to confirm the limiting factor.

Fresh, homogenised samples or long-term frozen whole pellets may produce information-rich spectra. Establishing pregnancy in herbivores using fresh whole pellets could still be possible. It would be valuable to repeat the study to distinguish between individuals since this investigation only included two individuals and to repeat with other species.

In conclusion, the results suggest some differences between pregnant and non-pregnant samples but not enough to establish a pregnancy test using REIMS. It could be that changes caused by pregnancy are not consistent enough to make a successful predictive model, or it could be an issue with sample preparation. The homogenisation or the long-term storage of these samples could have caused a reduction in spectra quality. A pregnancy test using REIMS may still be possible if a model was built using spectra produced from burning fresh whole pellets. Using samples from multiple individuals, half pregnant and the other half not pregnant, may improve the model rather than using pregnant and non-pregnant samples from the same individual. Using multiple individuals that are either pregnant or not should help the random forest model find the bins that change consistently due to pregnancy. Only using samples from two individuals meant that changes to the spectra could have been caused by other factors, such as age or environmental effects at the time of sample collection rather than pregnancy.

Chapter 4: REIMS Analysis of Stored Vole Pellets

The effect of storage time on the classification of RIEMS signatures using faecal samples from Bank and Field Voles

4.1 Introduction

The results from chapter three suggested that the long-term storage of samples caused a change in the faecal sample profile. Random forests could distinguish fresh whole samples by individual okapi but could not distinguish between the archived homogenised samples. It was unknown if the sample preparation, the long-term storage or both prevented their classification. Long-term monitoring of populations is a vital part of conservation research and management. Not being able to use archived samples would limit the usefulness of REIMS to long-term monitoring projects. Investigating the impact of long-term storage would help determine if pregnancy could still be determined in fresh faecal samples and if homogenisation of faeces was the limiting factor.

Long-term monitoring of populations can help determine if population fluctuations are due to cyclic changes or external factors, including climate change (Stenseth et al., 1997). A study from 1987 to 2003 demonstrated the pied flycatcher population had declined in the Netherlands by 90% due to climate change. The decline of the migratory bird was caused by their peak food abundance occurring earlier than the bird's breeding season due to increased temperatures (Both et al., 2006). The small mammal population of ten species was monitored in the Yukon Forest of Canada for 46 years using live-trapping methods. Most species (deer mouse, long-tailed vole, singing vole, western heather vole, meadow jumping mouse, West Siberian lemming and the Northern Bog lemming) populations decreased, but three species (red-backed vole, meadow vole and tundra vole) improved. Both the average winter and summer temperatures increased during this time. They may be responsible for the changes in the small mammal population, but predator numbers also increased during this time (Krebs et al., 2019).

Many research sites, such as the Kalahari Meerkat Project (Paniw et al., 2019) or the Botswana Predator Conservation Trust (Hofmann et al., 2021), have been set up to collect data routinely by volunteers. Multiple researchers can use this information to answer various questions at any time. Since the data was collected for multiple reasons, the data collected can be quite expansive. Long-term population monitoring can be achieved through direct observations and the presence of signs and markers recorded over many years or by using biological sample analysis. The biological samples may be analysed as they are collected or stored and analysed at the end of the project. The results from chapter two suggested that

classification analysis using REIMS is better when samples are analysed at the same time. REIMS would be a valuable tool to conservationists for long-term monitoring if it can be used for faecal samples that have been stored by freezing. Other types of storage are available, including being stored in alcohol but freezing at -20°C is the most common method.

This study aimed to determine how long-term storage (up to four years) would affect the classification of vole samples by species (Bank and Field voles) and sex (male and female). Samples were collected once a year from 2018 to 2021, and random forest accuracies were compared when samples from all collection years were in the model and used one year at a time. A random forest was built to distinguish between the collection year of the sample. A high random forest accuracy would suggest that the REIMS signature changes significantly over time. A random forest accuracy between 60% and 75% would suggest that changes were occurring, but they were small and may not affect other classification types. Even if random forests can identify collection year, that does not mean changes in the faeces affect the ability to classify other factors. A random forest was built to distinguish between bank and field voles from all collection years. The model was repeated using only the oldest samples in the model and again with the newest samples. If there was no significant difference between the random forest accuracies, then it would suggest that the time in storage has not affected the ability to classify species. Random forests were built to distinguish sex using all the samples and separated into the oldest and newer samples. If storage time does have an effect, then it would be expected that the random forest accuracy would be higher when using only recently stored samples in the model compared to using older samples.

4.2 Method

4.2.1 Wild Vole Trapping and Breeding

In the summer of 2020, 13 small mammal tube traps were placed in an area known as the meadow in Ness Gardens, Neston (Figure 4.1). The area was chosen as the grass had not been cut for a year, and signs, including runs and faecal pellets, indicated the presence of small rodents. Animals, when moving through their home range, will use the same paths and cause the grass to flatten over time; these animal paths are known as runs. The tube traps were filled with Harry Hamster, apple slices and hay (Figure 4.1c). The traps were placed for a week but were inactive to allow animals to become habituated to the presence of the traps. When the traps are inactive, animals can enter and leave the trap without being confined. After the first week, traps were switched to active, and traps were checked twice a day and topped up with food and bedding as needed. Any closed traps were opened into a large clear handling bag to determine the species of the caught animal. Bank voles, mice and any pregnant field voles were released in the location of the trap they were caught in. Field voles were transferred into an empty M3 cage and transported to the animal unit of the Mammalian Behaviour and Evolution group at the University of Liverpool's Leahurst campus (MBE). The caught individuals were placed into cages set up in the standard method. All individuals brought into MBE were treated with Hartz® UltraGuard® OneSpot® to remove the voles of ticks and fleas. Field voles were set up in large containers (bins) to be allowed to breed. Voles already present in MBE were set up for breeding first to allow the captive individuals to acclimatise to their new surroundings.

The breeding bins were filled to a depth of 15 cm with Lignocel Select Fine substrate, a water bottle, bedding, hay, cardboard, Harry Hamster and EURodent Diet were placed into the bin. Males and Females were placed into the bins simultaneously, and males were removed once offspring were detected or after five weeks. Offspring were removed from the bin after three weeks or earlier if the dam sired a second litter. Mothers were left in the bin for at least two weeks after the male was removed, as they were usually pregnant for a second time. Faecal pellets were collected from all offspring when they were weaned and once a week until they reached over 120 days old, and pellets were also collected at least once from captured voles. The animals were trapped and handled following international best practice

guidelines (Anonymous, 2018). Traps were checked twice daily, and animals in the lab were provided enrichment and handled using the tunnel method (Hurst and West, 2010).

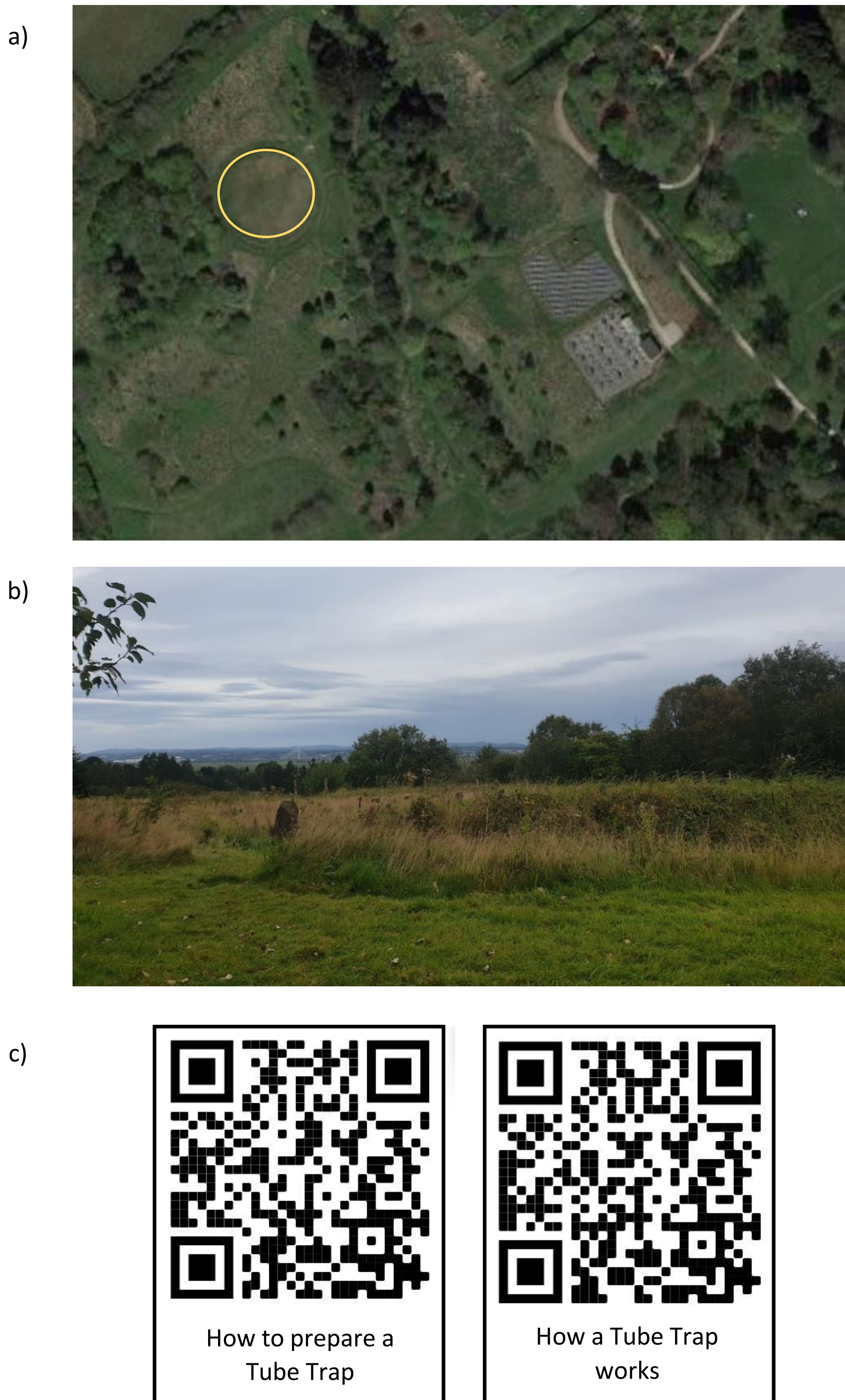


Figure 4.1 a) Overview of Ness Gardens in Neston, the 13 tube traps were placed in a circle approximately 10 paces apart as shown by the yellow circle (53°16'18.4"N 3°02'54.5"W) ©Google. b) The area used for trapping was surrounded by a trench and the grass had been left to grow for a year © Natalie Koch 02/09/20. c) Demonstration videos of how the tube traps worked and how they were set up before being placed in Ness Gardens

4.2.2. Sample Collection

Faecal samples were collected from captive and wild-caught bank voles (*Myodes glareolus*) and field voles (*Mictrotus agrestis*) (Table 4.1). Both vole species were housed in individual cages; field voles were kept in the larger MB1 cages, and bank voles in M3 cages. All animals had access to food, 5FL2 EURodent Diet (IPS Product Supplies Limited, London, UK) and water *ad libitum.* Both vole species were provided with Harry Hamster complete muesli (Supreme Petfoods Ltd., Ipswich, UK), and field voles were given fresh-cut grass once a day. Cardboard tubes and paper wool nest material for enrichment were provided to all animals, and voles also received hay. During sample collection, voles were moved to individual clean cages for a maximum of two hours or once the individual had produced at least five pellets. Pellets produced by an individual were removed from the cage using metal tweezers, placed in a 1.5 ml Eppendorf tube, and stored at -18°C. After the two hours, voles were returned to their home cage regardless of defaecation.

Table 4.1. The total number of voles faecal samples were collected from. Juveniles were only included in models built to distinguish age. Models build to distinguish sex or age only included captive born field voles.

Captive born or Wild Caught	Species	Sex	Age	Number of Individuals
Captive Born	Bank Vole	Female	Adult (>36 days)	28
			Juvenile	5
		Male	Adult (>36 days)	27
			Juvenile	5
	Field Vole	Female	Adult (>36 days)	96
			Juvenile	44
		Male	Adult (>36 days)	117
			Juvenile	37
Wild Caught	Bank Vole	Female		5
		Male		6
	Field Vole	Female		2
		Male		3
			Total	376

4.2.3 Using REIMS to Burn and Analyse Faecal Pellets of Bank and Field Voles

Three faecal pellets from ten different vole individuals caught in 2020 were burned using different electrode modes and voltages to establish the optimal settings for burning vole pellets. The burn signal was compared for each pellet burned to determine which settings gave the highest signal. Faecal pellets were burned using the cut mode at 35 V. Five faecal pellets were burned for each vole sample if possible, but at least three pellets were used. MilliQ was added as necessary (approximately 100 µl over all five pellets of each sample) to ensure sufficient conductivity. Leu-enkephalin was continuously ejected at 50 µL/min into the inlet capillary of the REIMS ionisation source to provide a lock mass. The sample cone and the heater bias were set to 60 V. The spectra were recorded in full-scan resolution, negative ion mode, at a scan rate of 1 scan per second from 50-1200 m/z. The spectra were uploaded to LiveID to be normalised lock mass corrected and binned to 0.05 Da; the mass range was reduced to 400 to 1100 m/z

Samples were analysed in two separate groups. Group one contained the samples of bank and field voles that were already in the lab for use by other PhD students and contained a mixture of wild and captive-born individuals. Group one had one sample (consisting of several pellets collected simultaneously) per individual. The ages of the individuals born in captivity varied between 15 and 490 days old. Group two contains captive-born field voles; some had captive-born parents, and others had wild-caught parents (Figure 4.2). Group two contains multiple samples from the same individuals at different ages, but one random sample was selected from each vole for building random forest models. All samples were frozen at -20°C on the collection day and burned over two months (Table 4.2).

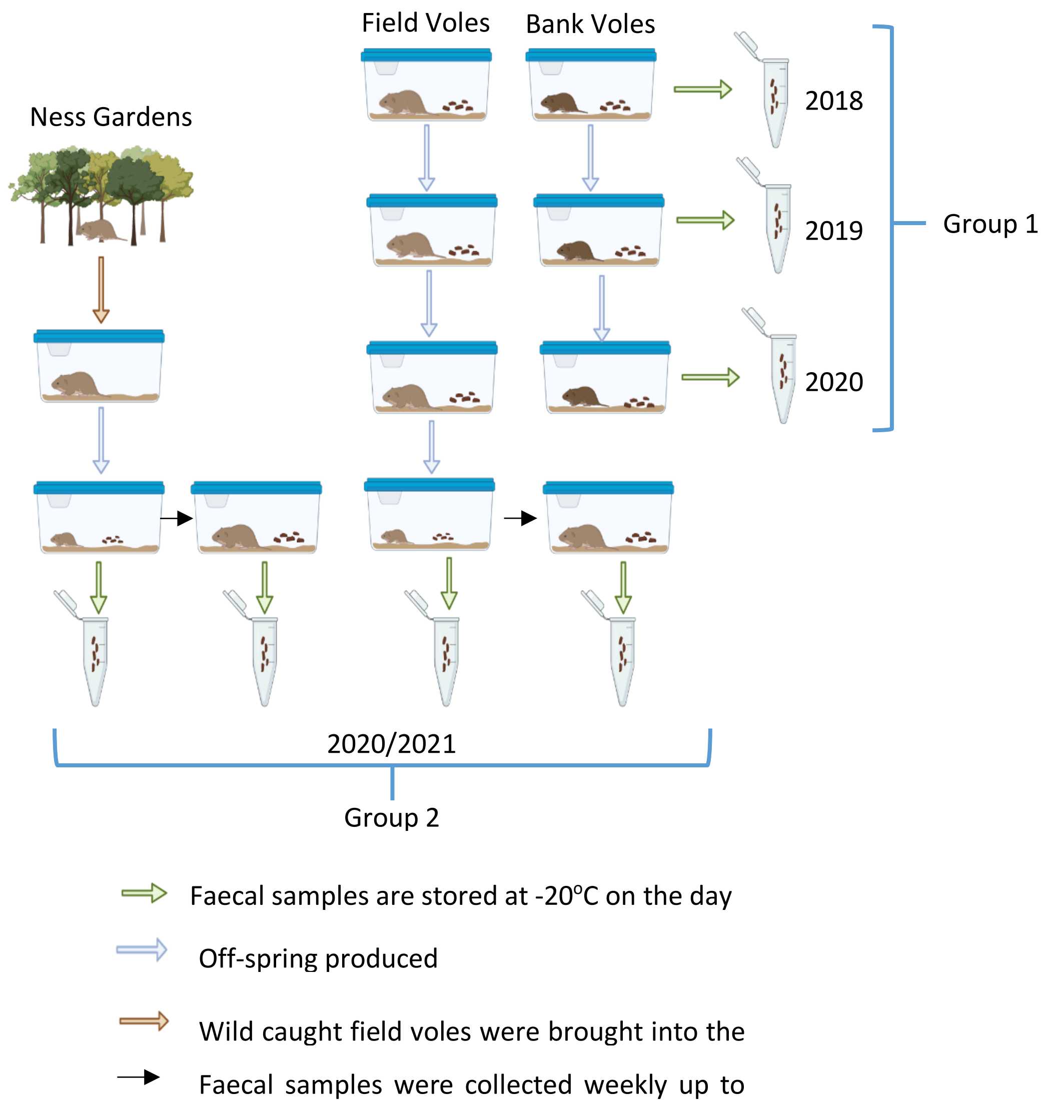


Figure 4.2 Faecal samples were split into two groups based on when they were collected. Group 1 contains the field (n=115) and bank (n=77) voles already present in the lab. Some of these had been captive born and some wild-caught by another PhD Student. Group 2 contains samples that were collected on a weekly basis from field voles (n=297) born from either recently caught voles or the field voles already in the lab. The figure was made using biorender.com

Table 4.2. The range of days faecal samples were in -20°C storage for the different collection year groups.

Collection Year	Minimum days in storage	Maximum days in storage	Number of Samples
2021	10	190	99
2020	200	470	121
2019	500	700	77
2018	750	900	74

4.3 Results

4.3.1 Discrimination of Vole Pellets by Storage Collection Year

A less than 60% classification accuracy would suggest that the spectra do not change with the length of storage time. Group one contained three different year groups; a random forest could classify the year groups with an accuracy of 65%. (Figure 4.3). Suggesting there are minor changes to the spectra caused by the length of time in storage. The top three mass bins that were most responsible for the differences between collection years did show a significant difference between the three storage years (Figure 4.4). The intensities of samples from the year 2020 diverged more compared to the other two groups. Group Two had four classifications years 2018 to 2021 and produced a random forest classification of 65%. Most samples were classified as either 2018 or 2021, suggesting no difference between the 2020 and 2021 samples or between 2018 and 2019. Significant differences existed between the intensities of field vole samples between each collection year. Suggesting there are minor changes to the spectra caused by the length of time in storage. The most significant differences between collection years occurred between 2019 and 2020, between two and three in storage (Figure 4.5). This suggests some changes occurred to the samples over time but after the first two years of storage.

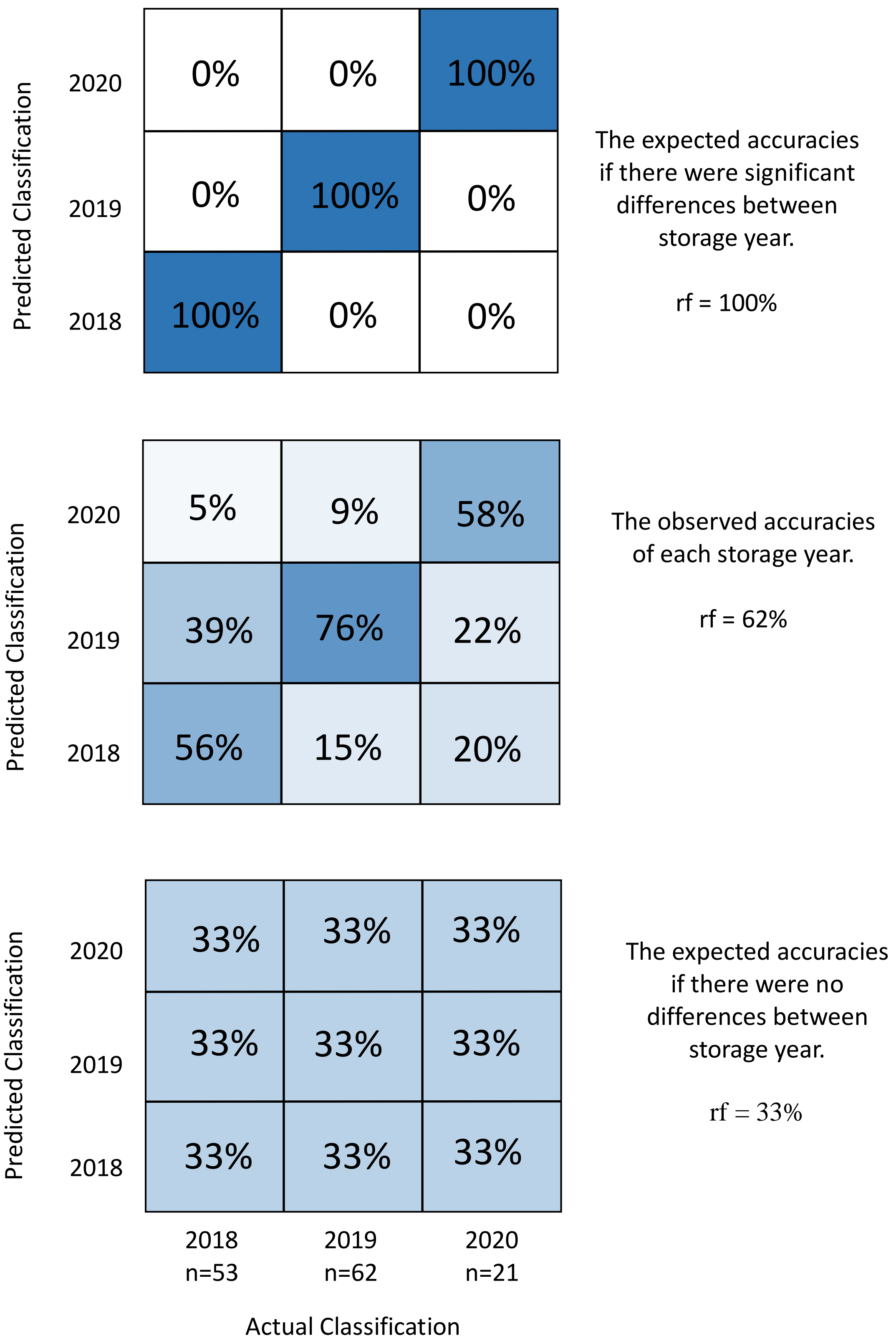


Figure 4.3 a) The expected accuracies for each collection year if there were significant differences between the spectra because of length of storage time. b) The random forest classification accuracy for each collection year using the group 1 data. c) The expected accuracies for each collection year if there were no significant differences between the spectra because of length of storage time. The observed accuracies lay between the two expected accuracies suggesting there are small changes between spectra due to storage time.

a)

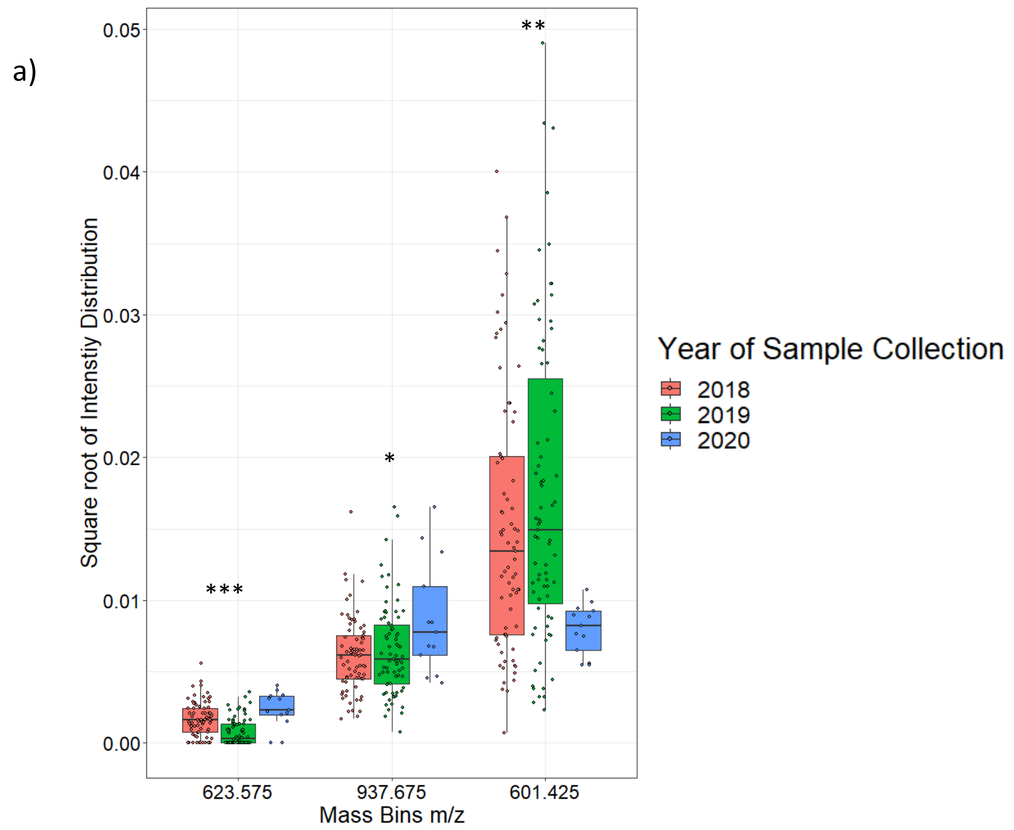


Figure 4.4 The data from Group 1. The relative intensities of each of the sample collection years for the top three mass bins that are most responsible for the differences between collection years. Samples from 2018 were stored at -20°C for over 750 days. Samples from 2019 were stored between 550 days and 700 days. Samples from 2020 were stored for less than 100 days.

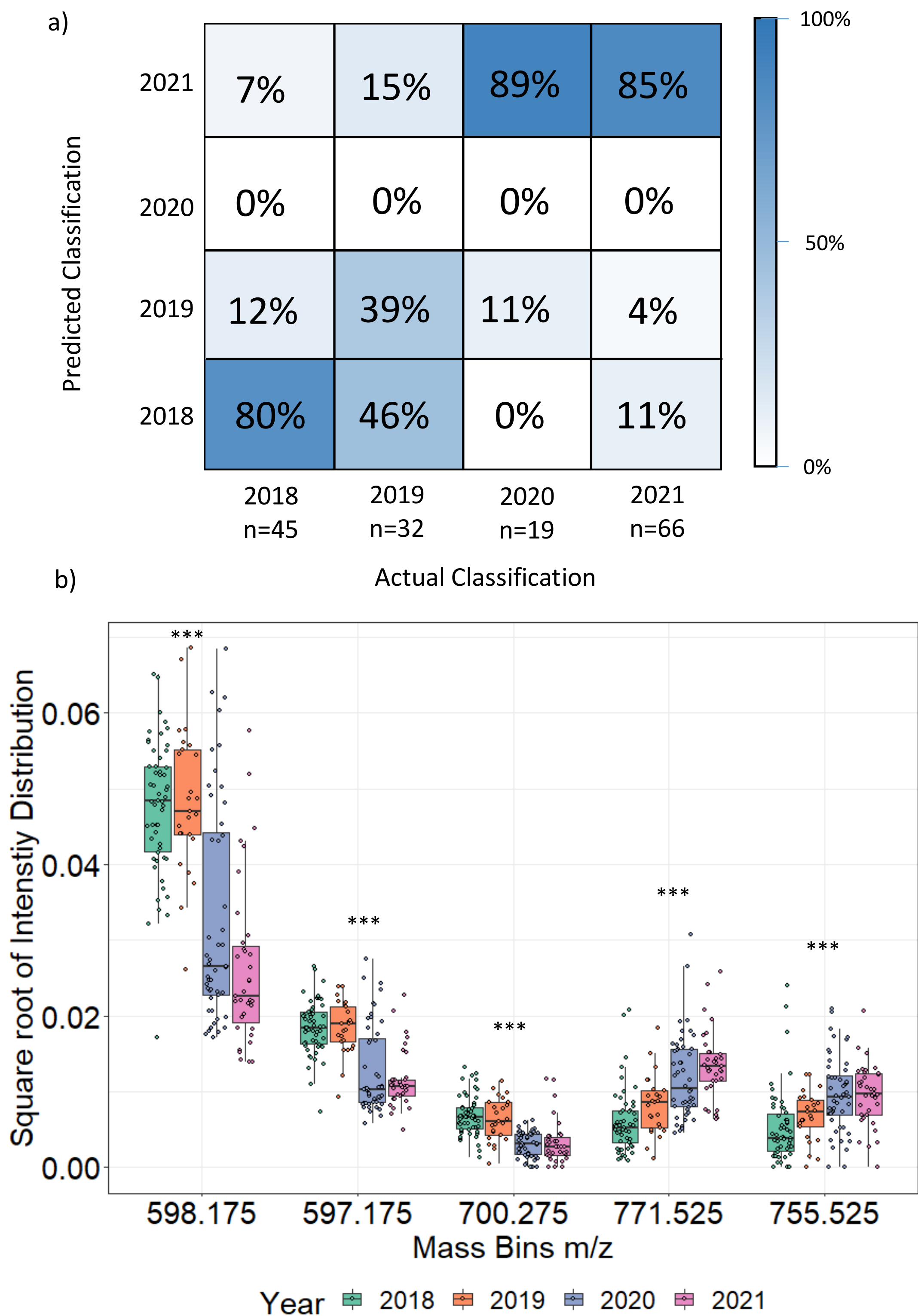
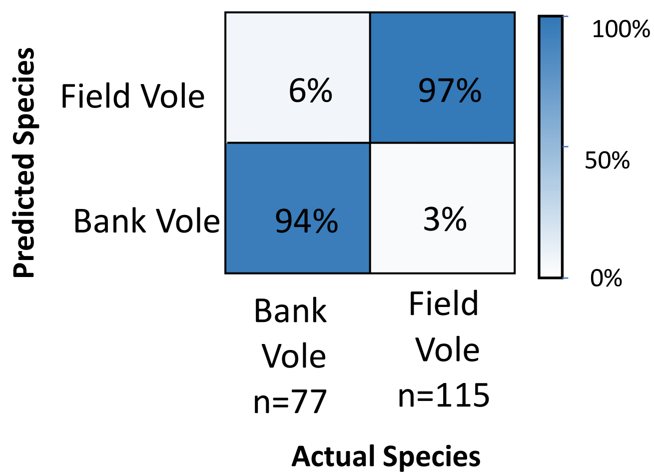


Figure 4.5 a) The random forest classification accuracy for each of the collection years from the Group 2 data. The samples were mostly classified as one of the two extreme years suggesting that the spectra change with a significant amount of storage time but does not significantly change from year to year. b) The intensities for the five most discriminant bins most responsible for differences between collection years. There is a significant difference between intensities with storage time but how they differ varies for each mass bin.

4.3.2 Discrimination of Vole Species of Faecal Pellets Stored For up to Four Years

Using all samples to create a random forest model to distinguish between bank and field voles gave a random forest accuracy of 96%. An LDA on the top 12 PCA components could distinguish between the intensities of field and bank voles (Figure 4.6). Even though some of the faecal pellets had been stored for up to three years, REIMS could distinguish species. Using samples from 2018 only gave a random forest accuracy of 95%, and using samples from 2019 only gave an accuracy of 99%. There was no significant difference between the accuracies of species classification between different sample collection years (Kruskal-Wallis, p-value > 0.8). Samples collected from 2020 were excluded as there were only a limited number of bank voles. The separation observed from an LDA on the top 12 PCA components between the intensities of field and bank voles from 2018 was very similar to the intensities from samples from 2019 (Figure 4.7). The results from analysing the two collection years suggest that samples are unaffected by the time left in storage.

The ten five most discriminant mass bins for species varied slightly between using all samples (model all-years), samples from 2018 only (model 2018) and 2019 only (model 2019). Three mass bins were the same for all three models, and five were found in at least two models (Table 4.3). The intensities of the three mass bins in the top ten most discriminant mass bins for each model showed significant differences between bank and field voles but only slightly between collection years. Two mass bins only appeared in the top ten most discriminant bins for the model using all years, but they showed a significant difference for each year group. Three mass bins were randomly selected, and there were no differences between the intensities for bank and voles or between collection years (Figure 4.8). Mass bins may still show significant differences between classes without appearing in the top ten. It may be just the ranking of mass bins that changes between models rather than what bins have a significant difference between species Differences between species were more significant than the differences between collection years, even for three randomly selected mass bins.



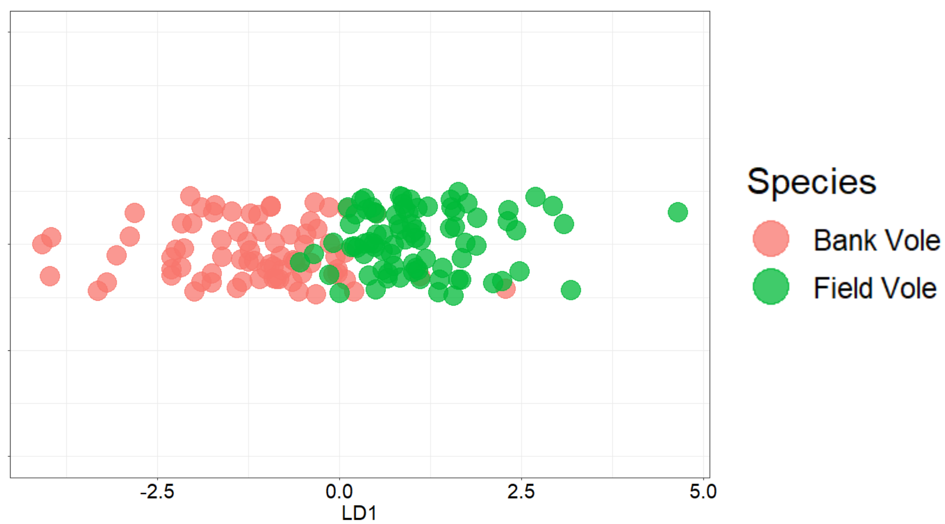


Figure 4.6 All the random forest classification accuracy for all samples. The mass spectra for bank and field voles from all samples collected. An LDA using the top 12 PCA components for all samples. The results show that REIMS can be used to discriminate samples that have been stored for up to three years.

Figure 4.7 All the random forest classification accuracy for bank and fields voles for when only using samples from the same collection year 2018, or 2019. The mass spectra for bank and field voles from samples collected in 2018 or 2019. An LDA using the top 12 PCA components for samples from 2018 or 2019. Using samples from one year only did not significantly change the classification accuracy or spectra of species, therefore storage time does not affect species classification.

Table 4.3 The top ten most discriminant mass bins for the random forest model using all the samples (model all-inclusive) when using samples collected in 2018 only (model 2018) and when using samples collected in 2019 only (model 2019).

Model All-Inclusive	Model 2018	Model 2019
	403.325	
		431.375
	431.425	
433.325	433.325	433.325
434.325	434.325	
459.425		459.425
	474.225	
482.275		482.275
	497.325	
		529.525
	604.475	
796.625		
811.525	811.525	
811.575		
812.525	812.525	812.525
812.575	812.575	812.575
		813.575
		825.525
826.525		826.525

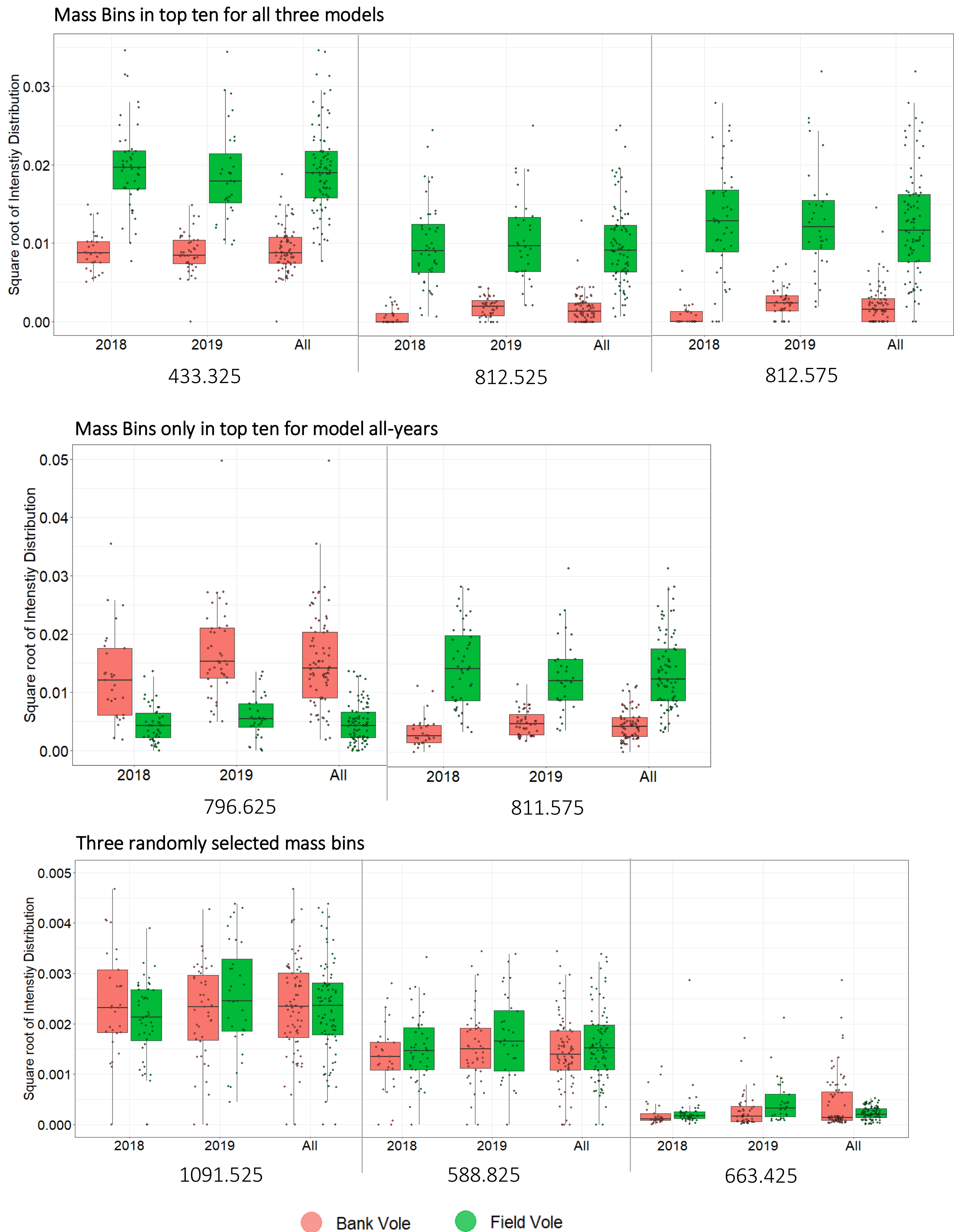


Figure 4.8 The relative intensities for bank voles and field voles for samples from 2018 only, 2019 only and all samples (2018-2020). a) The three mass bins that were in the top ten most discriminant mass bins for three models (model 2018, model 2019, model all-inclusive). b) The two mass bins that were only in the top ten most discriminant mass bins for model all-inclusive. c) Three randomly selected mass bins.

4.3.3 Using Older Samples to Predict Species of Newer Samples.

Samples collected in 2018 could predict the species classifications of samples collected in 2019 to an accuracy of 95%. Longer stored samples can therefore predict the species of newer samples when samples are burned simultaneously. Samples from 2018 and 2019 could correctly predict the field vole samples from 2020 with respective accuracies of 97% and 82%. Only four bank voles were collected in 2020, so longer stored samples could not predict bank vole classification (Figure 4.9). These results confirm that REIMS classify vole species using faecal samples that have been in storage for up to four years and they could be new to predict newer samples..

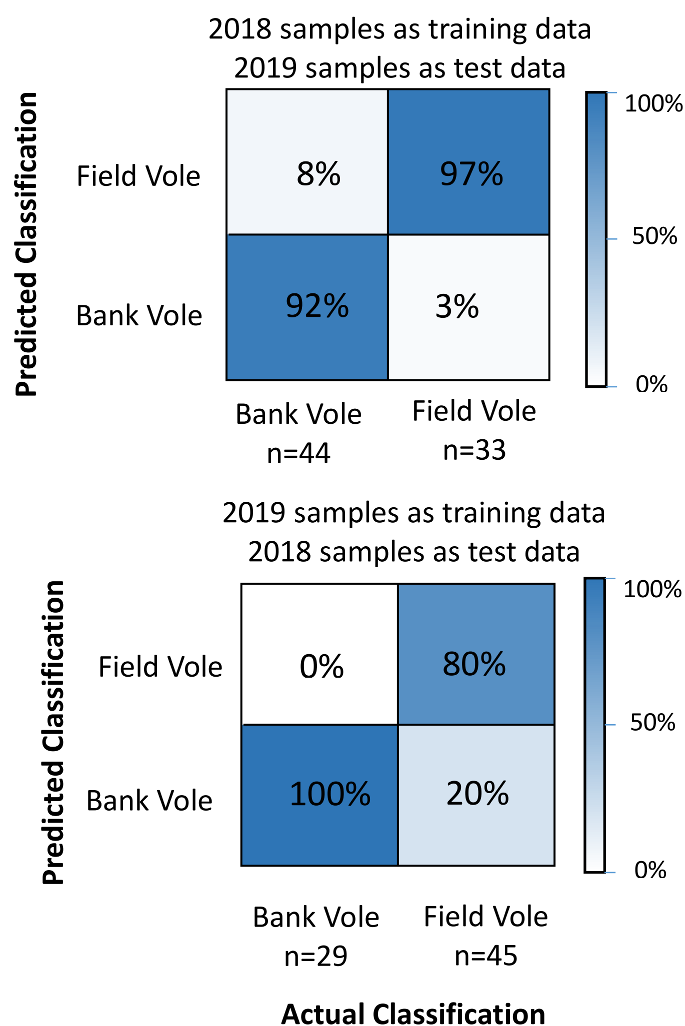


Figure 4.9 The random forest accuracies of using one year group to build the random forest model (training data) to predict the species of samples from a different year group (test data).

4.3.4 Discrimination of Sex of Vole Faecal Pellets Stored up to Four Years

Bank Voles

Group one was used to classify bank voles as male or female; a random forest accuracy of 73% was obtained, and the classification accuracy of females (73%) was higher than males (61%). There was no discrimination between males and females of an LDA using the top ten PCA components, suggesting no difference between the spectra of male and female bank voles. The random forests were repeated using the top 100 mass bins only, and the classification accuracy of the sex of bank voles increased to 90%. An LDA of the top 5 PCA components showed discrimination between male and female bank voles (Figure 4.10).

To determine if storage time influences sex discrimination, random forest models were built to discriminate between male and female bank voles from 2018 only and 2019 only. The random forest accuracy for using 2018 bank vole samples was 86%. The classification of females increased to 99%, but the classification of males decreased to 64%. This suggests that the faecal pellets of male bank voles may have been influenced by being in storage. The classification of male and female samples from 2019 remained at 90%, suggesting that if changes did occur to the male spectra, they occurred in storage between storage years three and four. An LDA using the top 5 PCA components also showed a slight increase in separation between males and females in 2018 compared to 2019, it was expected that 2019 would show an increased separation, but there were more samples from 2019 than 2018, which could be influencing the LD component (Figure 4.11).

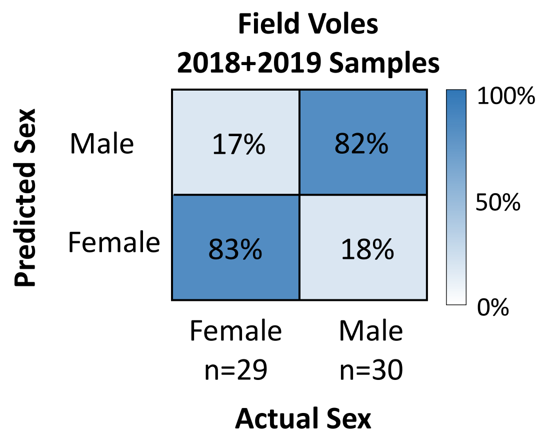
Figure 4.10 The classification accuracy of male and females using all bank vole samples. An LDA using the top 10 pca components using all bank vole samples. A random forest model was conducted with all mass bins for each species to determine the top 100 mass bins. A new random forest model was created with just the top 100 most discriminant mass bins. The confusion matrices for the classification accuracy of males and females all bank vole samples using the top 100 most discriminant mass bins. An LDA using the top five pca components was performed with the relative intensities of the top 100 mass bins.

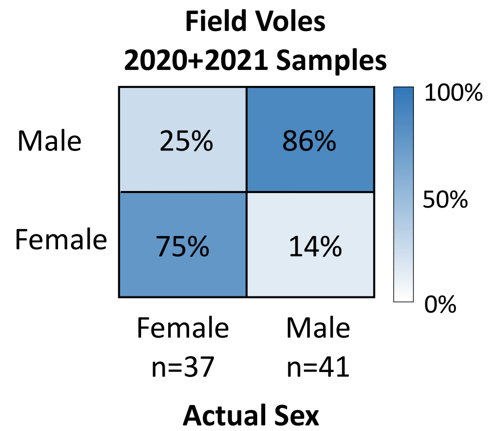
Figure 4.11 The classification accuracy of males and females using bank vole samples from 2018 or 2019. An LDA using the top 5 pca components using bank vole samples from 2018 or 2019

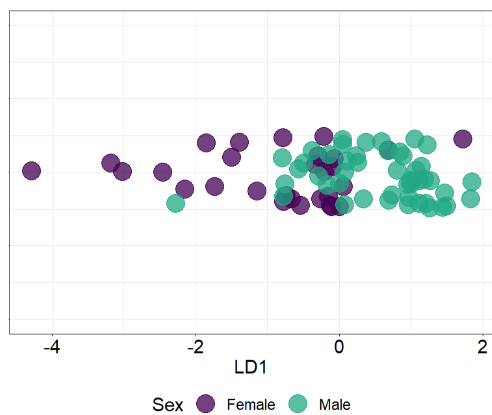
Field Voles

A random forest using all field vole samples from group 2 gave a random forest accuracy of 62%; this increased to 80% when using the top 100 mass bins (Figure 4.12). The field vole samples were split into two groups; the first group contained samples from 2018 and 2019 (longer stored samples) the second group contained samples from 2020 and 2021 (shorter stored samples). The random forest accuracy of sex discrimination using 2018 and 2019 field vole samples was 82% and 81% when using samples from 2020 and 2021. A Kruskal-Wallis test found no difference between the mass bins that discriminate sex for samples stored for more than two years compared to samples stored for less than two years ($p>0.432$). An LDA using the top 5 PCA components suggested that males and females of shorter stored samples were less distinguished than samples stored for longer (Figure 4.13).

Figure 4.12 The classification accuracy of male and females using all field vole samples. An LDA using the top 10 pca components using all field vole samples. A random forest model was conducted with all mass bins for each species to determine the top 100 mass bins. A new random forest model was created with just the top 100 most discriminant mass bins. The confusion matrices for the classification accuracy of males and females all field vole samples using the top 100 most discriminant mass bins. An LDA using the top five pca components was performed with the relative intensities of the top 100 mass bins.







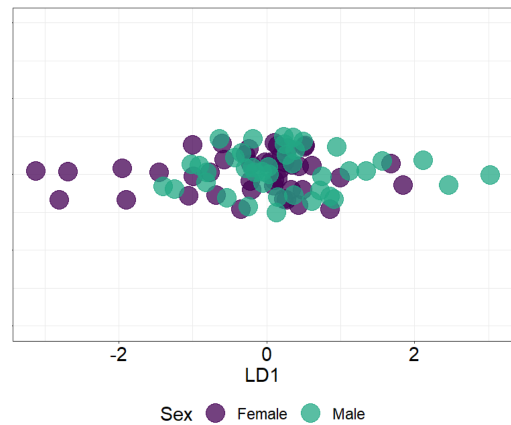


Figure 4.13 The classification accuracy of males and females using field vole samples from 2018 or 2019 The averaged spectra of 2018 or 2019 bank vole samples of each sex. An LDA using the top 5 pca components using bank vole samples from 2018 or 2019

Comparing the most discriminate mass bins of sex

The top five most discriminant mass bins of sex when using only longer stored were slightly different to samples stored for less (Table 4.4). The relative intensities were more similar between males and females of older samples. It could be an external factor in 2018 leads to faecal pellets of males and females being more similar or pellets changing due to storage time. For bank voles, there were still significant differences between males and females of older stored samples and shorter stored samples of the top five mass bins when using all samples (Figure 4.14). With field voles, two of the discriminant mass bins that showed a significant difference for short-termed stored samples did not for long-termed stored samples (Figure 4.15).

Table 4.4 The top five most discriminant mass bins for the random forest model using all the bank vole samples, when using samples collected in 2018 only and when using samples collected in 2019 only. The field vole samples, when using all samples, samples collected in 2018 and 2019 when using samples collected in 2020 and 2021.

Bank Voles

All Samples	2018 Samples	2019 Samples
		493.425
	620.575	
622.625		622.625
638.575	638.575	
655.575		655.575
	656.575	
	880.025	
		953.575
991.425		991.425
998.225	998.225	

Field Voles

All Sample	2018 & 2019 Samples	2020 & 2021 Samples
	534.875	
539.625		539.625
	553.325	
	562.575	
565.975		565.975
595.025		595.025
690.725		690.725
720.775	720.775	
		829.975

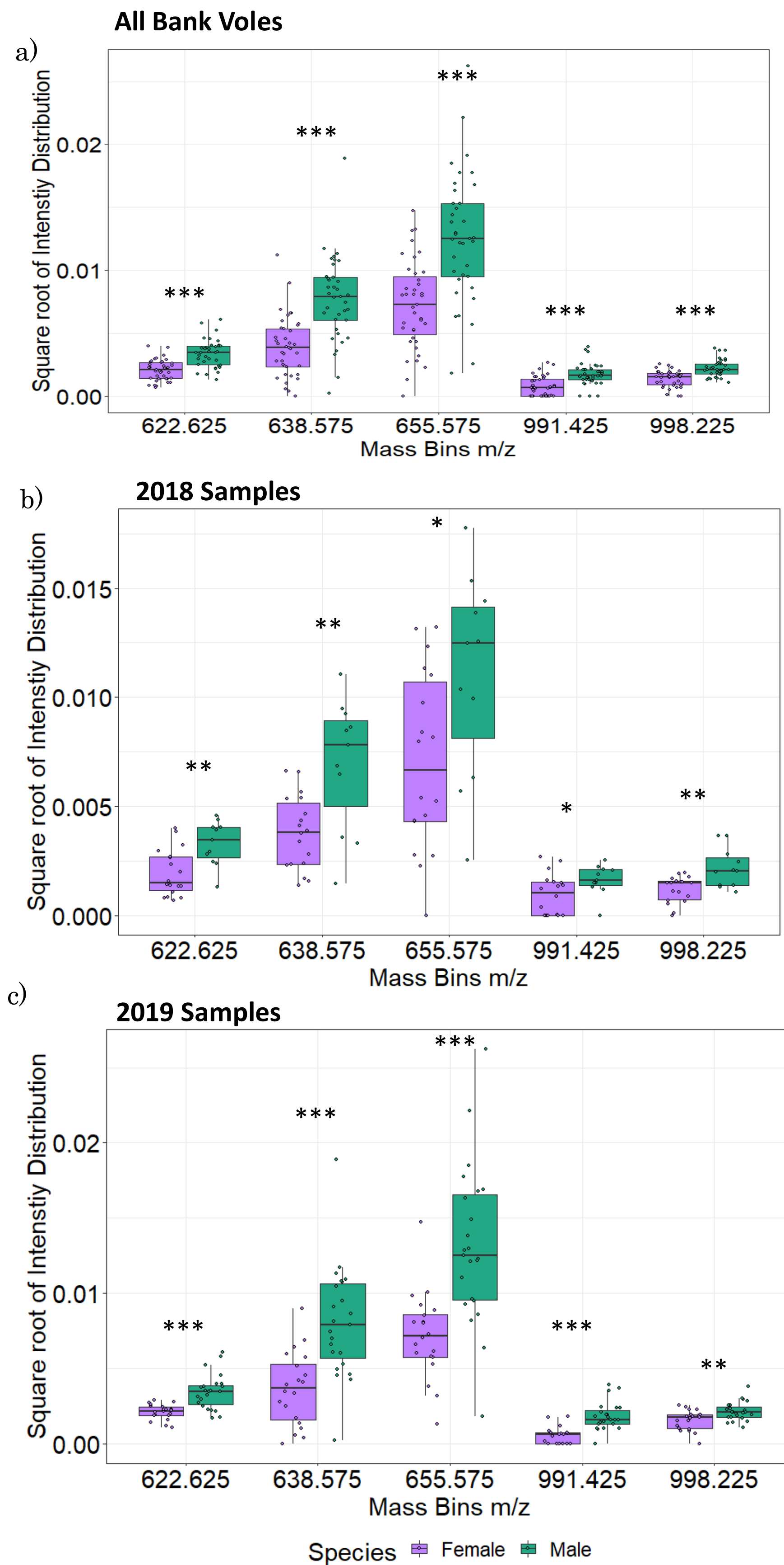


Figure 4.14 A comparison of how five discriminant mass bins of males and female bank voles vary with storage year. a) The relative intensities of the top five most discriminant mass bins of males and females for all bank vole samples. b) The relative intensities bank vole samples collected in 2018 and stored for four years c) The relative intensities of bank vole samples collected in 2019 and stored for three years.

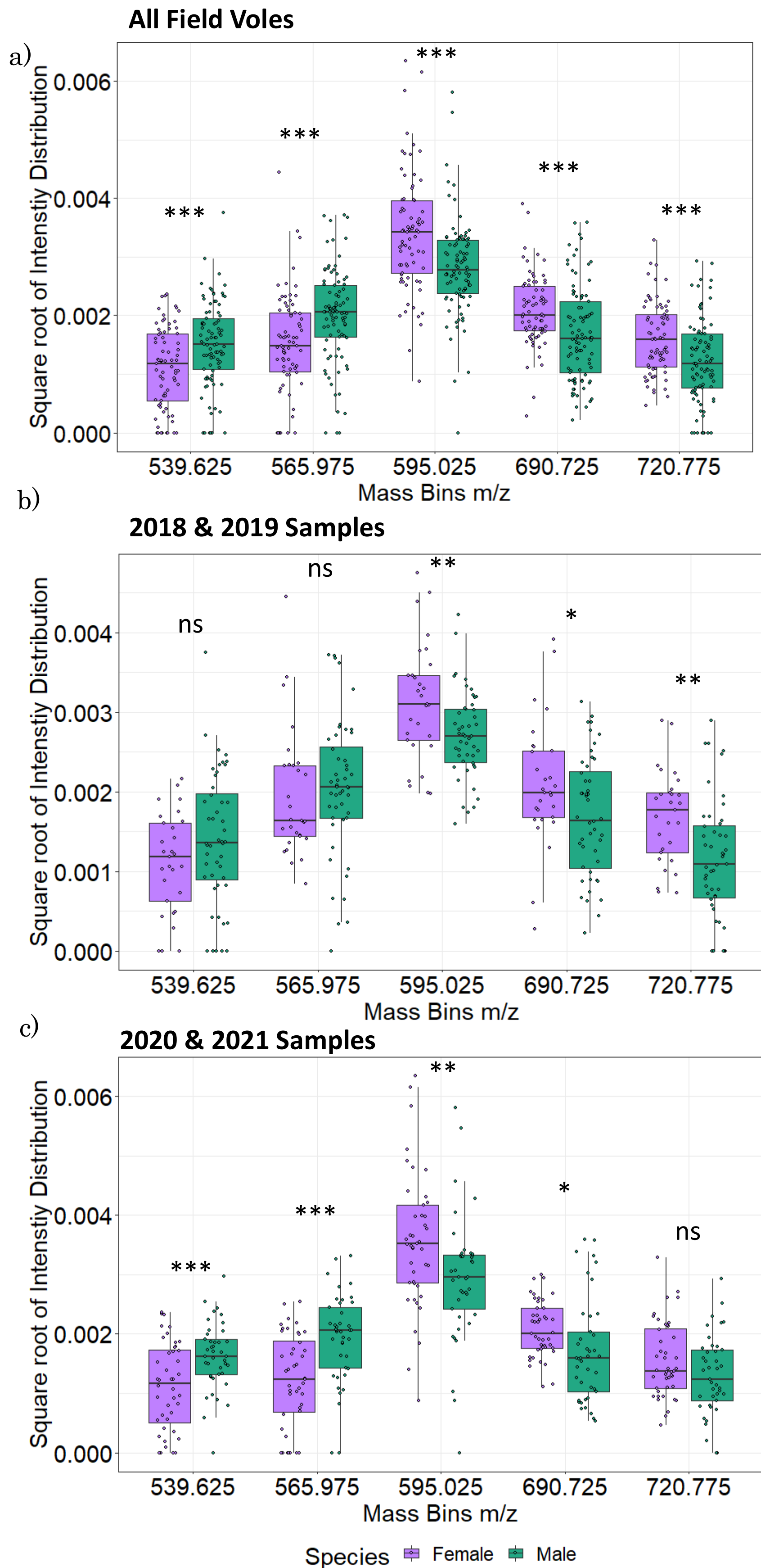


Figure 4.15 A comparison of how five discriminant mass bins of males and female field voles vary with storage year. a) The relative intensities of the top five most discriminant mass bins of male and females for all field vole samples, b) The relative intensities of field vole samples collected in 2018 and 2019 and stored for three to four years c) The relative intensities of field vole samples collected in 2020 and 2021 and stored up to two years.

4.3 Discussion

Random forests could distinguish between collection years, but this did not affect the classification of species or sex of bank and field voles. REIMS could only discriminate sex when the random forest models were built using the top 100 most discriminant mass bins. Storage time did not seem to affect the ability to classify the sex of bank voles, but more analysis with an increased sample number would be beneficial as only two storage years could be compared. REIMS could be used to classify the sex of field voles, but the range of intensities between males and females was closer in the longer stored samples. Bank voles showed a decrease in the sex classification accuracy of males when only using samples from 2018, but the classification of females increased. If the storage time affects sex classification accuracy, it would increase when using samples from 2019. Since this did not occur, the reduction in male accuracy using 2018 samples could be due to the lack of male samples rather than changes caused by storage. To confirm the effect of storage time on sex, the model needed to be repeated with more samples stored over 700 days.

Most of the top ten mass bins responsible for differences between years and sex were within the 600 – 900 m/z range, suggesting that glycerophospholipids (Edward et al., 2017) were responsible for the differences between these factors. The top ten mass bins for species separation were mainly in the fatty acid range <600 m/z. A study classifying types of meat using REIMS also observed that changes to the phospholipid profile caused by age and the animal's diet did not affect species classification (Balog et al., 2016). Fatty acids may be changed more due to storage time than phospholipids.

REIMS was used to determine how the lipidomics profiles changed with the storage time of dried sea cucumber (*Stichopus japonicas*). Sea cucumbers are freeze-dried routinely after being removed from seawater as it helps with transportation, but this can cause a decrease in nutritional value. The oxidation of fatty acids causes the deterioration. Multivariate analysis confirmed the lipidomic profiles (acquired through REIMS) signifcantly changed due to the amount of lipid oxidation caused by increased storage of the dried cucumber (Song et al., 2021). Therefore, the lipid profile of REIMS spectra of faecal samples may also be changing due to storage. The fatty acid profiles of 22 human serum samples did not change after being stored for up to ten years at -80°C (Matthan et al., 2010). The changes in microbiota composition of 24 human faecal samples due to five years in storage at -80°C

were less than the individual variation already present (Tap et al., 2019). The microbiota of human faecal samples that were freeze-dried before being stored at -20°C for 14 years was comparable to fresh samples suggesting microbial profiles are preserved during storage (Kia et al., 2016). Faecal glucocorticoid concentrations of elephants and bears have been found reliable for up to two years at -20°C. Samples that had been freeze-dried and then stored at room temperature were comparable to the control group for the elephant samples but not for the bear samples after two years (Hunt and Wasser, 2003). The spectral fingerprint produced by REIMS may be more stable with time if faecal samples were stored at -80°C rather than -20°C. Freeze-drying may also be an option, but it can be expensive and inaccessible, especially when working in the field. Most of the changes due to storage occurred between years one and two. This work would benefit from using samples stored for up to five or ten years to determine if changes continue with storage time or begins to stabilise.

The differences in intensities between different storage years may not result from storage. Environmental factors also influence the microbiome; the same strain of lab mice from different vendors will have different microbiomes (Hufeldt et al., 2010). Rodents obtained several species of bacteria after moving from the wild to the lab (Bowerman et al., 2021). The microbiomes of rodents kept in the same cage become more similar; it will also be influenced by genetics and diet (Hildebrand et al., 2013). Currently, no studies have been carried out using REIMS to investigate the microbiome, but REIMS has been used to identify seven different *E. coli* strains (Strittmatter et al., 2014) and to classify different *Candida* species (Cameron et al., 2016). Since REIMS can be used to find differences between bacteria, it could be possible to detect changes in the microbiome. Changes in the microbiome due to diets, storage or environmental factors may result in different faecal profiles and therefore REIMS spectra.

Changes to the fatty acid phospholipid profile may cause a change to the REIMS spectral fingerprint, but this may not affect the classification of spectra. Spectral differences caused by phenotypic differences, such as species, may always be more prominent than differences caused by storage. Therefore, the storage time of faecal samples may not have to be considered for all classification studies. Intra-species variations such as sex may be more prone to changes caused by storage as these factors are harder to classify. It was, therefore,

more challenging for random forests to find what bins change due to sex and which change due to storage. When carrying out species classification, storage time should be taken into consideration. Using samples from long-term storage for some factors may be possible, but they should not be mixed with new or short-term stored samples. In other cases, only fresh samples may be suitable.

In conclusion, to investigate species and sex, classification models can be built using samples that have been stored for up to four years and potentially longer. A small sample number affected the classification accuracies more than any other factor. With suitable sample numbers REIMS could be used for long-term monitoring projects.

Chapter 5: Application of REIMS to Species Profiling in Field Studies

5.1 Introduction

Monitoring small mammals is used to track the population of endangered species, measure the effect of climate change and human interference and the impact of small mammals on agriculture (Freeman et al., 2022). The monitoring of bank and field vole populations over six decades indicated the cyclic behaviour of damage to forests in Germany was parallel to the fluctuations in the vole population. Rodent abundance was measured using snap traps and the number caught per 100 trap nights over a few nights each year from 1952 to 2014 in Lower Saxony and 2002 to 2012 in Saxony was calculated. The damage to trees caused by rodents was recorded in May and December each year. The results showed that bank and field vole populations fluctuated every two to three years. This was the first study to confirm the cyclicity of voles in temperate forests. The vole population growth correlated with beech seed intensity; beech trees were the most common type in the surrounding area, and that year's climate influenced the density of beech seeds. The damage caused to the forests in Lower Saxony was in synch with the vole population cycles. An increased population increased the demand for nesting material and food resources. A higher density seed year was not always followed by a population increase in the Saxony forests suggesting that ten years was not long enough to track cyclic fluctuations. This study showed that the use of rodenticides could be reduced in years of lower populations. Trees could be planted in the year after peak population to reduce damage to young trees and increase their chance of survival (Imholt et al., 2017). The disadvantage of the method used in this study to monitor the vole population was lethal snap traps. In the UK, snap traps are not allowed outside due to the risk to non-target species (BPCA, 2018).

Small mammal monitoring is often used in rat control studies; methods for rat control should aim not to affect non-target species (Davidson and Hurst, 2019). Rodenticides were a prevalent method used to target rats in the 20th century, but they caused non-target species to become poisoned, including predators that consumed the contaminated prey. A study in Spain that examined 71 carcasses of great bustards (*Otis tarda*) collected from 1991 to 2010 found that birds with higher concentrations of chlorophacinone (an anticoagulant found in rodenticide) in their liver also had more pathogens and parasites (Lemus et al., 2011). In UK agriculture, wood mice, bank voles and field voles are not considered pest species but may be affected by methods used to target rats, house mice or grey squirrels (Brakes and Smith,

2005). The UK limited the use of certain anticoagulants because they affected non-target species, but this may have increased the resistance of Norway rats to anticoagulants. Farmers in the UK must choose between using older rodenticides that cause less harm to non-target species that rats have developed resistance to or using new toxic ones (Buckle, 2013). The populations of non-target species have been shown to decrease in areas with routine rat control. This could then negatively affect the predators such as owls, weasels and kestrels that feed on the declining rodent populations and increase their risk of secondary poisoning (Brakes and Smith, 2005).

Many small mammals are at risk of extinction; the water vole, hazel dormouse, Orkney vole, harvest mouse and the lesser-white-toothed shrew are all small mammals in the UK listed as near-threatened to endangered by the IUCN red list. The mammal society runs national surveys on the harvest mouse and the water vole (Coomber et al., 2023). Conservation projects can distribute their limited funds more efficiently with more information about population numbers and their distribution. REIMS could be a potential tool for conservationists and for research into rodent control methods. Faecal pellets could be collected from rat bait boxes and analysed using REIMS to determine what species are visiting the boxes. REIMS could also be used as a non-invasive method to determine the species distribution of an area and reduce the use of snap and live trapping.

This study aimed to determine if REIMS could be used to monitor the rodent population by analysing faecal samples collected from the field without trapping animals. This would be achieved by comparing the species distribution of field mice, bank and field voles using traditional live-trapping, REIMS analysis and camera traps. Faecal samples were collected from animals caught in tube traps (active traps) and from traps that remained open (inactive traps). The faeces were analysed using REIMS, and the active trap samples were used to build a random forest model to predict species of samples from the inactive traps. Camera traps were set up to watch the inactive traps so species could be confirmed. However, due to a large amount of camera trap footage, species of pellets collected from inactive traps could not be confirmed. The number of tube and camera traps was chosen based on the number of animals caught in the previous year collecting field voles [Chapter 4]. A slight increase in the rodent population was anticipated as the population gradually increased over the previous two years. However, there was an unanticipated “boom” in population, and the number of

animals entering the traps daily was much higher than expected. It was assumed that the videos from the camera traps would be used to check what animals visited the inactive traps each night and determine if REIMS had correctly identified the species. Based on the previous year and preliminary testing of the camera traps, there were expected to be 600 video events or around five hours of footage. However, the boom in rodent population meant there was a total of 19,000 video events collected over the four weeks, for 158 hours of footage. During the four weeks of fieldwork, 952 pellets were collected from the open traps, but not all of the camera trap footage could not be used to confirm the species classification of every pellet. Footage from one night was monitored for each inactive trap; this was used to establish species distribution via camera traps and could be compared to the species distribution of active and inactive traps. The species distribution was expected to vary for voles and be consistent for mice across the four sites. If there were no significant differences between species distribution between the inactive traps and using the camera, then it would be assumed the predicted species were correct.

Four different field types were used in the study, as rodents have different habitat preferences due to their diets. Bank voles live in woodlands and hedgerows and eat fruit, nuts and small insects. Field voles are found in grasslands and eat seeds, roots and leaves. Field mice are found in both woodland and grasslands as they eat berries and seeds (Flowerdew et al., 2004). The species distribution of each site was compared for each method. If similar results were found between the three methods, it would demonstrate that species can be established by collecting faecal pellets. The efficiency of using the two methods to establish species distribution was also compared to confirm whether the REIMS method is valuable to conservationists or ecologists.

5.2 Method

Fieldwork was carried out over four weeks to establish the species distribution of field mice, field and bank voles using three different methods. Fieldwork occurred at four sites, two at the University of Liverpool's Leahurst Campus and Ness Gardens (Figure 5.1). A total of 40 tube traps were used; they were baited with Harry Hamster, apple slices and bedding. Half of the traps were set as active for a week, and the other half were set as inactive and then were swapped the following week. The live-trapping method used active traps; they would trap an animal inside, the traps were checked daily, and the animal was released after recording the species. The species was identified by observing the animal's features, including ear size, tail length, colouring and face shape (Figure 5.2). Faecal pellets were collected into a single Eppendorf tube and stored at -20°C. The REIMS method involved inactive traps, they would not trap an animal, and therefore multiple animals could enter and leave the trap daily. Faecal pellets were collected from the inactive trap. Each pellet was placed into a tube and analysed as an individual sample. What species produced the sample or how many animals defecated in the tube was unknown. The camera trap method involved setting up cameras in front of the entrances of the inactive traps to record what animals entered the traps (Figure 5.3).

Ten tube traps, five active and five inactive, were placed at each site, but camera traps were not placed in Site IV as there was a limited number. The traps were placed in a transect with ten spaces between each trap, alternating between active and inactive. Four different sites (Site I, Site II, Site III, Site IV) were chosen. Two of the sites chosen had long grass, one was a woodland area, and one hedgerow area was chosen (Figure 5.4).

All faecal pellets were burned with REIMS in a randomised order. Up to five faecal pellets were burned using cut mode on 35 V for each sample collected from active traps. Up to ten faecal pellets were burned one at a time from the inactive trap using the same REIMS settings. The spectra produced were uploaded to LiveID, normalised, lock mass corrected, binned to 0.5 Da, and the mass range was reduced to 400 to 1100 m/z. The exported LiveID data was uploaded to R, and the data from the active traps were used to create a model to predict the species of the faecal samples collected from the inactive traps.

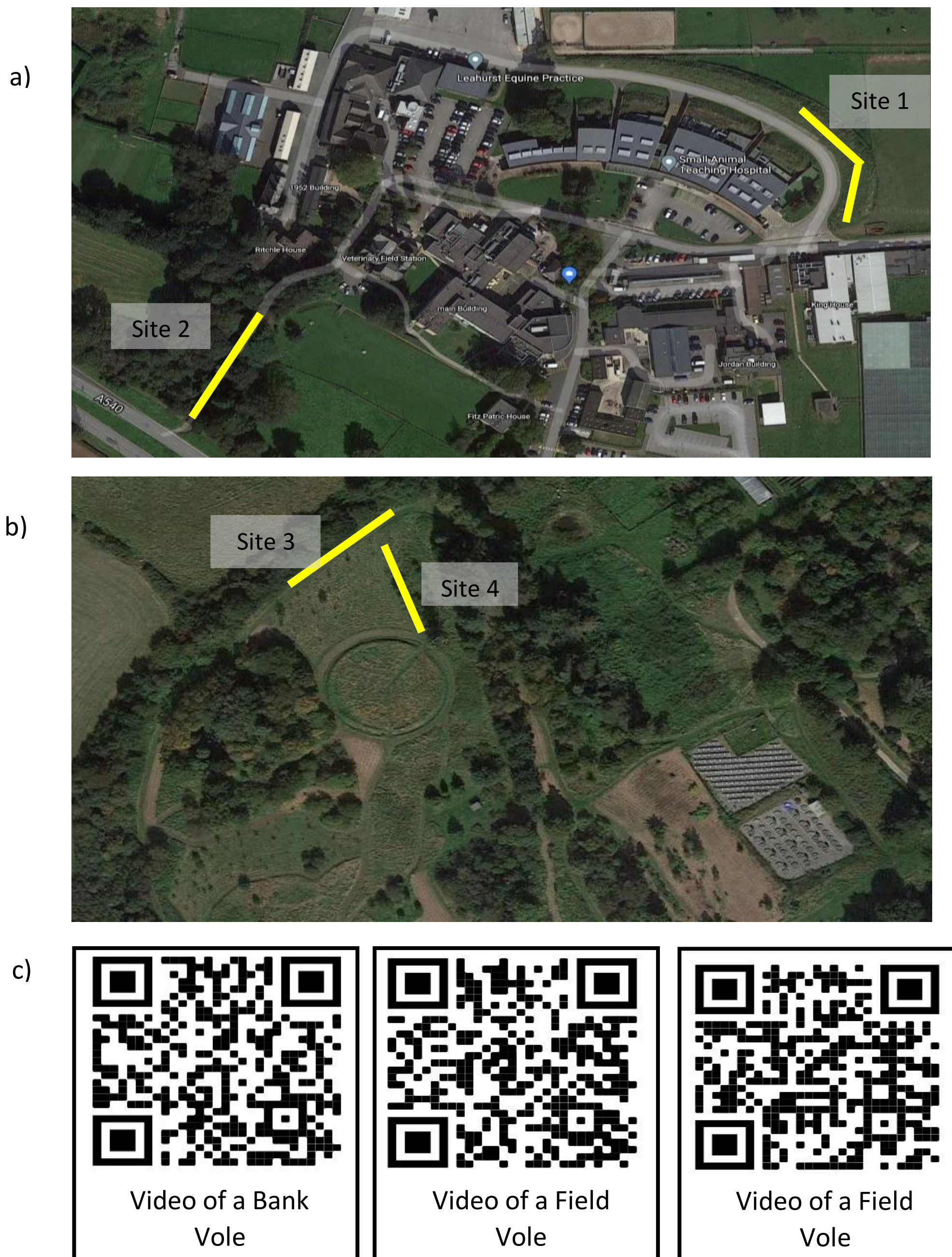
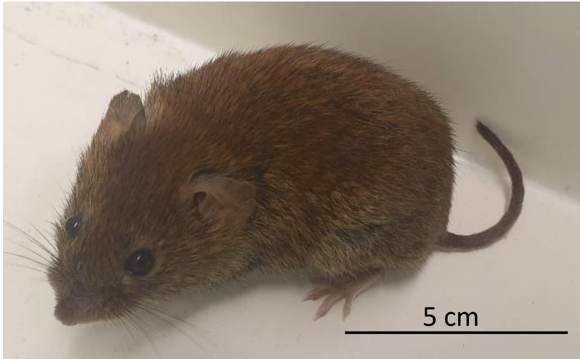


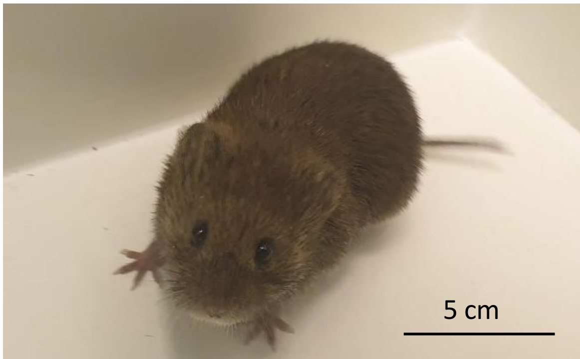
Figure 5.1 The placement of the tube traps, indicated by the yellow lines. Ten traps were placed along each transect with ten spaces between them. a) Two field sites were at the University of Liverpool Leahurst Campus, 10 traps were placed at each site ©Google. Site 1 53°17'24.1"N 3°01'30.3"W, site 2 53°17'22.1"N 3°01'45.3"W b) Two sites with ten traps each were in Ness Gardens, Neston ©Google. Site 3. 53°16'20.8"N 3°02'54.5"W Site 4, 53°16'20.3"N 3°02'54.5"W c) Video of three species of rodents interacting with the tube traps.

Bank Vole



Large ears, long tail (50-80% body length), red-like colouring, pointed face, fur on tail

Field Vole



Flat ears, short tail (30-50% body length), square shaped face, fur on tail

Field Mouse

Very large ears, tail much longer than body length, large eyes, no fur on tail

Figure 5.2. The identifying features of the three species of rodent caught in tube traps. Shrews may also enter the trap but a hole in the trap door allows them to escape. The bank and field voles were captive voles from MBE photographed in the summer 2020. The field mouse was caught during the summer of 2021 and released back into the wild. © Natalie Koch

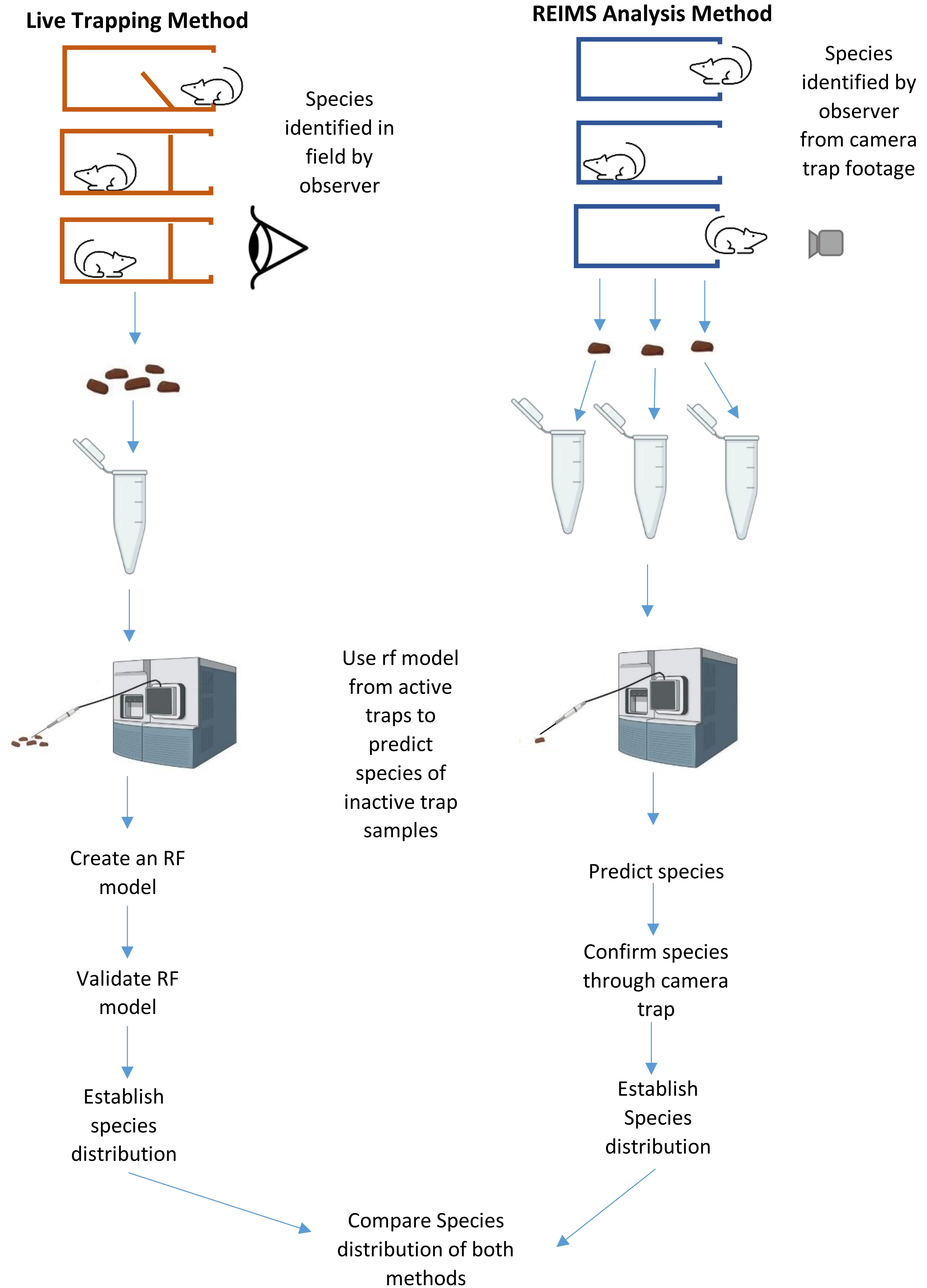
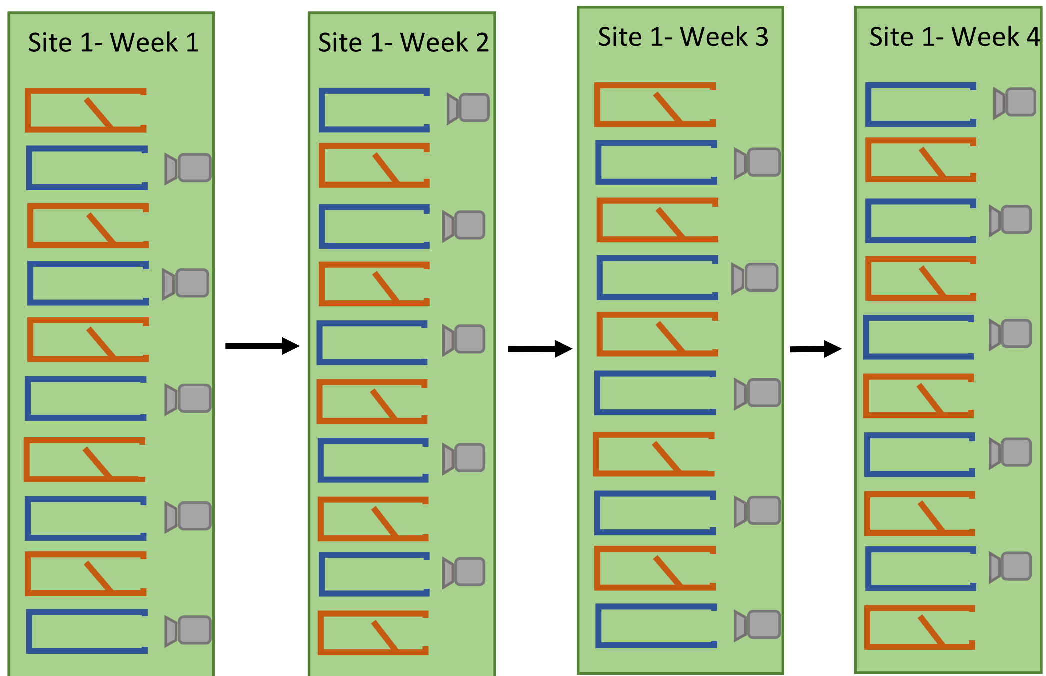


Figure 5.3 Both active and inactive traps were used during the field work. Active traps would trap an animal therefore all pellets in the trap came from the same individuals. Inactive traps would not trap the animal and therefore faecal pellets could be from multiple individuals and species.



Site 1 – Grassland: Ideal for field voles

Site 2 – Woodland: Ideal for bank voles

Site 3 – Hedgerows: Ideal for field voles

Site 4 – Grassland: Ideal for field voles

Figure 5.4. Four different sites were chosen of different habitat types as voles prefer difference areas due to their diets. Five of each trap were placed in a transect, every other trap was set to active and were swapped over each week.

5.3 Results

5.3.1 Determining The Species of Faecal Pellets Collected from Inactive Traps

A random forest built using pellets from the active traps gave an accuracy of 91±2% (training set, 70% of samples) and a prediction accuracy of 86±6% (test set). The intensities of the top five most discriminant mass bins significantly differed between the three species (Figure 5.5). This model was then used to predict the species of the faecal pellets collected from the inactive traps. The species distribution identified by camera traps and REIMS analysis of pellets collected from inactive traps was compared for thirteen traps (Figure 5.6). As expected, there were slight differences between the inactive and camera traps, as not all the animals visiting a trap will leave a faecal pellet. There was only one trap where REIMS identified an individual as a species not observed visiting the trap on the camera. An ANOVA confirmed no significant difference between the distribution of each species between the two methods (Bank vole $p=0.21$, Field voles $p=0.86$ and Field mice $p=0.22$). These results suggest that the random forest model has correctly identified the faecal pellets from the inactive traps.

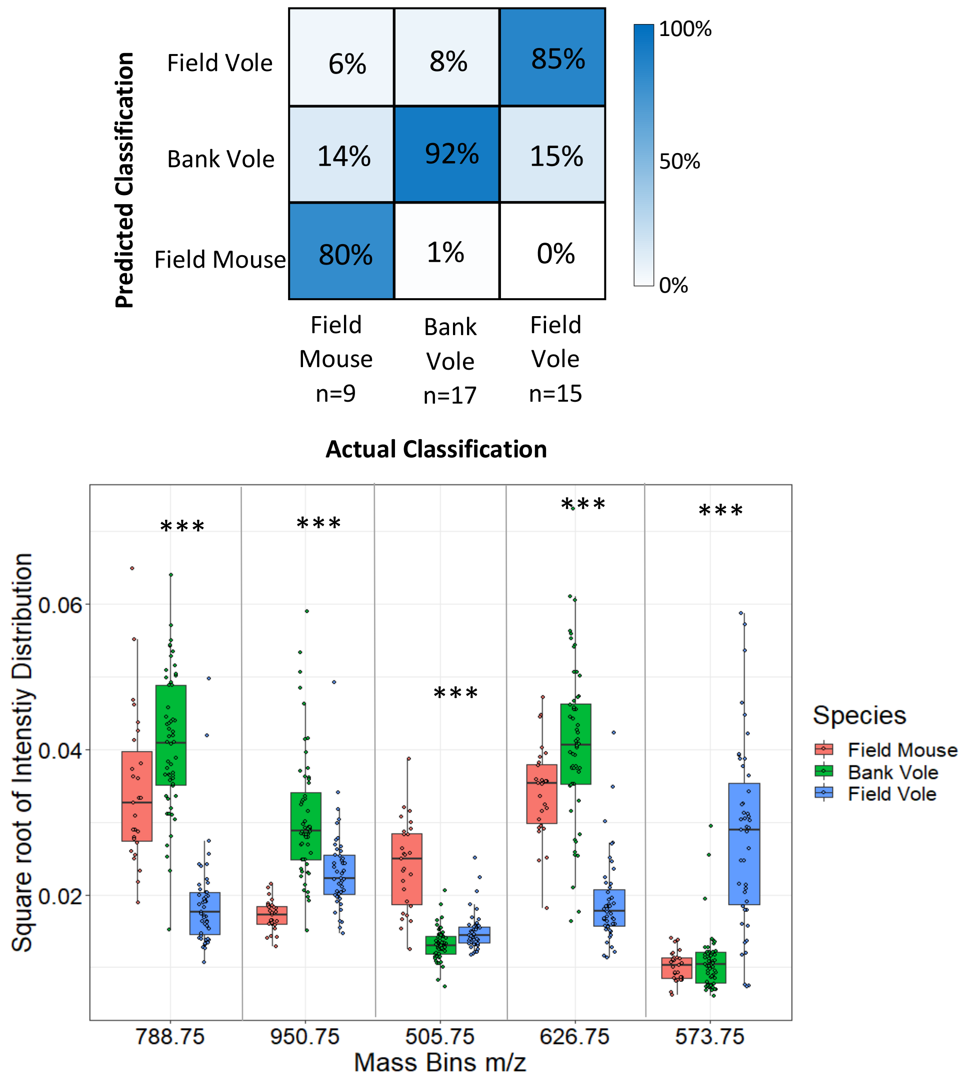


Figure 5.5 Four different sites were chosen of different habitat types as voles prefer difference areas due to their diets. Five of each trap were placed in a transect, every other trap was set to active and were swapped over each week.

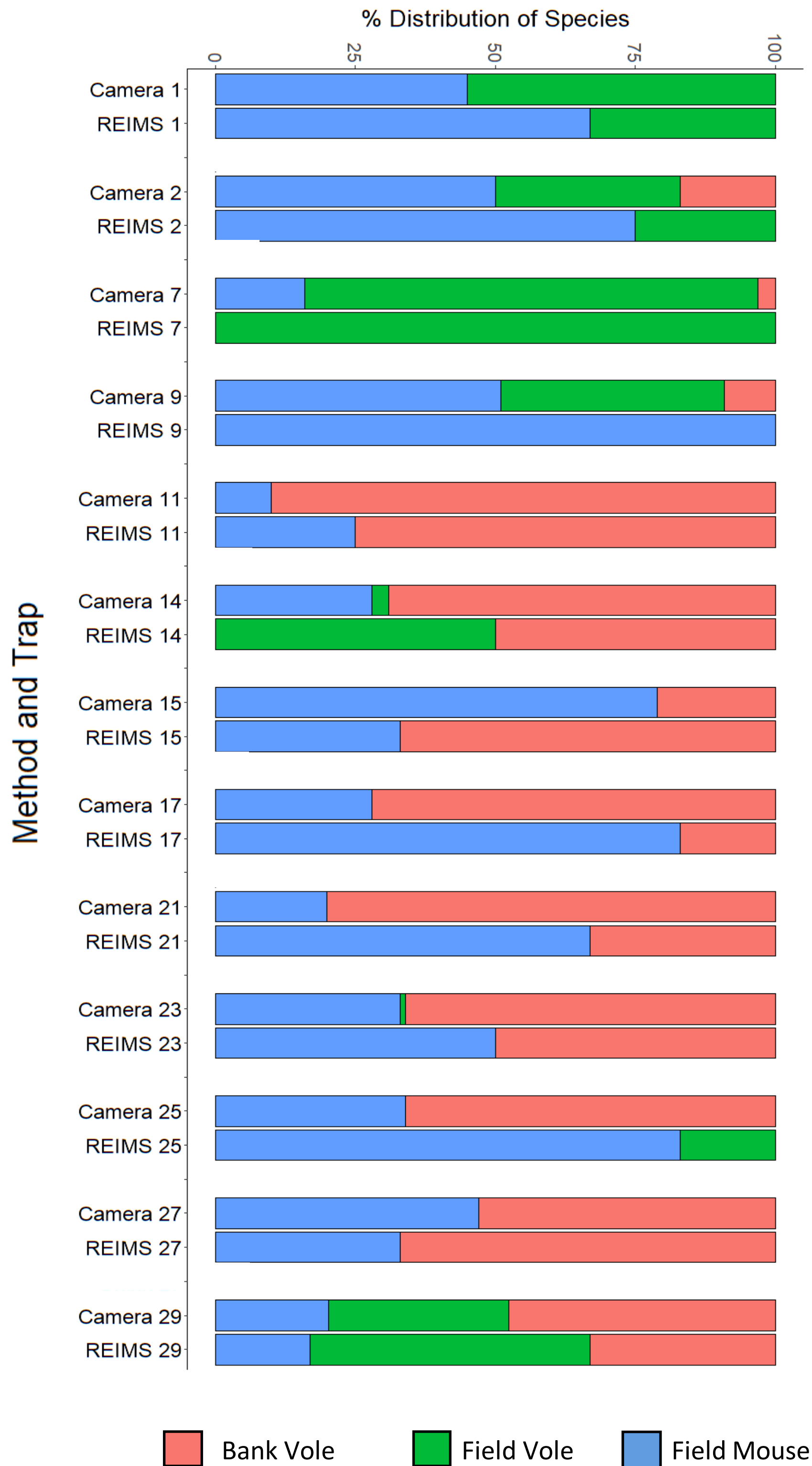


Figure 5.6. The species distribution of samples collected from day one from 13 traps. The camera data is the total number of times any animal was observed on camera. The faecal samples collected from the inactive traps were analysed using REIMS and the species predicted using the random forest built with all samples collected from the active traps. If REIMS was not able to distinguish between samples the data would show an even distribution of all three species. Trap 25 was the only trap that REIMS classified a sample as a species that was not observed on camera.

5.3.2 Species Distribution of Rodents Across Four Different Field Sites

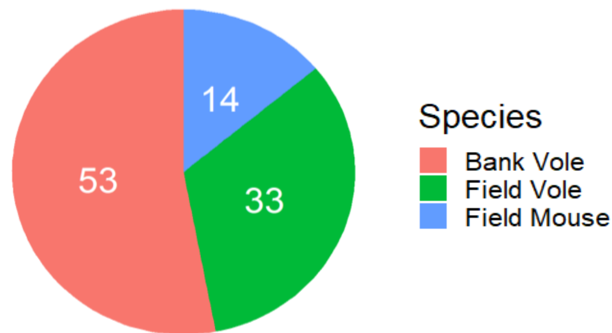
The total number of observed animals is summarised in Table 5.1. There was no significant difference between the species distribution using the REIMS method (inactive traps) and using camera traps (ANOVA $p > 0.19$). There was a significant difference between the species

Table 5.1. The total number of individuals caught in camera traps, observed on camera or faecal pellet classified with REIMS of each species for all four sites. No cameras were placed in site IV.

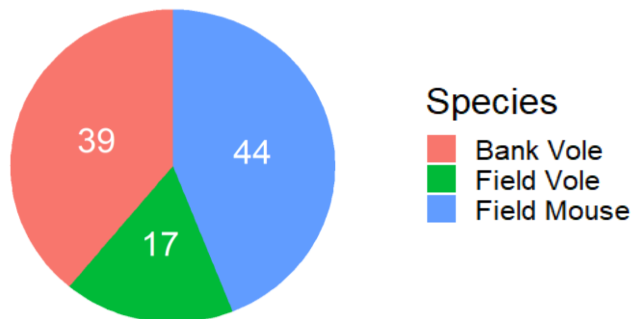
Method	Site	Bank Vole	Field Vole	Field Mouse
Active Trap	All Sites	115	72	30
Camera Trap	All Sites	481	202	535
Inactive Trap	All Sites	309	144	499
Active Trap	Site I	1	48	10
Camera Trap	Site I	18	180	156
Inactive Trap	Site I	20	51	125
Active Trap	Site II	42	2	9
Camera Trap	Site II	185	4	225
Inactive Trap	Site II	89	20	127
Active Trap	Site III	44	0	7
Camera Trap	Site III	278	21	154
Inactive Trap	Site III	135	17	171
Active Trap	Site IV	28	22	2
Inactive Trap	Site IV	65	56	76

distribution using the live trapping method (active traps) compared to cameras (Kruskal-Wallis, $p<0.003$) and inactive traps (Kruskal-Wallis, $p<0.0009$) (Figure 5.7). The distribution of species varied throughout the 24 hours of camera trap footage that was analysed, only bank voles were active during the day, and field voles became active before field mice (Figure 5.8). The camera and inactive traps analysis showed that field mice were the most abundant, followed by bank and field voles. The results for the active traps suggested that field mice were the least abundant; it did suggest that there were more bank voles than field voles, as with the other two methods. The species distribution of voles was as expected for three sites,

Species Distribution (%) for live trapping (Active Traps)



Species Distribution (%) for Camera Traps



Species Distribution (%) for REIMS analysis (Inactive Traps)

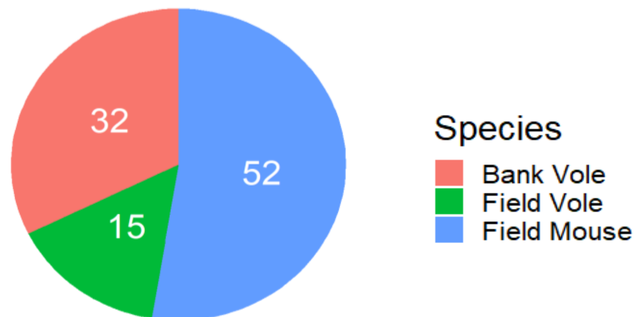


Figure 5.7. The total species distribution across all four sites and for all four weeks. The tube trap samples were the animals that were caught in the closed traps and identified through observation. The camera traps samples was the total number of times an animal of each species appeared in the camera trap footage over a 24 h period. There was a total of 15 cameras places in front of the open traps in sites 1, 2 and 3, the animals were identified through observation. The REIMS analysis samples were the faecal samples collected from the open traps over four weeks from four sites. The samples were analysed with REIMS and random forest model was used to predict the species.

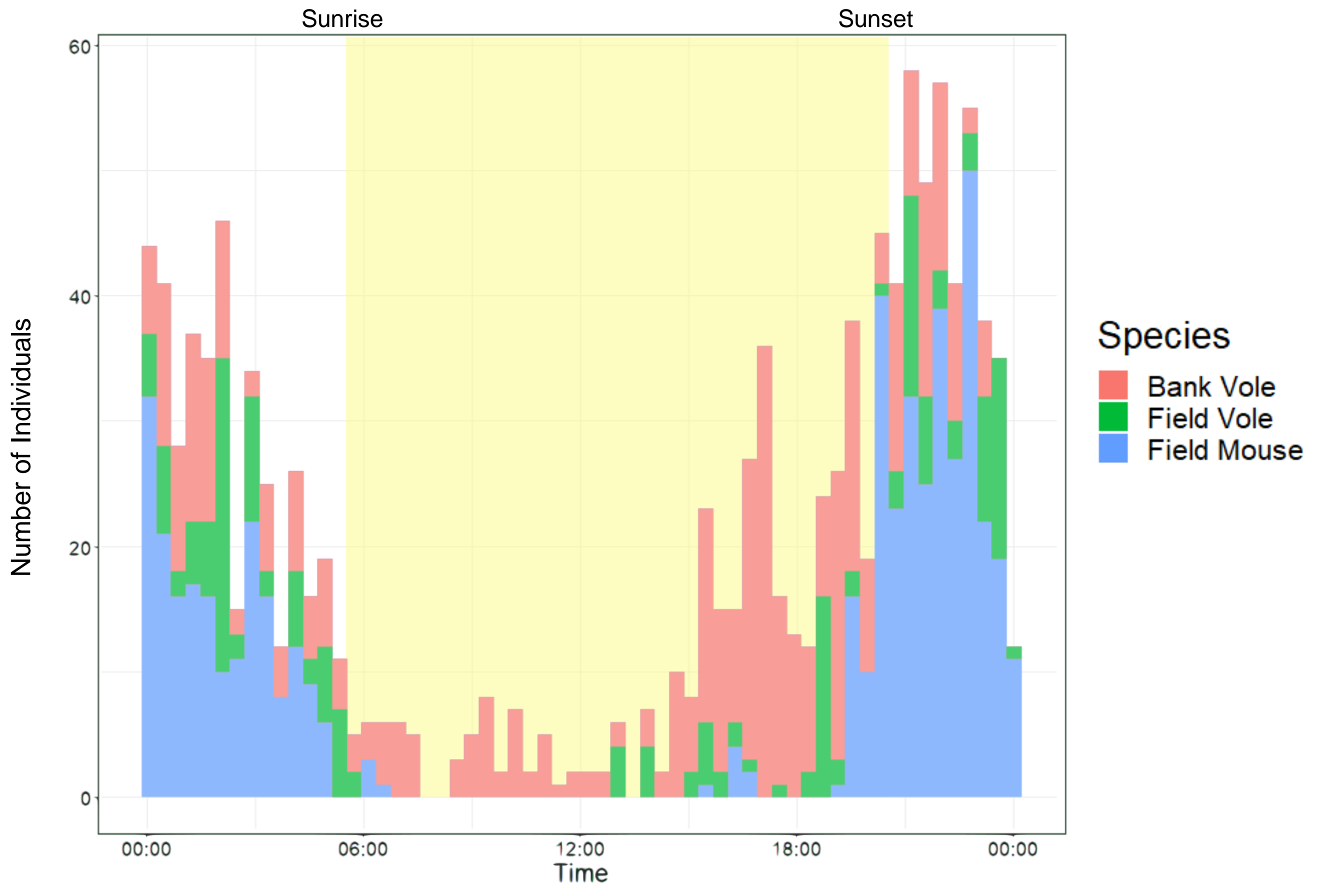


Figure 5.8. The number of individuals of each species that interacted with the open tube trap at a certain time from the camera trap footage (n=15). Field Mice were almost exclusively observed at night. Bank voles were observed more often at night but would also visit traps during the day. Field voles did not visit the traps during the day as much as bank voles but were observed earlier than field mice. Active and inactive traps were checked between 11 am and 6 pm.

but site IV had more bank voles than expected. The species distribution also varied between methods used for each site (Figure 5.9). The active traps showed fewer field mice for all four sites than the other two methods. In site III, no field voles were trapped in the active traps, but a small number was observed on the cameras and through REIMS analysis (inactive traps). The species distribution for the camera traps and REIMS analysis (inactive traps) were more like each other than compared live trapping (active traps) distribution.

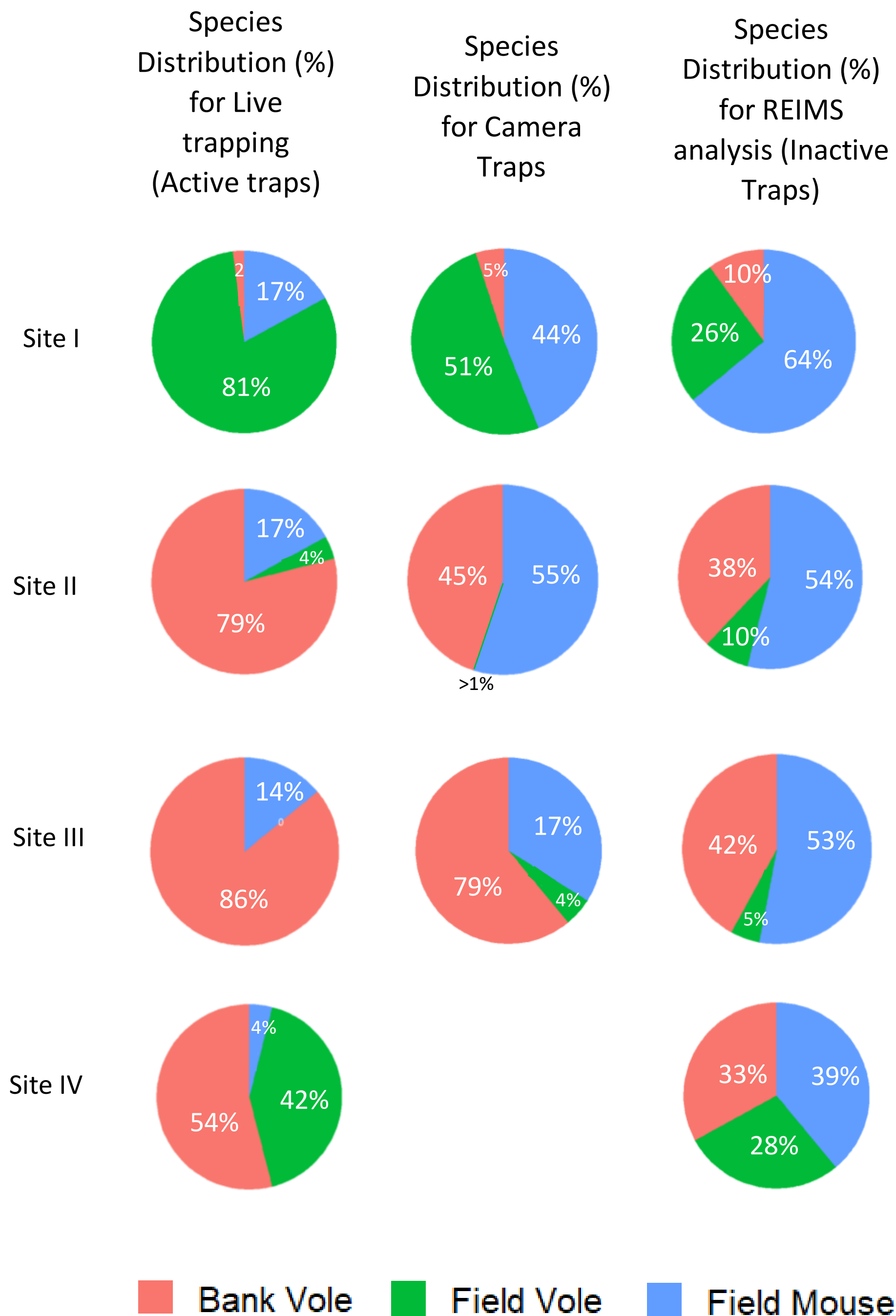


Figure 5.9. The total species distribution for each site from all four weeks, no cameras were placed in site 4. The species distribution varied for each site habitat type varied and suited different species. Sites one and four were grassland and preferred by field voles. Sites Two was a woodland and sites three was hedgerows both were preferred by bank voles. The active traps had a smaller number of field mice compared to the cameras and inactive traps. Site four had a higher number of bank voles than expected

5.4 Discussion

5.4.1 Comparison of the Three Different Methods Used to Determine Species Distribution

5.4.1.2 Live-Trapping (Active Traps)

The advantage of using active traps is observing the individual closely. It is the best method for determining other attributes about the animals, such as sex, health, pregnancy or lactating status. Mark-recapture methods can be used with live trapping to track an individual and determine their home range. The benefit of live trapping is that all the data was collected on the day of trapping; no additional analysis time was required, such as watching videos or lab work needed to establish species distribution. There are fewer costs; trapping requires the up-front cost of the traps, bait and bedding. Disadvantages include detrimental effects on the animal; this includes handling the animal (Gelling et al., 2009) and keeping the animal in the trap overnight (Fletcher and Boonstra, 2006). A study trapping bank voles and field mice found traditional handling methods caused a reduction in the animal's immune response compared to animals that were trapped but not handled (Gelling et al., 2009). Corticosterone concentrations increased in meadow voles caught in Longworth traps (Fletcher and Boonstra, 2006). The temperatures were relatively high during the fieldwork period, and covers made from grass were placed over any exposed traps to help keep the temperatures low. Apples were placed in the traps to help with hydration, but there were many incidents when the caught animal did not seem to eat them, which could have caused the animal to become dehydrated. However, it could be that the animals only eat apples when they are dehydrated. Using live trapping means animal identification depends on the person checking the traps. Therefore the species classification accuracy is reliant on the ability of the person to determine the differences between species. The species distribution of the tube traps greatly underestimated the number of field mice in the area. This is because field mice are nocturnal, whereas the voles were observed on the camera traps visiting traps throughout the day (Figure 5.8). Voles are, therefore, more likely to visit the trap first and set it off, suggesting there are more voles in the area than mice.

5.4.1.2 Camera Traps

The advantage of camera traps was the potential to obtain behavioural information and observe other species, including shrews and hedgehogs. How the animals used the traps

varied between species, bank voles would quickly enter the trap, take food and leave the trap area. Mice and field voles would take more time interacting with the trap before entering. Mice and field voles were more likely to spend time in the trap or take food from the trap and sit outside the trap for long periods. The advantage of cameras over live trapping is the information can be stored and investigated later for different research purposes. The camera traps also produced the most data, although this meant there was too much to be processed. Camera traps were able to observe every animal that visited the trap, unlike tube traps which only capture the first animal to visit, and REIMS can only identify animals that leave faecal pellets. Camera traps are non-invasive; there is no interaction with the animal, and the animal is not contained for any time. Camera traps also required the least amount of time in the field, the batteries were changed in the camera each day, but this took less time than collecting pellets from traps or handling animals. The camera traps' disadvantage was the time required to analyse the data. So much video was collected that data was only collected from one night for each trap. The amount of footage collected was higher than expected due to the population boom, so in some studies, too much footage may not be an issue. If it was known that there was such a high population, fewer cameras could have been used, and cameras could have been used in site IV. There were incidents of data being unusable from the camera due to the camera view being obstructed, usually due to bad weather. Heavy rain obscured the footage, so it could not determine what animals were present. It was difficult to identify species that visited during dusk when there was low vision due to lack of sunlight, but the camera had not yet switched to night vision. Species identification could be difficult even when the camera was not obscured. The camera trap method relies on human identification and is much harder to determine species than live trapping. Identifying features such as ears and tails may be challenging to see, and most of the videos were in black and white, so colour could not be used to determine between bank and field voles. The animals would move so fast that the images were blurred when stopping the video for identification. The species identification using camera traps likely had the lowest classification accuracy. To confirm this, other people would need to watch all the trap videos and compare results. The data from the cameras were based on the number of videos of a species, but likely many videos were of the same individual. It was probable that all three species would return to the trap, so the species distribution of all videos is representative of the number of individuals.

5.4.1.3 REIMS Analysis of Pellets from Inactive Traps

The advantage of using REIMS for population monitoring was identifying multiple animals in a trap, not just the first to visit. Out of the 20 inactive traps over the four weeks, samples could be collected 280 times. Samples were collected from tubes 227 times, with a successful collection rate of 81%. This was higher than the rate at which individuals were caught in the active traps; the active traps had a collection rate of 78%. Collecting faecal pellets from the tubes took about the same time as handling animals from closed traps. Unlike live trapping, the traps are not checked twice a day. Active tube traps need to be checked regularly to release any enclosed animals. Collecting faecal pellets means stored biological samples can be used for further testing, such as DNA analysis or providing diet information. The REIMS method is also non-invasive and does not rely on human observation for species identification. The random forest classification accuracy of samples from active traps was 92%. Therefore model likely identified 92% of the inactive trap samples as the correct species. The classification accuracy of the other methods will depend on the individual researchers' ability to distinguish between species. Even for a researcher with a high ability to classify rodent species, it is unlikely that the classification accuracy when using camera traps would be higher than 92% due to the number of videos that do not show a clear image. The disadvantage of REIMS was the extra analysis steps required; the burning of the samples took more time compared to live trapping. Access to a mass spectrometer was required; mass spectrometers are very expensive and inaccessible to some research groups. The cost of running samples on a mass spectrometer is relatively low. A known model is required to identify the unknown samples, which must be burned simultaneously with the unknown samples. However, the known samples do not need to be collected simultaneously with the unknown samples. As shown in chapter four, stored samples could predict the unknown samples. Despite the potential extra costs and analysis required, there are many advantages of using REIMS compared to the other two methods (Table 5.2)

Table 5.2 The advantages and disadvantages of using each method to establish species distribution of rodents. The REIMS analysis of faecal pellets collected from inactive traps had the most advantages, but extra costs may limit its use.

	Active Traps	Camera Traps	Inactive Traps
Accurate representation of species distribution	x	✓	✓
Non-invasive	x	✓	✓
Identification by observation	✓	✓	x
Additional analysis required	x	✓	✓
Additional costs	x	x	✓
Allowed for multiple sampling	x	✓	✓
Time-consuming	x (least)	✓ (Most)	✓

Other methods could be used to collect faeces besides leaving tube traps open. An undergraduate student explored other methods of collecting faeces and found that using an upside-down clear food container on top of a piece of wood with bait was sufficient for collecting faecal samples (Figure 5.10). Further work could be carried out to determine if REIMS could analyse pellets found by chance in the wild, as they would be a random sampling method that is not biased due to using bait. Previous large-scale monitoring programmes have relied on volunteers using live trapping techniques (Flowerdew et al., 2004). Using REIMS analysis would vastly reduce the amount of live trapping required, and it would be easier to obtain volunteers as they would not need to know how to identify species.

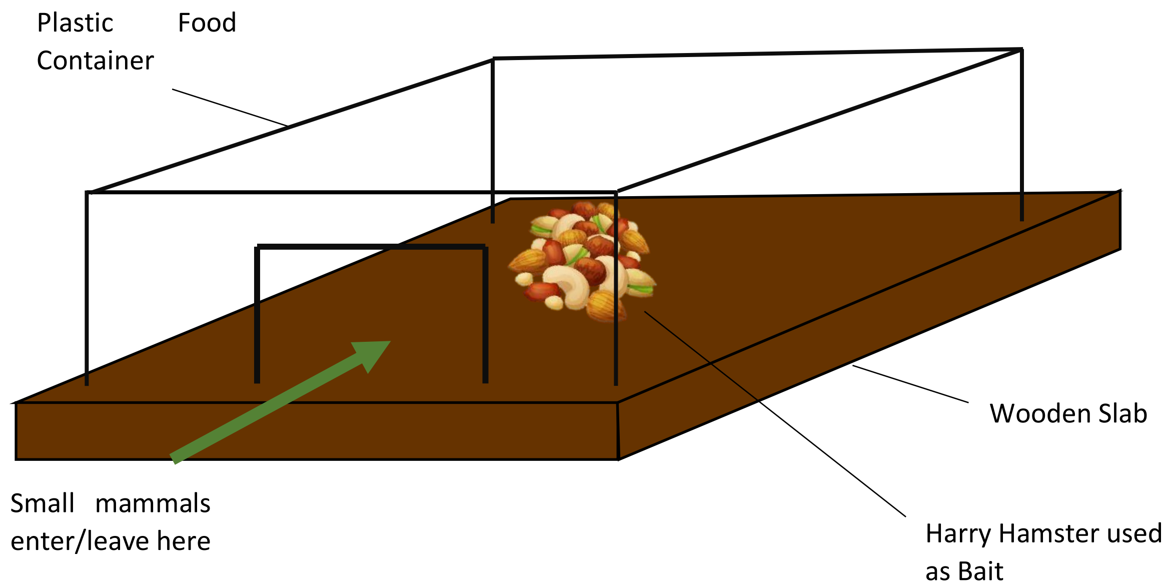


Figure 5.10. An alternative method to collecting faecal pellets rather than using traps. A plastic containers is attached to a piece of wood with food placed inside. The container can be lifted to remove faecal pellets. Animals can enter and leave, since the trap is clear the animal can see there is no risk of being trapped. These traps could be checked less often than traditional traps since there is no risk of trapping an animal. It is also easier to obtain the faecal samples when no bedding is required.

5.4.2 The Affect of the Population Boom

The identification of REIMS samples could not be confirmed using camera traps due to the population boom. The population increase was likely due to the COVID-19 pandemic and the relatively mild winter. Due to the pandemic, the grass and hedges were not cut back as usual. Rodents prefer enclosed spaces (Jensen et al., 2002), so the rodents may move to other areas when the grass is cut during the summer. Not cutting the grass would also lead to more food during the year than in pre-pandemic winters. Rodents only breed in the summer, and the population usually declines over winter. With a mild winter, more individuals are likely to survive to breed, and they are more likely to breed earlier (Kaikusalo, 1972). Site IV may have had more bank voles than expected because of the increased rodent population. Although site four was mainly grassland, it was between two woodland areas. Bank voles were more likely to be in the grassy areas to move between the two patches of woodland or because they had to create home ranges in these areas as there was no more room in the woods. The home ranges of voles do not usually overlap, although large males may tolerate younger males. Larger, more dominant individuals may force others to move into less preferable areas

when resources are scarce (Lin and Batzli, 2004). It is rare for such a high trapping rate. The highest daily trap rate the previous year was only 31%. This high trap rate probably contributed to the underestimation of field mice, as other studies have shown camera trapping to be comparable to live trapping (De Bondi et al., 2010).

The population boom meant there was too much camera trap data to process by one person. Advances in deep learning algorithms will allow computers to watch and record data from camera traps. Deep neural networks (DNN) were used to extract information from camera trap images from the Snapshot Serengeti dataset. The model was trained to detect if an image contained an animal, the species, the number of animals and if they performed one of six behaviours. Any images with more than one species were removed from the dataset. The DDN was unsuccessful at identifying rare classifications, including the presence of offspring, which only occurred in 1.8% of images (Norouzzadeh et al., 2018). Most of the literature and current projects (Carl et al., 2020, Loos et al., 2018) attempting to improve the automatic detection of species from camera trap videos are built to detect large mammals. Studies are being used in lab settings to track small animals (Bains et al., 2016, Singh et al., 2019), but it is unlikely that deep learning will be able to determine the differences between wild bank and field voles for some time yet.

An issue with any trapping method is that it is not considered random due to neo-phobia. Neo-phobia is the fear of novel items; therefore, small mammals may avoid traps. This could cause biased sampling based on health and age, as individuals needing food will take more significant risks and enter the trap. Tube traps may enforce neo-phobia, and animals may learn to avoid traps (Stryjek et al., 2019). Due to the population boom, it is unlikely that neo-phobia was causing a sample bias towards unhealthy animals, as most animals would lack resources in high-density populations. Other studies have shown that the number of times an individual rodent is willing to enter a trap can vary throughout a trapping session (Brehm and Mortelliti, 2018). The live trapping in this study was biased towards voles as they arrived at traps earlier. Collecting faeces through open traps is also less likely to increase neo-phobia as the animals would not associate the open traps with a negative experience.

5.4.3 Conclusion

The relative distribution was different for each species depending on the method used. Camera traps could identify nearly every individual that visited the trap; REIMS could only establish individuals in the trap long enough to defecate, and live trapping could only identify the first animal to find the trap. More samples were collected using the REIMS method than the tube traps. Camera traps recorded the most data, but until automated detection technology and algorithms are developed to the point of widespread usage, they are not the most practical choice for small mammal monitoring. REIMS was a more straightforward method to implement than camera traps, although this was due to the population boom. Unlike live trapping, it did not underestimate mice and is better for the animal's welfare. Although some live trapping would be required to build a random forest to predict the unknown samples, only a small number of known samples would be required to predict a large number of unknown samples. REIMS may be best utilised for large-scale, long-term monitoring projects. These projects would be more cost-effective as only one mass spectrometer would be required. Long-term projects are usually carried out in collection blocks rather than continuously each year. A few samples collected via live trapping during each block could be used to predict the species of all other faecal pellets collected. Further work would need to be required to determine the effect of location on prediction accuracy. But a training model with samples from all locations may be able to predict samples from all locations, or a training model would be required for each location. If a model of multiple locations could be used, REIMS could conduct monitoring studies in multiple countries with all samples sent back to one location for REIMS analysis. In conclusion, REIMS can be used to monitor wild populations of rodents as results are comparable to traditional methods such as live trapping and camera traps.

5.5 Ethical Statement

Fieldwork was carried out in accordance with international best practice guidelines, handling of animals was minimal, and animals were released at the site of capture. The fieldwork did not involve pain, suffering or lasting harm, and no specific licences were required to carry out the work.

Chapter 6: Discussion

6.1 Key Findings

REIMS can analyse faecal samples for population monitoring. Faecal pellets can be burned using a diathermy electrode to produce an 'aerosol' drawn into a TOF mass spectrometer, producing a spectral fingerprint unique to the sample. Machine learning or dimension-reducing analytic methods can be applied to discriminate the samples by various classification groups. The ease of discrimination depends on the factor being analysed, the classification accuracy of species was higher than differences within a species, and differences within an individual were the hardest to classify. Classification accuracy improved dramatically with sample numbers, especially for classification factors with more subtle differences and classes that required a balanced number of samples. The results suggest that analysis can work with samples that have been stored. Therefore for ongoing projects, samples can be stored in a freezer until an adequate number of samples are acquired. Sample preparation is an essential factor to consider before REIMS analysis for herbivores. REIMS could discriminate between whole pellets from two okapis but could not discriminate between homogenised pellets of two individuals. Okapis have particularly inefficient digestive systems, and the method used to prepare samples for storage may have reduced faecal content and increased foliage content. The ability to discriminant between individual herbivores is promising and suggests that REIMS analysis could be used to analyse faecal samples of other large mammals. The results suggest that the spectral fingerprints are influenced slightly by the time the samples are kept in storage but that these changes do not affect the ability to classify samples by species or sex.

The principles of the 3Rs (Replacement, Reduction and Refinement) were proposed in 1959 and suggest all scientists carrying out research with live animals should find ways of implementing the 3Rs to reduce the suffering of animals (Stratton and Burch, 1959). REIMS has the potential to refine experiments by providing an option to collect information about a species through the non-invasive collection of their faeces. The results using REIMS to discriminant species from wild samples confirmed that faecal pellets can be collected from small mammals without live-trapping. The method of analysing pellets collected from inactive traps was more accurate in species distribution than live trapping. As well as being better for animal welfare, collecting faecal pellets was more manageable and efficient than live trapping in the field. Although time is required to burn and analyse the pellets after collection, using

REIMS still provides many benefits. The time required to burn the pellets was less than required to analyse camera trap footage from the same collection period. The key findings are summarised in Table 6.1.

Table 6.1. A summary of the adavtages and disdavantages of each Factor REIMS was used to investigate in this study

Factor analysed by REIMS	Advantage	Disadvantage	Conclusion
Distinguish Sex	Classification accuracy of over 80% was achieved when using adult samples	Classification Accuracy was less than species Required a large number of samples	REIMS could be used to distinguish the sex of adult samples
Distinguish Age	Significant differences were observed between adult and juvenile samples for some mass bins	Could not obtain a classification accuracy of 80% for adults and juveniles	Could potentially classify age for animals with a longer life span as some differences observed
Distinguish Strain	Could distinguish between some strains of lab mice with over 80% accuracy	Could not distinguish between the more closely related strains	The classification was only achieved when class number was balanced. The results suggest REIMS could distinguish between sub-species
Distinguish Pregnancy	Significant differences were observed between pregnant and non-pregnant samples for some mass bins	Could not be used to determine pregnancy of long-termed stored homogenised samples	Without further investigation, it cannot be determined if REIMS is not able to distinguish pregnancy in any okapi samples or just not in archived homogenised samples
Faecal samples of ruminants	It could be used to determine between individuals when using whole pellets	It could not be used to determine individuals when using archived homogenised samples	REIMS can be used to analyse faecal pellets of ruminants, but more work is needed to investigate the impact of sample preparation and storage
Distinguish Individuals	REIMS could be used to determine between individuals when using whole pellets	Only two individuals of one species were compared	The study should be repeated with other species and more individuals to confirm that REIMS can be used to identify individuals
Archived Samples	Older samples could be used to make a predictive model to determine the species and sex of lab mice and voles of newer samples	Cleaning the STEP wave meant samples analysed before the cleaning had different relative intensities than those analysed after. The archived okapi samples could not be distinguished by pregnancy or individual	Archived samples may need to be considered on a case-by-case basis. The REIMS instrument needs to be monitored for changes, and spectra may need to be adjusted when working with archived samples. More work is needed to determine if sample preparation had more of an effect than storage length
Samples from the Field	REIMS could be used to determine the species of samples collected from the wild. It is less invasive than live trapping.	More time-consuming and expensive than live trapping alone.	REIMS analysis could be used as an additional non-invasive tool for conservationists and ecologists

6.2 Further Work

It would be a benefit if REIMS could detect rare species in the field as it can be easier to collect faeces of rare animals than to observe them. The hazel dormouse is endangered in the UK, and a national reintroduction programme is needed to help the species recover. The dormouse monitoring is challenging using live trapping as they primarily live in the canopy of woodland. Monitoring of the dormouse is vital to why some reintroduction schemes have worked while others have not. The UK National Dormouse Monitoring Programme (NDMP) have been using nest box recording schemes organised by volunteers since 1988. Nest boxes are set up at various sites, and volunteers check the nest boxes around once a month. The number of nest boxes provided has varied between 10 and 500. A review of the nest box recordings found a population decline of 72% between 1993 and 2014 (Goodwin et al., 2017). Although nest boxes help monitor long-term population trends, they can underestimate a population by two-thirds during short-term surveys compared to live trapping methods (Vogel et al., 2012). Another method of monitoring dormice is to use footprint tunnels. Footprint tunnels are long plastic open-ended tubes with a sheet of white cardboard inside. Ink is placed at the tunnel entrance, and as an animal walks through, they leave footprints on the white cardboard; the footprints can be used to identify the animal (Melcore et al., 2020). Tunnel footprints are highly cost-effective for monitoring small mammals but rely on an excellent clear footprint for identification. If too many animals pass through the tube, or the same animal moves back and forth through the tunnel, it can be complicated to distinguish footprints. REIMS could be used along with these methods; faecal pellets could be collected from nest boxes and footprint tunnels to identify animals. Faecal samples could identify what animals have been in the nest boxes, even if they were empty at the time of collection or if footprints are unclear. Faecal samples have already been collected from wild dormice using nest boxes to be analysed using PCR to evaluate their diets (Chanin et al., 2015). Samples from nest boxes with dormice present could be used to build the random forest model needed to predict the species of samples collected from empty boxes or tunnels. Samples could also be collected from dormice that are quarantined before being reintroduced. REIMS could also be used to identify faecal samples found opportunistically.

More work is required to establish the benefit of using REIMS in zoo and conservation research. The results from this study suggested that REIMS could be used to discriminant

individuals from whole faecal pellets from herbivores. It would be beneficial to determine if REIMS could be used to analyse faeces from other species. Being able to detect pregnancy earlier than they can with hormone assays could still possibly be achieved using REIMS. This investigation in this study was limited due to COVID restraints; only samples from one species were analysed and stored for a significant time. The samples used for pregnancy were also prepared differently than those used for individual discrimination, which may have had an effect. Therefore, repeating the pregnancy study, using whole pellets and samples collected simultaneously from multiple individuals, would be beneficial. This would mean obtaining samples from more than one zoo. It may be easier to begin with a different species, such as zebras, as there are more of them in zoos than okapis. A preliminary study could be carried out using samples from lab mice already being used for breeding. Collecting samples from multiple individuals without storing them for a long time would be easier. If pregnancy can be discriminated in the lab mice but not in zoo animals, it would suggest that herbivore pellets were too difficult to analyse. It may also be beneficial to investigate species of herbivores as species was the easiest factor for REIMS to discriminate. If REIMS cannot be used to identify species of some herbivores, then it is unlikely REIMS could analyse any factor of herbivores. REIMS should be able to discriminate species of herbivores as it could discriminate between two okapi individuals. If REIMS can be used to investigate large mammals in captivity, it could then be used to analyse samples of wild animals. This would work best if samples from captive animals could be used to build the random forest models to predict phenotypes of wild-collected samples.

One of the most significant benefits REIMS could bring to zoos would be to detect sex in the excrement (guano) of birds. Identification of sex through observation can be very difficult and almost impossible for some adult species and all chicks. It is vital in zoos to determine sex early as chicks of a different sex may have different husbandry needs. If the bird is to be relocated for breeding purposes, it will allow the animal more time to settle in if it is moved as soon as possible. It would also increase the accuracy of matching breeding pairs (Griffiths, 2000). Sex identification of chicks can be necessary for research purposes. It has been found that temperature affects the sex ratio of the Australian brush-turkey. The sex ratio was due to embryo mortality at particular temperatures rather than temperature-dependent sex determination found in reptiles (Goth and Booth, 2005). Faecal samples can

be collected easily from captive and wild birds using a plastic box, a vinyl-coated cloth and a plastic tray. The tray with plastic cloth is placed at the bottom of the box with a mesh covering to keep the bird raised above the collection tray. The bird can be placed in the box, and after defaecation, the contents in the tray can be poured into a plastic tube to be stored until analysis (Borrelli et al., 2020). The sample could be burned using either the diathermy electrode or bipolar forceps. The composition of bird faeces is different to mammals, so the samples may need to be prepared differently, such as by being dried before burning. If the samples are capable of being burned and producing a REIMS signature, then discrimination of bird sex could be analysed using the same method as the lab mice.

If REIMS can be used to discriminate pregnancy in samples, it would suggest that REIMS may have the potential to discriminant other physiological variations such as stress. Good animal welfare is defined by the five freedoms: freedom from hunger and thirst, discomfort, pain, injury and disease, freedom to express normal behaviour and freedom from fear and stress. This last freedom is the hardest to measure. An animal’s response, especially behaviour to a stressful event, will depend on species, age, life history, pregnancy or presence of offspring and could vary due to “personality”. Therefore, it is better to have a quantitative measurement of stress that anthropomorphism cannot influence (Wolfensohn et al., 2018). The stress response has been measured using heart rates, respiration rate, blood pressure and temperature (von Borell et al., 2007). The problem with these techniques is that the method themselves are invasive and may produce a stress response. The other commonly used method is to measure hormones, particularly cortisol. Cortisol concentrations have been shown to increase compared to the basal cortisol concentration when an individual is exposed to a short-term stressor, such as being transported (Tozlu Çelik et al., 2021). However, when an animal is exposed to a long-term stressor, such as poor housing conditions, cortisol concentration may not increase. Cortisol fluctuations also occur for factors other than a stress response, which may not accurately indicate stress (Wolfensohn et al., 2018). REIMS could potentially measure stress and detect the presence of cortisol, but the REIMS spectra may also contain other stress indicators. The spectral fingerprint may vary for animals before and after a stressful event and therefore randomForest models could be used to discriminant between a stressed and non-stressed animal. This could be used by zoos and farms to determine how stressful a particular experience is such as transport or a medical procedure.

For this to be investigated it would require purposely stressing multiple animals which have several ethical considerations. It would therefore be better to investigate REIMS's ability to discriminant stress by collecting faeces from animals already being exposed to a stressful event for other purposes. This could involve collecting faeces from lab mice that are already being researched for stress-related reasons are animals that are exposed to invasive procedures for medical reasons.

REIMS could be used as a quick preliminary study before investigating longer, more complicated mass spectrometry methods. Faecal samples have been analysed using other mass spectrometers, and a spectral fingerprint of bovine faeces was achieved using LC-HRMS (High-Resolution Mass Spectrometry) (Cesbron et al., 2017). MALDI-TOF analysis was used to establish pregnancy bio-markers in the faeces of polar bears that were unique to pregnant samples. Faecal pellets were collected from three pregnant and three non-pregnant polar bears housed in a zoo. Protein was extracted from the samples by suspending samples with phosphate-buffered saline (PBS), then agitated for 30 min, and centrifuged for 20 min. The supernatant was then filtered and added to ammonium sulphate, mixed for 30 min, centrifuged for 30 mins, resuspended in PBS, desalted, and then a Bradford assay was to evaluate protein concentration. The samples were then precipitated, resuspended in lysis buffer, mixed with dye and kept in the dark on ice for 30 minutes. Lysine was added to stop the labelling reaction and left for another 15 minutes in the dark. Buffers were added, and samples were separated using 2D-gel electrophoresis. The bands present due to pregnancy were extracted, and the peptides were then identified using MALDI-TOF. Two proteins had much higher concentrations in pregnant samples than in non-pregnant ones (Curry et al., 2012). This method required multiple sample preparation steps before MALDI-TOF could analyse the samples. REIMS and random forest models could be used to identify if there are any ions present that can discriminant between samples before being identified by other MS methods. REIMS is a much faster method that can be used to confirm an investigation has potential before using the much longer methods required by running gels and mass spectrometry.

It would be interesting to investigate what influences the spectral fingerprint produced by REIMS. If the microbiome produces a REIMS spectrum, it will benefit research in multiple disciplines. A host's microbiome is influenced by the microbes obtained from parents

and their environment and can affect the individual's health. It is unknown to what extent the distribution of the microbiota present in the environment influences an individual's microbiota (Cullen et al., 2020). There are many reasons to investigate the microbiome; some bacteria in the gut have been found to increase cancer development, and manipulating the microbiome could improve treatment (Bhatt et al., 2017). Changing the diets of mice from low-fat and plant-based to high-fat, high-sugar diets altered the microbiome after one day and increased adiposity in the mice (Turnbaugh et al., 2009). Some neurobiological diseases, including Parkinson's and Alzheimer's, have been affected by the microbiome and could be used to establish a biomarker (Tremlett et al., 2017). The microbiome is more diverse between species than individuals of the species. If the microbiome influences the REIMS spectrum, it would account for factors such as species having higher classification accuracies than sex. REIMS might not be used to determine a microbiome's diversity or the exact composition but could be used to track changes. One method to determine if the microbiome influences the spectral fingerprint would be to compare the faeces of germ-free mice to mice of the same species, sex and age but exposed to the environment. If REIMS could be used to discriminate between the germ-free and non-germ groups, then the microbiome may influence the spectra. Sibling groups that were just born could be split between two nursing mothers. If REIMS can discriminate between the nursing groups but not their sibling group, this would also suggest a change in the spectrum due to the microbiome. The microbiome could be changed by providing probiotics or antibiotics to some mice and not others to be distinguished by REIMS.

6.3 Conclusion

REIMS is a non-invasive technique for analysing animal faeces for population monitoring. REIMS can be used to discriminate species, sex, and age of faecal pellets from rodents and has the potential to be used for other species and to discriminate physiological differences. REIMS could help conservation by helping breeding programmes in zoos, establishing species distribution, and identifying rare species. Further work would establish if REIMS could be used to monitor the welfare of animals by measuring stress or the microbiome. The ability of REIMS to discriminate depends on the samples being used and the investigated factor. However, extra steps and refinements can be made to increase classification accuracy, including

increasing the sample number, changing the sample type (whole pellet) and reducing the number of mass bins used in random forests. More advanced machine learning techniques like neural networks may also increase classification accuracy. This study has examined some of the possible uses of REIMS for faecal analysis, but more research would demonstrate the true potential of this method.

Chapter 7: Publications

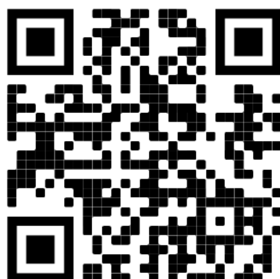
Publication contributions completing this PhD

7.1 Rapid identification of species, sex and maturity by mass spectrometric analysis of animal faeces

DAVIDSON, N. B., KOCH, N. I., SARSBY, J., JONES, E., HURST, J. L. & BEYNON, R. J. 2019. Rapid identification of species, sex and maturity by mass spectrometric analysis of animal faeces. BMC Biology, 17, 1-14.

I collected and analysed the faecal samples by REIMS using the method established in chapter 2. I used random forests to discriminate the sex, maturity and strain origin of the faecal samples and determined that samples labelled as juveniles above 30 days old were always classified by randomForest as adults (Figure 5). I was able to establish what bins were most responsible for the differences between males and females and between adults and juveniles using RandomForestExplainer (Figure 6). I repeated this for the different lab strains and carried out LDA-PCA to create kernel density and scatter plots to highlight the separation between strains (Figure 7). The software provided by Waters to normalise and bin the data was updated between the initial species discrimination analysis and my analysis. I, therefore, reanalysed all the work by Nicola Davidson using the New Software and all my work using the old software to ensure there were no differences. There was no significant difference between the random forest accuracies using either software therefore in the paper, species discrimination used the results from the old software and sex, age and strain used the results from the new software.

7.2 The application of rapid evaporative ionization mass spectrometry in the analysis of Drosophila species—a potential new tool in entomology



WAGNER, I., KOCH, N. I., SARSBY, J., WHITE, N., PRICE, T. A. R., JONES, S., HURST, J. L. & BEYNON, R. J. 2020. The application of rapid evaporative ionization mass spectrometry in the analysis of Drosophila species—a potential new tool in entomology. Open Biology, 10, 200196

The paper aimed to determine if REIMS could be used to discriminant different species of Drosophila by burning the whole insect. The study used the same random forest analysis I had developed for this PhD and I wrote the R scripts that were used to analyse the spectra produced by burning whole insects. This included random forests that were used to discriminate five different species of *Drosophila* (Figure 4) and the associated boxplot figures (Figure 5). The random forests were used to discriminant the sex of one species *D. melanogaster* and all five species combined and the associated kernel density and scatterplots (Figure 6).

7.3 Social status and ejaculate composition in the house mouse.

BAYRAM, H. L., FRANCO, C., BROWNRIDGE, P., CLAYDON, A. J., KOCH, N., HURST, J. L., BEYNON, R. J. & STOCKLEY, P. 2020. Social status and ejaculate composition in the house mouse. Philosophical Transactions of the Royal Society B: Biological Sciences, 375, 20200083.

The paper aimed to determine if the house mice changed the composition of seminal vesicle secretions due to social status. The protein content of the secretions was analysed using HPLC-MS; an electrospray ionisation source and an LTQ Orbitrap Velo mass spectrometer were used. The male house mice were divided into three groups, dominant, subordinate and control based on their scent-marking behaviour, dominant individuals will deposit more. The abundance of 29 proteins that were selected based on previous studies was used to compare the three groups. I performed random forest analysis to determine if the abundances of these proteins varied between dominant, subordinate and control individuals. I was able to establish which proteins varied the most between the three groups. I also performed an LDA-PCA analysis to help visualise the results (Figure 5). I investigated the use of cforests instead of random forests, cforests consist of extra steps when creating the decision trees. The classification accuracy of the training data was generally higher than using random forests, but the training data was much lower. It could be the extra steps in the cforests cause overfitting and I, therefore, determine cforests would not be a sufficient method to use for predicting unknown samples which would be required for most studies involving REIMS.

References

ABU-RABIE, P., SHEELAN, D., LAURES, A., SPAULL, J. & DOWELL, S. 2021. Increasing the discrimination power of rapid evaporative ionisation mass spectrometry (REIMS) in analytical control tissue quality screening and cell line sample identification. *Rapid Communications in Mass Spectrometry,* 35**,** 10.

AFRICA GEOGRAPHIC STORIES. 2018. *Okapi: The shy forest dwellers of central Africa* [Online]. Private Travel & Conservation Club. Available: https://africageographic.com/stories/okapi-shy-forest-dwellers-central-africa/ [Accessed 21/03/2022].

ANIMALS (SCIENTIFIC PROCEDURES) ACT 1986 2012. Animals (Scientific Procedures) Act 1986. 1st ed. London.

ANONYMOUS 2018. Guidelines for the treatment of animals in behavioural research and teaching. *Animal Behaviour,* 135**,** I-X.

ARIF, I. A., KHAN, H. A., BAHKALI, A. H., AL HOMAIDAN, A. A., AL FARHAN, A. H., AL SADOON, M. & SHOBRAK, M. 2011. DNA marker technology for wildlife conservation. *Saudi Journal of Biological Sciences* 18**,** 219-225.

ARZMI, M. H., JOHN, A., RISMAYUDDIN, N. A. R., KENALI, N. M. & DARNIS, D. S. 2021. LC-MS Data set on the Malayan Deer (Cervus timorensis) Antler Velvet and its antibiofilm activity against Candida species LC-MS Data set on the Malayan Deer (Cervus timorensis) Antler Velvet and its antibiofilm properties against Candida species. *Data in Brief,* 35.

BAINS, R. S., CATER, H. L., SILLITO, R. R., CHARTSIAS, A., SNEDDON, D., CONCAS, D., KESKIVALI-BOND, P., LUKINS, T. C., WELLS, S., AROZENA, A. A., NOLAN, P. M. & ARMSTRONG, D. 2016. Analysis of Individual Mouse Activity in Group Housed Animals of Different Inbred Strains Using a Novel Automated Home Cage Analysis System. *Frontiers in Behavioral Neuroscience* 10**,** 106.

BALOG, J., PERENYI, D., GUALLAR-HOYAS, C., EGRI, A., PRINGLE, S. D., STEAD, S., CHEVALLIER, O. P., ELLIOTT, C. T. & TAKATS, Z. 2016. Identification of the Species of Origin for Meat Products by Rapid Evaporative Ionization Mass Spectrometry. *Journal of Agrticultural and Food Chemistry* 64**,** 4793-4800.

BALOG, J., SZANISZLO, T., SCHAFER, K. C., DENES, J., LOPATA, A., GODORHAZY, L., SZALAY, D., BALOGH, L., SASI-SZABO, L., TOTH, M. & TAKATS, Z. 2010. Identification of Biological Tissues by Rapid Evaporative Ionization Mass Spectrometry. *Analytical Chemistry,* 82**,** 7343-7350.

BELLEMAIN, E., SWENSON, J. E., TALLMON, D., BRUNBERG, S. & TABERLET, P. 2005. Estimating population size of Elusive Animals with DNA from Hunter-Collected Feces: Four methods for Brown Bears. *Conservation Biology,* 19**,** 150-161.

BEREZOWSKA-CNOTA, T., LUQUE-MÁRQUEZ, I., ELGUERO-CLARAMUNT, I., BOJARSKA, K., OKARMA, H. & SELVA, N. 2017. Effectiveness of different types of hair traps for brown bear research and monitoring. *PLOS ONE,* 12**,** e0186605.

BHATT, A. P., REDINBO, M. R. & BULTMAN, S. J. 2017. The role of the microbiome in cancer development and therapy. *CA: A Cancer Journal for Clinicians,* 67**,** 326-344.

BLACK, C., CHEVALLIER, O. P., HAUGHEY, S. A., BALOG, J., STEAD, S., PRINGLE, S. D., RIINA, M. V., MARTUCCI, F., ACUTIS, P. L., MORRIS, M., NIKOLOPOULOS, D. S., TAKATS, Z. & ELLIOTT, C. T. 2017. A real time metabolomic profiling approach to detecting fish fraud using rapid evaporative ionisation mass spectrometry. *Metabolomics,* 13.

BORRELLI, L., MINICHINO, A., PACE, A., DIPINETO, L. & FIORETTI, A. 2020. Fecal Sample Collection Method for Wild Birds-Associated Microbiome Research: Perspectives for Wildlife Studies. *Animals (Basel),* 10.

BOSSON, C. O., ISLAM, Z. & BOONSTRA, R. 2012. The impact of live trapping and trap model on the stress profiles of North American red squirrels. *Journal of Zoology,* 288**,** 159-169.

BOTH, C., BOUWHUIS, S., LESSELLS, C. M. & VISSER, M. E. 2006. Climate change and population declines in a long-distance migratory bird. *Nature* 441**,** 81-83.

BOWERMAN, K. L., KNOWLES, S. C. L., BRADLEY, J. E., BALTRŪNAITĖ, L., LYNCH, M. D. J., JONES, K. M. & HUGENHOLTZ, P. 2021. Effects of laboratory domestication on the rodent gut microbiome. *ISME Communications,* 1**,** 49.

BPCA 2018. Code of Best Practice: Snap back traps. *British Pest Control Association*

BRAKES, C. R. & SMITH, R. H. 2005. Exposure of non-target small mammals to rodenticides: short-term effects, recovery and implications for secondary poisoning. *Journal of Applied Ecology,* 42**,** 118-128.

BREHM, A. M. & MORTELLITI, A. 2018. Mind the trap: large-scale field experiment shows that trappability is not a proxy for personality. *Animal Behaviour,* 142**,** 101-112.

BUCKLE, A. 2013. Anticoagulant resistance in the United Kingdom and a new guideline for the management of resistant infestations of Norway rats (Rattus norvegicus Berk.). *Pest Manag Sci,* 69**,** 334-41.

CAMERON, S. J. S., BOLT, F., PERDONES-MONTERO, A., RICKARDS, T., HARDIMAN, K., ABDOLRASOULI, A., BURKE, A., BODAI, Z., KARANCSI, T., SIMON, D., SCHAFFER, R., REBEC, M., BALOG, J. & TAKATS, Z. 2016. Rapid Evaporative Ionisation Mass Spectrometry (REIMS) Provides Accurate Direct from Culture Species Identification within the Genus *Candida*. *Scienctific Reports,* 6**,** 36788.

CAMPBELL, J. F., MULLEN, M. A. & DOWDY, A. K. 2002. Monitoring Stored-Product Pests in Food Processing Plants with Pheromone Trapping, Contour Mapping, and Mark-Recapture. *Journal of Economic Entomology,* 95**,** 1089-1101.

CARAGIULO, A., PICKLES, R. S. A., SMITH, J. A., SMITH, O., GOODRICH, J. & AMATO, G. 2015. Tiger (*Panthera tigris*) scent DNA: a valuable conservation tool for individual identification and population monitoring *Conservation Genet Resour,* 7**,** 681-683.

CARAVAGGI, A., BANKS, P. B., BURTON, A. C., FINLAY, C. M. V., HASWELL, P. M., HAYWARD, M. W., ROWCLIFFE, M. J. & WOOD, M. D. 2017. A review of camera trapping for conservation behaviour research. *Remote Sensing in Ecology and Conservation* 3**,** 109-122.

CARL, C., SCHONFELD, F., PROFFT, I., KLAMM, A. & LANDGRAF, D. 2020. Automated detection of European wild mammal species in camera trap images with an existing and pre-trained computer vision model. *European Journal of Wildlife Research,* 66**,** 62.

CEBALLOS, G., EHRLICH, P. R. & DIRZO, R. 2017. Biological annihilation via the ongoing sixth mass extinction signalled by vertebrate population losses and declines. *Proc Natl Acad Sci U S A,* 114**,** E6089-E6096.

CESBRON, N., ROYER, A. L., GUITTON, Y., SYDOR, A., LE BIZEC, B. & DERVILLY-PINEL, G. 2017. Optimization of fecal sample preparation for untargeted LC-HRMS based metabolomics. *Metabolomics,* 13**,** 99.

CHANIN, P., O'REILLY, C., TURNER, P., KERSLAKE, L., BIRKS, J. & WOODS, M. 2015. Insects in the diet of the hazel dormouse (*Muscardinus avellanarius*): a pilot study using DNA barcoding. *Mammal Communications,* 1**,** 1-7.

CHERNUSHEVICH, I. V., LOBODA, A. V. & THOMSON, B. A. 2001. An introduction to quadrupole – time-of-flight mass spectrometry. *Journal of Mass Spectrometry,* 36**,** 849-865.

CHESTER ZOO 2021. Preventing Extinction in a changing World: A conservation Masterplan for Chester Zoo. North of England Zoological Soiety

CLAUSS, M., HUMMEL, J., VOLLM, J., LORENZ, A. & HOFFMANN, R. R. 2006. The allocation of a ruminant feeding type to the okapi (Okapia johnstoni) on the basis of morphological parameters. *Zoo animal nutrition,* 3**,** 253-270.

CLAUSS, M., LECHNER-DOLL, M. & STREICH, W. J. 2002. Faecal particle size distribution in captive wild ruminants an approach to the browser/grazer dichotomy from the other end. *Oecologia,* 131**,** 343-349.

CODY, R. B., DANE, A. J., DAWSON-ANDOH, B., ADEDIPE, E. O. & NKANSAH, K. 2012. Rapid Classification of White Oak (Quercus alba) and Northern Red Oak (*Quercus rubr*a) by using pyrolysis direct

analysis in real time (DART™) and time-of-flight mass spectrometry. *Journal of Analytical and Applied Pyrolysis,* 95**,** 134-137.

COLLEN, B., MCRAE, L., DEINET, S., DE PALMA, A., CARRANZA, T., COOPER, N., LOH, J. & BAILLIE, J. E. M. 2011. Predicting how populations decline to extinction. *Philosophical Transactions of the Royal Society B,* 366**,** 2577-2586.

COOMBER, F. G., LE MARQUAND, C., CRAWLEY, D., GILFORD, E., KIRK, S., LLOYD, A. J., WEBB, S. & GURNELL, J. 2023. The Mammal Society's National Harvest Mouse Survey:Results from the 2021-2022 Survey Season. *A report produced by the Mammal Society*.

CROMSIGT, J. P. G. M., VAN RENSBURG, S. J., ETIENNE, R. S. & OLFF, H. 2009. Monitoring large herbivore diversity at different scales: comparing direct and indirect methods. *Biodiversity Conservation,* 18**,** 1219-1231.

CULLEN, C. M., ANEJA, K. K., BEYHAN, S., CHO, C. E., WOLOSZYNEK, S., CONVERTINO, M., MCCOY, S. J., ZHANG, Y., ANDERSON, M. Z., ALVAREZ-PONCE, D., SMIRNOVA, E., KARSTENS, L., DORRESTEIN, P. C., LI, H., SEN GUPTA, A., CHEUNG, K., POWERS, J. G., ZHAO, Z. & ROSEN, G. L. 2020. Emerging Priorities for Microbiome Research. *Frontiers in Microbiology,* 11.

CURRY, E., SKOGEN, M. & ROTH, T. L. 2021. Evaluation of an odour-detection dog for non-invasive pregnancy diagnosis in polar bears Ursus maritimus: Considerations for training sniffer dogs for biomedical investigations in wildlife species. *Journal of Zoo and Aquarium Research,* 9**,** 1-7.

CURRY, E., STOOPS, M. A. & ROTH, T. L. 2012. Non-invasive detection of candidate pregnancy protein bio-markers in the feces of captive polar bears (Ursus maritimus). *Theriogenology,* 78**,** 308-314.

DAVIDSON, N. B. & HURST, J. L. 2019. Testing the potential of 50 kHz rat calls as a species-specific rat attractant. *PLOS ONE,* 14**,** e0211601.

DAVIDSON, N. B., KOCH, N. I., SARSBY, J., JONES, E., HURST, J. L. & BEYNON, R. J. 2019. Rapid identification of species, sex and maturity by mass spectrometric analysis of animal faeces. *BMC Biology,* 17**,** 1-14.

DE BARBA, M., ADAMS, J. R., GOLDERBERG, C. S., STANSBURY, C. R., ARIAS, D., CISNEROS, R. & WAITS, L. P. 2014. Molecular species identification for multiple carnivores. *Conservation Genetics Resources,* 6**,** 821-824.

DE BARBA, M., WAITS, L. P., GENOVESI, P., RANDI, E., CHIRICHELLA, R. & CETTO, E. 2010. Comparing opportunistic and systematic sampling methods for non-invasive genetic monitoring of a small translocated brown bear population. *Journal of Applied Ecology* 47**,** 172-181.

DE BONDI, N., WHITE, J. G., STEVENS, M. & COOKE, R. 2010. A comparison of the effectiveness of camera trapping and live trapping for sampling terrestrial small-mammal communities. *Wildlife Research,* 37**,** 456-465.

DE HOFFMANN, E. & STROOBANT, V. 2007. *Mass Spectrometry Principles and Applications,* Sussex, England, John Wiley & Sons Ltd.

DING, B., ZHANG, Y.-P. & RYDER, O. A. 1998. Extraction, PCR amplification, and sequencing of mitochondrial DNA from scent mark and feces in the giant panda. *Zoo Biology,* 17**,** 499-504.

DIXON, R. & COATES, D. 2009. Near infrared spectroscopy of faeces to evaluate the nutrition and physiology of herbivores. *Journal of Near Infrared Spectroscopy,* 17**,** 1-31.

EDWARD, R. S. J., JULIA, B., JAMES, S. M., MERJA, R., APRIL, C., LAURA, M., ZSOLT, B., FRANCESCA, R., ABIGAIL, V. M. S., SAMI, S., RATHI, R., ARA, D., ZOLTAN, T. & DANIEL, R. L. 2017. Rapid evaporative ionisation mass spectrometry of electrosurgical vapours for the identification of breast pathology: towards an intelligent knife for breast cancer surgery.

EDWARDS, K. L., TROTTER, J., JONES, M., BROWN, J. L., STEINMETZ, H. W. & WALKER, S. L. 2016. Investigating tempoary acylicity in a captive group of Asian elephants (Elephas maximus): Relationship between managment, adrenal acticity and social factors. . *General and Comparative Endocrinology,* 225**,** 104-116.

EGGERT, L. S., EGGERT, J. A. & WOODRUFF, D. S. 2003. Estimating population sizes for elusive animals: the forest elephants of Kakum National Park, Ghana. *Molecular Ecology,* 12**,** 1389-1402.

ENARI, H., ENARI, H. S., OKUDA, K., MARUYAMA, T. & OKUDA, K. N. 2019. An evaluation of the efficiency of passive acoustic monitoring in detecting deer and primates in comparison with camera traps. *Ecological Indicators,* 98**,** 753-762.

ERNEST, H. B., PENEDO, M. C. T., MAY, B. P., SYVANEN, M. & BOYCE, V. M. 2000. Molecular tracking of mountain lions in the Yosemite Valley region in California: genetic analysis using microsatellites and faecal DNA. *Molecular Ecology* 9 433-441.

ESPINOZA, E. O., LANCANSTER, C. A., KREITALS, N. M., HATA, M., CODY, R. B. & BLANCHETTE, R. A. 2014. Distinguishing wild from cultivated agarwood (Aquilaria spp.) using direct analysis in real time and time of flight mass spectrometry *Rapid Communications in Mass Spectrometry,* 28**,** 281-289.

FAHLMAN, A., LOVERIDGE, A., WENHAM, C., FOGGIN, C., ARNEMO, J. M. & NYMAN, G. 2005. Reversible anaesthesia of free-ranging lions (*Panthera leo*) in Zimbabwe. *Journal of the South African Veterinary Association,* 76.

FINCH, K., ESPINOZA, E., JONES, F. A. & CRONN, R. 2017. Source identification of western Oregon Douglas-fir wood cores using mass spectrometry and random forest class ification. *Applications in Plant Sciences* 5**,** 1600158.

FLETCHER, Q. E. & BOONSTRA, R. 2006. Impact of live trapping on the stress response of the meadow vole (Microtus pennsylvanicus). *Journal of Zoology,* 270**,** 473-478.

FLOWERDEW, J. R., SHORE, R. F., POULTON, S. M. C. & SPARKS, T. H. 2004. Live Trapping to monitor small mammals in Britain. *Mammal Review,* 34**,** 31-50.

FLURKEY, K., CURRER, J. M. & HARRISON, D. E. 2007. The Mouse in Aging Research. *In:* MEDICINE, A. C. L. A. (ed.) *The Mouse in Biomedical Research* 2ed. Burlington, MA: Elsevier.

FREEMAN, C. M., BARTHMAN-THOMPSON, L., KLINGER, R., WOO, I. & THORNE, K. M. 2022. Assessing small-mammal trapping design using spatially explicit capture recapture (SECR) modeling on long-term monitoring data. *PLOS ONE,* 17**,** e0270082.

GAUTHIER-CLERC, M., GENDNER, J. P., RIBIC, C. A., FRASER, W. R., WOEHLER, E. J., DESCAMPES, S., GILLY, C., LE BOHEC, C. & LE MAHO, Y. 2004. Long-term effects of flipper bands on penguins. *Proceedings of the Royal Society of London B,* 271**,** 423-426.

GELLING, M., MCLAREN, G. W., MATHEWS, F., MIAN, R. & MACDONALD, D. W. 2009. Impact of trapping and handling on Leukocyte Coping Capacity in bank voles (Clethrionomys glareolus) and wood mice (Apodemus sylvaticus). *Animal Welfare,* 18**,** 1-7.

GOMPPER, M. E., KAYS, R. W., RAY, J. C., LAPOINT, S. D. & BOGAN, D. A. 2006. A Comparison of Noninvasive Techniques to Survey Carnivore Communities in Northeastern North America. *Wildlife Society Bulletin,* 34**,** 1142-1151.

GONZALEZ, F. J. & JOHNSON, S. 2017. Standard operating procedures for UAV or drone based monitoring of wildlife. *Proceedings of the UAS4RS 2017* 1-7.

GONZALEZ, L. F., MONTES, G. A., PUIG, E., JOHNSON, S., MENGERSEN, K. & GASTON, K. J. 2016. Unmanned Aerial Vehicles (UAVs) and Artificial Intelligence Revolutionizing Wildlife Monitoring and Conservation. *Sensors (Basel),* 16.

GOODWIN, C. E. D., HODGSON, D. J., AL-FULAIJ, N., BAILEY, S., LANGTON, S. & MCDONALD, R. A. 2017. Voluntary recording scheme reveals ongoing decline in the United Kingdom hazel dormouse Muscardinus avellanarius population. *Mammal Review,* 47**,** 183-197.

GOTH, A. & BOOTH, D. T. 2005. Temperature-dependent sex ratio in a bird. *Biol Lett,* 1**,** 31-3.

GOUVEIA, K. & HURST, J. L. 2019. Improving the practicality of using non-aversive handling methods to reduce background stress and anxiety in laboratory mice. *Sci Rep,* 9**,** 20305.

GRIFFITHS, J. 2008. A brief history of mass spectrometry. *Anal Chem,* 80**,** 5678-83.

GRIFFITHS, P. R. 2000. Sex Identification in Birds. *Seminars in Avian and Exotic Pet Medicine,* 9**,** 14-26.

GROTTA-NETTO, F., PERES, P. H. F., PIOVEZAN, U., PASSOS, F. C. & DUARTE, J. M. B. 2021. Camera Trap Feasibility for Ecological Studies of Elusive Forest Deer. *Wildlife Society Bulletin,* 44**,** 640-647.

GUSSET, M. & DICK, G. 2010. Building a Future for Wildlife? Evaluuating the contribution to *in situ* conservation *International Zoo Yearbook,* 44**,** 183-191.

HAAHR, M. 1998. *True Random Number Service* [Online]. Dublin, Ireland. Available: https://www.random.org [Accessed 20/03/2022].

HAMMERSCHLAG, N., COOKE, S. J., GALLAGHER, A. J. & GODLEY, B. J. 2014. Considering the fate of electronic tags: interactions with stakeholders and user responsibility when encountering tagged aquatic animals. *Methods in Ecology and Evolution,* 5**,** 1147-1153.

HAUSKNECHT, R., GULA, R., PIRGA, B. & KUEHN, R. 2007. Urine — a source for noninvasive genetic monitoring in wildlife. *Molecular Ecology Notes,* 7**,** 208-212.

HEDGES, L., MORRANT, D. S., CAMPOS-ARCEIZ, A. & CLEMENTS, G. R. 2015. Feasibility of using scent-baited hair traps to monitor carnivore populations in Peninsular Malaysia. *Tropical Conservation Science,* 8**,** 975-982.

HEIN, A., PALME, R., BAUMGARTNER, K., FERSEN, L., WOELFING, B., GREENWOOD, A. D., BECHSHOFT, T. & SIEBERT, U. 2020. Faecal glucocorticoid metabolites as a measure of adrenocortical activity in polar bears (Ursus maritimus) *Conservation Physiology,* 8**,** coaa012.

HILDEBRAND, F., NGUYEN, T. L. A., BRINKMAN, B., YUNTA, R. G., CAUWE, B., VANDENABEELE, P., LISTON, A. & RAES, J. 2013. Inflammation-associated enterotypes, host genotype, cage and inter-individual effects drive gut microbiota variation in common laboratory mice. *Genome Biology,* 14**,** R4.

HILL, A. P., PRINCE, P., PIÑA COVARRUBIAS, E., DONCASTER, C. P., SNADDON, J. L. & ROGERS, A. 2018. AudioMoth: Evaluation of a smart open acoustic device for monitoring biodiversity and the environment. *Methods in Ecology and Evolution,* 9**,** 1199-1211.

HO, C. S., LAM, C. W. K., CHAN, M. H. M., CHEUNG, R. C. K., LAW, L. K., LIT, L. C. W., NG, K. F., SUEN, M. W. M. & TAI, H. L. 2003. Electrospray Ionisation Mass Spectrometry: Principles and Clinical Applications. *Clinical Biochemist Review* 24**,** 3-12.

HOFFMANN, E. & STROOBANT, V. 2007. *Mass Spectrometry Principles and Applications* Chichester, West Sussex, England, Wiley.

HOLLEMEYER, K., ALTMEYER, W. & HEINZLE, E. 2007. Identification of furs of domestic dog, raccoon dog, rabbit and domestic cat by hair analysis using MALDI-ToF mass spectrometry. *Spectroscopy Europe,* 19**,** 8-15.

HSIEH, Y. H., PETERSON, C. M., RAGGIO, A., KEENAN, M. J., MARTIN, R. J., RAVUSSIN, E. & MARCO, M. L. 2016. Impact of Different Fecal Processing Methods on Assessments of Bacterial Diversity in the Human Intestine. *Front Microbiol,* 7**,** 1643.

HSU, C. C., CHOU, P. T. & ZARE, R. N. 2015. Imaging of Proteins in Tissue Samples Using Nanospray Desorption Electrospray Ionization Mass Spectrometry. *Analytical Chemistry,* 87**,** 11171-11175.

HUFELDT, M. R., NIELSEN, D. S., VOGENSEN, F. K., MIDTVEDT, T. & HANSEN, A. K. 2010. Variation in the gut microbiota of laboratory mice is related to both genetic and environmental factors. *Comp Med,* 60**,** 336-47.

HUNT, K. E. & WASSER, S. K. 2003. Effect of Long-Term Preservation Methods on Fecal Glucocorticoid Concentrations of Grizzly Bear and African Elephant. *Physiological and Biochemical Zoology,* 76**,** 918-928.

HUNTER, M. E., HOBAN, S. M., BRUFORD, M. W., SEGELBACHER, G. & BERNATCHEZ, L. 2018. Next-generation conservation genetics and biodiversity monitoring. *Evolutionary Applications,* 11**,** 1029-1034.

HURST, J. L. & WEST, R. S. 2010. Taming anxiety in laboratory mice. *Nat Methods,* 7**,** 825-6.

IMHOLT, C., REIL, D., PLAŠIL, P., RÖDIGER, K. & JACOB, J. 2017. Long-term population patterns of rodents and associated damage in German forestry. *Pest Management Science,* 73**,** 332-340.

IUCN. 2017. *The IUCN Red List of Threatened Species. Version 2017-3* [Online]. Available: http://www.iucnredlist.org [Accessed 12 December 2017].

JACKSON, J., MAR, K. U., HTUT, W., CHILDS, D. Z. & LUMMAA, V. 2020. Changes in age-structure over four decades were a key determinant of population growth rate in a long-lived mammal. *Journal of Animal Ecology,* 89**,** 2268-2278.

JACOME, L. F., BURKET, J. A., HERNDON, A. L. & DEUTSCH, S. I. 2011. Genetically inbred Balb/c mice differ from outbred Swiss Webster mice on discrete measures of sociability: relevance to a genetic mouse model of autism spectrum disorders. *Autism Res,* 4**,** 393-400.

JENSEN, S. P., GRAY, S. J. & HURST, J. L. 2002. How does habitat structure affect activity and use of space among house mice? *Animal Behaviour,* 66**,** 239-250.

JUNG, T. S., BOONSTRA, R. & KREBS, C. J. 2019. Mark my words: experts' choice of marking methods used in capture-mark-recapture studies of small mammals. *Journal of Mammalogy,* 101**,** 307-317.

KAIKUSALO, A. 1972. Population turnover and wintering of the bank vole, Clethrionomys glareolus (Schreb.), in southern and central Finland. *Annales Zoologici Fennici,* 9**,** 219-224.

KASSAMBARA, A. 2021. Pipe-Friendly Framework for Basic Statistical Tests. *CRAN,* v0.7.0.

KELLY, R. T., TOLMACHEV, A. V., PAGE, J. S., TANG, K. & SMITH, R. D. 2010. The ion funnel: theory, implementations, and applications. *Mass Spectrom Rev,* 29**,** 294-312.

KETTEL, E. F., PERROW, M. R. & READER, T. 2016. Live-trapping in the stalk zone of tall grasses as an effective way of monitoring harvest mice (Micromys minutus). *European Journal of Wildlife Research,* 62**,** 241-245.

KIA, E., MACKENZIE, B. W., D., M., LAU, A., WAITE, D. W., LEWIS, G., CHAN, Y., SILVESTRE, M., COOPER, G. J. S., POPPITT, S. D. & TAYLOR, M. W. 2016. Integrity of the Human Faecal Microbiota following Long-Term Sample Storage. *PLoS one,* 11**,** e0163666.

KOEPFLI, K. P., POLLINGER, J., GODINHO, R., ROBINSON, J., LEA, A., HENDRICKS, S., SCHWEIZER, R. M., THALMANN, O., SILVA, P., FAN, Z., YURCHENKO, A. A., DOBRYNIN, P., MAKUNIN, A., CAHILL, J. A., SHAPIRO, B., ALVARES, F., BRITO, J. C., GEFFEN, E., LEONARD, J. A., HELGEN, K. M., JOHNSON, W. E., O'BRIEN, S. J., VAN VALKENBURGH, B. & WAYNE, R. K. 2015. Genome-wide Evidence Reveals that African and Eurasian Golden Jackals Are Distinct Species. *Curr Biol,* 25**,** 2158-65.

KOJU, N. P., BASHYAL, B., PANDEY, B. P., SHAH, S. N., THAMI, S. & BLEISCH, W. V. 2020. First camera-trap record of the snow leopard *Panthera uncia* in Gaurishankar Conservation Area, Nepal. *Oryx,* 55**,** 173-176.

KREBS, C. J., BOONSTRA, R., GILBERT, B. S., KENNEY, A. J. & BOUTIN, S. 2019. Impact of climate change on the small mammal community of the Yukon boreal forest. *Integrative Zoology,* 14**,** 528-541.

KUSUDA, S., MORIKAKU, K., KAWADA, K., ISHIWADA, K. & DOI, O. 2007. Excretion patterns of fecal progestagens, androgen and estrogens during pregnancy, parturition and postpartum in okapi *(Okapia johnstoni*). *Journal of Reproduction and Development,* 53**,** 143-150.

LACY, R. C. 2013. Achieving true sustainability of zoo populations. *Zoo Biol,* 32**,** 19-26.

LANCASTER, C. & ESPINOZA, E. 2012. Analysis of select *Dalbergia* and trade timber using direct analysis in real time and time-of-flight mass spectrometry for CITES enforcement. *Rapid Communications in Mass Spectrometry,* 26**,** 1147-1156.

LANDAU, S., GIGER-REVERDIN, S., RAPETTI, L., DVASH, L., DORLÉANS, M. & UNGAR, E. D. 2008. Data mining old digestibility trials for nutritional monitoring in confined goats with aids of fecal near infra-red spectrometry. *Small Ruminant Research,* 77**,** 2-3.

LANDE, R. 1988. Genetics and demography in biological conservation. *Science,* 241**,** 1455-60.

LAPLANCHE, C., MARQUES, T. A. & THOMAS, L. 2015. Tracking marine mammals in 3D using electronic tag data. *Methods in Ecology and Evolution,* 6**,** 987-996.

LARA-RUIZ, P. & CHIARELLO, A. G. 2005. Life-history traits and sexual dimorphism of the Atlantic forest maned sloth Bradypus torquatus (Xenarthra: Bradypodidae). *Journal of Zoology,* 267**,** 63-73.

LEMUS, J. A., BRAVO, C., GARCIA-MONTIJANO, M., PALACIN, C., PONCE, C., MAGANA, M. & ALONSO, J. C. 2011. Side effects of rodent control on non-target species: Rodenticides increase parasite and pathogen burden in great bustards. *Sci Total Environ,* 409**,** 4729-34.

LIAW, A. & WEINER, M. 2002. Classification and Regression by randomForest. *R News,* 2**,** 18-22.

LIN, Y. K. & BATZLI, G. O. 2004. Movement of Voles Across Habitat Boundaries: Effects of Food and Cover. *Journal of Mammalogy,* 85**,** 216-224.

LOOS, A., WEIGEL, C. & KOEHLER, M. 2018. Towards Automatic Detection of Animals in Camera-Trap Images. *2018 26th European Signal Processing Conference (EUSIPCO),* IEEE**,** 1805-1809.

LUBA, C. N., KLUYBER, D., MASSOCATO, G. F., ATTIAS., N., FROMME, L., LUISM A., RODRIGUES, R., FERREIRA, A. M. R. & DESBIEZ, A. L. J. 2020. Size matters: penis size, sexual maturity and their consequences for giant armadillo conservation planning. *Mammalian Biology,* 100**,** 621-630.

LUCAS, T. C. D., MOORCROFT, E. A., FREEMAN, R., ROWCLIFFE, J. M. & JONES, K. E. 2015. A generalised random encounter model for estimating animal density with remote sensor data. *Methods in Ecology and Evolution,* 6**,** 500-509.

LUNARDON, N., MENARDI, G. & TORELLI, N. 2013. ROSE: Random Over-Sampling Examples.

MACAULAY, L. T., SOLLMANN, R. & BARRETT, R. H. 2019. Estimating Deer Populations Using Camera Traps and Natural Marks. *The Journal of Wildlife Management,* 84**,** 301-310.

MALHERBE, G. P., MAUDE, G. & BASTOS, A. D. S. 2009. Genetic clues from olfactory cues: brown hyaena scent marks provide a non-invasive source of DNA for genetic profiling. *Conservation Genetics,* 10**,** 759-762.

MALLINSON, J. C. 1995. Conservation breeding programmes: an important ingredient for species survival. *Biodiversity and Conservation* 4**,** 617-635.

MAMYRIN, B. A. 2001. Time-of-flight mass spectrometry (concepts, achievements, and prospects). *International Journal of Mass Spectrometry,* 206**,** 251-266.

MARCH, R. E. 1997. An Introduction to Quadrupole Ion Trap Mass Spectrometry. *Journal of Mass Spectrometry,* 32**,** 351-369.

MASON, K. A., LOSOS, J. B., SINGER, S. R., RAVEN, P. & JOHNSON, G. B. 2011. *Biology,* New York, USA, McGraw-Hill.

MATTHAN, N. R., IP, B., RESTEGHINI, N., AUSMAN, L. M. & LICHENSTEIN, A. H. 2010. Long-term fatty acid stability in human serum cholesterylester, triglyceride, and phospholipid fractions. *Journal of Lipid Research,* 51**,** 2826-2832.

MCKELVEY, K. S., VON KIENAST, J., AUBRY, K. B., KOEHLER, G. M., MALETZKE, B. T., SQUIRES, J. R., LINDQUIST, E. L., LOCH, S. & SCHWARTZ, M. K. 2006. DNA Analysis of Hair and Scat Collected Along SnowTracks to Document the Presence of Canada Lynx. *Wildlife Society Bulletin,* 34**,** 451-455.

MELCORE, I., FERRARI, G. & BERTOLINO, S. 2020. Footprint tunnels are effective for detecting dormouse species. *Mammal Review,* 50**,** 226-230.

MINTEER, B. A. & COLLINS, J. P. 2013. Ecological Ethics in Captivity: Balancing Values and Responsibilities in Zoo and Aquarium Reseach under Rapid Global Change *Institute for Laboratory Animal Research* 54**,** 41-51.

MORIN, D. J., KELLY, M. J. & WAITS, L. P. 2016. Monitoring Coyote Population Dynamics with Fecal DNA and Spatial Capture-Recapture. *The Journale of Wildlife Management* 80**,** 824-836.

MOSTL, E., MAGGS, J. L., SCHROTTER, G., BESENFELDER, U. & PALME, R. 2002. Measurement of cortisol metabolites in faeces of ruminants. *Veterinary Research Communications,* 26**,** 127-139.

MUKHOPADHYA, I., MARTIN, J. C., SHAW, S., MCKINLEY, A. J., GRATZ, S. W. & SCOTT, K. P. 2022. Comparison of microbial signatures between paired faecal and rectal biopsy samples from healthy volunteers using next-generation sequencing and culturomics. *Microbiome,* 10**,** 171.

MURTAGH, R., BEHRINGER, V. & DESCHNER, T. 2013. LC-MS as a method for non-invasive measurement of steroid hormones and their metabolites in urine and faeces of animals. *Wiener Tierarztliche Monatsschrift,* 100**,** 247-254.

NIER, A. O. 1947. A Mass Spectrometer for Isotope and Gas Analysis. *Review of Scientific Instruments,* 18**,** 398-411.

NIXON, S. C. & LUSENGE, T. 2008. Conservation status of okapi in Virunga National Park, Democratic Republic of Congo. *ZSL Conservation Report.* London: The Zoological Society of London.

NOROUZZADEH, M. S., NGUYEN, A., KOSMALA, M., SWANSON, A., PALMER, M. S., PACKER, C. & CLUNE, J. 2018. Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning. *PNAS,* 115**,** E5716–E5725.

OKAPI CONSERVATION PROJECT 2020. 2020 Annual Report. Florida, USA: Wildlife Conservation Global.

OWEN, A., WILKINSON, R. & SÖZER, R. 2014. *In situ* conservation breeding and the role of zoological institutions and private breeders in the recovery of highly endangered Indonesian passerine birds. *International Zoo Yearbook,* 48**,** 199-211.

PALUSZYNSKA, A. & BIECEK, P. 2017. randomForestExplainer: Explaining and Visualizing Random Forests in Terms of Variable Importance.

PAULI, J. N., MENDOZA, J. E., STEFFAN, S. A., CAREY, C. C., WEIMER, P. J. & PEERY, M., Z., 2014. A syndrome of mutualism reinforces the lifestyle of a sloth. *Proceedings of the Royal Society of Biology,* 281**,** 20133006.

REGAL, P., VAZQUEZ, B. I., FRANCO, C. M., CEPEDA, A. & FENTE, C. 2009. Quantitative LC–MS/MS method for the sensitive and simultaneous determination of natural hormones in bovine serum. *Journal of Chromatography B,* 877**,** 2457-2464.

ROFFE, S. 2021. *We're celebrating the birth of a rare baby okapi.* [Online]. Chester: Chester Zoo. [Accessed 1st February 2022].

ROPERT-COUDERT, Y. & WILSON, R. P. 2005. Trends and perspectives in animal-attached remote sensing. *Frontiers in Ecology and Evolution,* 25**,** 437-444.

ROWCLIFFE, J. M., KAYS, B. K., CARBONE, C. & JANSEN, P. A. 2014. Quantifying levels of animal activity using camera trap data. *Methods in Ecology and Evolution* 5**,** 1170-1179.

SCHAFER, K. C., DENES, J., ALBRECHT, K., SZANISZLO, T., BALOG, J., SKOUMAL, R., KATONA, M., TOTH, M., BALOGH, L. & TAKATS, Z. 2009. In Vivo, In Situ Tissue Analysis Using Rapid Evaporative Ionization Mass Spectrometry. *Angewandte Chemie International Edition,* 48**,** 8240-8242.

SCHRADER, C., SCHIELKE, A., ELLERBROEK, L. & JOHNE, R. 2012. PCR inhibitors - occurrence, properties and removal *Journal of Applied Microbiology,* 113**,** 1010-1026.

SCHROEDER, N. M., PANEBIANCO, A., GONZALEZ MUSSO, R. & CARMANCHAHI, P. 2020. An experimental approach to evaluate the potential of drones in terrestrial mammal research: a gregarious ungulate as a study model. *R Soc Open Sci,* 7**,** 191482.

SCHWARZENBERGER, F., MOSTL, E., PALME, R. & BAMBERG, E. 1996. Faecal steroid analysis for non-invasive monitoring of reproductive status in farm, wild and zoo animals. *Animal Reproduction Science,* 42**,** 515-526.

SCHWARZENBERGER, F., RIETSCHEL, W., B., M., SCHAFTENAAR, W., BIRCHER, P., VAN PUIJENBROECK, B. & LEUS, K. 1999. Noninvasive reproductive monitoring in the okapi (*Okapia johnstoni*) *Journal of Zoo and Wildlife Medicine,* 30**,** 497-503.

SHESTAK, A. G., BUKAEVA, A. A., SABER, S. & ZAKLYYAZMINSKAYA, E. V. 2021. Allelic Dropout Is a Common Phenomenon That Reduces the Diagnostic Yield of PCR-Based Sequencing of Targeted Gene Panels. *Frontiers in Genetics,* 12**,** 620337.

SINGH, S., BERMUDEZ-CONTRERAS, E., NAZARI, M., SUTHERLAND, R. J. & MOHAJERANI, M. H. 2019. Low-cost solution for rodent home-cage behaviour monitoring. *PLOS one,* 14**,** e0220751.

SMITH, O. & WANG, J. 2014. When can noninvasive samples provide sufficient information in conservation genetics studies? . *Molecular Ecology Resources* 14**,** 1011-1023.

SNEDDEN, W. & PARKER, R. B. 1971. Determination of Volatile Constituents of Human Blood and Tissue Specimens by Quantitative High Resolution Mass Spectrometry *Analytical Chemistry,* 43**,** 1651-1656.

SOBKOWIAK, M., KOCHAN, J. I. & KRUSZYŃSKI, W. 2021. Assessing the efficiency of using passive hair traps as a method for non-invasive sampling from European beavers (*Castor fiber L.*). *Journal of Vertebrate Biology,* 71**,** 21053.

SOLAZZO, C. 2017. Follow-up on the characterization of peptidic markers in hair and fur for the identification of common North American species. *Rapid Communications in Mass Spectrometry,* 31**,** 1375-1384.

SONG, G., ZHAO, Q., DAI, K., SHUI, R., LIU, M., CHEN, X., GUO, S., WANG, P., WANG, D., GONG, J., FENG, J. & SHEN, Q. 2021. In Situ Quality Assessment of Dried Sea Cucumber (*Stichopus japonicus*) Oxidation Characteristics during Storage by iKnife Rapid Evaporative Ionization Mass Spectrometry. *Journal of Agricultural and Food Chemistry,* 69**,** 14699-14712.

STENSETH, N. C., FALCK, W., BJØRNSTAD, O. N. & KREBS, C. J. 1997. Population regulation in snowshoe hare and Canadian lynx: Asymmetric food web configurations between hare and lynx. *Proceedings of the National Academy of Sciences,* 94**,** 5147-5152.

STRATTON, R. W. M. & BURCH, R. L. 1959. *The principles of humane experiment technique*, Methuen.

STRITTMATTER, N., REBEC, M., JONES, E. A., GOLF, O., ABDOLRASOULI, A., BALOG, J., BEHRENDS, V., VESELKOV, K. A. & TAKATS, Z. 2014. Characterization and Identification of Clinically Revelevant Microorganisms Using Rapid Evaporative Ionisation Mass Spectrometry *Analytical Chemistry,* 86**,** 6555-6562.

STRYJEK, R., KALINOWSKI, A. & PARSONS, M. H. 2019. Unbiased Sampling for Rodents and Other Small Mammals: How to Overcome Neophobia Through Use of an Electronic-Triggered Live Trap—A Preliminary Test. *Frontiers in Ecology and Evolution* 7**,** 11.

SULTANA, R., OGUNDELE, O. M. & LEE, C. C. 2019. Contrasting characteristic behaviours among common laboratory mouse strains. *R Soc Open Sci,* 6**,** 190574.

TABERLET, P., MATTOCK, H., DUBOIS-PAGANON, C. & BOUVET, J. 1993. Sexing free-ranging brown bears *Ursus avctos* using hairs found in the field *Molecular Ecology,* 2**,** 399-403.

TAN, M. P., WONG, L. L., RAZALI, S. A., AFIQAH-ALENG, N., MOHD NOR, S. A., SUNG, Y. Y., VAN DE PEER, Y., SORGELOOS, P. & DANISH-DANIEL, M. 2019. Applications of Next-Generation Sequencing Technologies and Computational Tools in Molecular Evolution and Aquatic Animals Conservation Studies: A Short Review. *Evol Bioinform Online,* 15**,** 1176934319892284.

TAP, J., COOLS-PORTIER, S., PAVAN, S., DRUESNE, A., ÖHMAN, L., TÖRNBLOM, H., SIMREN, M. & DERRIEN, M. 2019. Effects of the long-term storage of human fecal microbiota samples collected in RNAlater. *Scientific Reports,* 9**,** 601.

TELLA, J. L. 2001. Sex ratio theory in conservation biology. *Trends in Ecology & Evolution*.

TENA, A., STOUTHAMER, R. & HODDLE, M. S. 2017. Effect of host deprivation on the foraging behavior of the Asian citrus psyllid parasitoid Tamarixia radiata: observations from the laboratory and the field. *Entomologia Experimentalis et Applicata***,** 51.

THE JACKSON LABORATORY. 2013. *Mouse Genome Database (MGD) Mouse Genome Informatics* [Online]. Bar Harbor, Maine. Available: http://www.informatics.jax.org [Accessed 15/03/2022].

TOLLESON, D. R., RANDEL, R. D., STUTH, J. W. & NEUENDORFF, D. A. 2005. Determination of sex and species in red and fallow deer by near infrared reflectance spectroscopy of the faeces. *Small Ruminant Research,* 57**,** 141-150.

TOZLU ÇELIK, H., ASLAN, F. A., US ALTAY, D., KAHVECI, M. E., KONANÇ, K., NOYAN, T. & AYHAN, S. 2021. Effects of transport and altitude on hormones and oxidative stress parameters in sheep. *PloS one,* 16**,** e0244911-e0244911.

TREMLETT, H., BAUER, K. C., APPEL-CRESSWELL, S., FINLAY, B. B. & WAUBANT, E. 2017. The gut microbiome in human neurological disease: A review. *Annals of Neurology,* 81**,** 369-382.

TROLLIET, F., HUYNEN, M. C., VERMEULEN, C. & HAMBUCKERS, A. 2014. Use of camera traps for wildlife studies. A review. *Biotechnology, Agronomy, and Society and Environment* 18**,** 446-454.

TURNBAUGH, P. J., RIDAURA, V. K., FAITH, J. J., REY, F. E., KNIGHT, R. & GORDON, J. I. 2009. The Effect of Diet on the Human Gut Microbiome: A Metagenomic Analysis in Humanized Gnotobiotic Mice. *Science Translational Medicine,* 1**,** 6ra14.

VAN DER WEYDE, L. K., MARTIN, G. B. & PARIS, M. C. J. 2016. Monitoring stress in captive and free-ranging African wild dogs (*Lycaon pictus*) using faecal glucocorticoid metabolites. *General and comparative Endocrinology,* 226**,** 50-55.

VOGEL, P., WEY, A. & SCHUBNEL, É. Evaluation of Muscardinus avellanarius population density by nest box and by trap checking. 2012.

VOM SAAL, F. S. & DHAR, M. G. 1992. Blood flow in the uterine loop artery and loop vein is bidirectional in the mouse: implications for transport of steroids between fetuses. *Physiology & Behavior,* 52**,** 163-171.

VON BORELL, E., LANGBEIN, J., DESPRÉS, G., HANSEN, S., LETERRIER, C., MARCHANT-FORDE, J., MARCHANT-FORDE, R., MINERO, M., MOHR, E., PRUNIER, A., VALANCE, D. & VEISSIER, I. 2007. Heart rate variability as a measure of autonomic regulation of cardiac activity for assessing stress and welfare in farm animals — A review. *Physiology & Behavior,* 92**,** 293-316.

WAGNER, I., KOCH, N. I., SARSBY, J., WHITE, N., PRICE, T. A. R., JONES, S., HURST, J. L. & BEYNON, R. J. 2020. The application of rapid evaporative ionization mass spectrometry in the analysis of Drosophila species—a potential new tool in entomology. *Open Biology,* 10**,** 200196.

WALTON, Z., SAMELIUS, G., ODDEN, M. & WILLEBRAND, T. 2018. Long-distance dispersal in red foxes *Vulpes vulpes* revealed by GPS tracking. *European Journal of Wildlife Research,* 64**,** 64-64.

WEARN, O. R. & GLOVER-KAPFER, P. 2019. Snap happy: camera traps are an effective sampling tool when compared with alternative methods. *Royal Society of Open Science,* 6**,** 181748.

WEAVER, J. L., WOOD, P., PAETKAU, D. & LAACK, L. L. 2005. Use of scented hair snares to detect ocelots. *Wildlife Society Bulletin,* 33**,** 1384-1391.

WEDEKIND, C. 2012. Managing Population Sex Ratios in Conservation Practice: How and Why? *Topics in Conservation Biology***,** 82-96.

WELDEN, H. L. A., STELVIG, M., NIELSEN, C. K., PURCELL, C., ECKLEY, L. B., M. F. & HVILSOM, C. 2019. The contributions of EAZA zoos and aquaria to peer-reviewed scientific research. *Journal of Zoo and Aquarium Research,* 8**,** 113-138.

WELTERING, A., SCHAEBS, F. S., PERRY, S. E. & DESCHNER, T. 2012. Simultaneous measurement of endogenous steroid hormones and their metabolites with LC-MS/MS in faeces of a New World primate species, Cebus capucinus *Physiological and Biochemical Zoology,* 105**,** 510-521.

WELTRING, A., SCHAEBS, F. S., PERRY, S. E. & DESCHNER, T. 2012. Simultaneous measurement of endogenous steroid hormones and their metabolites with LC-MS/MS in faeces of a New World primate species, Cebus capucinus. *Physiology & Behavior,* 105**,** 510-521.

WICKHAM, H. 2016. ggplot2: Elegant Graphics for Data Analysis.

WICKHAM, H., FRANCOIS, R., HENRY, L. & MULLER, K. 2021. dplyr: A Grammar of Data Manipulation. *R package,* v1.0.7.

WITMER, G. W. 2005. Wildlife population monitoring: some practical considerations. *Wildlife research,* 2005 v.32 no.3**,** pp. 259-263.

WOLFENSOHN, S., SHOTTON, J., BOWLEY, H., DAVIES, S., THOMPSON, S. & JUSTICE, W. S. M. 2018. Assessment of Welfare in Zoo Animals: Towards Optimum Quality of Life. *Animals (Basel),* 8.

WONG, C. H. F., LEUNG, D. K. K., TANG, F. P. W., WONG, J. K. Y., YU, N. H. & WAN, T. S. M. 2012. Rapid screening of anabolic steroids in horse urine with ultra-high-performance liquid chromatography/tandem mass spectrometry after chemical derivatisation *Journal of Chromatography A,* 1232**,** 257-265.

YANG, Y. X., ZHENG, N., YANG, J. H., BU, D. P., WANG, J. Q., MA, L. & SUN, P. 2014. Animal species milk identification by comparison of two-dimensional gel map profile and mass spectrometry approach. *International Dairy Journal,* 35**,** 15-20.

ZEALE, M. R., BUTLIN, R. K., BARKER, G. L., LEES, D. C. & JONES, G. 2011. Taxon-specific PCR for DNA barcoding arthropod prey in bat faeces. *Mol Ecol Resour,* 11**,** 236-44.