Next Generation Sequencing as a genomic used tool to detect species substitution and mislabelling in meat products sold in South Africa

by

NYARADZO STELLA CHAORA

submitted in accordance with the requirements

DOCTOR OF PHYLOSPHY

in the subject

LIFE SCIENCES

at the

UNIVERSITY OF SOUTH AFRICA

SUPERVISOR: DR FC MUCHADEYI

CO-SUPERVISORS: PROF FT TABIT

31 January 2023

DECLARATION

Name:	Nyaradzo Stella Chaora
Student number:	55794955
Degree:	Doctor of Philosophy in Life Sciences (Full Dissertation) (TFLSC01)
Title of a thesis:	Next Generation Sequencing as a genomic tool used to detect species substitution and mislabeling in meat products sold in South Africa

I declare that the above thesis is my own work and that all the sources that I have used or quoted have been indicated and acknowledged by means of complete references.

I further declare that I submitted the thesis to originality checking software and that it falls within the accepted requirements for originality.

I further declare that I have not previously submitted this work, or part of it, for examination at Unisa for another qualification or at any other higher education institution.

_____ _____31/01/2023_____

SIGNATURE DATE

ACKNOWLEDGEMENTS

I would like to give my utmost gratitude and praise to God the Creator and our heavenly Father. I am grateful for each day I have been given on earth and the ability to complete this degree, that had many challenging periods.

I would like to give a big thank you to my main supervisor Dr FC Muchadeyi, who has been present in my academic life since my Masters degree up to the completion of my PhD. Thank you for your guidance and never giving up on me, even during the time when I was struggling. Thank you for opening up my mind so that I can think scientifically. Thank you for demonstrating to me that women can conduct exceptional research in science that can be placed on the map. I will always be entirely grateful for your presence in my life. Dr RE Pierneef for his knowledge in bioinformatics that had an undeniable impact on my research. I would also like to thank Prof FT Tabit for his guidance, patience, and support during my research.

To the Agricultural Research Council – Biotechnology platform (ARC-BTP) staff and students, I am forever grateful for the friendships that I built during my time in the lab. To Miss L Mpye, thank you for your tough love. I did not understand it then, but it molded me to be a strong individual that never gives up. To Mr K Khanyile and Dr P Malatji, thank you for your mentorship and helping me settle in well when I joined the lab. To my lab mates, Dr J Bison, Dr P Hove, Dr K Hadebe, Miss M Ranketse and Miss Z Nocqobo, thank you for making the lab a fun environment and bearable when things were tough. To Mr G Mafireyi, you have truly been a blessing in my life.

I would also like to thank my family for your continued support and encouragement. I could not have concluded this journey without you as well. To my mother, for always encouraging me to go to school, so that I can make a better life for myself. To my father, for encouraging me to continue with school by studying for a PhD and speaking with pride to people when you mentioned that I

am studying for a PhD. That pride reminded me to stay strong when I felt like giving up. To my brothers Arthur, Anesu and Takudzwa, the levels you have reached in life pushed me to work towards the completion of this degree. To my cousin sister Linda and my sisters in-law Sarudzai, Michelle and Belinda, you have each played a pivotal role during the period that I studied this degree.

My gratitude also goes to the University of South Africa (UNISA), the Agricultural Research Council Professional Development Program (ARC – PDP) and Department of Forestry and Fisheries (DAFF) for financial assistance. Finally, I would like to thank my fiancé Mr BF Oluwasanmi for your continued support and for never losing faith in me. Thank you always wanting the best for me.

DEDICATION

I would like to dedicate this thesis to my mother Faith Julian Chaora for her never-ending support. This one is for you Mama.

I would also like to dedicate this thesis to my first-born child Ashirai Jeremy Chaora and my unborn children. Always finish what you have started my children. Never doubt yourself. The sky is the limit.

Philippians 4:13

I can do all this through him who gives me strength.

ABSTRACT

Meat species adulteration is a subject that has brought a lot of controversy to the meat industry and has led to meat adulteration studies being conducted worldwide. Correct labelling of meat products promotes fair trade and ensure that consumers make well informed decisions when choosing the meat they want to consume. Unfortunately, the unabated increase in the cost of resources has resulted in an increase in meat species substitution and/or mislabeling in the global meat industry. Deoxyribonucleic acid-based methods have previously been used in the past for species identification, however, they are limited to a few targeted species. Next Generation Sequencing is a universal technology that can be used to identify meat species. The main objective of this study was to evaluate genomics and bioinformatics pipelines that will facilitate utilization of genetic markers and Next Generation Sequencing (NGS) technology to identify and discriminate meat species in the South African meat industry.

The first study investigated the discriminatory potential of *16S rRNA*, *12S rRNA*, *COX3* and *ATP6* mitochondrial genes in mammalian species identification and differentiation. To achieve this, a phylogenetic analysis was conducted using the entire database of 262 mammalian species for each of the above genes downloaded from Genbank (http://www.ncbi.nlm.nih.gov/nuccore). All four genes managed to separate distantly related species and group closely related species with a common ancestry, however, the *ATP6* and *COX3* genes grouped some species that were not closely related together. Overall, the *16S rRNA* gene performed the best with bootstrap values of 97 – 100% in all clades observed whilst the *ATP6* gene performed the least. There was no improvement in performance when the *12S rRNA*, *COX3* and *ATP6* genes were individually combined with the *16S rRNA* gene. There were high bootstrap values of 100% observed in most clade groupings when all four genes were combined. Consequently, the study recommends the use of the *16S rRNA* gene on its own for species identification, as it performed well in comparison to the *12S rRNA*, *COX3* and *ATP6* genes. The use of four different genes in a species identification experiment will be more expensive and time consuming albeit yielding relatively higher bootstrap values.

The second experiment developed a molecular and bioinformatics diagnostic pipeline that utilizes the mitochondrial *16S* ribosomal RNA (rRNA) barcoding gene, to determine processed meat product mislabelling through Next Generation Sequencing. Pure meat samples were artificially mixed at different ratios, to verify the sensitivity and specificity of the pipeline. Processed meat samples (n = 155) namely, minced meat (n = 49), biltong (n = 28), burger patties (n = 35), and sausages (n = 43) were collected from across South Africa and sequenced using the Illumina MiSeq sequencing platform. All the species used in the artificially mixed pure samples were identified, confirming the specificity and sensitivity of the pipeline. Processed meat samples had reads that mostly mapped to the *Bos* (90% and above) genus, with traces of the *Ovis* and *Sus* (2 – 5%) genus. This confirmed that the majority of the processed meat samples were from beef. Amongst all processed meat samples, sausages had the highest level of contamination, with 46% of the samples having mixtures of beef, pork or mutton in one sample, which was in contrast to the labelling, as the only labelling provided was of samples labelled as beef sausages. The pipeline further demonstrated its specificity by identifying species with percentages as low as 0,1% in both the artificially mixed pure samples and processed meat samples. Overall, the developed pipeline can be used with confidence to authenticate meat products and furthermore, investigate and manage any form of mislabelling in the meat industry.

The third and last experiment investigated the presence of breed-specific Single Nucleotide Polymorphisms (SNPs) using the entire mitochondrial genome of 13 European and Indigenous cattle breeds reared in South Africa for use in breed assignment and traceability. Whole genome sequencing was performed on 13 European and Indigenous cattle breeds reared in South Africa. A total of 42 animals were used from Afrikaner (n = 4), Beefmaster (n = 4), Boran (n = 4) Charolais (n = 2), Hereford (n = 2), Nguni (n = 2), Simbra (n = 3), Bonsmara (n = 4), Brahman (n = 4), Drakensberger (n = 4), Limousin (n = 2), Santa (n = 3) and Simmentaler (n = 4) breeds. Whole genome sequencing was performed using the Illumina HiSeq 2500 (Illumina, San Diego, CA, United States) at 10X coverage. A total of 12 996 variants were identified and of these 12 633 were SNPs and 363 were Indels. The highest number of variants were identified in the European Brahman breed (n = 2 066) and the lowest in the Indigenous Nguni breed (n = 340). The SNPs were also divided into homozygous and heterozygous SNPs. The highest number of homozygous

SNPs were found in the Limousin breed (n = 534) and the highest number of heterozygous SNPs were found in the Brahman breed (n = 1 872). To identify breed-specific SNPs we used all homozygous SNPs identified that have the same alleles. A total of 125 breed-specific SNPs were identified in all breeds except for the Charolais breed that did not contain any breed-specific SNPs. The Limousin breed had the highest number of breed specific SNPs (n = 59) and the lowest were found in the Nguni breed (n = 1). The *COX3* mitochondrial gene had the highest number of breed specific SNPs (n = 22), followed by the *16S* mitochondrial gene (n = 19). Nineteen of the breed-specific SNPs were shared amongst breeds and the *ND5* gene contained the highest number of shared SNPs. This study provides an insight on the presence of SNPs within the mitochondrial genome of cattle breeds reared in South Africa. The breed specific SNPs identified provided an understanding of the regions within mitochondrial genes that are unique in each breed and can be used in the authentication of beef meat in the meat industry.

In conclusion the study illustrated that NGS can be used as a genomic tool to detect meat species mislabelling in meat products in South Africa and worldwide. Mitochondrial sequences were used to determine the discriminatory potential of four genes (*16S* rRNA, *12S* rRNA, *COX3* and *ATP6*) in mammalian species. The *16S* rRNA gene demonstrated the highest discriminatory potential amongst the four genes. Using the *16S* rRNA gene, NGS technology was further used to identify meat species in both artificially mixed pure and processed meat samples. The NGS technology proved that it can be used as a universal tool in meat species identification. Finally, using NGS cattle breeds reared in South Africa were sequenced and the sequences were used to identify breed specific SNPs in the mitochondrial genome of the cattle breeds. The mitochondrial genes that contained the most breed-specific SNPs were within the *ND5*, *COX3* and *16S* rRNA genes.

Key Terms: Meat adulteration, Meat industry, Processed meat, Mitochondrial genes, *16S rRNA* mitochondrial gene, Next Generation Sequencing, Whole genome sequencing, Cattle breeds, Breed-specific SNPs.

REASEARCH OUTPUTS

RESEARCH ARTICLES

Nyaradzo Stella Chaora, Khulekani Sedwell Khanyile, Kudakwashe Magwedere, Rian Pierneef, Frederick Tawi Tabit and Farai Catherine Muchadeyi. A *16S* Next Generation Sequencing Based Molecular and Bioinformatics Pipeline to Identify Processed Meat Products Contamination and Mislabeling. *Animals* **2022**, *12*, 416. https://doi.org/10.3390/ani12040416

Nyaradzo Stella Chaora, Godwin Mafireyi, Frederick Tawi Tabit and Farai Catherine Muchadeyi. Development of SNP marker panel for discrimination and assignment of cattle breeds reared in South Africa using next generation sequencing technologies. In preparation.

Conference Presentations

Chaora, NS, Tabit, FT, Pieneef, RE., Khanyile, KS, Muchadeyi, FC. A *16S* Next Generation Sequencing Based Molecular and Bioinformatics Pipeline to Identify Processed Meat Products Contamination and Mislabeling. BIO Africa Convention, Durban, 26 - 28 August 2019.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER FOUR

LIST OF TABLES

CHAPTER TWO

CHAPTER THREE

CHAPTER FOUR

CHAPTER FIVE

LIST OF APPENDICES

CHAPTER THREE

CHAPTER FOUR

LIST OF ABBREVIATIONS

nDNA	nuclear deoxyribonucleic acid
rRNA	ribosomal ribonucleic acid
mtDNA	mitochondrial deoxyribonucleic acid
DAFF	Department of Agriculture, Forestry and Fisheries
USD	United States Dollar
USDA	United States Department of Agriculture
FAO	Food and Agriculture Organization
RFLP-PCR	restriction fragment length polymerase chain reaction
dd PCR	droplet digital polymerase chain reaction
FTIR	fourier transform infrared spectroscopy.
STRs	short tandem repeats
PCR	polymerase chain reaction
NGS	Next Generation Sequencing
DNA	deoxyribonucleic acid
DTI	Department of Trade and Industry
COX1	*cytochrome c oxidase I*
COX2	*cytochrome c oxidase II*
COX3	*cytochrome c oxidase III*
SNPs	single nucleotide polymorphisms
bp	base pair
UNISA	University of South Africa
ASF	African Swine Fever
USA	United States of America
UK	United Kingdom
EMA	Economically Motivated Adulteration
RASFF	Rapid Alert for Food and Feed
ELISA	Enzyme-linked immunosorbent assay
d-loop	displacement loop

ND5	*NADH-ubiquinone oxidoreductase chain 5*
Cyt b	*cytochrome b*
RNA	ribonucleic acid
ATP6	*ATPase subunits 6*
ATP8	*ATPase subunits 8*
MAFFT	Multiple alignment using Fast Fourier Transform
GATK	Genome analysis toolkit
BWA	Burrows Wheeler Aligner
Dalrrd	Department of Agriculture Land Reform and Rural Development
CNVs	Copy Number Variations
DoH	Department of Health
EFSA	European Food Safety Authority
LIBS	Laser Induced Breakdown Spectroscopy
AFLP	Amplified fragment length polymorphism

CHAPTER ONE

1. INTRODUCTION

1.1. BACKGROUND

Meat is defined as the dressed flesh of animals that is edible (Rout, 2018). It includes the skeletal muscle, fat, edible tissue, offal and organs (Rout, 2018). Amongst the various agricultural sectors worldwide, the meat industry is expanding economically. The industry is powered by high income and reinforced by improvements in technology (DAFF, 2017). A total of R127.3 billion was contributed to the Gross Domestic Product in 2016 – 2017 from the South African livestock industry with a growth of 11.3% during that period (DAFF, 2017). Overall animal products contributed 46% of income amongst all agricultural activities in South Africa. Of the meat products, the largest contribution was from poultry meat (DAFF, 2017). Globally, the meat industry was expected to rise from USD714 billion in 2016 and contribute USD1,5 trillion by 2022 due to the prediction of an increase in the global population. Similar, to the market trend in South Africa, the largest global contributor is from poultry meat (Shahbendeh, 2019). Currently, the livestock industry provides 40% of the world's agricultural outputs and is a source of income for the public. Livestock provide 15% of energy in food and 25% of protein in diets (DAFF, 2017).

Meat processing is modifying the properties of fresh meat using procedures such as mincing, grinding, chopping, salting, curing, addition of seasoning and heat treatment (Larsson and Orsini, 2014). These procedures extend the shelf life of meat, improve its intake and quality, and add flavor to its original composition (Shahbendeh, 2019). Processed meat can be categorized based on the type of meat that includes beef, pork, poultry, mutton and wildlife. Depending on the meat type, processed meat can further be categorized into chilled, frozen and canned or preserved processed meat. In addition, meat can be processed into value-added products that include, burger patties, sausages, cured meat, corned meat and biltong. The market value of processed meat worldwide was approximately USD 670,5 billion in 2019 and is predicted to rise to USD 11140,42 trillion by 2023 (Zion Market Research, 2017). North America holds the highest market share of processed meat, mainly due to the U.S. market, followed by Europe (France, Germany and U.K.), Asia Pacific (China, Japan and India), Latin America (Brazil) and lastly The Middle East and Africa (Zion Market Research, 2017). The

rise in revenue from processed meat is attributed to an increase in income from consumers in emerging countries (Zion Market Research, 2017). In addition, there have been an increase in the variety of processed meats that are sold at lower prices and have resulted in an increased demand for processed meat (Zion Market Research, 2017). Furthermore, there is a rise in consumer awareness on the need for high protein content in meat coupled by an increase in the demand for quick meals (Shahbendeh, 2019). Poultry is the most prevalent type of processed meat worldwide and accounts for 38 % of the global market, followed by red meat (beef and pork) which account for 33% (Shahbendeh, 2019).

The price of meat in South Africa is constantly fluctuating, which puts pressure on a large part of the population (USDA, 2015). South Africans spent approximately R165 billion on meat products in 2014. This was a 28% increase from what South Africans spent in 2004 which was R46 billion (USDA, 2015). The consumption of white meat in South Africa has increased significantly more than that of red meat between the years 2000 – 2014. In the year 2000, South Africans were consuming 21.5kg/capita/year of white meat. This amount increased by about 80% to 38.5kg/capita/ year in 2014 (USDA, 2015). The consumption of red meat however, increased moderately at 22.4kg/capita/year to 26.6kg/capita/year between 2000 and 2014. This was a 19% increase over those 14 years (USDA, 2015). Poultry meat is now a common protein source in South Africa, seeing as that it is relatively cheaper and readily available. Trends in the retail price of various meat products sold in South Africa indicates that chicken is the cheapest source of protein (R/kg) (USDA, 2015). In 2015 the cost per kg of animal protein from highest to lowest for lamb, beef T-bone, beef mince, pork chops, fresh chicken portions, fresh whole chicken and frozen chicken portions were R109.30, R81.15, R65.36, R62.25, R51.12, R39.96 and R29.29, respectively (USDA, 2015). Unfortunately, there are no recent published scholarly articles on the cost of meat in South Africa. However, a report on South African meat market analysis reported a meat market value of $7,5 billion in 2020, which is expected to grow by more than 3% during the 2021 – 2025 forecast (Global Data, 2021). The average price for beef per kg between 2017 – 2022 has increased from R129,82 to R180,32; pork per kg from R77,79 to R101,56; lamb per kg from R140,14 to 185,75 and whole chicken from R43,41 to R54,34 (South African Market Insights, 2022).

Meat intake is affected by various factors, namely, income, geography, season, religion culture, ethnicity, social networks, and the type of consumer (FAO, 2019). The global amount of meat

consumption is 43.22kg. Australia has the highest level of meat consumption, consuming 116.23kgs per capita. This is followed by the Americas, where North America consumes 112.72kgs and South America consumes 81.49kgs. Europe consumes 77.34kgs and Asia and Africa consume 32.55kgs and 19.01kgs, respectively (Erasmus, 2017). Meat is an essential component in the diet of South Africans. South Africans are popular for having barbeque meat (known as braai in Afrikaans) and biltong (a dried preserved meat) (FAO, 2019). Total meat consumption was projected to be 64.92kgs kg per capita in South Africa, which is higher when compared to the global consumption average of 43.22 kg per capita (Rout, 2018).

The upsurge in the human population has resulted in an increase in the demand for meat. This has resulted in an escalation in meat product costs, leading to its susceptibility to fraudulent adulteration, substitution and mislabeling. Meat adulteration can be defined as degrading or making a meat product impure by adding or mixing it with inferior or harmful substances (di Pinto *et al.*, 2015). Meat adulteration is more prevalent in processed meat because it is difficult to identify the physical properties of a meat type morphologically after it has been processed. Meat species substitution and fraud in processed meat products is influenced by the rise in meat imports, price of meat commodities, growth of food trade and the rise of prices for processed meat (Doosti *et al.,* 2014). The substitution of meat species can either be intentional or unintentional. An example of intentional substitution is adding pork, which costs less, into beef or mutton products for economic gain (Ha *et al.*, 2017). Incidences where species substitution in meat may be unintentional, is through cross contamination from using equipment that has not been cleaned properly after grinding meat from different species. Ideally, different equipment should be used for different meat species to avoid cross-contamination (Singh and Neelam, 2011).

Deoxyribonucleic acid (DNA) is the most applicable technology for identification of species (Yang *et al.*, 2018). The advantages of using DNA are that it is stable at relatively high temperatures and can be analyzed in processed, mixed, frozen, and fresh products (Cermakova *et al.*, 2023). Deoxyribonucleic acid is present in all tissue cells and can differentiate species that are closely related (Cermakova *et al.*, 2023). The success of DNA methods in identifying species depends on, extracting the correct amount of DNA from the tissue of interest, the quality of DNA and developing databases that have reference DNA sequences from different

species where unknown sequences can be compared (Böhme *et al.*, 2019). The types of DNA that are evaluated are nuclear DNA (nDNA), ribosomal DNA (rDNA) and mitochondrial DNA (mtDNA). Mitochondrial DNA is highly discriminative and has a higher copy number than nDNA. Nuclear DNA only has a single copy per cell, whilst mtDNA is found in numerous copies per cell and exists with a double membrane organelle, making it more stable over time and conditions (Unajak *et al.*, 2011; Böhme *et al.*, 2019). Phylogenetic applications are possible with the use of mtDNA, owing to its high mutation rate that increases 5-10 times more compared to a single copy nuclear gene. As a result, detection of different animal species is enabled using various mtDNA genes (Zhang *et al.*, 2020). However, although mtDNA genes can identify unknown species in a meat sample, quantification of the amount of DNA of that particular species in a sample is often difficult and inaccurate, mainly due to mtDNA having a variable content of mitochondria in different mammalian tissues (Zhang *et al.*, 2020). Generally, quantification results should be based on genome: genome and not weight: weight proportions, due to the differences in tissue composition, species genome size, DNA degradability and extractability (Floren *et al.*, 2015a). Therefore, the quantification of species in processed food requires the use of a nuclear gene because it contains a single copy per cell and therefore, avoids over or underestimation of results (Ha *et al.*, 2017).

The most common DNA methods being used for meat species identification are species-specific PCR (Doosti *et al.,* 2014), restriction fragment length PCR (RFLP-PCR) (Kurniawati *et al.,* 2014), droplet digital PCR (Floren *et al.*, 2015a), fourier transform infrared spectroscopy (FTIR) (Omran *et al.*, 2019) and PCR-based short tandem repeats (STRs) (Rasmussen *et al.,* 2011). Most of the DNA based methods for species identification concentrate on single-species DNA sources, that are established from amplification of PCR using species-specific primers. In meat species identification the use of species-specific PCR methods is inefficient in cases where mixed samples need to be analyzed and no prior information of the samples is available (Tillmar *et al.*, 2013). A species-specific PCR method is unable to concurrently detect different DNA components in a mixture. Information on what to look for is required for the results to be accurate (Cottenet *et al.*, 2020). In instances where different species are present, specific primers would be needed for each different species, thereby making the work time consuming and expensive. There is therefore a need to establish a universal typing method in circumstances where previous information about the species is not obtainable or when there is

more than one species in the sample. There may be cases where prior knowledge of the species is not available and requiring a universal typing method that can detect various species in mixed samples.

There are two new DNA based methods for species identification that have emerged in the past few years. These are microarrays or chips-based methods and Next Generation Sequencing (NGS) technologies (Rasmussen *et al.,* 2011). Both these methods can identify species using high throughput sequence information and enable the analysis of large amounts of samples concurrently. However, DNA microarrays contain probes that are species specific and therefore prior information/ hypothesis on the species is needed. The use of NGS is universal, allowing for advanced automation and use of high-throughput information in identification of species (Tillmar *et al.*, 2013). After an NGS analysis, many individual sequences are produced that may be present in food containing various species. With NGS each ingredient in the sample produces a single and unique DNA sequence of the relevant species. The sequence can be compared with different databases, resulting in a list of all species present in the sample including the scientific names. Next Generation Sequencing has improved the capacity of gaining sequenced data from single molecules with low quality or degraded DNA sources (Gupta and Verma, 2019). The use of NGS in meat species identification is advantageous because it is a universal method that requires no prior information on suspected species and allows simultaneous detection of different species in a mixed source (Liu *et al.*, 2021).

Various studies have been conducted on the use of NGS in meat species identification in processed meat samples. Xing *et al.*, (2019) used NGS as an untargeted tool to identify meat species in ground meat and detected up to 1% adulteration of undeclared species. High throughput metabarcoding successfully identified adulteration in processed meat and poultry products sold in China (Pan *et al.*, 2020) including low frequency animal species in meat products. In the same study (Pan *et al.*, 2020), a combination of a cytochrome oxidase 1 mini-barcode with NGS identified adulteration in processed meat products. The use of NGS was superior in accuracy, sensitivity and detection efficiency in meat species identification of processed meat samples sold in South Africa by revealing undeclared species in processed meat samples (Cawthorn *et al.,* 2013). In the study by Cawthorn *et al.,* (2013), Pork, Asian water buffalo and chicken DNA were detected in samples labeled as ‘beef’ whilst goat and cattle

DNA were detected in samples labeled as ‘mutton’ and ‘venison’, and gluten and soya were detected at exceeding high levels of over 1 000mg kg^{-1} in samples that were not labelled to consist of such species. In the same study, common games species (kudu, gemsbok, ostrich, impala, and springbok) farmed or hunted in the South Africa market were substituted with domestic species like cattle. Furthermore, wildlife species (giraffe, waterbuck, bush buck, duiker and mountain zebra) not common in the South African market were found to be present and undeclared on the product label (D’Amato *et al.*, 2013a).

Mitochondrial DNA barcoding genes can be sequenced without knowledge of which species are present and without targeting specific species using the NGS method. This then enables the identification of every species present in the sample and not just those suspected. Such a method will be of utility in South Africa where meat adulteration has been cited and meat products are often not accurately labeled. There is a need to evaluate emerging genomics and bioinformatics pipelines that can be used in species identification, streamline data interpretation and can therefore, be adopted by the South African meat industry as diagnostic methods.

1.2. PROBLEM STATEMENT

Processed meat products in some cases include meat from different animal and plant species. Once meat is processed into a value-added product, it is difficult to confirm the authenticity of product composition morphologically. Adulterated meat products are often sold on the market for financial gain. Meat consumers on the other hand have the right to be informed on the composition of the product to make informed decisions and choices of meat products based on their health, religious, ethical and environmental reasons amongst other factors. Identification of species that are present in meat products is important for health, religious, economic and legal reasons. Cases of meat species adulteration are to date still being reported in the meat industry (Cottenet *et al.*, 2020; Zhang *et al.*, 2020; Chaora *et al.*, 2022).

The intentional substitution of meat products was first reported in South Africa through (Wiener, 2011). In 2011 it was reported that the Orion Cold Storage in Cape Town was bringing in different meat products and knowingly relabeling these as food-grade and specifying it to

conform to the standards of being Halal (Wiener, 2011). These results sent panic to the Muslim community and food industry. South African studies following this scandal reported that 68% of samples analyzed comprised of species that were not mentioned on the label. The highest occurrence was found in deli meats, burger patties and sausages. The addition of non-declared plant proteins like soya and gluten and animal species like pork and chicken was also detected (Cawthorn, Steinman and Hoffman, 2013). In another South African study on game meat, 76.5% adulteration with domestic species was reported. Domestic species like lamb, cattle, pigs and horses; popular game species like ostrich, springbok, gemsbok, impala and kudu; rare game species like mountain zebra, water buck, duiker, giraffe and bushbuck; and extra-continental animals like kangaroo were found mislabeled (D'Amato *et al.*, 2013a). In yet another South African study carried out in the Durban Metropolitan area, 40 processed meat samples analyzed showed that 65% were contaminated by other meat types not labelled. The contamination was mostly found in sausages followed by patties (Tembe *et al.,* 2018). Furthermore, amidst the reports of falsely labeled or mixed meat products, there were no clearly defined methods to verify the authenticity of the products. There is therefore a need for robust and scientific methods to verify meat products sold in the South African industry.

Food labelling regulations require that ingredients in food products are accurately declared to consumers. The governing organizations in South Africa have issued new legislations to encourage clarity and accurate explanation of food products, in response to the consumer's needs. These are, the controls linked to Advertising and Labeling of Foodstuffs (R. 146/2010) comprising of a compulsory ingredient list on food labels (DoH, 2012) and the Consumer Protection Act (R. 147/2009), which safeguards individuals from manipulation in the sale and marketing of goods (DTI, 2009). Food labeling regulations are common beyond South Africa. The Food Regulation standard in Europe requires that meat products should be accurately labeled with information that includes the composition and percentage of ingredients included in the products (European Commission, 2002).

There are instances where meat is sold as a specific high value product or most preferred breed, for example Angus Beef in the USA (Bass, 2016). In Ireland, a DNA technique called IndentiGen was developed to trace the origin of domesticated cattle and trace a piece of steak to the animal and farm origin. This technique was adopted by the United States (US) beef industry and restaurants in the US now serve Braveheart Black Angus Beef, that is verified to

be from all-vegetation, corn-fed Angus cattle (Pant *et al.*, 2012). The meat industry generally seeks to develop high quality breeds of animals. This has resulted in some breed names being marketed as brand names, making authentication of breeds vital to ensure food safety in both global and domestic markets (Cheong *et al.*, 2013). For example, the Hanwoo breed is a Korean native breed, favoured for its marbling and carcass quality. Meat labelled as 'Hanwoo beef' in Korea costs more than meat from other local and imported breeds. Such market trends and consumer demands require a species/breed identification system, that can differentiate among the beef breeds and authenticate the meat sold (Cheong *et al.*, 2013). A similar need exists in South Africa that boosts of multiple indigenous and exotic beef breeds in a highly commercialised industry.

1.3. RELEVANCE OF THE STUDY

The determination of species substitution or mislabeling in processed meat is vital for (i) ensuring that fraudulent traders do not gain economically and (ii) the rights of consumers for either health, religious or ethical reasons are met. Furthermore, determination of species substitution or mislabeling in processed meat ensures fair trade and compliance with legislation. Meat adulteration is problematic in the meat industry in South Africa and globally, although the true extent is not known. Coupled to this is globalization that has led to the possibility of processed meat containing ingredients from different regions of the world.

To date, only a few recent studies have been carried out to verify the occurrence of species substitution in the South African meat industry (Cawthorn *et al.,* 2013; D'Amato *et al.*, 2013; Tembe *et al.,* 2018). These previous studies have confirmed that meat adulteration and mislabeling is present in the South African meat industry. These studies undertaken to date were however species specific and can only identify desired species in a sample and not those unknown or unsuspected. The current study will develop a universal method for species identification, that can identify any species present in a sample. The study will develop a diagnostic molecular and bioinformatics protocol that can be used for routine laboratory testing of meat samples within the meat industry. This protocol can be used to authenticate meat products and will instill confidence in both manufacturers and consumers on the meat products sold within the industry.

1.4. AIMS AND OBJECTIVES

The overall aim of this study is to:
Evaluate and develop genomics and bioinformatics pipelines that will facilitate utilization of next generation sequencing technology to identify and discriminate meat species in the South African meat industry.

The fulfillment of these aims will be achieved by the following objectives:

1. To conduct a multi-locus phylogenetic analysis of mammalian species to determine the discrimination power of the *16S rRNA*, *12S rRNA*, *ATP6* and *COX3* mitochondrial genes.

2. To develop a universal molecular and bioinformatics pipeline, that can utilize the 16S rRNA mitochondrial gene to identify processed meat product mislabeling/contamination using NGS.

3. To investigate the presence of breed-specific SNPs using the entire mitochondria of 13 European and indigenous cattle breeds reared in South Africa for use in breed assignment and traceability.

1.5. RESEARCH QUESTIONS

The study's main research question is:

How can NGS be used as a universal method for meat species identification in processed meat samples?

To answer this question the following sub-questions are addressed:

Which gene amongst the mitochondrial genes *16S rRNA*, *12S rRNA*, *ATP6* and *COX3* have the discriminatory potential to be used in meat species identification?

Can the sequencing of the mitochondrial *16S* rRNA gene be used in a universal genomic and bioinformatics pipeline for meat species identification?

Can breed specific SNPs in beef cattle breeds reared in South Africa be used to trace and assign breeds?

1.6. DISSERTATION LAYOUT

This study is made up of six (6) chapters, organised as follows:

1.6.1. Chapter 1: Introduction

This is an introductory chapter to the study, which gives the background and overview of the research. Included in this chapter are the background, problem statement, relevance of the study, aim and objectives, research questions and dissertation layout.

1.6.2. Chapter 2: Literature Review

This chapter gives an overview of existing literatures on the global meat industry, processed meat products, livestock production trends, meat species substitution and methods used to identify meat species substitution.

1.6.3. Chapter 3: A multi locus sequence analysis of the *16S* rRNA, *12S* rRNA, *COX3* and *ATP6* mitochondrial genes to determine their discriminatory potential in mammalian species

This chapter is the first experimental chapter that investigated the discriminative potential of the mitochondrial *16S* rRNA, *12S* rRNA, *COX3* and *ATP6* genes using mammalian sequences downloaded from an online database.

1.6.4. Chapter 4: A *16S* Next Generation Sequencing Based Molecular and Bioinformatics Pipeline to Identify Processed Meat Products Contamination and Mislabelling

This is the second experimental chapter that developed and demonstrated the utility of a universal molecular and bioinformatics pipeline, that utilizes the *16S* rRNA mitochondrial gene to identify processed meat product mislabeling and / or contamination.

1.6.5. Chapter 5: Development of discrimination SNP markers for assignment of cattle breeds reared in South Africa

This is the third and last experimental chapter that investigated SNPs that can be used to assign South African beef breeds. The identified SNPs can be used to trace the type and origin of beef breeds sold in the South African market.

1.6.6. Chapter 6: General Discussion, Conclusions and Recommendations

This chapter discusses the findings from all experimental chapters, provides conclusions on the study objectives and makes recommendations for future studies.

CHAPTER TWO

2. LITERATURE REVIEW

2.1. Abstract

Meat species mislabelling/substitution is prevalent globally and is a topic that needs to be further addressed in South Africa. Understanding the meat market trend in South Africa and the world around will aid in finding out why meat species fraud occurs, and which meat types are most affected. This information will help to find ways on how to prevent meat species mislabelling/substitution from occurring. Even though there are studies that have been conducted on identification of meat species substitution/mislabelling, cases of meat fraud are still being observed. There is therefore, a need to find permanent solutions to prevent meat fraud and put stringent measures towards those that are still committing meat fraud. Amid prevalence of meat species adulteration and contamination, there is lack of both diagnostic and monitoring tools for meat species identification to support the sector.

This review explores the global meat production and trade, meat products that are sold worldwide, the South African meat industry, food fraud reporting systems, commonly substituted species and methods that can be used to identify meat species identification/substitution. The review further discusses the available and upcoming molecular and genomic tools and their potential in meat species identification. Understanding the meat industry, species that are commonly substituted and the current methods for species identification, will inform the development of methods and pipelines that can be implemented to facilitate precise identification of species in meat products and ensure that (i) legislation related to meat products are enforced and adhered to, (ii) the standards of meat products maintained, (iii) unfair competition in the meat industry is avoided, (iv) consumer rights and religious and social traditions are upheld and (v) poaching of wildlife animals and protected species is minimised.

Keywords: Meat species substitution, meat production and trade, South African meat industry, Next generation sequencing

2.2. Introduction

The first evolutionary change in the human diet was the incorporation of meat and bone-marrow from large animals, which occurred about 2.6 million years ago (Wyness, 2016). Meat is a valuable source of protein and other micro and macro nutrients such as vitamin A, Vitamin D, Vitamin B, Vitamin B12, niacin and bioavailable zin and iron (FAO, 2012). The importance of meat is derived from its high-quality protein, which contains the essential amino acids. Despite the aforementioned advantages of meat, consumption of excess red meat can result in the development of cardiovascular disease, cancer, high cholesterol and dementia (Giromini and Givens, 2022).The common sources of meat are domesticated species such as, cattle, poultry, pigs, sheep and to a lesser extent goats (Cawthorn *et al.,* 2013). Other animal species like camels, horses, ostriches and game animals are a source of meat (D'Amato *et al.*, 2013). To a limited extent, meat is also derived from reptile animals such as crocodiles, snakes, and lizards. Worldwide, pigs are the most consumed type of meat species, accounting for 36.3% of consumption, followed by poultry, cattle, sheep/goats accounting for 35.2%, 22.2% and 4.6% respectively (USDA, 2015). In 2015, consumption of poultry, beef and pork in South Africa was recorded at 2.9 million tons per year and poultry meat accounted for 60% of the total meat consumption (Ritchie, 2017). The global meat production records indicate that Asia was the largest meat producer in 2014 with 40 – 50% of total meat production followed by Europe and North America at 19% and 15%, respectively. This was a drastic change from previous decades, were in 1961 Europe and North America accounted for 42% and 25%, respectively of the global meat production and Asia recorded 12% (Ritchie, 2017). The contribution from Africa towards the global meat production remained stable between 1961 and 2015, ranging from 5.55 to 5.8% (FAO, 2020b).

2.3. Global trends in meat consumption, production, and trade

2.3.1. Global Meat production and trade

The global meat output in 2019 was estimated at 335 million tonnes, however, this was 1% lower than the output in 2018. This decline was caused by the spread of African Swine Fever

(ASF), which originated from China and spread through the East Asian countries. The impact of ASF resulted in the end of a 20-year stable increase of global meat output. China recorded an 8,8% decline in meat production, which is equivalent to 7.8 million tonnes. This was due to a 21% decline in pig meat (FAO, 2020b). Due to ASF causing a decline on pig meat output, China was forced to import meat and countries such as the USA, Brazil, India, Mexico, Russia, Canada, and Argentina benefited. Global meat exports were recorded at 36 million tonnes in 2019. The largest growth was seen in pork, followed by beef and poultry. Mutton, which accounts for 3% of global meat exports, remained stable in 2019 due to a decline in production in Zealand. However, there was an increase in exports from Australia, resulting in the stability experienced in 2019. China imported 37 % (approximately 2 million tonnes) more meat in 2019 than in 2018, due to high meat shortages. In contrast, several countries, such as South Africa, Angola, Iraq, Saudi Arabia, Vietnam, Russia, and the USA, reduced importation of meat. Reasons relating to the decline were increases in domestic meat production, matters relating to meat certificate requirements and concerns over animal diseases. Despite increased domestic production in the USA, trade disputes inhibited export of meat from the USA, which is the second largest meat producer in the world. The regions that had high imports of meat in 2019 were Europe, Brazil, Argentina, Mexico, Ukraine and Thailand (FAO, 2020a). Unfortunately, there is no recent statistics on South Africa's meat trade, however, in 2015 FAO predicted a 38% increase in poultry consumption, a 28% increase in beef consumption and a 33% increase in pork consumption in the next 10 years (USDA, 2015). It was then recommended for South Africa to import meat, as the current production would not meet consume demands in the future (USDA, 2015). South Africa has been globally ranked 8th in terms of poultry consumption and 16th in terms of beef consumption (Murcott, 2021). The table below illustrates global meat statistics worldwide between 2018 and 2019.

Table 2. 1: Global meat statistics (000 tonnes, carcass weight equivalent)

	[1]Production		**Imports**		**Exports**	
	2018	**2019**	**2018**	**2019**	**2018**	**2019**
Africa						
South Africa	3241	3309	641	600	137	135
Algeria	806	810	66	68	2	2
Angola	262	265	630	518	0	0
Egypt	2152	2210	304	353	8	4
Nigeria	1451	1449	2	3	1	1
Asia						

China	88105	80340	5485	7496	637	603
India	7454	7747	1	2	1472	1451
Saudi Arabia	762	781	866	827	78	87
Viet Nam	5228	4845	1024	956	30	19
Thailand	2933	3012	32	29	1304	1371
The Americas						
Mexico	7051	7314	2231	2288	473	565
Argentina	5930	6114	85	63	748	1072
Brazil	29341	30244	58	54	6942	7493
USA	46870	48262	2178	2141	7901	7953

FAO 2020: [1]Production refers to carcass weight. The data above includes official, non-official and estimates at the time the report was written.

2.3.2. Meat Products

Processed meats are meats that have undergone a further processing step such as grinding smoking, salting, curing or adding any chemical preservatives (Shan *et al.*, 2016). These steps may change the appearance texture or taste of the processed meat. Various number of processed and semi-processed meat products are produced from meat. These products vary in taste and shape depending on the method of processing (Sen, Antara and Sen, 2021). Meat products are characterized into the following categories:

2.3.2.1. Fresh processed meat products

These meat products are made from raw meat and fatty tissues that are ground, sliced, diced and were sometimes spices and common salt is added to them. They are sold raw but cooked before consumption to make them more appetizing. Examples of these are burger patties, raw sausages and minced meat (FAO, 2020a).

2.3.2.2. Cured meat products

These meat products consist of muscle pieces and can either be cured raw or cured cooked meat. The curing procedure involves dry salting the meat or submerging and/or injecting in a curing salt solution. Cured meats are normally consumed raw and an example of these are sandwich ham (FAO, 2020a).

2.3.2.3. Raw cooked meat products

These products include muscle meat, fat, and non-meat ingredients. Non-meat ingredients need to be grinded, chopped, or mixed beforehand. The mixture is then used to make sausages or meat loaves that are heated treated resulting in protein coagulation, aiding in palatability and bacterial stability (FAO, 2020a). Examples of the products are hotdogs, frankfurters, viennas and meat loaf.

2.3.2.4. Pre-cooked meat products

These products consist of a mixture of various constituents like lower-grade muscle trimmings, animal skin, fatty tissues, blood, head meat, edible body parts and liver. Processing of precooked-cooked products has two manufacturing processes that involve precooking the raw meat followed by cooking of the final mixture. These meat products make use of non-meat ingredients, animal by-products and a wide variety of meat. Examples include corned beef, liver pates and blood sausages (FAO, 2020a).

2.3.2.5. Raw-fermented sausages

These types of meat products comprise of a mixture of lean meat, fatty tissue, curing salt, sugars spice and other non-meat ingredients. All these are stuffed into casings. The flavor, color and texture come about from a fermentation process together with a reduction in moisture. Final products are consumed raw, and examples include salami and sausages (FAO, 2020a).

2.3.2.6. Dried meat products

These types of products come about from drying lean meat. The lean meat is cut into pieces that are of the same shape and gradually dried. Dried meat does not easily spoil because a significant part of the natural tissue will have evaporated from the drying process. Examples are jerkey and biltong (DAFF, 2018a).

Processed meat products are favored especially in working households due to their convenience of being easy to prepare, ready to cook and full of flavor. They are also easy to prepare because they may have undergone a further processing step that may involve pre-cooking and heating. Some processed meats are ready to eat and do not need to be cooked further (Shan *et al.*, 2016; Sen, Antara and Sen, 2021). These favorable attributes of processed meat result in the meat types being easily contaminated or substituted by other species, because they will have undergone changes during processing that will change the appearance of their original appearance. In some cases, by-products of different meat species may be used. For instance when making sausages the casing used to make the sausages may be a by-product of a different species and there is no way of the consumer know this (Hossain *et al.*, 2021). Other forms of substitution in processed meat samples include the addition of nonmeat protein additives to sausage, burger patties or minced meat for economic gain (Tanabe *et al.*, 2007; Cawthorn, Steinman and Hoffman, 2013; Ha *et al.*, 2017a).

2.4. The meat industry in South Africa

2.4.1. Description of the industry

Livestock farming is a viable agricultural activity in a large part of South Africa, where approximately 80 % of the country's agricultural land is suitable for extensive grazing (DAFF, 2018a). This industry is powered by income growth and supported by technological and structural change (USDA, 2015). Apart from providing income and food, livestock are vital in preserving wealth and can be used as collateral for credit. Livestock are produced throughout South Africa; however, the types of species, breeds and numbers depend on the environment and production systems that can be implemented within the area. The main types of domestic livestock kept for meat in South Africa are cattle, sheep, pigs, goats and chickens (USDA, 2015). In South Africa, the total meat consumption is estimated at 2,9 million tonnes per year of beef, poultry, mutton and pork per year (Murcott, 2021), of which 58kg is consumed per capita.

2.4.2. Beef industry

The beef industry is the fastest growing sector in agriculture after the broiler industry (DAFF, 2019). Beef production holds the largest proportion of cattle heads at 80% and dairy production holds 20% (DAFF, 2019). Cattle producers range from commercial farmers who depend on advanced technologies, to communal farmers who depend on indigenous knowledge and limited technologies. There are three major groups of beef cattle farmers in South African and these are:

- The commercial beef producer where production is relatively high and is compared to production in developed countries. This type of production is characterised with crossbreeding Indicus/Sanga types and their crosses as dams.
- The emerging beef cattle farmer who owns or lease land. The cattle belonging in this group are mainly indigenous crossbred or exotic type of animals.
- The communal beef farmer who farms on communal grazing land and their cattle are mainly indigenous type.

In South Africa 60% of cattle is owned by commercial farmers, while emerging and communal farmers own 40%. The average gross value of beef produced increased from R13,6billion in 2008/09 to R37 billion in 2017/18 and the gross value of beef produced during this period was R21 billion per year. The gross value of beef depends on the number of cattle slaughtered and the prices received from abattoirs (DAFF, 2019) . The main beef breeds reared in South Africa are categorized into Sanga (Drakensberger, Afrikaner and Tuli), British (Angus South Africa, Hereford, Sussex), European (Charolais, Braunvieh, Pinzgauer, Limousin, Simmentaler and Brangus), Composite (Bonsmara, Beefmaster, Santa Gertrudis, Simbra and Braford and Bos Indicus (Brahman) (SA Stud Book, 2016).

Beef is produced throughout South Africa and the number of cattle reared depends on the feedlots and abattoirs available in an area. Mpumalanga Province holds the greatest amount of beef produced in South Africa, accounting for 20% of the beef produced in 2017/18, followed by Free State, Gauteng, KwaZulu Natal, Northwest 19%, 17%, 10%, and 9%, respectively. Northern Cape, Northwest, and Eastern Cape account for 25%. Western Cape and Limpopo account for 5%. South Africa has around 430 abattoirs that slaughter cattle, pigs, sheep, and game meat yearly. Forty percent of all slaughtering is carried out by abattoirs with the capacity

to slaughter an unlimited number of animals (Class A) and 60% of cattle are slaughtered by highly regulated abattoirs (Class A and B). A greater percentage of the abattoirs are linked to feedlots, making the process more convenient. There was a 36% increase in production and 28% increase in cattle slaughtered in 2016/17. The total amount of beef produced between 2008/09 and 2017/18 amount to 9 million tonnes (DAFF, 2019).

Statistics show that South Africa is self-sufficient in terms of beef production because the beef consumed matches the amount of beef produced in the country. The consumption of beef has increased by 34% in South Africa, and this may be due to population increase, increase in affordability by consumers and consumers shifting from a plant to a meat diet (DAFF, 2019). A portion of cattle abattoirs slaughter animals and sell them off as whole cuts to retailers, whereas some abattoirs process slaughtered meat into value added products such as sausages, minced meat, burger patties and even biltong. The slaughtering of more than one species in an abattoir may lead to cross contamination of equipment that is used on more than one species and not cleaned properly, thereby resulting leading to unintentional meat species substitution (Singh and Neelam, 2011; Tembe, Mukaratirwa and Zishiri, 2018). The processing of whole cut beef meat into processed meat samples, can lead adulteration of beef meat which costs more with cheaper meat types such as pork for economic gain (Ha *et al.*, 2017a).

2.4.3. Broiler industry

The broiler industry in South Africa is the largest fraction of the South African agricultural sector in production value. In the year 2018/19 the broiler industry contributed R46,2 billion gross value, which amounted to 16,2% of the total gross value of agricultural products in South Africa. Broiler production is the cheapest protein source in South Africa followed by beef and broilers accounting for 34% of all animal products in South Africa in terms of the Rand value. South Africa is the largest broiler producer in Southern Africa and accounts for 80% of the production in the region. Broiler meat accounts for 90% of poultry meat in South Africa. The remainder constitutes of backyard chickens, mature chicken slaughter and other specialized broiler meat products such as, geese, turkey, ducks, and guinea fowl. The gross value of broiler meat depends on the quantity produced and prices received by producers. Between the years 2009/10 – 2018/19, the average gross values of broiler production added up to R 35 billion per year (DAFF, 2020). In South Africa chicken breeds kept for meat consumption are mainly from

exotic breeds brough into the country, such as, Ross (UK), Cobb (USA), Hubbard (USA), Arbor Acres (USA) and Hybro (Netherlands) (DAFF, 2020b).

Broiler meat is produced throughout South Africa and largest producing provinces are Northwest, Western Cape, Mpumalanga and Free State, accounting for approximately 74% of the total production. In 2019 North West province produced 22% of the broiler meat in South Africa, followed by Mpumalanga with 21%, Western and Northern Cape with 19%, Free State with 12%, KwaZulu-Natal with 6% and Limpopo with 3%. In 2019 the number of birds slaughtered increased by 14% compared to 2010, whilst production increased by 23%. This was due to an increase in demand, resulting from a slow economic growth and consumers opting to resort to chicken as a cheaper protein source (DAFF, 2020b) .

Statistics have shown that South Africa consumes more broiler meat than what is produced, resulting in South Africa becoming a net importer of broiler meat. The gap between production and consumption of chicken continues to increase, resulting in South Africa constantly importing chicken. Reason of this may be due to the growth of disposable income and health reasons. The amount of broiler meat consumer per person per year increased from 39,19 kg in 2009 to 39,85 kg in 2019. Generally, broiler meat has the highest per capita consumption than all other meats consumed in South Africa, mainly because chicken is the most affordable source of animal protein (DAFF, 2020b). The abattoirs that slaughter chickens in South Africa slaughter poultry meat only and not red or game meat. The poultry abattoirs can be found in all provinces of South Africa (Dalrrd, 2021). In some of the abattoirs whole chickens are processed into value added products such as poly, minced meat, sausages, just to mention a few (Dalrrd, 2021). In terms of meat adulteration, chicken is usually the meat type that is adulterated, because it is more reasonably priced than beef and lamb. It has been reported to be a commonly adulterated meat in other meat types of higher value, due to its low cost and readily availability (Wang *et al.*, 2020). The adulterers gain economically by replacing higher priced beef and lamb with cheaper chicken duck and pork (Ha *et al.*, 2017a; Qin *et al.*, 2019).

2.4.4. Pork industry

The South African pork industry is relatively small in terms of the overall South African agricultural sector. The industry contributes 2.,45% to the primary agricultural sector in South

Africa. The quantity produced and the price farmers receive after sale determine the gross value of production. The gross value of pigs slaughtered between 2009 – 2019 amounted to R4,2 billion per annum. Between 2009 – 2018 the gross value of pork increased, mainly due to the decrease in the price of inputs and an increase in the price of beef, which is the main competitor. In 2019 the gross value of pork decreased slightly by 2 %, due to the outbreak of listeriosis in 2018 (DAFF, 2020d) . The common pig breeds produced in South Africa are Landrace, Duroc, Large White, Kolbroek, Windsnyer and Piestran (DAFF, 2020d). Pork is produced throughout South Africa with Limpopo and Northwest provinces being the largest producers accounting for 24% and 21%, respectively, of total production in 2019. Western Cape and Gauteng accounted for 11% each and KwaZulu-Natal accounted for 10% of the total pork production in 2019. The lowest pork producing provinces are Northern Cape and Eastern Cape, which account for 1% and 6%, respectively (DAFF, 2020d). There are approximately 400 commercial producers and 19 stud breeders in South Africa. The total number of pigs in South Africa was estimated at 1,389 million in 2019. However, this was a decrease of 2,8% compared to 2018. During 2009 – 2019, over 28,5 million pigs were slaughtered, resulting in almost 2,3 million tons of pork meat.

Pork consumption in South Africa is influenced by religious and cultural beliefs (DAFF, 2020d). In the Islam religion the consumption of pork is not allowed, therefore the adulteration of halal processed meat product with pork is a cause of a concern in the Muslim community (Doosti, Ghasemi Dehkordi and Rahimi, 2014; Mohd-Hafidz *et al.*, 2020). The consumption of pork is generally higher than the production, making the country a net importer of pork meat. However, in 2014 South African became self-sufficient in pork, producing 236 000 tons of pork meat and having consumption slightly lower. The decrease in consumption may have been due to an increase in the price of pork meat during this period (DAFF, 2017a). It is worth mentioning that the number of live pigs in South Africa has decreased over the years and this may be attributed to an increase in pork consumption (DAFF, 2020d). The increase in pork consumption may be attributed to the increase in the price of beef and mutton meat (DAFF, 2019, 2020e). The abattoirs in South Africa that slaughter pigs also slaughter sheep cattle and game meat (Dalrrd, 2021). Pork meat has been reported to have been intentionally substituted in beef and mutton products for economic gain (Cawthorn, Steinman and Hoffman, 2013; Tembe, Mukaratirwa and Zishiri, 2018; Chaora *et al.*, 2022). Surprisingly, there have been cases where a pork meat product has been mislabelled/substituted with beef meat (Jane K

Njaramba *et al.*, 2021). This may have been a case of mistakenly labelling pork as beef in a set up that processes different types of meat species, or addition of off-cuts/low value beef meat into a pork product for economic gain. From a religious perspective, the consumption of pork is prohibited in the Muslim and other Christian communities (Hossain *et al.*, 2021; Islam *et al.*, 2021).

2.4.5. Mutton and lamb industry

Sheep farming is practiced throughout South Africa, but is it more prominent in Northern Cape, Eastern Cape, Western Cape, Free State and Mpumalanga. The total number of sheep farms are estimated at 8 000 commercial sheep farms and 5 800 communal farmers. In South Africa the total number of sheep is estimated at 21,9 million (DAFF, 2020c). The Dorper sheep breed is a popular South African breed, mainly suited in the more arid areas of South Africa. It is mainly reared for mutton and has a good carcass quality when taking fat distribution and conformation into consideration. The gross value of mutton production is dependent on the price and quantity of meat produced. During the period of 2009 – 2019, the average gross production value amounted to R 5,31 billion per year. There was a 13% decrease in the gross value of sheep in 2019 due to a decline in production and stabling of prices A decrease in sheep numbers and increase in population growth in South African has resulted in an increase in demand and subsequently, shortages in the supply of mutton. The decline in sheep numbers is predominantly due to predation and stock theft (DAFF, 2020c). The meat breeds reared in South Africa are Damara, Namaqua Afrikaner, Zulu, Pedi, South African meat Merino, Dorper, Blackhead Persian, Dohne Merino, Dormer, Afrikaner, Afrikaner, Afrino, Meatmaster, Van Rooy, Sufflok and Ile de France (DAFF, 2020c).

In South Africa, all nine provinces produce sheep, and the total number is estimated at 21,9 million. There are approximately 30% of the South African sheep in Eastern Cape followed by Northern Cape, Free State and Western Cape at 24%, 20% and 12%, respectively. These provinces account for 86% and the other five provinces (Gauteng, Mpumalanga, Northwest, Limpopo, and KwaZulu-Natal) share the remaining 14%. Sheep and lamb are slaughtered in abattoirs around the country and most of the mutton in South Africa is consumed locally (DAFF, 2020c). During the period of 2009 to 2019, the amount of mutton consumed was more than what was produced. The amount of pork consumed in 2014 was 193 000 tonnes and was

184 000 tonnes. There was a decline in sheep production between 2014 and 2019 due to drought experience in 2015/16, followed by ongoing thefts increase in the price of mutton. This lead to mutton becoming more expensive than beef, pork and chicken (DAFF, 2020c). The abattoirs that slaughter sheep in South Africa also slaughter cattle, pigs and some game animals (Dalrrd, 2021). Cases of substitution of mutton (sheep) meat with cheaper meat species like pork, duck and chicken have been reported (Jane Kagure Njaramba *et al.*, 2021; He *et al.*, 2022) and these meat samples will have been from abattoirs and retail outlets. In China mutton meat has been reported to be substituted with *Murine* (rat and mouse) species (Ouso *et al.*, 2020). Whilst the consumption of rodents may be acceptable in other countries like China (Ouso *et al.*, 2020), it is not the case for South African consumers. There is therefore a need to track any form of substitution in abattoirs and retail outlets that process mutton meat, to avoid consumers from consuming undesired species, as well as unscrupulous traders gaining economically at the expense of consumers.

2.4.6. Chevon industry

South Africa is a small goat producing country possessing approximately 3% of Africa's goat and less than 1% of the world's goat population (DAFF, 2020a). The breeds that are reared commercially in South Africa are Boer, Savanna and Kalahari Red and these breeds are reared for meat, skin, and small quantities of cashmere. Goat meat from adults is referred to as chevon and as cabrito when it is from young animals. In comparison to cow's milk, goat milk is highly priced, probably because it is less likely to cause allergies in humans than cow's milk. Indigenous goats are those that are reared by small-scale producers and mainly provide for their family needs of meat and to a less extent milk. Indigenous goats comprise of 65% of the goats reared in South Africa and are mainly found in Eastern Cape, Limpopo, Northwest and KwaZulu-Natal provinces. The Boer goat, that is indigenous to South Africa, is mainly reared in South Africa. It has less fat and calories and contains high levels of protein and iron when compared to beef, pork, mutton and broiler meat. Similarly, to mutton the gross value of chevon depends on the quantity produced and prices received by producers. Between the periods of 2009 – 2019 the average gross value of chevon was R422 million per year, and this was due to an increase in the price per kg of chevon. The gross value of chevon increased by 234% between 2009 – 2019 (DAFF, 2020a). The goat breeds that are reared in South Africa are South

African Indigenous (Nguni, Tswana, Venda, Tankwa, Xhosa and Zulu) Kalahari Red, Boer and Savannah (DAFF, 2020a).

Goats are produced throughout South Africa with Eastern Cape, Limpopo and KwaZulu - Natal being the largest producers, accounting for 68% of the total live goats. The Eastern Cape produces the highest number of goats, accounting for 38%, followed by Limpopo, KwaZulu-Natal, and Northwest at 17%, 13% and 13%, respectively. These provinces account for 81% of South Africa's total goat numbers and the other five provinces (Western Cape, Northern Cape, Mpumalanga, and Free State) share 19% (DAFF, 2020a). Goat produces meat and milk, but the primary reason for keeping goats is for meat. During the years 2009 – 2019 , the average number of goats slaughtered was 823 108 goats per year and chevon production was averaged at 11,6 million kg per year (DAFF, 2020a). South Africa is self-sufficient in goat production and does not need to import chevon meat to meet consumption levels in the country. Production and consumption of chevon has been increasing between 2009 – 2019. Production increased by 10% in 2010 and 11% in 2019. The highest numbers were reached in 2019, where production was 11 590 tonnes an consumption was 11 510 tonnes (DAFF, 2020a).

Goats in South Africa are predominantly sold as live animals through informal transactions, since goats are mainly produced and consumed in smallholder communal farming systems (DAFF, 2020a). Therefore, there are no published reports of goat meat being substituted in the meat industry in South Africa, however, it does not mean such cases do not exist in the country. In Kenya, there have been reports of chevon meat being substituted with wildlife meat (Ouso *et al.*, 2020). In yet another Kenyan study, Jane Kagure Njaramba *et al.*, 2021 reported that goat meat had the highest substitution, with mutton and beef meat being fraudulently added to goat meat. In some countries such as Kenya goat meat fetches a higher price (USD5 – USD6 per kg) than beef and mutton (USD3 – USD4,5 per kg). This is due to goat meat being used to prepare an Eastern African delicacy meal called "Nyama Choma" (Jane Kagure Njaramba *et al.*, 2021). Nyama Choma means "roasted meat" that usually consists of beef or chevon meat, where consumption of it is related to a higher social status (Gorski *et al.*, 2016). Some consumers do not favour the distinct odour that mutton meat has and prefer to consume chevon meat. Furthermore, consumption of mutton and beef can cause health risks through an allergic reaction called 'midnight anaphylaxis' and these cases were reported in South Africa (Gray,*et al.*, 2016). It is possible for unintentional substitution to occur through cross-contamination of

equipment in establishments that process more than one species. This is the case in abattoirs that slaughter goats in South Africa, where other species such as cattle, sheep and game are also processed there. It is important therefore to have strict surveillance measure in abattoirs and retail outlets that process more than one species.

2.4.7. Game industry

The South African game farming industry has progressed in becoming a world leader in extensive sustainable utilization of game species and has also penetrated the local and international market (Sommerville M *et al.*, 2021). The game industry uses bushveld land that is not favourable to the hardiest of cattle breeds, thereby producing venison, which is a popular meat alternative. Similarly, to domestic animals, the gross value of wildlife meat production is dependent on the quantity produced and the prices received by abattoirs (DAFF, 2018b). The average gross value of revenue from game meat was estimated at R611 million in 2016 (Sommerville, M *et al.*, 2021). Game meat consumption is gradually increasing and in South Africa game meat comes from springbok, kudu, gemsbok, impala, eland, wilder beast and ostrich species (D'Amato *et al.*, 2013b; Sommerville, M *et al.*, 2021). These apart from ostrich that is farmed intensively are all farmed in extensive farming systems (D'Amato *et al.*, 2013b). Most consumers are health conscious thereby, making game meat popular, not only in the traditional form of biltong or dried sausage, but also in the form of game recipes in restaurants and lodges where game farming is practiced. Consumption of game meat and ostrich is considered healthy due to their low cholesterol and the assumed absence of antibiotics, anabolic steroids, hormones, and other additives (DAFF, 2018c). This is quite crucial to a health-aware market. This, therefore, makes the price of game meat and ostrich expensive and increases the chances of meat substitution occurring. Furthermore, game meat is sold as processed meat products such as sausages, beef patties, minced meat and biltong (Sommerville, M *et al.*, 2021). This also raises the chances of meat substitution, as it will be difficult to tell the authenticity of the meat product using the naked eye after it has undergone morphological change. Most abattoirs that slaughter game and ostrich meat also slaughter other meat types like beef, mutton, and poultry. Given the high price of game meat, cases of substitution of processed game meat with cheaper red meat such as beef and pork have been reported in South Africa (Cawthorn, Steinman and Hoffman, 2013; D'Amato *et al.*, 2013b; Tembe, Mukaratirwa and Zishiri, 2018). A study in the USA that collected whole cut game meat, observed that 18,5% of the game meat

was mislabeled (Quinto, Tinoco and Hellberg, 2016). These results indicate that mislabeling/substitution in game meat not only occurs in processed meat, but also in whole cut meat.

South Africa is a top ostrich producer and holds 75 % of the global market share (DAFF, 2018c). There are three products derived from ostrich farming namely, meat, leather and feathers, and the main source of income is meat and leather. In South Africa, the value of a slaughtered ostrich is broken down into 45% skin, 45% meat and 10% feather. Whereas, in Europe it is 75% meat and 25%, due to the healthy attributes of low fat, low cholesterol and high protein and iron which ostrich meat carries. Previously, South Africa used to be the only producers of farmed ostriches in the world. This resulted in jealously that led to laws preventing the export of live birds. Fortunately, there came a period when these restrictions were lifted and several countries began to import Ostriches from South Africa, to start their own industries. Ostrich farming requires dry climates, mainly for breeding. Therefore, ostrich farming is more prominent in the western drier parts of the country or in winter rainfall areas. The ostrich industry is mainly practiced in the Western Cape in the Klein Karoo and Southern Cape regions. The gross value of production of ostrich products depends on the quantity produced and the prices received by producers (DAFF, 2018a)

Game meat production in South Africa is relatively small compared to other meats. Furthermore, game meat is produced mainly in winter. The province with the largest percentage of game meat is Limpopo at 50%, followed by Northwest,, Mpumalanga, Northern Cape, Eastern Cape, Free State, Western Cape, KwaZulu-Natal and Gauteng at 11%, 10%, 9%, 8%, 6%, 3%, 2%and 1%, respectively (Taylor *et al.*, 2020) . South Africa is ranked 20 in world production and 15 in Africa. There is an estimated production of 18 000 tons (DAFF, 2018d). Ostriches are produced in four provinces in South Africa, namely Western Cape, Eastern Cape, Northern Cape, and Limpopo. Western Cape is the largest ostrich producer at 83%, followed by Eastern Cape at 10%. The remaining 6% are produced in Limpopo and Northern Cape. Oudtshoorn in the Klein Karoo of the Western Cape is considered as the ostrich capital of the world. Approximately 70% of the world's ostriches are found in South Africa. South Africa has around 588 registered export ostrich farms. Out of these 453 are in the Western cape, 102 in the Eastern Cape and 33 in the rest of the country. Most ostriches are slaughtered at 10 – 14

months of age, producing about 27 kg of meat, 4.2 m^2 of leather and 1 kg of feathers (DAFF, 2019).

2.5. Initial reports of market substitution of food products

Food fraud is the act of defrauding consumers of food and food ingredients for predominantly economic gain, and this has resulted in problems within the food industries worldwide. The earliest reports of food fraud dating back thousands of years involved olive oil (Schinto, 2010), tea (Foster, 2011), wines and spices (Renée Johnson, 2014). Generally, the food and/or ingredients that were usually associated with fraud were fruit juices, alcoholic beverages, spices, milk, dairy products, fish, grain-based foods, olive oil, wine, organic foods, honey, highly processed food, coffee, tea and meat products (Hovda *et al.*, 2005). However, most cases go unnoticed as consumers may not notice any problems in the quality of products and some fraud cases may not cause any health risks. There have been cases reported in the past that resulted in public health risks. Between 2002 – 2004 a liquor adulteration in Norway killed 9 people and hospitalised 51 people from methanol poisoning. The liquor contained 20 % methanol and 80 % ethanol (Becker, 2007). In 2007, it was reported that adulterated pet food ingredients from China caused the death of cats and dogs in the United States (Xiu and Klein, 2010). In another case that involved health risks to human lives, baby formula milk to increase the protein content, resulted in approximately 300 000 Chinese children becoming ill and 6 infant fatalities (Buck, 2007). In another disturbing report, it was discovered that in 2007 that 150 tonnes of rotten meat were knowingly distributed across German markets (Bosley, 2007).

The adulteration of fish and seafood was also previously reported in the United States. Cases involved the mislabelling and substitution of high value species with inferior species that could have been associated with some sort of food poisoning or allergens (Stanciu, 2015). In 2013, the Food Safety Authority of Ireland (FSAI) reported on incidences of mislabelling in beef burgers. Beef burger products tested positive for horse DNA, where samples were found to contain 80 – 100% horsemeat (Bouzembrak and Marvin, 2016). Subsequently, in China it was reported that an unauthorised colorant called Sudan Red was added to beef meat to enhance the colour of meat (Jia and Jukes, 2013). A food fraud case that uncovered bribery and corruption in the Brazilian government (including inspectors, politicians and the president), revealed that rotten meat was distributed into the food chain and the practice of hygiene was intentionally overlooked (Haynes, B. and Spagnoul, 2017). In Belgium expiry dates were

fabricated and non-organic was reported to be labelled and sold under the pretence of it being organic meat. Furthermore, it was reported that rotten meat meant for animal feed ingredients, found its way into the Belgian human food chain (Green, 2018; Grobe, 2018). These cases mentioned are just a fraction of food fraud cases that have been revealed. It is predicted that many other cases could have occurred but never noticed or reported. Food fraud is affecting the food supply chain drastically. It is predicted that global food fraud costs the food supply chain $10- 15 billion per year (Johnson, 2014). PriceWaterhouseCoopers, a major global accounting firm estimated a higher loss of $30 – 40 billion per year (PricewaterhouseCoopers, 2017).

2.6. Food fraud reporting systems

Developments have taken place around the world to create food fraud systems that can identify, document, track and understand food fraud incidences. Three major databases have been established, which include HorizonScan, Economically Motivated Adulteration (EMA) and Rapid Alert for Food and Feed (RASFF). The most common food fraud cases reported in RASFF, EMA and HorizonScan between 2014-2015 are illustrated in Table 2.2.

2.6.1. HorizonScan

HorizonScan was developed in the UK (https://www.fera.co.uk/horizonscan-food-safety-at-your-fingertips). It pays special attention to cases that involve global and feed integrity like adulteration, substitution, fraud, microbial contaminants, allergens, pesticides, and drug residues (Bouzembrak *et al.*, 2018). The company set up a dataset that includes information on adulteration, fraudulent health certificates, unapproved premises, expiry date changes and unauthorized transport. It tracks over 500 commodities, in 200 countries from official websites of 65 countries. There are more than 100 data sources that are searched daily and on average 30 new reports are added daily (FERA, 2018).

2.6.2. Economically Motivated Adulteration (EMA) database

This Economically Motivated Adulteration (EMA) database was established by the USDA in USA (https://www.fda.gov/food/compliance-enforcement-food/economically-

motivatedadulteration-food-fraud). This database contains food fraud cases since 1980, including information such as food product, fraud incident year, adulterant, types of fraud, health consequences and country of origin. The EMA database includes information on substitution, artificial enhancement, dilution, mislabeling and counterfeit. The EMA database concentrates on intentional adulteration for economic gain. The information in the database is gathered from media searches of economically motivated cases since 1980 (Bouzembrak *et al.*, 2018).

2.6.3. Rapid Alert for Food and Feed (RASFF)

This food fraud reporting system was developed in Europe. This database was used to exchange information and for traceability purposes, after it was reported that horse meat was fraudulently included in beef products (Bouzembrak *et al.*, 2018). The RASFF database includes information on improper, fraudulent, missing, or absent health certificates; illegal importation; tampering; improper, expired, fraudulent or missing common entry documents; expiration date and mislabeling. The public is allowed access to RASFF to retrieve information on any food fraud notifications that are sent and received. The information gathered is compiled from official control on markets, border control and consumer complaints (EFSA, 2010; Djekic, Jankovic and Rajkovic, 2017).

Table 2. 2: The most common food fraud cases reported in RASFF, EMA and HorizonScan between 2014-2015

System	Product Category		Country of origin		Type of fraud	
	Product	%	Country	%	Fraud type	%
RASFF	Nuts and seeds	38	China	28	Health certificates	63
	Fruits and vegetables	16	India	12	Illegal importation	25
	Fish and fish products	8	Turkey	11	Fraudulent [1]CED	10
	Mixed	6	Nigeria	11	Expiry date	1
EMA	Meat	27	India	22	Tampering	73
	Dairy	20	United States	18	Origin labelling	22
	Alcohol	9	Unknown	13	Theft and resale	4
HorizonScan	Wines	20	Czech Republic	27	Adulteration	72
	Meat	8	United States	8	Health certificates	11
	Honey	7	Maldova	5	No inspection	10
	Potatoes	5	Slovak Republic	4	Expiry date	4

[1] – Common Entry Documents

In a recent review on the beef supply chain by (Robson *et al.*, 2020), food fraud reports between 1997 – 2017 were downloaded from RASFF and HorizonScan. The most reported fraud types were counterfeit (42.9%), adulteration (41.9%), diversion (9.4%) and tampering (5.8%). The top 5 countries with the most reports were Brazil (20.1%), USA (15.2%), Germany (8.2%), Poland (6.5%) and United Kingdom (5.8%), accounting for 65.2% of the reports. Reports on counterfeit referred to trademark or patent violation. Adulteration referred to products that included the addition of unknown ingredients and dilution of products with a wide range of unknown contaminants. Diversion referred to reports on products that were illegally imported

or exported and illegal products were sold in a legitimate market. Tampering referred to products that had fraudulent expiry dates, labels, and countries of origin (Robson *et al.*, 2020).

2.7. Factors leading to meat species adulteration in South Africa

Meat is highly priced worldwide and more especially in South Africa (Tembe, Mukaratirwa and Zishiri, 2018). This places financial pressure on the population in general, considering that over 50% of South Africans are found below the poverty line (Tembe *et al.,* 2018). The prices of beef and mutton have significantly increased since 2000, such that, they are now termed as luxury goods in South Africa (Cawthorn, *et al.,*, 2013). The prices of beef and mutton were previously reported to sell at a price that is twice more than that of chicken and 1.5-fold more than pork (Bhat *et al.*, 2015). In South Africa the price of beef and mutton have increased due to an increase in production costs (DAFF, 2019, 2020c). Due to the increase in meat prices, there has been an increase in the use of cheap meat ingredients in processed products for economic gain. Meat substitution may involve the use of cheaper ingredients from the same declared species, but from different body parts like blood, connective tissue, and offal. Adulteration may be in the form of substituting an expensive meat species with a different cheaper species (substitution of beef with horsemeat) or through non-meat ingredients like plants and dairy sources. In earlier times, meat was not commonly connected with adulteration, because meat was marketed as fresh with easily recognizable meat cuts (DoH, 2012). However, with the increase in processing of meat into value-added products, the incidence of meat adulteration has become common (Hossain *et al.*, 2021; Islam *et al.*, 2021). The flavour and texture of meat differ slightly after it is processed. Therefore, it is difficult to differentiate meat species based on visual appearance and taste. For example, the taste of crocodile meat is similar to that of chicken and fish, although crocodile meat is more expensive (Cawthorn, *et al.*, 2013). As soon as meat is processed into value added products, identification using appearance and sensory parameters becomes difficult and even impossible.

2.8. Commonly substituted or mislabelled species in South Africa

The intentional substitution of meat products was first reported through media in South African (Wiener, 2011). In 2011 it was reported that the Orion Cold Storage in Cape Town, was importing various meat products and knowingly relabelling these as food-grade and Halaal

(Rafudeen, 2013). The Orion Cold Storage was reported to be importing pork products from Belgium and Ireland and labelling them as Halaal sheep or beef products, kangaroo from Australia and water buffalo from India and labelling them as beef products, importing non-Halaal poultry from Spain through the UK and labelling them as Halaal and importing non-food-grade milk powder for animal feed and labelling it as Halaal skim milk powder fit for human consumption (Rafudeen, 2013). These results sent panic in the Muslim community and food industry. The first South African study following this scandal revealed that 95 of 139 (68%) samples contained species that were not declared on the product labelling. The highest occurrence was found in sausages, burger patties and deli meats (Cawthorn, *et al.*, 2013). The addition of non-declared plant proteins like soya and gluten and animal species like pork and chicken was detected (Cawthorn, *et al.,* 2013). Thereafter, a second South African study on game meat samples, 76.5% substitution was reported (D'Amato *et al.*, 2013b). Domestic species like cattle, horse, pig and lamb; common game species like kudu, gemsbok, ostrich, impala and springbok; uncommon game species like giraffe, waterbuck, bushbuck, duiker and mountain zebra; and extra-continental species like kangaroo were substituted (D'Amato, *et al.*, 2013b). Tembe, *et al.*, 2018 conducted a study was carried out in the Durban Metropolitan. Forty processed meat samples were examined and 65 % were found to be contaminated with unlabelled species. Contamination was found to be higher in sausages than patties, where 80% of sausages analysed were contaminated as compared to 50% of patties. Beef, mutton, chicken and pork were analysed and of the four meat types analysed, beef was the most common contaminant in the meat species (Tembe, *et al.*, 2018).

2.9. Consequences of meat species substitution or mislabelling

Consumers rely on precise and thorough declaration of food ingredients, so that they can choose products fitting with their lifestyles. If these standards are not maintained, then brand loyalty can be compromised. Meat species substitution poses health risks to consumers who may be allergic to certain foods. For example, soya and gluten are common allergens for some consumers (Cawthorn, *et al.,* 2013). Food allergies to meats like beef, chicken, pork, turkey, mutton and rabbit are not strange and may be intensified in individuals that may be young, old, ill or immune-compromised (Cawthorn, *et al.*, 2013; Doosti, *et al.*, 2014; Gray, *et al.*, 2016; Hossain *et al.*, 2021). Therefore, it is important to accurately declare food ingredients in-order

to avoid potential health risks. The South Africa food labelling regulations lists soya and gluten-containing cereals as common allergens (The Commission of the European Communities, 2003). Soya and gluten are among foods that cause 90% potentially fatal food allergic reactions (Cawthorn, *et al.*, 2013). Nowadays consumers are more aware of their health and require information with regards to the composition of foods they consume. Incidences of undeclared species posing health risks to a consumer could also be in the form of the substitution of cheaper chicken flesh or fat for more expensive beef or mutton constituents (Ha *et al.*, 2017b; Njaramba *et al.*, 2021; He *et al.*, 2022).

Pork is one of the meat species that is substituted in processed meat products because it is a less expensive ingredient. This is a cause of concern from a religious point of view. In the dietary laws for Muslims (Halal) and Jews (Kashrut), there are restrictions on the consumption of pork and its associated products. In countries such as Bangladesh, China, Japan and Korea where beef is expensive, beef products are intentionally adulterated with pork for economic gain (Hossain *et al.*, 2021; Islam *et al.*, 2021). Therefore, individuals following these cultures rely on the accurate labelling of foodstuffs because adulteration goes against their ethical rights and religious beliefs. Some consumers choose to consume certain meat types as a lifestyle choice. For instance, certain beef cattle breeds are sold as branded or certified breeds. Examples of breeds sold as certified or branded in South Africa are Angus, Wagyu, Bonsmara and Afrikaner (Coleman, 2017; Lombard *et al.,* 2017). Incidences of meat species substitution for economic gain result in unfair competition in meat trade. Consumers are concerned about the quality of meat products they purchase and would much rather prefer high quality meat, making meat speciation important in maintaining food safety to consumers. Overall, the presence of undeclared species, whether through cross contamination or deliberate substitution, results in significant financial, religious, ethical, and public health consequences. Public health consequences include exposure to toxins, pathogens or allergens (Magiati *et al.*, 2019). These practices also go against legislation in South Africa and weaken fair trade on the local meat market. In addition, these practices result in concerns on the functioning of the meat supply chain in South Africa (Cawthorn, *et al.*, 2013). Although efforts have been made to authenticate meat products that include, protection origin of meat, protected geographic location and health certificates, meat adulteration continues to occur, and it is impractical to certify all meat products from adulteration.

2.10. Methods used for species identification

There is need to develop pipelines that can determine whether meat substitution is intentional or unintentional, be it in abattoirs or in the retail market. According to previous studies if adulteration is less than 0,1% then it is considered as unintentional (Xu *et al.*, 2022). Therefore, there is also a need to develop molecular pipelines that can measure the level of adulteration and be able to differentiate between intentional and unintentional contamination. The precise identification of species in meat products is vital (i) to ensure that legislation related to meat products are enforced and adhered to, (ii) to maintain the standards of meat products, (iii) to avoid unfair competition in the meat industry, (iv) to uphold consumer rights religious and social traditions and (v) also to limit the poaching of wildlife animals (Lenstra, 2003). Consequently, there is a need to adopt techniques that can accurately authenticate species in meat products. There are several methods available for meat speciation. The methods depend on the physical, chemical, and biochemical properties of the meat product.

2.10.1. Protein based methods for species identification

In the past species identification in food involved the detection of species-specific proteins, to determine the origin of food ingredients for human consumption (Rasmussen *et al.,* 2011). Protein is the main component of meat and the specific protein composition and three-dimensional structure of certain proteins, allow for detection of meat adulteration (Li *et al.*, 2020). Other protein molecules are tissue-specific and can be used to detect the adulteration of less valuable molecules such as, blood plasma or connective tissues (Ofori and Hsieh, 2015; Jiang *et al.*, 2018). Examples of protein-based techniques that have been implemented are immunoassays on antibody specific reactions, electrophoretic analysis of protein band characteristics and mass spectrometric analysis of proteins or short peptides(Li *et al.*, 2020). However, protein-based methods do not have a high detection sensitivity and specificity (Li *et al.*, 2020). Heating during the processing of foodstuffs can result in denaturation of the proteins, which has now shifted attention to DNA as a source of species identification (Li *et al.*, 2020).

2.10.1.1. Enzyme-linked immunosorbent assay

There are two types of immunoassay techniques used in the detection of meat adulteration and these are Enzyme-linked immunosorbent assay (ELISA) and immunosensors. When performing meat adulteration detection using ELISA a conjugated enzyme is prepared using a known antibody (Li *et al.*, 2020) . The conjugated enzyme binds to samples with that antibody. The samples are washed, incubated and a coloured substrate is added. The extent of colour development will determine the amount of antibody in the samples being tested (Li *et al.*, 2020). The commonly used ELISA methods for meat adulteration detection are direct ELISA (Mandli *et al.*, 2018; Seddaoui and Amine, 2020), sandwich ELISA (Hsieh and Ofori, 2014; Zvereva *et al.*, 2015) and indirect competitive ELISA (Hsieh and Ofori, 2014; Zvereva *et al.*, 2015). ELISA methods require simple sample preparation, are low costing and less time consuming. However, ELISA methods cannot be used in multispecies detection and can only detect meat species for which specific antibodies have already been developed (Zvereva *et al.*, 2015). ELISA methods can also give false positives caused by cross-reactivity and proteolysis from heat processing (Li *et al.*, 2020).

2.10.1.2. Immunosensors

Immunosensors were developed as a more sensitive, low cost and less time-consuming protein-based methods for meat detection and have previously been used in meat adulteration detection (Mandli *et al.*, 2018). The use of an electrochemical competitive immunosensor based on an anti-pig IgG antibody, managed to identify pork adulteration as low as 0.01% in 20 minutes (Mandli *et al.*, 2018). The use of a lateral flow device managed to identify 0,01%, 0,1% and 1% pork DNA in raw meat, beef meatballs and cooked meat, respectively (Masiri *et al.*, 2016; Kuswandi *et al.,* 2017). The disadvantages of immunosensors are that they require the use of antibodies and antigens, pre-treatment and have low specificity because of the potential of species cross reacting (Lu *et al.*, 2021).

2.10.1.3. Mass spectrometry analysis

Mass spectrometry methods based on protein and peptide analysis have been applied in meat species identification, because the amino acid sequence of peptides is more stable than DNA during meat processing, especially in highly processed and similar meat species (Prandi *et al.*,

2017; Naveena *et al.*, 2018). Mass spectrometry methods can also identify species by simultaneously monitoring multiple specific peptides and reduces the probability of false positives (Li *et al.*, 2018). In previous studies the authenticity of cooked and smoked sausages was tested using label-free quantification method that utilized high resolution mass spectrometry (Montowska *et al.*, 2015). In the study by Montowska *et al.*, (2015) an LOD of 5% (w/w) for pork and beef and 1% for horse meat was reported. Mass spectrometry can also be used in identification of similar species, due to its ability to achieve multi-marker detection (Montowska *et al.*, 2015). A mass spectrometry-based method to detect chicken, duck and goose meat in processed meat demonstrated high levels of qualitative and quantitative data under low matrix interference (Fornal and Montowska, 2019).

2.10.2. Metabolite profiling

Meat contains small molecules called metabolites that can assist in meat identification, through the comparison of metabolite profiles in the samples (Lim *et al.*, 2017). This technology can uncover the physiological and biochemical status of meat samples and reveal small variations in metabolites (Lim *et al.*, 2017). Meat species contain specific quantities of fatty acids and different flavours depending on the species type. Therefore, lipids can be used to distinguish meat species that have been adulterated (Ballin, 2010). (Trivedi *et al.*, 2016) developed a method to detect pork adulterated in beef meat using a metabolomics and lipidomics method. The results demonstrated that 23 metabolites were significantly correlated with adulteration of pork in beef meat. The flavour of meat has different characteristics depending on the species type. Therefore, volatile compounds can be used as a method to detect meat adulteration. (Haddi *et al.*, 2015) and (Zhang *et al.*, 2015) used an electric nose and multivariate analysis to identify meat adulteration. The disadvantage of using metabolic methods is that other factors such as growth environment, meat processing and meat storage can affect the metabolic content of the meat product (Li *et al.*, 2020). In addition, metabolite profiling cannot achieve quantitative results and is therefore not commonly used in meat identification analysis (Li *et al.*, 2020).

2.10.3. Spectroscopic techniques for species identification

Meat identification techniques that involve the use of DNA, protein and metabolite techniques require sample pre-treatment such as target extraction and purification (Wang *et al.*, 2018)). These processes are invasive and time consuming. Researchers have tried to come up with methods that are non-invasive and spectroscopic techniques are such examples (Wang *et al.*, 2018). Compounds such as moisture, protein, fatty acids, lipids, and elements in meat products produce various spectra at different wavelengths. Spectroscopic techniques have been introduced as methodologies in detecting meat adulteration, because they are less time consuming, have simple sample preparation and do not require sample pre-treatment (Wang *et al.*, 2018). Such non-destructive techniques that have been used in meat identification include infrared spectroscopy (IRS) (Hu *et al.*, 2017; Wu, Zhong and Yang, 2018), raman spectroscopy (RS) (Hu *et al.*, 2019), laser-induced breakdown spectroscopy (LIBS) (Casado-Gavalda *et al.*, 2017; Chu *et al.*, 2018) and hyperspectral imaging (HSI) (Zheng *et al.*, 2019).

2.10.4. DNA based methods for species identification

DNA is a major material that is used in replication, transmitting, and storing genetic information. DNA is present in all animal species and more conserved as compared to proteins (Kumar *et al.*, 2015). The advantages of the use of DNA include (i) ability to discriminate different animal species and breeds (Rasmussen *et al.,*, 2011; Kumar *et al.*, 2015; Xu *et al.*, 2018), (ii) DNA has a higher thermal stability than proteins (Rasmussen *et al.*, 2011; Kumar *et al.*, 2015; Xu *et al.*, 2018), (iii) DNA is present in the majority of cells and enables the same information to be obtained from one animal, regardless of the origin (Rasmussen *et al.*, 2011; Kumar *et al.*, 2015; Xu *et al.*, 2018). Polymerase Chain Reaction (PCR) methods based on the use of DNA are the most common techniques used in meat species identification and adulteration, due to their sensitivity and reliability(Xu *et al.*, 2018). The main DNA-based PCR methods used in meat species identification include direct PCR, Real-Time PCR, Restriction Fragment Length Polymorphism PCR (RFLP-PCR) and Droplet Digital PCR (ddPCR) that are reviewed below.

2.10.4.1. Direct PCR

Direct PCR has high sensitivity, resolution and specificity, hence, and is used in a number of authenticity and traceability studies (Ha *et al.*, 2017a). In a species-specific experiment that

was developed using the mitochondrial *D-loop* to detect pork adulteration in commercial beef and pork managed to identity pork levels as low as 1% in beef and chicken products (Ha *et al.*, 2017). The use of multiplex PCR has also been used in meat species identification by using species-specific primers for detection of multiple targets in a single reaction (Ha *et al.*, 2017a). Ali *et al.*, 2015 designed species-specific primers that targeted the ATPase 6 gene, *ND5* and *cyt b* mitochondrial genes to detect rat, monkey, dog, cat, and pig in Islamic food and applied in commercial samples as well. Other meat species such as horse, buffalo, chicken and ostrich have also been authenticated using multiplex PCR (Li *et al.*, 2019).

2.10.4.2. Real-Time PCR

Real-Time PCR is more sensitive and specific compared to direct PCR (Xu *et al.*, 2018). Furthermore, real-time PCR reduces the levels of PCR contamination and can give quantitative results by looking at the amount of DNA template and measuring its Ct value(Xu *et al.*, 2018). This technology monitors the fluorescence signals during PCR, enabling the tracing of initial quantity of target genes without any additional steps (Xu *et al.*, 2018). Real-time PCR was used to identity beef, chicken, lamb, pork and turkey from heat treated processed samples in quantities less than 0,1% (Kumar *et al.*, 2015). The mitochondrial *12S* rRNA gene was selected as a target gene to design primers that could identify 75bp of horse DNA, using TaqMan Real-Time PCR for amplification and the results demonstrated high specificity and sensitivity and no cross-reaction with other species (Pegels *et al.*, 2015). The contamination of mutton meat by murine DNA was revealed by amplification of the *cyt b* gene using TaqMan probes (Fang and Zhang, 2016). The method managed to detect levels of murine DNA as low as 0,1% in low DNA concentrations of 1pg per reaction (Fang and Zhang, 2016) implying that the TaqMan Real-time PCR had high levels of sensitivity and specificity of TaqMan. The disadvantages of real-time PCR are that the methodology requires the use of DNA that is of high quality (Li *et al.*, 2020). The use of single-stranded DNA and RNA can give false positive or negative results (Li *et al.*, 2020)l. The cost of reagents and equipment used in real-time PCR are also more expensive as compared to conventional PCR (Li *et al.*, 2020).

2.10.4.3. Restriction fragment length polymorphism – polymerase chain reaction

The RFLP-PCR technique involves the use of restriction endonuclease digestion to identify conserved regions of DNA (Ali *et al.*, 2018). The method is sensitive, accurate and can be used in authenticity studies (Ali *et al.*, 2018) RFLP-PCR is also much simpler and less time consuming as compared to real-time PCR (Ali *et al.*, 2018). The use of RFLP-PCR in combination with a lab-on-a-chip technology managed to identify dog meat (0,1% w/w) in beef and chicken burgers (Rahman *et al.*, 2015). RFLP-PCR can also be used in the identification of closely related species. Donkey and horse DNA were identified in a Halal meat product using RFLP-PCR and *AluI* restriction enzyme (Doosti *et al.,*, 2014). The PCR amplification sizes of horse and donkey using species-specific primers and mtDNA are the same, therefore, direct PCR could not be used in this experiment (Doosti *et al.,* 2014). The disadvantages of using RFLP-PCR are that (i) it needs to be used in a lab that is suitably equipped, (ii)the enzymes are expensive, (iii) it cannot be used for quantification and (iv)the process can lead to incomplete digestion leading to unreliable results (Kumar *et al.*, 2015; Hossain *et al.*, 2017).

2.10.4.4. Droplet digital PCR

Droplet digital PCR is a new molecular method that is used in nucleic acid detection and quantification. The method performs PCR on reactors in the form of droplets containing a single copy of the target molecule in each reactor. The number of copies of the target sequence are measured by the number of positive reactions that respond to the florescent signal (Li *et al.*, 2020). The ddPCR technique has been used in food adulteration technique and was advantageous over real-time PCR with a limit of detection of different meat products of 0,001% and 0,01%, respectively (Floren *et al.*, 2015a). The ddPCR technique was also used to detect adulteration in processed meat samples of turkey in bovine and pork in chicken and the limits of detection were as low as 0,05% and 0,01%, respectively (Shehata *et al.*, 2017). Despite these good results ddPCR cannot in some instances be converted from gene copy number to meat mass ratio (Shehata *et al.*, 2017). The conversion ratios in different animal species can be inaccurate because of the differences in genome size, copy numbers and cell density (Ren *et*

al., 2017). Furthermore, ddPCR experiments need to be highly accurate as the droplet partition and volume may negatively influence the detection results (Demeke and Dobnik, 2018).

2.11. New and emerging methods for species identification

The traditional methods used for meat species identification that make use of DNA are generally species-specific. However, in meat adulteration detection there may be unknown species that need to be identified. There is therefore a need for a universal method that is can be applied to products of unknown and multiple species composition. There are emerging methods that are universal and demonstrate the ability to deliver high-throughput information in species identification. These include DNA barcoding, DNA microarrays and next generation sequencing (NGS) as described below.

2.11.1. DNA microarrays

DNA microarrays also known as DNA chips, comprise between ten and tens of thousands of different oligonucleotide probes that are immobilized on the surface of a glass slide or microscopic beads. Despite the high throughput advantages of DNA microarrays, there are a few disadvantages, that have prevented the widespread use of microarrays in species identification. Generally, the use of DNA microarrays is time consuming, requires the design and experimental testing of all probes, and is costly compared to other DNA based methods (Wadapurkar and Vyas, 2018).

2.11.2. DNA barcoding

DNA barcoding is a method that does not target specific species and can be used in meat species identification. This method uses PCR amplification and sequencing of specific barcode gene fragments. DNA barcoding a fragment of a genetic target that is common in identifying most animal species (Kane and Hellberg, 2016). The mitochondrial *COI* contains a standard genetic target that can identify different animal species (Kane and Hellberg, 2016). The sequences are then identified by searching for the sequences in databases such as the Barcode of Life Data (BOLD), National Centre for Biotechnology Information and BLAST (Fiorino *et al.*, 2018). DNA barcoding is a fast and accurate method in species identification and has therefore, been

used in meat (Xing *et al.*, 2019; Cottenet *et al.*, 2020) and fish species identification (Fiorino *et al.*, 2018). The initial DNA barcoding methods used Sanger sequencing of a 650bp fragment size. In cases where there were multiple adulterated species in meat samples, Sanger sequencing would produce overlaying sequencing peaks resulting in false positives (Yang *et al.*, 2018).

2.11.3. Next Generation sequencing

Next Generation Sequencing is defined as a collection of methods in which different sequencing reactions occur simultaneously, resulting in a large amount of sequenced data at a low cost (Pabinger *et al.*, 2014). The advantages of the NGS method are that it has a short PCR amplicon, facilitating the analysis of degraded and poor-quality DNA samples; uses universal PCR-primers, therefore, there is no need for prior species information; uses deep sequencing, so it is possible to detect DNA in minute amounts; and it is not species specific, but can universally identify multiple species contained in one sample (Xing *et al.*, 2019). The NGS technique can be used in identification of nucleotide sites that can be targeted in specie-specific PCR assays or for species identification using direct sequencing of fragments(Xing *et al.*, 2019).

There are several platforms available for NGS and these include Roche/454 Life Sciences (Indianapolis, IN), Illumina HiSeq/MiSeq (San Diego, CA), Ion Torrent and the Applied Biosystems/SOLiD System (Wadapurkar and Vyas, 2018). Each of these platforms have their own unique enzyme system, sequencing chemistry, hardware, and software engineering. The sequenced reads obtained from NGS technique, and the total sequencing output vary from one platform to another. The Roche/ 454 platform was the first commercial NGS platform introduced in 2005. Some of its major advantages are its speed and can complete a sequencing run in 10 hours. It can generate from several hundred thousand to 1 million reads of 200-300 bp DNA fragments per run. The Illumina/Solexa and Applied Biosystems/SOLiD platforms was introduced in 2006 and can generate tens of millions of short reads of about 30-40 bp per run. It is the cheapest sequencing platform, costing $0.02 per million bases. In 2007 the Life Technologies SOLiD was introduced, and this platform has the highest accuracy of 99.94% since errors are corrected with a two base coding system. It produces short reads per run that are 35bp (Wadapurkar and Vyas, 2018). The Ion Torrent platform became commercial in 2010.

Its focus is on monitoring pH change using a unique silicon conductor. It produces reads that are 200bp and is a fast technique that is suitable for laboratories that are working with small data sets (Langmead and Salzberg, 2012). Next generation sequencing has previously been successfully used in meat species identification (Xing *et al.*, 2019; Cottenet *et al.*, 2020; Liu *et al.*, 2021). The technology was reported to be an untargeted tool that can identify closely related species and even exotic species (Cottenet *et al.*, 2020). Meat species that were in in mixtures were identified even species that were as low as 1% (w/) (Cottenet *et al.*, 2020). Liu *et al.*, 2021 managed to identify pig, chicken, duck, cattle, and sheep meat that was mixed at different proportions and the detection of the meat samples was as low as 0,5%. There is need to test NGS in a South African study and possibly measure the sensitivity and specificity of the technology to as low as 0,1%.

2.12. Workflow for NGS data analysis

When running an NGS experiment, a DNA library needs to be prepared before the samples are sequenced using the sequencing platform of choice. Thereafter, quality assessment of the sequenced reads is performed, followed by aligning the reads to a reference genome. Variant identification and annotation are performed before visualization (Li and Durbin, 2010).

2.12.1. Quality assessment of raw reads

The first step after completing a sequence run is to assess the quality of raw reads to remove, trim or correct reads that do not meet the desired standards. Trimming is required to avoid drawing incorrect biological tools. There are different tools that can be used to perform various stages of quality assessment: Fast QC, FASTX-Toolkit, Galaxy, NGSQC, PRINSEQ and SolexaQA, just mention a few (Li *et al.,* 2008). The stand-alone tools such as NGSQC (Alkan *et al.*, 2009) toolkit and PRINSEQ (Larkin *et al.*, 2007) can handle FASTQ and 454 (SFF) files, produce summary reports and also filter and trim reads. The FastQC tool is compatible with all the main sequencing platforms and can give output summary graphs and tables to quickly assess the data quality. Galaxy tool is able to create FASTQ summary statistics and carryout trimming and filtering tasks. SOLEXAQA performs quality assessment, processing and visualization functions (Li *et al.*, 2009). Furthermore, there are software tools that have been published that support only the Illumina platform, and these are FASTX-Toolkit, PIQA and

TileQC (Lee and Schatz, 2012).

2.12.2. Alignment of sequenced reads to a reference genome

Following cleaning of reads to meet a certain quality standard, alignment to an existing reference genome occurs. There are several alignment programs that have been developed in order to process millions of short reads and include among others, Bowtie/Bowtie2 (Pabinger *et al.*, 2014), BWA (Wadapurkar and Vyas, 2018), MAQ (Botstein and Risch, 2003), mrFAST (Kathiresan and Srivastava, 2012), Clustal W (Meindl *et al.*, 2010) and SOAP (Margulies *et al.*, 2005). In addition to the selection of an alignment program here are three factors that need to be considered. Firstly, to avoid the problem of ambiguity when mapping short reads to a reference genome, paired end reads have shown to be a valuable solution and are highly recommended for whole genome sequencing (Kathiresan and Srivastava, 2012). Secondly, reads that can only be mapped with many mismatched should not be considered and mutations that are backed with such reads should be discarded from further analysis. Thirdly, current NGS technologies incorporate PCR steps in their library preparations, multiple reads originating from only one template might be sequenced, thereby interfering with variant calling statistics. Therefore, it is common to remove PCR duplicates after alignment in whole genome sequencing studies (Meindl *et al.*, 2010).

2.12.3. Variant identification of NGS data

The tools for variant identification can be grouped in four groups: (i) germline callers, (ii) somatic callers, (iii) Copy number variation (CNV) identification and (iv) Structural variation (SV) identification (Cibulskis *et al.*, 2011). Examples of tools that can perform germline callers are Atlas 2 (Tatusova and Madden, 1999), GATK (Pabinger *et al.*, 2014), SAMtools (Pabinger *et al.*, 2014) and SOAPPindel (Meindl *et al.*, 2010). For somatic callers, GATK (Pabinger *et al.*, 2014), SAMtools (Pabinger *et al.*, 2014) and Somatic call (Cline and Kent, 2009) may be used. Another tool that can be used for gene and protein identification, using sensitive, selective and rapid similarity searches of protein and nucleotide sequence databases is BLAST (Bataille *et al.*, 1999).

2.12.4. Variant annotation of NGS data

The tools that are available for variant annotation often implement different methods. A large proportion of them concentrate on the annotation of Single Nucleotide Polymorphisms (SNPs), since they can be easily identified and analyzed. INDELs are also performed by some tools, however, annotation of structural variants is limited to CNVs and only performed by recently developed programs. The most popular way of annotation is providing database links to different public variant databases such as dbSNP. Examples of tools that can be used for variant annotation are Ann tools, Absolute and SNP hunter Top (Li *et al.*, 2020).

2.12.5. Visualization of NGS data

Visualization representation of data is useful for interpretation of results obtained. Consequently, NGS visualization tools should support users by displaying aligned reads, mapping quality and identified mutations combined with annotations from various public resources. The tool should also be user friendly, intuitive and responsive. Visualization tools can be divided in three: (i) finishing tools supporting the interpretation of sequence data of de novo or re-sequencing experiments, (ii) genome browsers that allow users to browse mapped experimental data in combination with different types of annotation and (ii) comparative viewers that facilitate the comparison of sequences from multiple organisms or individuals (Li *et al.*, 2020). Examples of genome browsers are Samtool tview (Yang *et al.*, 2014), Apollo and Abrowse. Some visualization tools that have been developed enable visualization of CNVs and SVs. Examples of these are Circos and Germlin (Karabasanavar *et al.*, 2017).

Genome browsers can be divided into two groups: Web-based applications running on a dedicated web server (Wang) and stand-alone tools that are mostly Graphical Guided Interface (GUI). The advantages of web-based genome browsers are that they support a variety of annotations. Therefore, the user is able to browse reference genomes and different types of genomic annotation from a variety of public databases. In addition, users do not need to install new applications with numerous dependencies and computationally intensive calculations are performed on the server. One disadvantage of web-based genome browsers is the need to upload data on a remote server, thereby posing security and legal issues. The advantages of stand-alone browsers are that they offer interactive browsing and zooming features that may

not be available in web-based browsers and there is no need for uploading the data to websites. The disadvantages are that there is a need to download annotation files and keep the annotations up to date. Furthermore, complex calculations must be performed by regular PCs, which may not be powerful enough to handle such workloads (Floren *et al.*, 2015a). When interpreting aligned sequences using a genome browser: (i) reads that could not be mapped with many mismatched should not be trusted and mutations that are only backed by a small fraction of reads should be discarded and (ii) reads should only be trusted for further processing if they align at a unique starting position (Céspedes *et al.*, 2000).

2.13. Conclusion

In conclusion this review has shown that the consumption of meat has increased worldwide due to an increase in the global population. The leading countries in meat production and consumption are China, USA and Brazil. In Africa, South Africa are the leading producers of meat followed by Egypt and Nigeria. Beef meat is the most common produced meat type in South Africa followed by broiler meat and pork. The increase in the cost of beef and mutton has resulted in an increase in the consumption of chicken and pork because they are more affordable and available.

Meat adulteration cases were first reported in Europe which led to cases from around the globe being reported. The main methods that have previously been used to detect meat species adulteration are DNA and protein-based methods. However, these methods are species specific and have various drawbacks. Next generation sequencing is an upcoming technology that can be used as a universal method in meat species identification. This study will use the NGS technology to (i) identify genes/genomic regions that can be used as universal markers for meat species identification; (ii) use the identified gene(s) to develop a molecular and bioinformatics tool that can identify artificially mixed and processed meat samples; (iii) develop a pipeline that can be used to identify breed specific SNPS in cattle breeds.

CHAPTER THREE

3. A multi locus sequence analysis of the *16S* rRNA, *12S* rRNA, *COX3* and *ATP6* mitochondrial genes to determine their discriminatory potential in mammalian species

3.1. Abstract

The accurate labelling of meat products has become a concern to consumers, particularly after reports of fraudulent activities in the meat industry were revealed. Processed meat products are most susceptible to adulteration. A universal molecular and bioinformatic pipeline is required to investigate meat adulteration and support the meat industry. This study compares the discriminatory potential of *16S* rRNA,*12S* rRNA, *COX3* and *ATP6* mitochondrial genes of species, to determine which gene has the highest discriminatory potential for use in meat species identification. An entire database of 263 species for each gene were downloaded from Genbank (http://www.ncbi.nlm.nih.gov/nuccore). The sequences for the respective genes were extracted from the databases using FeatureExtract 1.2 Server (http://www.cbs.dtu.dk/services/FeatureExtract). All sequences were exported into MAFFT v.7 (https://mafft.cbrc.jp/alignment/server/) for a multiple sequence analysis. The phylogenetic analysis was run using Maximum Likelihood (ML) trees in IQ-Tree v.1.6.10 (http://www.iqtree.org) at 1000 bootstrap replications. The tree was visualised in iTOL (Interactive tree of life) v.3 (http://itol.embl.de). Eight phylogenetic trees were constructed, with four trees of each individual gene being initially constructed, followed by three trees with two gene combinations, then finally one tree with all four genes combined. The phylogenetic analysis of each individual mitochondrial gene demonstrated that all four genes managed to separate distantly related species and group closely related species with a common ancestry. However, some discrepancies were observed in the *12S* rRNA, *ATP6* and *COX3* genes. The *12S* rRNA and *COX3* individual trees failed to group all of the bat species (*Rhinolophidae, Phyllostomidae, Vespertilionidae and Hipposideridae*) together. Furthermore, in the *COX3* individual tree, the *Halichoerus Grypus* (Seal) species of the *Phocidae* family grouped with the *Bubalus Bubalis* (Buffalo) species of the *Bovidae* family. Finally, in the *ATP6* individual tree the *Soricidae* (Shrew) family did not group with *Talpidae* family, which are part of the

shrew family. Monophyletic clades were identified in most of the species' groupings and one paraphyletic clade was identified in the *Aves* and *Reptalia* species. Overall, the *16S* rRNA gene performed well on its own with high bootstrap values in all clades observed (97 – 100%). The *ATP6* gene performed the least with low bootstrap values as low as 45% in some groupings. The same clades were observed when the *16S* rRNA gene was paired with the other genes. However, there was no improvement in the performance of the three other genes when they were paired with the *16S* rRNA gene. The combination of all four genes in one tree resulted in bootstrap values of 100% in most of the clades. Based on these findings, the use of the *16S rRNA* gene in species identification molecular experiments is recommended as this gene has a high discriminatory potential on its own. A multiplex of all four genes will be more expensive and time consuming, without necessarily improving the discriminatory potential of the pipeline.

Keywords: Species identification; discriminatory potential, mitochondrial genes, mammalian species, processed meat

3.2. Introduction

Species identification in processed meat has become prominent and vital in meat industries worldwide. This is because processed meat products are more vulnerable to meat adulteration because they are difficult to identify morphologically after they are processed into value added products. Processing changes the physical and chemical make-up of meat product through heating, addition of salts and processing aids. However, processing of meat also increases the value of a meat product through improving the taste and flavour, decreasing preparation time and decreasing preparation steps. Meat species identification is mainly carried out in processed meat samples, that undergo a lot of heating during processing into value added products. Common methods that have traditionally been used for meat species identification are protein based and these methods include isoelectric focusing (IEF) and immunological methods (Céspedes *et al.*, 2000). Isoelectric focusing is not well suited for heated samples, as accurate results are not guaranteed. Immunological methods pose the risk of cross-reaction with closely related proteins because the methods are based on the use of antibodies (Carrera *et al.*, 2000). The setbacks of the above-mentioned identification methods brought about the introduction of Deoxyribonucleic Acids (DNA) methods in species identification. Nucleic acids have proven

to be sensitive, reliable and specific (Li *et al.*, 2020). Various DNA based techniques have been implemented and these include Polymerase Chain Reaction (PCR), DNA hybridization, species-specific PCR, restriction enzyme digestion and DNA sequencing (Yang *et al.*, 2014). The validity of nucleic acid-based techniques depends on the discriminatory potential of the method. Morden day molecular biology has allowed for rapid screening of DNA sequences, producing accurate and reliable results (Carrera *et al.*, 2000). DNA based methods for species identification include use of specifically targeted mitochondrial DNA genes (mtDNA) (Aranishi, 2005) and nuclear single copy genes (Morán and Garcia-Vazquez, 2006).

In comparison to nuclear DNA, mitochondrial DNA has a maternal inheritance, fast evolution rate, simple structure, small size and limited recombination (Javonillo *et al.*, 2010; Luo *et al.*, 2011). These characteristics play an active role in understanding the phylogenetic distribution of species. Mitochondrial DNA has been used in species identification and phylogeny for the past 30 years (Rubinoff and Holland, 2005; Tillmar *et al.*, 2013; Yang *et al.*, 2014). Mitochondrial DNA has one non-coding region, 13 protein-coding genes, namely, *cytochrome subunit 1*, *2* and *3* (*COX1*, *COX2* and *COX3*), *cytochrome b subunit* (*cytb*), *NADH dehydrogenase subunits 1*, *2*, *3*, *4*, *4L*, *5* and *6* (*ND1*, *ND6* and ND4L), *ATPase subunits 6* and *8* (*ATP6* and *ATP8*). The mitochondria genome also contains 2 ribosomal RNA (rRNA) genes, namely, *16S* rRNA and *12S* rRNA and 22 transfer RNA genes (Carrera *et al.*, 2000; Klinbunga *et al.*, 2003; Luo *et al.*, 2011). The use of mtDNA in species identification is advantageous because; it is easy to extract using the correct DNA extraction kit and methodology; it has several copy numbers in each species – thereby preventing sequencing ambiguity from heterozygous genotypes; it does not contain large non-coding sequences and; it does not have any recombination that cause genetic rearrangements. (Floren *et al.*, 2015b; Ren *et al.*, 2017). Mitochondrial DNA has been used extensively in genetic research and therefore, various universal primers have been designed that can be used in identification of meat and fish species (Biswas and Rajkumar, 2009). Some of the common mtDNA genes that have previously been used as universal markers in meat and fish species identification and have proven to show accurate results are *cytb*, and *COX1* (Unajak *et al.*, 2011; Cawthorn, Steinman and Hoffman, 2013; Dai *et al.*, 2015; Zhang *et al.*, 2020). A few studies have previously been conducted using the *16S* rRNA and *12S* rRNA genes with results demonstrating potential utility of these genes (Mane *et al.*, 2013; Tillmar *et al.*, 2013; He *et al.*, 2018). A more recent study (Spychaj *et al.*, 2021) looked at species identification in six processed white meat products using specific

primers covering the *12S* rRNA, *COX3*, *5S* rRNA, *Pan I*, *ATP6*, *D-Loop* and *COX2* genes. However, the genes used in this study were not tested in all species and the primers used were not universal. A gene labelled as a universal marker for accurate species identification, needs to accurately identify a species at close to 100% thresholds (Dupuis, Julius *et al.,* 2012). This will result in the gene being able to be used for identification of a wide variety of inter- and intra-species.

Often a gene selected as a universal marker in species identification may not accurately identify all species within a population. Furthermore, it is essential to select alternative universal markers for species identification and avoid constantly relying on the commonly used *COX1* and *cyt b* genes. It is therefore important to conduct a thorough testing of the alternative genes, determine the effectiveness of gene for potential use in meat species investigations. Before selecting a gene of choice in meat species identification, it is also recommended to use a multi-locus phylogenetic analysis to determine the best gene that can accurately differentiate a wide variety of mammalian species, with potential to be used as a universal marker. Closely related species are more difficult to differentiate than distantly related species and a good gene region is one that can discriminate both closely related and distantly related species. The aim of this study was to conduct a multi-locus phylogenetic analysis of mammalian species to determine the discrimination power of the *16S rRNA*, *12S* rRNA, *ATP6* and *COX3* mitochondrial genes and determine their potential for use in processed meat species identification studies.

3.3. Materials and methods

3.3.1. Extraction of sequences from the Genbank

An entire database of sequences for 258 *Mammalia*, species, 4 *Aves* species and 1 *Reptalia* species were downloaded from GenBank (https://www.ncbi.nlm.nih.gov/genbank/) for the mitochondrial *16S rRNA*, *12S rRNA*, *COX3* and *ATP6* genes. The species belonged to 76 taxonomic families whose scientific names, common names and families are found in Appendix 3.1. Appendix files 3.2 – 3.5 contain the sequences of each species for each of the mitochondrial genes. The DNA sequences were extracted using Feature Extract 1.2 Server (http://www.cbs.dtu.dk/services/FeatureExtract/).

3.3.2. Phylogenetic analysis for validation of discriminating power across species

All sequences were exported into MAFFT v.7 (https://mafft.cbrc.jp/alignment/server/) for a multiple sequence alignment. Multiple sequence alignment of these was performed by using the default settings in MAFFT which included MAFFT v7.471 with the mafft-linsi option (An alias for an accurate option (L-INS-i) for an alignment of up to ~200 sequences × ~2,000 sites) specified. To visualize the ability of the genes to separate different mammalian species, Maximum Likelihood (ML) trees were constructed using the GTR + I + G model in IQ-Tree v.1.6.10 (http://www.iqtree.org). This software performs a phylogenetic analysis on a multiple sequence alignment file developed in MAFFT. The software uses default settings and can automatically detect the type of sequence and most appropriate substitution model.

Ultrafast Bootstrap and SH – aLRT branch tests were performed. The number of bootstrap and SH – aLRT branch test replications were set at 1 000. The bootstrap analysis shows how well supported a tree is, taking into consideration the data input and the method used to construct the tree. The horizontal length of branches indicates the evolutionary distance between organisms revealing the number of nucleotide substitutions per site along the branch from the node to the endpoint (Alkan *et al.*, 2009). A total of 8 ML trees were constructed as follows: (i) Four ML phylogenetic trees of each individual gene: *16S* rRNA gene tree, *12S* rRNA gene tree, *COX3* gene tree and *ATP6* gene tree. (ii) Three ML phylogenetic trees with combinations of genes: *16S* rRNA and *12S* rRNA genes tree, *16S* rRNA and *COX3* genes tree and *16S* rRNA and *ATP6* genes tree and finally (iii) one ML phylogenetic tree with a combination of all four genes was constructed. Visualisation of the ML phylogenetic trees was performed in iTOL (Interactive tree of life) v.3 (http://itol.embl.de).

3.4. Results

3.4.1. Species analysed

The 258 *Mammalia* species, 4 *Aves* species and 1 *Reptalia* species were downloaded from the database and analysed. The species belonged to 35 families as presented in Table 3.1.

Table 3. 1: Groupings of species used in the phylogenetic analyses

Family	Species
Delphinidae	Dolphins
Balaenidae, Eschrichtiidae, Meobalaenidae and Physteridae	Whales
Cervidae and Moschidae	Deers
Bovidae	Cattle, Sheep, Goats, Buffalos, Kudus, Bisons and Antelopes
Suidae	Domestic Pig, Wild Boar and Warthog
Camelidae	Dromedary, Camels, and Vicunas
Rhinocerotidae	Rhinoceroses
Felidae	Domestic Cats, Lions, Leopards, Tigers, and Cougars
Ursidae	Bears and Pandas
Mustelidae	Badgers, Otters, Martens, and Sables
Phocidae and Otariidae	Seals and Sea Lions
Rhinolophidae, Phyllostomidae, Vespertilionidae and Hipposideridae	Bats
Talpidae and Soricidae	Moles and Shrews
Elephantidae	Elephants and Mammoths
Leporidae and Ochotoridae	Hares and Pikas
Lemuridae, Lepilemuridae and Indriidae	Lemurs and Sikakas
Cebidae	Capuchin and Squirrel Monkey

Homidae	Orangutan, Gorilla, Chimpanzee and Bonobo
Hylobatidae	Gibbon and Saimang
Cercopithecidae and Rhinolophodae	Baboon, Marcaque, Grivet, Surili and Duoc
Cricetidae	Voles and Hamsters
Aves and Reptalia	Ducks, Turkey, Chickens, Ostriches, and Crocodiles

3.4.2. Phylogenetic analysis

We initially carried out an individual phylogenetic analyses on the four genes to see how they would perform in discriminating species. The phylogenetic analysis of each individual mitochondrial gene, namely, *16S* rRNA, *12S* rRNA, *COX3* and *ATP6*, indicated that all four genes managed to separate distantly related species and group closely related species with a common ancestry as illustrated in Figure S1 – S4 and described in the details below.

3.4.2.1. Maximum likelihood *16S* rRNA phylogenetic tree

In the *16S* rRNA phylogenetic tree a total of 22 monophyletic clades and one paraphyletic were formed (Table 3.2, Appendix 3.6). The *Delphinidae* (Dolphins) species were at the top of the tree and formed a clade with 73 – 97% bootstrap values. The *Balaenidae*, *Eschrichtiidae*, *Meobalaenidae* and *Physteridae* (Whales) species followed and formed a clade with bootstrap values of 97 – 100%. This was the highest value amongst the four genes for this grouping. The *Cervidae* and *Moschidae* (Deer) species formed a clade below the whales with a bootstrap value of 97 – 100%. The *Bovidae* (Cattle, Sheep, Goats, Antelopes, Kudus, Bisons and Buffalos) family also formed a clade with bootstrap values of 80 – 100%. The *Bos taurus* (Cattle) grouped with the *Bos javanicus* (Banteng) species with a bootstrap value of 100%. The *Capra hircus* (Goat) species formed a clade with the *Pseudois schaeferi* (Dwarf Blue Sheep) and *Ammotragus lervia* (Barbary Sheep) with a bootstrap value of 100% within the *Bovidae*

family clade. The *Ovis* species (Sheep and Argali) formed a clade within with *Bovidae* family with the *Oryx* (Gemsbok and Scmitar Horned Oryx) with a bootstrap value of 100%.

The *Suidae* (Domestic Pig, Wild Boar and Warthog) species followed below with a bootstrap value of 100%. The *Camelidae* (Dromedary, Camels, and Vicunas) species followed below with a bootstrap value of 100 %. The *Rhinocerotidae* (Rhinoceros) grouped together with a bootstrap value of 87 – 100%. The *Felidae* (Domestic Cats, Lions, Leopards, Tigers, and Cougars) species formed a clade with bootstrap values of 78 – 100%, which was the highest value for this clade amongst the four genes. The *Ursidae* (Bears and Pandas) species also formed a clade with bootstrap values of 80 – 100%. The *Mustelidae* (Badgers, Otters, Martens and Sables) species all grouped together with a bootstrap value 100%. This was also the highest bootstrap value amongst the four genes. Thereafter, the *Phocidae* and *Otariidae* (Seals and Sea Lions) species grouped together with bootstrap values of 100%. This was the highest value amongst the four genes for this grouping. The *Rhinolophidae*, *Phyllostomidae*, *Vespertilionidae* and *Hipposideridae* (Bats) species followed and formed a clade with 96 – 100% bootstrap values. Similarly, this was the highest value amongst the four genes for this grouping.

The *Talpidae* and *Soricidae* (Moles) species followed with a bootstrap value of 100%, which was the highest bootstrap value once again, amongst the four genes. The *Elephantidae* (Mammoths and Elephants) grouped together with 100% bootstrap values. The *Leporidae* and *Ochotoridae* (Hares and Pikas) species formed their own clade with 100% bootstrap values. The *Lemuridae*, *Lepilemuridae* and *Indriidae* (Lemurs and Sikaka) species followed and formed a clade with 100% bootstrap values. The *Cebidae* (Capuchin and Squirrel Monkey) species grouped together with a 100% bootstrap value. This was followed by the *Homidae* (Orangutan, Gorilla, Chimpanzee and Bonobo) species that formed a clade with values of 96 – 100%. The *Hylobatidae* (Gibbon and Saimang) species also formed a clade with 99 – 100% bootstrap values. This was the highest value amongst the four genes for this grouping. The *Cercopithecidae* and *Rhinolophodae* (Baboon, Marcaque, Grivet, Surili and Duoc) species all grouped together with bootstrap values of 100%. Similarly, this was the highest value amongst the four genes for this grouping. The *Cricetidae* (Voles and Hamsters) species grouped together with bootstrap values of 99 – 100%. The *Muridae* (Rats and Mice) species formed a clear distinct clade with bootstrap values of 98 – 100%), which was the highest bootstrap value

amongst the four genes. At the bottom of the tree, the *Aves* (Ducks, Turkey, Chickens and Ostriches) species formed a paraphyletic clade with the R*eptalia* (Crocodiles) species with a bootstrap value of 100%.

Table 3. 2: Discrimination of species and families based on *16S* rRNA phylogeny

Clade	Species	Bootstrap Values
Delphinidae	Dolphins	73 - 97%
Balaenidae, Eschrichtiidae, Meobalaenidae and Physteridae	Whales	97 - 100%
Cervidae and Moschidae	Deers	97 - 100%
Bovidae	Cattle, Sheep, Goats, Antelopes, Kudus, Bisons and Buffalos	80 - 100%
Suidae	Domestic Pig, Wild Boar and Warthog	100%
Camelidae	Dromedary, Camels, and Vicunas	100%
Rhinocerotidae	Rhinoceros	87 - 100%
Felidae	Domestic Cats, Lions, Leopards, Tigers, and Cougars	78 - 100%
Ursidae	Bears and Pandas	80 - 100%
Mustelidae	Badgers, Otters, Martens, and Sables	100%
Phocidae and *Otariidae*	Seals and Sea Lions	100%
Rhinolophidae, *Phyllostomidae*, *Vespertilionidae* and *Hipposideridae*	Bats	96 - 100%
Talipidae and *Soricidae*	Moles	100%
Elephantidae	Mammoths and Elephants	100%

Leporidae and *Ochotoridae*	Hares and Pikas	100%
Lemuridae, *Lepilemuridae* and *Indriidae*	Lemurs and Sikakas	100%
Cebidae	Capuchin and Squirrel Monkey	100%
Homidae	Orangutan, Gorilla, Chimpanzee and Bonobo	96 - 100%
Hylobatidae	Gibbon and Saimang	99 - 100%
Cercopithecidae and Rhinolophodae	Baboon, Marcaque, Grivet, Surili and Duoc	100%
Cricetidae	Voles and Hamsters	99 - 100%
Muridae	Rats and Mice	98 - 100%
Aves and Reptalia	Ducks, Turkey, Chickens, Ostriches, and Crocodiles	100%

3.4.2.2. Maximum likelihood *12S* rRNA phylogenetic tree

In the *12S* rRNA tree a total of 22 monophyletic clades and one paraphyletic were formed (Table 3.3, Appendix 3.7). The Delphinidae (Dolphins) species formed a clade and were also at the top of the tree with 67% bootstrap values, which was the lowest amongst the four individual trees (Table 3.3, Figure S2). The *Balaenidae*, *Eschrichtiidae*, *Meobalaenidae* and Physteridae (Whales) species followed with bootstrap values of 86 – 100%. The *Camelidae* (Dromedary, Camels, and Vicunas) species followed below the whales with a bootstrap value of 100 %. The *Cervidae* and *Moschidae* (Deer) species formed a clade below the whales with a bootstrap value of 61 – 100%. The *Bovidae* (Cattle, Sheep, Goats, Antelopes, Kudus, Bisons and Buffalos) family formed a clade below with bootstrap values of 78 – 100%. Within the *Bovidae* clade, the *Bos taurus* (Cattle) grouped with the *Bos javanicus* (Banteng) species with a bootstrap value of 99%. The *Ovis* (Sheep and Argali) species grouped together with 100% bootstrap value and formed a clade with the *Capra hircus* (Goat), *Naemorhedus Caudatus* (Long tailed goral) and *Capricornis Swinihoei* (Taiwan serow) with a bootstrap value of 99 – 100%.

The *Suidae* (Domestic Pig, Wild Boar and Warthog) species followed below with a bootstrap value of 100 %. The *Rhinolophidae*, *Phyllostomidae* and *Hipposideridae* (Bats) species followed and formed a clade with 53 – 100% with bootstrap values, but the other bat species (*Vespertilionidae*) grouped alone lower further below in the tree after the *Talpidae* (Shrews and Moles) species. It seems the *12S* rRNA tree failed to discriminate them. The *Rhinocerotidae* (Rhinoceros) grouped together with a bootstrap value of 99 – 100%. This was the highest value amongst the four genes for this grouping. The *Felidae* (Domestic Cats, Lions, Leopards, Tigers, and Cougars) species followed with bootstrap values of 93 – 100%. This was also the highest value amongst the four genes for this grouping like the 16S gene. The *Ursidae* (Bears and Pandas) species also formed a clade with bootstrap values of 60 – 100%. The *Mustelidae* (Badgers, Otters, Martens, and Sables) species all grouped together with a bootstrap value 100%, which was the highest bootstrap value like the *16S* rRNA gene. Thereafter, the *Phocidae* and *Otariidae* (Seals and Sea Lions) species grouped together with bootstrap values of 97 – 100%.

The *Talpidae* and *Soricidae* (Moles) species followed with a bootstrap value of 42 – 100%. The *Lemuridae*, *Lepilemuridae* and *Indriidae* (Lemurs and Sikaka) species followed and formed a clade with 99 – 100% bootstrap values. The *Cebidae* (Capuchin and Squirrel Monkey) species grouped together with a 100% bootstrap value. This was followed by the *Homidae* (Orangutan, Gorilla, Chimpanzee and Bonobo) species that formed a clade with 100% bootstrap value. This was the highest value amongst the four genes for this grouping. The *Hylobatidae* (Gibbon and Saimang) species also formed a clade with 49 – 100% bootstrap values. The *Cercopithecidae* and *Rhinolophodae* (Baboon, Marcaque, Grivet, Surili and Duoc) species all grouped together with bootstrap values of 62 – 100%. The *Elephantidae* (Mammoths and Elephants) grouped together with 100% bootstrap values. The *Aves* (Ducks, Turkey, Chickens and Ostriches) species formed a paraphyletic clade with the R*eptalia* (Crocodiles) species with a bootstrap value of 100%. This clade was surprisingly formed amongst mammalian species clades. The *Leporidae* and *Ochotoridae* (Hares and Pikas) species formed their own clade with 100% bootstrap values. The *Cricetidae* (Voles and Hamsters) species grouped together with bootstrap values of 94 – 100%. At the bottom of the tree, the *Muridae* (Rats and Mice) species with bootstrap values of 86 – 100%.

Table 3. 3: Discrimination of species and families based on *12S* rRNA phylogeny

Clade	Species	Bootstrap Values
Delphinidae	Dolphins	67%
Balaenidae, Eschrichtiidae, Meobalaenidae and Physteridae	Whales	86 - 100%
Camelidae	Dromedary, Camels, and Vicunas	100%
Cervidae and Moschidae	Deers	61 - 100%
Bovidae	Cattle, Sheep, Goats, Antelopes, Kudus, Bisons and Buffalos	78 - 100%
Suidae	Domestic Pig, Wild Boar and Warthog	100%
Rhinolophidae, Phyllostomidae and Hipposideridae	Bats	53 - 100%
Rhinocerotidae	Rhinoceros	99 - 100%
Felidae	Domestic Cats, Lions, Leopards, Tigers, and Cougars	78 - 100%
Ursidae	Bears and Pandas	60 - 100%
Mustelidae	Badgers, Otters, Martens, and Sables	100%
Phocidae and *Otariidae*	Seals and Sea Lions	97 - 100%
Talipidae and *Soricidae*	Moles	42 - 100%
Lemuridae, *Lepilemuridae* and *Indriidae*	Lemurs and Sikakas	99 - 100%
Cebidae	Capuchin and Squirrel Monkey	100%
Homidae	Orangutan, Gorilla, Chimpanzee and Bonobo	100%

Hylobatidae	Gibbon and Saimang	49 - 100%
Cercopithecidae and Rhinolophodae	Baboon, Marcaque, Grivet, Surili and Duoc	62 - 100%
Elephantidae	Mammoths and Elephants	100%
Aves and Reptalia	Ducks, Turkey, Chickens, Ostriches, and Crocodiles	100%
Leporidae and *Ochotoridae*	Hares and Pikas	100%
Cricetidae	Voles and Hamsters	94- 100%
Muridae	Rats and Mice	86 - 100%

3.4.2.3. Maximum likelihood *COX3* phylogenetic tree

In the *COX3* phylogenetic tree a total of 20 monophyletic clades and one paraphyletic were formed (Table 3.4, Appendix 3.8). the *Delphinidae* (Dolphins) species were at the top of the tree and formed a clade with 99 % bootstrap values, which was the highest amongst the individual gene trees (Table 3.4, Figure S3). The *Balaenidae*, *Eschrichtiidae*, *Meobalaenidae* and *Physteridae* (Whales) species followed with bootstrap values of 94 – 100%. The *Camelidae* (Dromedary, Camels, and Vicunas) species followed below the whales with a bootstrap value of 100%. The *Cervidae* and *Moschidae* (Deer) species formed a clade below the whales with a bootstrap value of 87 – 100%. The *Bovidae* (Cattle, Sheep, Goats, Antelopes, Kudus, Bisons and Buffalos) family also formed a clade with bootstrap values of 46 – 100%. However, within this clade it was surprising to see the *Halichoerus Grypus* (Seal) species of the *Phocidae* family grouped with the *Bubalus Bubalis* (Buffalo) species of the *Bovidae* family. This indicates that the *COX3* gene incorrectly grouped the *Halichoerus Grypus* (Seals).

The *Bos taurus* (Cattle) species grouped with the *Bos javanicus* (Banteng) species with a bootstrap value of 100% within the *Bovidae* clade. The *Ovis* (Sheep) species also grouped together with a 100% bootstrap value and further went on to form a monophyletic clade with

the with *Capra hircus* (Sheep), *Ammotragus lervia* (Barbary sheep) and *Pseudois schaeferi* (Dwarf blue sheep) species with a bootstrap value of 99 – 100%.

The *Suidae* (Domestic Pig, Wild Boar and Warthog) species followed below with a bootstrap value of 100%. The *Rhinocerotidae* (Rhinoceros) grouped together with a bootstrap value of 87 – 100%. The *Felidae* (Domestic Cats, Lions, Leopards, Tigers, and Cougars) species followed with bootstrap values of 52 – 100%. Thereafter, the *Phocidae* and *Otariidae* (Seals and Sea Lions) species grouped together with bootstrap values of 89 – 100%. The *Mustelidae* (Badgers, Otters, Martens and Sables) species all grouped together with a bootstrap value 97 – 100%. The *Ursidae* (Bears and Pandas) species also formed a clade with bootstrap values of 84 – 100%. The *Rhinolophidae* and Hipposideridae (Bats) formed a clade with 100% bootstrap values, but the other bat species (*Vespertilionidae* and *Phyllostomidae*) grouped lower down the tree with the *Sciuridae* (Red Squirrel) and *Erinaceidae* (European Hare) species. The *Phyllostomidae* and *Vespertilionidae* (Bats) species are not of the same order with the *Sciuridae* (Red Squirrel) and *Erinaceidae* (European Hare) species. The species belong to the *Chiroptera*, *Rodentia* and *Insectivora* orders, respectively.

The *Lemuridae*, *Lepilemuridae* (Lemurs) did not form a distinct clade, but simply settled one after the other in the *COX3* tree. However, in the *16S* rRNA, *12S* rRNA and *ATP6* trees, a distinct clade was formed. The *Elephantidae* (Mammoths and Elephants) grouped together with 100% bootstrap values. The *Leporidae* and *Ochotoridae* (Hares and Pikas) species formed their own clade with 100% bootstrap values. The *Aves* (Ducks, Turkey, Chickens and Ostriches) species formed a paraphyletic clade with the R*eptalia* (Crocodiles) species with a bootstrap value of 99 – 100%. Similarly, to the *12S* rRNA tree, this clade was surprisingly formed amongst mammalian species clades. The *Hominidae* and *Hylobatidae* (Gorillas, Chimpanzees, Orangutans and Gibbons) formed a clade with 50 – 100% bootstrap values. The *Cebidae*, *Cercopothecidae* and *Rhinolophidae* (Capuchin, Squirrel Monkey, Grivet, Marcaque, Monkeys and Duoc) species all formed a distinct clade with bootstrap values of 100%. This was the highest value amongst the four genes for this grouping. The *Talpidae* and *Soricidae* (Moles) species followed with a bootstrap value of 98 – 100%. At the bottom of the tree the *Muridae* (Rats and Mice) species grouped with the *Cricetidae* (Vole and Hamsters) species, with bootstrap values from 30 – 100%.

Table 3. 4: Discrimination of species and families based on *COX3* phylogeny

Clade	Species	Bootstrap Values
Delphinidae	Dolphins	99%
Balaenidae, Eschrichtiidae, Meobalaenidae and Physteridae	Whales	94 - 100%
Camelidae	Dromedary, Camels, and Vicunas	100%
Cervidae and Moschidae	Deers	87 - 100%
Bovidae	Cattle, Sheep, Goats, Antelopes, Kudus, Bisons and Buffalos	46 - 100%
Suidae	Domestic Pig, Wild Boar and Warthog	100%
Rhinocerotidae	Rhinoceros	87 - 100%
Felidae	Domestic Cats, Lions, Leopards, Tigers, and Cougars	52 - 100%
Phocidae and *Otariidae*	Seals and Sea Lions	89 - 100%
Mustelidae	Badgers, Otters, Martens, and Sables	97 - 100%
Ursidae	Bears and Pandas	84 - 100%
Rhinolophidae and Hipposideridae	Bats	100%
Elephantidae	Mammoths and Elephants	100%
Leporidae and *Ochotoridae*	Hares and Pikas	100%
Aves and Reptalia	Ducks, Turkey, Chickens, Ostriches, and Crocodiles	100%
Phocidae and *Otariidae*	Seals and Sea Lions	97 - 100%
Homidae and *Hylobatidae*	Orangutan, Gorilla, Chimpanzee, Bonobo, Gibbon and Saimang	50 - 100%

Cebidae	Capuchin and Squirrel Monkey	100%
Cercopithecidae and Rhinolophodae	Baboon, Marcaque, Grivet, Surili and Duoc	100%
Talpidae and *Soricidae*	Moles	98 - 100%
Muridae and *Cricetidae*	Rats, Mice, Voles and Hamsters	30 - 100%

3.4.2.4. Maximum likelihood *ATP6* phylogenetic tree

In the *ATP6* phylogenetic tree a total of 19 monophyletic clades and one paraphyletic were formed (Table 3.5, Appendix 3.9). The *Delphinidae* (Dolphins) species were at the top of the tree and formed a clade, also with 93 – 94% bootstrap values The *Balaenidae*, *Eschrichtiidae*, *Meobalaenidae* and *Physteridae* (Whales) species followed with bootstrap values of 94 – 100%, like the *COX3* tree. The *Suidae* (Domestic Pig, Wild Boar and Warthog) species followed below the whales with a bootstrap value of 100%. The *Camelidae* (Dromedary, Camels and Vicunas) species followed below the *Suidae* family with a bootstrap value of 100%. The *Bovidae* (Cattle, Sheep, Goats, Antelopes, Kudus, Bisons and Buffalos) family also formed a clade with bootstrap values of 85 – 100%. This was the highest value amongst the four genes for this grouping. The *Bos taurus* (Cattle) species grouped with the *Bos javanicus* (Banteng) species with a bootstrap value of 100% within the *Bovidae* clade. Interestingly, the *Capra hircus* species did not group with any other species in the *ATP6* tree, but instead settled between the *Ammotragus lervia* and *Pseudois schaeferi* species with a bootstrap value of 85%. The *Ovis* (Sheep and Argali) species grouped together with 100% bootstrap value but did not form a clade with any other species.

The *Cervidae* and *Moschidae* (Deer) species formed a clade below the *Bovidae* family with a bootstrap value of 99 – 100%. This was the highest value amongst the four genes for this grouping. The *Rhinocerotidae* (Rhinoceros) grouped together with a bootstrap value of 60 – 100%. Thereafter, the *Talpidae* (Moles) species followed with bootstrap values of 84 – 100% but the other shrew family – *Soricidae* were not part of the clade. The *Rhinolophidae*, *Phyllostomidae*, *Vespertilionidae* and *Hipposideridae* (Bats) all grouped together and formed a clade with 45 – 100% bootstrap values.

The *Felidae* (Domestic Cats, Lions, Leopards, Tigers and Cougars) species formed a clade with bootstrap values of 74 – 100%. The *Mustelidae* (Badgers, Otters, Martens and Sables) species all grouped together with a bootstrap value 92 – 100%. The *Ursidae* (Bears and Pandas) species also formed a clade with bootstrap values of 85 – 100%. This was the highest value amongst the four genes for this grouping. Thereafter, the *Phocidae* and *Otariidae* (Seals and Sea Lions) species grouped together with bootstrap values of 98 – 99%. The *Lemuridae*, *Lepilemuridae* (Lemurs) and the *Indriidae* (Sikaka) species followed and formed a clade with 100% bootstrap values.

The *Leporidae* and *Ochotoridae* (Hares and Pikas) species formed their own clade with 91 – 100% bootstrap values. The *Cebidae* (Capuchin and Squirrel Monkey) species grouped together with a 100% bootstrap value. This was followed by the *Homidae* (Orangutan, Gorilla, Chimpanzee and Bonobo) and *Hylobatidae* (Gibbon and Saimang) species with 82 – 100% bootstrap values. The *Cercopithecidae* and *Rhinolophodae* (Baboon, Marcaque, Grivet, Surili and Duoc) species all grouped together with bootstrap values of 88 – 100%. The *Elephantidae* (Mammoths and Elephants) grouped together with 100% bootstrap values. The *Aves* (Ducks, Turkey, Chickens and Ostriches) species formed a paraphyletic clade with the R*eptalia* (Crocodiles) species with a bootstrap value of 100%. Similar, to the *COX3* tree, this clade was surprisingly formed amongst mammalian species clades. The *Cricetidae* (Vole and Hamsters) formed a clade with 100% bootstrap value. This was the highest value amongst the four genes for this grouping. At the bottom of the tree *Muridae* (Mice and Rats) species did not form a distinct clade. These species simply settled on the tree one after the other without forming a clade.

Table 3. 5: Discrimination of species and families based on *ATP6* phylogeny

Clade	Species	Bootstrap Values
Delphinidae	Dolphins	93 - 94%
Balaenidae, Eschrichtiidae, Meobalaenidae and Physteridae	Whales	94 - 100%

Suidae	Domestic Pig, Wild Boar and Warthog	100%
Camelidae	Dromedary, Camels, and Vicunas	100%
Bovidae	Cattle, Sheep, Goats, Antelopes, Kudus, Bisons and Buffalos	85 - 100%
Cervidae and Moschidae	Deers	99 - 100%
Rhinocerotidae	Rhinoceros	60 - 100%
Talpidae	Moles	84 - 100%
Rhinolophidae, *Phyllostomidae*, *Vespertilionidae* and *Hipposideridae*	Bats	45 - 100%
Felidae	Domestic Cats, Lions, Leopards, Tigers, and Cougars	74 - 100%
Mustelidae	Badgers, Otters, Martens and Sables	92 - 100%
Ursidae	Bears and Pandas	85 - 100%
Phocidae and *Otariidae*	Seals and Sea Lions	98 - 99%
Lemuridae, *Lepilemuridae* and *Indriidae*	Lemurs and Sikakas	100%
Leporidae and *Ochotoridae*	Hares and Pikas	91 - 100%
Cebidae	Capuchin and Squirrel Monkey	100%
Homidae and *Hylobatidae*	Orangutan, Gorilla, Chimpanzee, Bonobo, Gibbon and Saimang	82 - 100%
Cercopithecidae and Rhinolophodae	Baboon, Marcaque, Grivet, Surili and Duoc	88 - 100%
Elephantidae	Mammoths and Elephants	100%
Aves and Reptalia	Ducks, Turkey, Chickens, Ostriches, and Crocodiles	100%
Cricetidae	Voles and Hamsters	100%

Overall, the *16S* rRNA and *12S* rRNA genes had the highest number of clades of 23 (Table 3.2 and Table 3.3), followed by the *COX3* that had 21 (Table 3.4) and the *ATP6* that had 20 (Table 3.5). The *16S* rRNA gene performed better as compared to the other three genes. When looking at the observed clades that were the same in all four genes, the *16S* rRNA gene had nine clades with the highest bootstrap values. The *12S* rRNA and *ATP6* had four and the *COX3* gene had two. The *16S* rRNA gene grouped all bat species **(***Rhinolophidae*, *Phyllostomidae*, *Vespertilionidae* and *Hipposideridae*) (Appendix 3.6), whereas that was not the case in *12S* rRNA (Appendix 3.7) and *COX3* genes (Appendix 3.8). Furthermore, the *Aves* and *Reptalia* species grouped at the bottom of the *16S* rRNA tree and not amongst the mammalian clade, as was observed in the *12S* rRNA tress and *COX3* tree. In the *COX3* tree the *Halichoerus Grypus* (Seal) species of the *Phocidae* family grouped with the *Bubalus Bubalis* (Buffalo) species of the *Bovidae* family. This was evident that the *COX3* incorrectly grouped these species. In the *ATP6* tree the *Soricidae* (Shrew) family did not group with *Talpidae* family (Appendix 3.9), which are also part of the shrew family. Furthermore, the *Muridae* family did not form a distinct clade in the *ATP6* tree. Given these overall results, we paired the *16S* rRNA gene with each of the other three genes to see how they would perform in a phylogeny. We also combined all four genes in a phylogeny to see how they would perform together.

3.4.2.5. Maximum likelihood *16S* rRNA and *12S* rRNA combined phylogenetic tree

We combined the *16S* rRNA gene with the *12S* rRNA gene to see how both genes combined would perform. In the *16S* rRNA and *12S* rRNA tree a total of 22 monophyletic clades and one paraphyletic were formed (Table 3.6, Appendix 3.10). Similarly, as in the individual gene trees, the *Delphinidae* (Dolphins) species were at the top of the tree and formed a clade with 61 – 74% bootstrap values (Table 3.6, Appendix 3.11). These bootstrap values were lower than those that were observed in the individual *16S* rRNA and *12S* rRNA trees. The *Balaenidae*, *Eschrichtiidae*, *Meobalaenidae* and *Physteridae* (Whales) species followed with bootstrap values of 97 – 100%. This was the highest bootstrap value amongst the combined genes. Thereafter, the *Cervidae* and *Moschidae* (Deer) species formed a clade below the *Bovidae*

family with a bootstrap value of 96 – 100%. The *Bovidae* (Cattle, Sheep, Goats, Antelopes, Kudus, Bisons and Buffalos) species also formed a clade with bootstrap values of 89 – 100%. The *Bos taurus* (Cattle) grouped with the *Bos javanicus* (Banteng) species with a bootstrap value of 100%. The *Ovis* species (Sheep and Argali) formed a clade with the *Capra hircus* (Goat), *Pseudois Schaeferi* (Dwarf blue sheep), *Naemorhedus Caudatus* (Long tailed goral) and *Capricornis Swinihoei* (Taiwan serow) with 98 – 100% bootstrap. This was then followed with the *Suidae* (Domestic Pig, Wild Boar and Warthog) family with a bootstrap value of 100%.

The *Camelidae* (Dromedary, Camels, and Vicunas) species followed below the *Suidae* family also with a bootstrap value of 100%. The *Felidae* (Domestic Cats, Lions, Leopards, Tigers, and Cougars) species formed a clade with bootstrap values of 79 – 100%. The *Ursidae* (Bears and Pandas) species also formed a clade with bootstrap values of 55 – 100%. These values are lower than those observed in the *16S* rRNA individual tree, that had a bootstrap value of 80 – 100%. Thereafter, the *Mustelidae* (Badgers, Otters, Martens and Sables) species all grouped together with a bootstrap value 99 – 100%. Thereafter, the *Phocidae* and *Otariidae* (Seals and Sea Lions) species grouped together with bootstrap values of 89 – 100%. These bootstrap values were lower that both the *16S* rRNA and *12S* rRNA trees, that both had bootstrap values of 100% for the *Phocidae* and *Otariidae* grouping. The *Rhinocerotidae* (Rhinoceros) species grouped together with a bootstrap value of 99 – 100%. This was the highest bootstrap value amongst the combined phylogenies.

Similarly, to the *16S* rRNA individual tree, the *Rhinolophidae*, *Phyllostomidae*, *Vespertilionidae* and *Hipposideridae* (Bats) species all grouped together and formed a clade with 98 – 100% bootstrap values. However, in the *12S* rRNA tree, the other bat species (*Vespertilionidae*) grouped alone lower further below in the tree after the *Talpidae* (Shrews and Moles) species. The *Talpidae* and *Soricidae* (Moles) species followed with a bootstrap value of 99 – 100%. Thereafter, The *Lemuridae*, *Lepilemuridae* (Lemurs) and *Indriidae* (Sikaka) species all grouped together with 100% bootstrap values. The *Cebidae* (Capuchin and Squirrel Monkey) species grouped together with a 100% bootstrap value. This was followed by the *Homidae* (Orangutan, Gorilla, Chimpanzee and Bonobo) species that formed a clade with values of 100%. The *Hylobatidae* (Gibbon and Saimang) species also formed a clade with 100% bootstrap values. The *Cercopithecidae* and *Rhinolophodae* (Baboon, Marcaque, Grivet,

Surili and Duoc) species all grouped together with bootstrap values of 100%. The *Elephantidae* (Mammoths and Elephants) species grouped together with 100% bootstrap values. The *Leporidae* and *Ochotoridae* (Hares and Pikas) species formed their own clade with 100% bootstrap values. The *Cricetidae* (Voles and Hamsters) species grouped together with bootstrap values of 100%. Thereafter, the *Muridae* (Rats and Mice) species formed a clear distinct clade with bootstrap values of 81 – 100%). At the bottom of the tree, the *Aves* (Ducks, Turkey, Chickens and Ostriches) species formed a paraphyletic clade with the R*eptalia* (Crocodiles) species with a bootstrap value of 100%.

Table 3. 6: Discrimination of species and families based on *16S rRNA and 12S rRNA* combined phylogeny

Clade	Species	Bootstrap Values
Delphinidae	Dolphins	61 - 74%
Balaenidae, Eschrichtiidae, Meobalaenidae and Physteridae	Whales	97 - 100%
Cervidae and Moschidae	Deers	96 - 100%
Bovidae	Cattle, Sheep, Goats, Antelopes, Kudus, Bisons and Buffalos	89 - 100%
Suidae	Domestic Pig, Wild Boar and Warthog	100%
Camelidae	Dromedary, Camels, and Vicunas	100%
Felidae	Domestic Cats, Lions, Leopards, Tigers, and Cougars	79 - 100%
Ursidae	Bears and Pandas	55 - 100%
Mustelidae	Badgers, Otters, Martens, and Sables	99 - 100%
Phocidae and *Otariidae*	Seals and Sea Lions	89 - 100%
Rhinocerotidae	Rhinoceros	99 - 100%
Rhinolophidae, *Phyllostomidae*, *Vespertilionidae* and *Hipposideridae*	Bats	98 - 100%

Talpidae and *Soricidae*	Moles	99 - 100%
Lemuridae, *Lepilemuridae* and *Indriidae*	Lemurs and Sikakas	100%
Cebidae	Capuchin and Squirrel Monkey	100%
Homidae	Orangutan, Gorilla, Chimpanzee and Bonobo	100%
Hylobatidae	Gibbon and Saimang	100%
Cercopithecidae and Rhinolophodae	Baboon, Marcaque, Grivet, Surili and Duoc	100%
Elephantidae	Mammoths and Elephants	100%
Leporidae and *Ochotoridae*	Hares and Pikas	100%
Cricetidae	Voles and Hamsters	99 - 100%
Muridae	Rats and Mice	81 - 100%
Aves and Reptalia	Ducks, Turkey, Chickens, Ostriches, and Crocodiles	100%

3.4.2.6. Maximum Likelihood *16S* rRNA and *COX3* combined phylogenetic tree

Looking at *16S* rRNA and *COX3* genes combined, a total of 21 monophyletic clades and one paraphyletic were formed (Table 3.7, Appendix 3.11) the *Delphinidae* (Dolphins) species were at the top of the tree and formed a clade with 89 – 100% bootstrap values (Table 3.7, Figure S6). The *Balaenidae*, *Eschrichtiidae*, *Meobalaenidae* and *Physteridae* (Whales) species followed with bootstrap values of 95 – 100%. Thereafter, the *Cervidae* and *Moschidae* (Deer) species formed a clade above the *Bovidae* family with a bootstrap value of 95 – 100%. The *Bovidae* (Cattle, Sheep, Goats, Antelopes, Kudus, Bisons and Buffalos) species also formed a clade with bootstrap values of 80 – 100%. Kudus, Bisons and Buffalos) species also formed a clade with bootstrap values of 89 – 100%. The *Bos taurus* (Cattle) grouped with the *Bos javanicus* (Banteng) species with a bootstrap value of 100%. The *Ovis* (Sheep and Argali) species formed a clade with the *Capra hircus* (Goat), *Ammotragus Lervia* (Barbary sheep) and *Pseudois Schaeferi* (Dwarf blue sheep) with a bootstrap value of 100%.

The *Suidae* (Domestic Pig, Wild Boar and Warthog) family followed with a bootstrap value of 100%. The *Camelidae* (Dromedary, Camels and Vicunas) species formed a clade below the *Suidae* family also with a bootstrap value of 100%. The *Rhinocerotidae* (Rhinoceros) grouped together above the *Felidae* species with a bootstrap value of 85 – 100%. The *Felidae* (Domestic Cats, Lions, Leopards, Tigers and Cougars) species formed a clade with bootstrap values of 94 – 100%. The *Ursidae* (Bears and Pandas) species also formed a clade with bootstrap values of 61 – 100%. These values are lower than those observed in the *16S* rRNA and *COX3* individual trees, that both had bootstrap value of 80 – 100%. Thereafter, the *Mustelidae* (Badgers, Otters, Martens and Sables) species all grouped together with a bootstrap value 100%. This was the highest bootstrap valued amongst the combined genes. The *Phocidae* and *Otariidae* (Seals and Sea Lions) species grouped together with bootstrap values of 71 – 100%. These bootstrap values were lower that of the *16S* rRNA individual tree, that had a bootstrap value of 100% for the *Phocidae* and *Otariidae* grouping. Similarly, to the *16S* rRNA tree, the *Rhinolophidae*, *Phyllostomidae*, *Vespertilionidae* and *Hipposideridae* (Bats) species all grouped together and formed a clade with 98 – 100% bootstrap values. However, in the *COX3* individual tree, the other bat species (*Vespertilionidae* and *Phyllostomidae*) grouped alone lower further below in the tree with the *Sciuridae* (Red Squirrel) and *Erinaceidae* (European Hare). The *Talpidae* and *Soricidae* (Moles) species followed with a bootstrap value of 99 – 100%.

The *Elephantidae* (Mammoths and Elephants) species grouped together with 100% bootstrap values. Thereafter, the *Lemuridae* and *Lepilemuridae* (Lemurs) and *Indriidae* (Sikaka) did not form a distinct clade, but simply settled one after the other species all grouped together. This was like the *COX3* individual tree. A similar order of clades was followed in this tree as was observed in the *16S* rRNA and *12S* rRNA combined tree. The *Cebidae* (Capuchin and Squirrel Monkey) species grouped together with a 100% bootstrap value. This was followed by the *Homidae* (Orangutan, Gorilla, Chimpanzee and Bonobo) species that formed a clade with values of 100%. The *Hylobatidae* (Gibbon and Saimang) species also formed a clade with 100% bootstrap values. The *Cercopithecidae* and *Rhinolophodae* (Baboon, Marcaque, Grivet, Surili and Duoc) species all grouped together with bootstrap values of 100%. The *Leporidae* and *Ochotoridae* (Hares and Pikas) species formed their own clade with 100% bootstrap values. The *Cricetidae* (Voles and Hamsters) species grouped together with bootstrap values of 98 – 100%. Thereafter, the *Muridae* (Rats and Mice) species formed a clear distinct clade

with bootstrap values of 98 – 100%. This was the highest bootstrap value amongst the combined phylogenies. Similarly, to the *16S* rRNA and *12S* rRNA combined tree, the *Aves* (Ducks, Turkey, Chickens and Ostriches) species formed a paraphyletic clade with the R*eptalia* (Crocodiles) species at the bottom of the tree with a bootstrap value of 100%.

Table 3. 7: Discrimination of species and families based on *16S rRNA and COX3 rRNA* combined phylogeny

Clade	Species	Bootstrap Values
Delphinidae	Dolphins	89 - 100%
Balaenidae, Eschrichtiidae, Meobalaenidae and Physteridae	Whales	95 - 100%
Cervidae and Moschidae	Deers	95 - 100%
Bovidae	Cattle, Sheep, Goats, Antelopes, Kudus, Bisons and Buffalos	80 - 100%
Suidae	Domestic Pig, Wild Boar and Warthog	100%
Camelidae	Dromedary, Camels, and Vicunas	100%
Rhinocerotidae	Rhinoceros	85 - 100%
Felidae	Domestic Cats, Lions, Leopards, Tigers, and Cougars	94 - 100%
Ursidae	Bears and Pandas	61 - 100%
Mustelidae	Badgers, Otters, Martens, and Sables	100%
Phocidae and *Otariidae*	Seals and Sea Lions	71 - 100%
Rhinolophidae, *Phyllostomidae*, *Vespertilionidae* and *Hipposideridae*	Bats	98 - 100%
Talpidae and *Soricidae*	Moles	99 - 100%
Elephantidae	Mammoths and Elephants	100%

Cebidae	Capuchin and Squirrel Monkey	100%
Homidae	Orangutan, Gorilla, Chimpanzee and Bonobo	100%
Hylobatidae	Gibbon and Saimang	100%
Cercopithecidae and Rhinolophodae	Baboon, Marcaque, Grivet, Surili and Duoc	100%
Leporidae and *Ochotoridae*	Hares and Pikas	100%
Cricetidae	Voles and Hamsters	98 - 100%
Muridae	Rats and Mice	98 - 100%
Aves and Reptalia	Ducks, Turkey, Chickens, Ostriches, and Crocodiles	100%

3.4.2.7. Maximum Likelihood *16S rRNA* and *ATP6* combined phylogenetic tree

In the *16S* rRNA and *ATP6* genes combined tree, a total of 22 monophyletic clades and one paraphyletic were formed (Table 3.8, Appendix 3.12). The *Delphinidae* (Dolphins) species were at the top of the tree and formed a clade with 76 – 89% bootstrap values. These values were lower than both the *16S* rRNA and *ATP6* individual trees, that had values of 73 – 97% and 93 – 94%, respectively. The *Balaenidae*, *Eschrichtiidae*, *Meobalaenidae* and *Physteridae* (Whales) species followed with bootstrap values of 97 – 100%. This was the highest bootstrap values amongst the combined genes. This was the highest bootstrap value together amongst the combined phylogenies. The *Bovidae* (Cattle, Sheep, Goats, Antelopes, Kudus, Bisons and Buffalos) species also formed a clade with bootstrap values of 64 – 100%. The *Bos taurus* (Cattle) grouped with the *Bos javanicus* (Banteng) species with a bootstrap value of 100%. Similarly, as in the *16S* and *COX3* combined tree, the *Ovis* (Sheep and Argali) species formed a clade with the *Capra hircus* (Goat), *Ammotragus Lervia* (Barbary sheep) and *Pseudois Schaeferi* (Dwarf blue sheep) with a bootstrap value of 97 – 100%. Thereafter, the *Cervidae* and *Moschidae* (Deer) species formed a clade below the *Bovidae* family with a bootstrap value of 96 – 100%. The *Suidae* (Domestic Pig, Wild Boar and Warthog) species grouped together

with a bootstrap value of 100%. Similarly, the *Camelidae* (Dromedary, Camels and Vicunas) family also had a bootstrap value of 100% in their clade. The *Rhinocerotidae* (Rhinoceros) grouped together above the *Felidae* species with a bootstrap value of 99 – 100%. The *Felidae* (Domestic Cats, Lions, Leopards, Tigers and Cougars) species grouped together with bootstrap values of 98 – 100%.

The *Ursidae* (Bears and Pandas) followed and formed a clade with bootstrap values of 81 – 100%. These values were similar to those observed in the *16S* rRNA and *ATP6* individual trees (85 – 100%). Thereafter, the *Mustelidae* (Badgers, Otters, Martens and Sables) species all grouped together with a bootstrap value of 86 – 100%. The *Phocidae* and *Otariidae* (Seals and Sea Lions) species grouped together with bootstrap values of 87 – 100%. These bootstrap values were similar to that of the *16S* rRNA (87 – 100%) and lower than those for the *ATP6* (98 – 99%) individual trees. Similarly, to the *16S* rRNA and *ATP6* individual tree, the *Rhinolophidae*, *Phyllostomidae*, *Vespertilionidae* and *Hipposideridae* (Bats) species all grouped together and formed a clade with 99 – 100% bootstrap values. This was the highest bootstrap value amongst the combined genes. However, in the *ATP6* individual tree, the bootstrap value was lower (45 – 100%). The *Talpidae* and *Soricidae* (Moles) species followed with a bootstrap value 100%. This was also the highest bootstrap value amongst the combined genes. However, in the *ATP6* individual tree the other shrew family – *Soricidae* were not part of the clade. The *Elephantidae* (Mammoths and Elephants) species grouped together with 100% bootstrap values.

Similar to the *16S rRNA* and *12S rRNA* combine tree, this was the highest bootstrap value amongst the combined genes. The *Leporidae* and *Ochotoridae* (Hares and Pikas) also grouped together with 100% bootstrap values. Thereafter, the *Lemuridae*, *Lepilemuridae* (Lemurs) and *Indriidae* (Sikaka) species formed a well-defined clade with 100% bootstrap values. A similar order of clades was followed in this tree as was observed in the *16S* rRNA and *COX3* combined tree. The *Cebidae* (Capuchin and Squirrel Monkey) species grouped together with a 100% bootstrap value. This was followed by the *Homidae* (Orangutan, Gorilla, Chimpanzee and Bonobo) species that formed a clade with values of 76 – 100%. The *Hylobatidae* (Gibbon and Saimang) species also formed a clade with 98 – 100% bootstrap values. The *Cercopithecidae* and *Rhinolophodae* (Baboon, Marcaque, Grivet, Surili and Duoc) species all grouped together with bootstrap values of 100%. The *Cricetidae* (Voles and Hamsters) species grouped together

with bootstrap values of 100%. Thereafter, the *Muridae* (Mice) species formed a clade on their own with bootstrap values of 97 – 100%. This was the highest bootstrap value for mice amongst the combined genes. The rat (*Muridae*) species did not form a distinct clade. They settled on the tree one after the other without forming a clade. Similarly, to the *16S* rRNA and *12S* rRNA combined tree, the *Aves* (Ducks, Turkey, Chickens and Ostriches) species formed a paraphyletic clade with the R*eptalia* (Crocodiles) species at the bottom of the tree with a bootstrap value of 78 – 100%.

Table 3. 8: Discrimination of species and families based on *16S rRNA and ATP6 rRNA* combined phylogeny

Clade	Species	Bootstrap Values
Delphinidae	Dolphins	76 - 89%
Balaenidae, Eschrichtiidae, Meobalaenidae and Physteridae	Whales	97 - 100%
Bovidae	Cattle, Sheep, Goats, Antelopes, Kudus, Bisons and Buffalos	64 - 100%
Cervidae and Moschidae	Deers	96 - 100%
Suidae	Domestic Pig, Wild Boar and Warthog	100%
Camelidae	Dromedary, Camels, and Vicunas	100%
Rhinocerotidae	Rhinoceros	99 - 100%
Felidae	Domestic Cats, Lions, Leopards, Tigers, and Cougars	98 - 100%
Ursidae	Bears and Pandas	81 - 100%
Mustelidae	Badgers, Otters, Martens, and Sables	86 - 100%
Phocidae and *Otariidae*	Seals and Sea Lions	87 - 100%

Rhinolophidae, *Phyllostomidae*, *Vespertilionidae* and *Hipposideridae*	Bats	99 - 100%
Talpidae and *Soricidae*	Moles	100%
Elephantidae	Mammoths and Elephants	100%
Leporidae and *Ochotoridae*	Hares and Pikas	100%
Lemuridae, Lepilemuridae and Indriidae	Lemurs and Sikakas	100%
Cebidae	Capuchin and Squirrel Monkey	100%
Homidae	Orangutan, Gorilla, Chimpanzee and Bonobo	76 - 100%
Hylobatidae	Gibbon and Saimang	98 - 100%
Cercopithecidae and Rhinolophodae	Baboon, Marcaque, Grivet, Surili and Duoc	100%
Cricetidae	Voles and Hamsters	100%
Muridae	Rats and Mice	97 - 100%
Aves and Reptalia	Ducks, Turkey, Chickens, Ostriches, and Crocodiles	78 - 100%

3.4.2.8. Maximum likelihood phylogenetic tree with a combination of all four genes

Finally, we combined all four genes to see how they would perform combined. A total of 21 monophyletic clades and one paraphyletic were formed (Table 3.9, Figure S8). Similarly, as in the individual gene trees, the *Delphinidae* (Dolphins) species were at the top of the tree with 96 – 100% bootstrap values (Table 3.9, Appendix 3.13). The *Balaenidae*, *Eschrichtiidae*, *Meobalaenidae* and *Physteridae* (Whales) species followed with bootstrap values of 94 – 100%. Thereafter, the *Cervidae* and *Moschidae* (Deer) species formed a clade above the *Bovidae* family with a bootstrap value of 99 – 100%. The *Bovidae* (Cattle, Sheep, Goats, Antelopes, Kudus, Bisons and Buffalos) species also formed a clade with bootstrap values of 98 – 100%. The *Bos taurus* (Cattle) grouped with the *Bos javanicus* (Banteng) species with a

bootstrap value of 100%. Similarly, as in the *16S* and *ATP6* combined tree, the *Ovis* (Sheep and Argali) species formed a clade with the *Capra hircus* (Goat), *Ammotragus Lervia* (Barbary sheep) and *Pseudois Schaeferi* (Dwarf blue sheep) with a bootstrap value of 99 – 100%. The *Suidae* (Domestic Pig, Wild Boar and Warthog) species also formed a clade with a bootstrap value of 100%. The *Camelidae* (Dromedary, Camels and Vicunas) species grouped below the *Suidae* family with a bootstrap value of 100%. The *Rhinocerotidae* (Rhinoceros) species grouped together with a bootstrap value of 99 – 100%. The *Felidae* (Domestic Cats, Lions, Leopards, Tigers and Cougars) species formed a distinct clade with bootstrap values of 99 – 100%. The *Ursidae* (Bears and Pandas) species grouped together with bootstrap values of 72 – 100%. Thereafter, the *Mustelidae* (Badgers, Otters, Martens and Sables) formed a distinct clade with a bootstrap value 100%.

The *Phocidae* and *Otariidae* (Seals and Sea Lions) species grouped together with bootstrap values of 100%. Similarly, to the *16S* rRNA individual tree and all other tree combinations that included the *16S* rRNA tree, the *Rhinolophidae*, *Phyllostomidae*, *Vespertilionidae* and *Hipposideridae* (Bats) species all grouped together and formed a clade with 58 – 100% bootstrap values. The *Talpidae* and *Soricidae* (Moles) species followed with a bootstrap value of 99 – 100%. The *Elephantidae* (Mammoths and Elephants) species grouped together with 100% bootstrap values. The *Leporidae* and *Ochotoridae* (Hares and Pikas) species formed a clade with 100% bootstrap values. Thereafter, The *Lemuridae*, *Lepilemuridae* (Lemurs) and *Indriidae* (Sikaka) species all grouped together with 100% bootstrap values. The *Cebidae* (Capuchin and Squirrel Monkey) species grouped together with a 100% bootstrap value. The *Homidae* (Orangutan, Gorilla, Chimpanzee and Bonobo) and *Hylobatidae* (Gibbon and Saimang) species also formed their own distinct clades with 100% bootstrap values. The *Cercopithecidae* and *Rhinolophodae* (Baboon, Marcaque, Grivet, Surili and Duoc) species all grouped together with bootstrap values of 100%. Surprisingly, the *Aves* (Ducks, Turkey, Chickens and Ostriches) species formed a paraphyletic clade with the R*eptalia* (Crocodiles) species with a bootstrap value of 100% amongst mammalian species. The *Cricetidae* (Voles and Hamsters) species grouped together with bootstrap values of 59 – 100%. Thereafter, the *Muridae* (Rats and Mice) species formed a clear distinct clade with a bootstrap value of 100%. The four genes combined had nine clades with the highest bootstrap values. These were *Delphinidae*, *Cervidae and Moschidae*, *Bovidae*, *Felidae*, *Ursidae*, *Mustelidae*, *Phocidae* and *Otariidae*, *Rhinocerotidae* and *Muridae*.

Table 3. 9 Discrimination of species and families based on *16S rRNA*, *12S* rRNA, *COX3* and *ATP6 rRNA* combined phylogeny

Clade	Species	Bootstrap Values
Delphinidae	Dolphins	96 - 100%
Balaenidae, Eschrichtiidae, Meobalaenidae and Physteridae	Whales	94 - 100%
Cervidae and Moschidae	Deers	99 - 100%
Bovidae	Cattle, Sheep, Goats, Antelopes, Kudus, Bisons and Buffalos	98 - 100%
Suidae	Domestic Pig, Wild Boar and Warthog	100%
Camelidae	Dromedary, Camels, and Vicunas	100%
Rhinocerotidae	Rhinoceros	99 - 100%
Felidae	Domestic Cats, Lions, Leopards, Tigers, and Cougars	99 - 100%
Ursidae	Bears and Pandas	72 - 100%
Mustelidae	Badgers, Otters, Martens, and Sables	100%
Phocidae and *Otariidae*	Seals and Sea Lions	100%
Rhinolophidae, *Phyllostomidae*, *Vespertilionidae* and *Hipposideridae*	Bats	58 - 100%
Talpidae and *Soricidae*	Moles	99 - 100%
Elephantidae	Mammoths and Elephants	100%
Leporidae and *Ochotoridae*	Hares and Pikas	100%
Lemuridae, *Lepilemuridae* and *Indriidae*	Lemurs and Sikakas	100%

Cebidae	Capuchin and Squirrel Monkey	100%
Homidae and *Hylobatidae*	Orangutan, Gorilla, Chimpanzee and Bonobo	100%
Cercopithecidae and Rhinolophodae	Baboon, Marcaque, Grivet, Surili and Duoc	100%
Aves and Reptalia	Ducks, Turkey, Chickens, Ostriches, and Crocodiles	100%
Cricetidae	Voles and Hamsters	59 - 100%
Muridae	Rats and Mice	98 - 100%

Overall, the phylogeny which had all four genes combined performed the best amongst all the gene combinations. The phylogeny with all genes combined had nine clades with the highest bootstrap values, followed by the *16S* rRNA and *ATP6* combination that had five clades and lastly both the *16S* rRNA and *12S* rRNA combination and *16S* rRNA and *COX3* combination that each had two clades. The results also indicated that the four genes combined in one tree produced the most discriminatory potential with bootstrap values of 100% in most groupings (Appendix 3.13). There was an improvement in discrimination when the *16S* rRNA gene was paired with the other three genes. The bat species (*Rhinolophidae*, *Phyllostomidae*, *Vespertilionidae* and *Hipposideridae*) were all grouped together in the phylogeny trees that paired with the *16S* rRNA gene. Whereas the bat species were not all grouped together in the *12S* rRNA (Appendix 3.7) and *COX3* (Appendix 3.8) trees. In the *16S* rRNA and *COX3* combined tree the *Halichoerus Grypus* (Seal) species of the *Phocidae* family did not group with the *Bubalus Bubalis* (Buffalo) species of the *Bovidae* family, as observed in the *COX3* individual tree. The *Aves* species were found at the bottom of all groupings in all the combined trees and not amongst the mammals, as in the *12S* rRNA, *COX3* and *ATP6* individual trees. We observed more clades when the *COX3* (23) and *ATP6* (22) genes were paired with the *16S* gene, as opposed to their individual phylogenies that each had twenty-one and twenty, respectively. When the *ATP6* gene was grouped with the *16S* rRNA gene, the *Soricidae* (Shrew) species was also grouped with the other shrew family – *Talpidae*.

3.5. Discussion

The choice of an ideal marker for meat species identification depends on several factors. Firstly, the chosen marker should contain conserved regions that enable the development of universal markers (Farag *et al.*, 2015). Secondly, the marker should be variable enough to identify closely related species (Farag *et al.*, 2015). Thirdly, the marker should contain regions that are short and informative to allow sequencing in a single reaction (Farag *et al.*, 2015). Lastly, the chosen marker should allow for direct PCR sequencing without cloning (Nicolas *et al.*, 2012). Some of the factors above require experimental trials to be conducted on meat samples to conclude on the right choice of marker. Before we ventured into a meat species identification experimental trial that would cost time, money, resources and may potentially not give the desired results, we conducted a multi-locus phylogenetic analysis using published sequences to choose markers that will accurately identify species in processed meat samples. Phylogenetic tress are diagrams that illustrate evolutionary descendants of species from a common ancestor. Phylogenetic trees give a better understanding of evolutionary events that occurred in the past and different clades that are formed within a phylogenetic tree. A clade is a group of species within a phylogeny that shows ancestral lineage and all its descendants (Baum, 2008).

There are different types of clades, namely, monophyletic, paraphyletic, and polyphyletic clades. Monophyletic clades are made up of a single clade consisting of one ancestor and all its descendants. For example, the class mammals are a monophyletic group of species with mammary glands (Baum, 2008). A paraphyletic clade is made up of all the descendants of a common ancestor but does not include all descendants from the common ancestor. For example, reptiles and birds share a common ancestor, however, reptiles form a paraphyletic clade to birds (Baum, 2008). Lastly, a polyphyletic clade comprises of unrelated descendants from more than one ancestor. For example, elephants, rhinoceroses, and hippopotamuses are mammals that originated from different ancestors (Baum, 2008).The discriminatory potential of the *16S* rRNA, *12S* rRNA, *COX3* and *ATP6* mitochondrial genes was analysed by a phylogenetic analysis of 263 mammalian species. To the best of our knowledge, this is the first multi-locus analysis involving these four mitochondrial genes that was conducted in South Africa.

3.5.1. Maximum Likelihood phylogenetic analysis of individual genes.

We initially performed a phylogenetic analysis of each individual tree to see how each gene would perform on its own. All four genes managed to group species according to their respective ancestry origins, with few species being incorrectly grouped in some of the genes. With the *16S* rRNA individual tree, twenty-three clades were observed that consisted of twenty-two monophyletic clade sand one paraphyletic clade. A total of twenty-three clades were also observed in the *12S* rRNA individual tree, twenty-two monophyletic clades and one paraphyletic clade. In the *COX3* individual tree, twenty-one clades were observed, where twenty were monophyletic and one was paraphyletic. Finally, in the *ATP6* individual tree a total of twenty clades were formed, where nineteen were monophyletic and one was paraphyletic. The *Bovidae* family included cattle, sheep, goats, antelopes, gazellas, buffalos and Bisons and these species formed a monophyletic clade in the *16S* rRNA, *12S* rRNA, *COX3* and *ATP6* individual trees. All four individual genes managed to delineate these species according to their respective groupings. The *Bos* (cattle) species grouped correctly in all four individual trees, with the *16S* rRNA and *ATP6* having the highest bootstrap values in this grouping.

The *Bubalus bubalis* (Water Buffalo) settled above the *Bos* species in all four gene individual trees with strong branch support, however, this species was grouped with the *Halichoerus grypus* (Grey Seal) species from the *Phocidae* family in the *COX3* gene tree, even though these two species are not closely related. Previous studies on meat species identification have reported meat from water buffalo being adulterated in beef mince, burger patties and sausages (Dantas *et al.,* 2019; Cruz-Monterrosa *et al.,* 2020; Hossain *et al.,* 2021). Cases of water buffalo being adulterated in beef meat products could have been a case of cheaper offcuts from buffalo meat being adulterated in beef, as buffalo meat is more expensive than beef meat (Wang *et al.,* 2018). In another incident me commercial buffalo meat was found to be adulterated with beef, pork and duck meat for economic gain in China (Wang *et al.,* 2018). Meat from water buffalo is like that of beef, when nutritional levels, flavour profiles and physiochemical factors are considered. This makes it difficult to differentiate the two based on sensory evaluation only (Springer *et al.,* 2001) and instead would need the use of molecular identification. Therefore, it is vital to select a gene that can correctly differentiate species that are not closely related.

According to our results the *COX3* gene will not be an ideal gene to use as a universal marker for meat species identification, since it grouped water buffalos with grey seals.

The mitochondrial genes used in our study also managed to discriminate antelope species, namely *Tragelaphus imberbis* (Lesser Kudu), *Antilope cervicapra* (Blackbuck) and *Antidorcas marsupialis* (Springbok) form the *Bos* species (Cattle and Aurochs), with high bootstrap values across all four genes. These game species have been reported in meat adulteration studies. D'Amato *et al.*, (2013) conducted a South African study on adulteration in game meat and found that commercial samples labelled as kudu and springbok were adulterated with other wildlife meat types such as gemsbok, kangaroo, hartebeest, eland, and domestic meat such as beef and lamb. Therefore, all four genes can be used in meat species identification that includes beef and game meat. *Oryx gazella* (Gemsbok) which was previously reported to be adulterated (D'Amato *et al.,* 2013b) is closely related to sheep. A phylogenetic analysis by maximum likelihood (PAML) was conducted by Farré *et al.,* (2019) found gemsbok to be more closely related to sheep than to cattle and yak. In our phylogenetic study, the *Oryx* (Gemsbok) species formed a monophyletic clade with the *Ovis* (Sheep) species in the *16S* rRNA and *ATP6* gene trees. However, this was not observed in the *12S* rRNA and *COX3* gene trees, where the Oryx species did not group with the *Ovis* species and settled closer to the *Bos* (Cattle) species in both trees. These results demonstrate that the *12S* rRNA and *COX3* genes, did not manage to accurately delineate the *Oryx* species and cannot be used in such investigations. Similarly, Khan *et al.,* (2008) conducted a phylogenetic analysis of *Oryx* species and found that the frequency of the polymorphic sites and the average evolutionary divergence of *12S* rRNA gene were comparatively less than the *16S* rRNA gene. The *12S* rRNA gene could differentiate the *Oryx* species from closely related *Addax* species, which differ from the *Oryx* species at *genus* level.

The *Capra hircus* (Goat) species formed a monophyletic clade with the *Ovis* (Sheep) species in the *16S* rRNA and *COX3* gene trees. These results demonstrated a close relation of goats to sheep and that the *16S* rRNA and *COX3* can be used to investigate species identification in a meat mixture with such species suspected. However, quantification of the species may not be accurate given the close relation of the two species, as the mitochondrial genes can over or underestimate the quantity of sheep and goat DNA that may be contained in a meat sample. It has been reported that mitochondrial genes may not be the most suitable for quantification of

meat species due to a five-fold inter-tissue variation in mitochondrial DNA content (Floren *et al.*, 2015). This may lead to a -70% underestimation or +160% overestimation of the DNA content of species (Floren *et al.*, 2015). The *Capra hircus* (Goat) species grouped with the *Pseudois schaeferi* (Dwarf Blue Sheep) and *Ammotragus lervia* (Barbary Sheep) with a bootstrap value of 100% and formed a monophyletic clade with the *Budorcas*, *Oryx* and *Ovis* species in the *16S* rRNA tree. Similar phylogenetic analysis results were reported by Bibinu *et al.,* (2016) and Pawar *et al.,* (2013), where sheep and goats were found to have similar lineage. The *ATP6* gene tree did not show a clear clade that formed between the *Capra hircus* and *Ovis* species, indicating that this gene did not manage to accurately delineate the goat with the sheep species. The *Bovidae* family formed a separate clade from the *Moschidae* and *Cervidae* species (Deer) in all four individual trees. The *ATP6* gene had the highest bootstrap value amongst the four genes, followed by the *16S* rRNA gene. This indicated that the mitochondrial genes that we used in this study would be able to identify deer species in a meat identification experiment. Previous studies (Amaral *et al.*, 2014) have reported on the adulteration of game meat specifically deer meat with beef in order to reduce production costs.

Looking at the *Suidae* (Warthogs and Pigs) family, all four genes managed to delineate the species from the *Bovidae* family to form a monophyletic clade of their own just below the *Bovidae* family with high bootstrap values. This classification was correct as cattle, sheep and goats that are members of the *Bovidae* family, share the order *Artiodactyla* with the *Suidae* family. A phylogenetic tree considering 1000 bootstrap values drawn in Mega 4.1 (Tamura *et al.*, 2007) found that monogastric and ruminants are from different clusters with a close evolutionary relationship, however, pigs diverged early from the bovid ancestors. Pawar *et al.,*(2013), further went on to confirm these findings by running a codon-based Z test using the Nei Gojobori method (5% significance level), indicating that that selecting pressure attributed to the early divergence of pigs from bovids. Adulteration of pork meat in higher priced meat types such as beef and mutton have been reported (Cawthorn *et al.,* 2013; Ha *et al.*, 2017; Tembe *et al.,* 2018; Yang *et al.*, 2018; Han *et al.*, 2020). These reports are of concern to consumers who may be allergic to pork or do not consume pork due to religious regions. For example, the Islamic and Jewish laws do not allow the consumption of pork (Nakyinsige *et al.,* 2012; Hossain *et al.*, 2021; Islam *et al.*, 2021). Therefore, the mitochondrial genes we chose in our study can be used in meat species identification of beef meat samples that may be suspected to contain pork in them.

One of the monophyletic clades that was formed in all four individual gene trees comprised of bat species which include the *Rhinolophidae*, *Phyllostomidae*, *Vespertilionidae* and *Hipposideridae* families. Within our results we observed that the *12S* rRNA and *COX3* genes failed to group all of the bat species together. In the *12S* rRNA gene tree the *Rhinolophidae, Phyllostomidae and Hipposideridae* all grouped together and the *Vespertilionidae* grouped alone lower further below in the tree after the *Talpidae* (Shrews and Moles) species. In the *COX3* gene tree the *Rhinolophidae and Hipposideridae* species grouped together, and the other bat species (*Vespertilionidae* and *Phyllostomidae*) grouped lower down the tree with the *Sciuridae* (Red Squirrel) and *Erinaceidae* (European Hare) species, even though they are not of the same order with the *Sciuridae* (Red Squirrel) and *Erinaceidae* (European Hare) species. These results observed amongst the bat species indicate that the *12S* rRNA and *COX3* gene are not ideal genes to select as a universal marker in meat species identification, but rather as species specific markers.

The *Talpidae* and *Soricidae* (Moles) species formed a monophyletic in the *16S* rRNA, *12S* rRNA and *COX3* gene individual trees. However, in the *ATP6* gene individual tree, the *Talpidae* species did not group together with the *Soricidae* species. The *Muridae* (Rats and Mice) species formed distinct monophyletic clades in the *16S* rRNA, *12S* rRNA and *COX3* gene individual trees, with the highest bootstrap values being observed in the *16S* rRNA gene. The *ATP6* gene, however, did not manage to put the rats and mice in well-defined separate clades and instead these species looked like outgroups of the *Cricetidae* clade that was above them. Some reports (Raharjo *et al.*, 2019; Suryawan *et al.,* 2020) on unscrupulous traders who adulterate beef, chicken or pork meat samples with rats or mice as a means of reducing production costs have been published. Rats or mice are not considered as halal by Muslims are also found to be unhygienic. The presence of these species in meat samples, therefore, pose an ethical violation and health risk to consumers. Given these discrepancies, the *ATP6* gene would not be an ideal gene to select as a universal marker in meat species identification.

The *Aves* species (Chicken, Duck, Turkey and Ostrich) grouped on their own and formed a paraphyletic clade with *Crocodylus niloticus* (Nile Crocodile) that are *Reptilia* species with high branch support values in all four mitochondrial gene trees. Crocodile meat is an alternative protein source that has a similar taste to chicken but more expensive to purchase (Unajak *et al.,* 2011). It is therefore, considered as a delicacy. There is a concern on the exploitation of

wildlife, particularly that which is traded as bushmeat. For example, in the Philippines, wildlife species such as deer, warthog, land snail, crocodile, pangolin and ducks are illegally traded and sold in local markets (Fortajada *et al.,* 2021). Enforcement of laws against these illegal activities becomes difficult once the meat can no longer be visually identified. Hence, why there is a need to choose genes that will clearly separate crocodiles, ducks, chickens, turkeys and ostriches that usually fall within the same clade in a phylogenetic analysis. Despite high bootstrap values being observed in the paraphyletic clades for the *Aves* and *Reptalia* species in all four phylogenies of the individual genes, only the *16S* rRNA individual gene tree place the clade at the bottom of the tree. The *12S* rRNA, *COX3* and *ATP6* trees placed this clade amongst mammals and yet the *Aves* and *Reptalia* species are not the same class as *Mammalia*. These results further indicate that the *12S* rRNA, *COX3* and *ATP6* genes are not ideal to be used as universal markers for meat species identification.

Overall, the *16S* rRNA gene showed the highest discriminatory potential, the accurate groupings and high bootstrap values across all species. The *16S* rRNA individual phylogenetic tree had nine clades with the highest bootstrap values, followed by the *12S* rRNA and *ATP6* genes that each had four and the lowest being the COX3 genes that had two. Our results showed that the mitochondrial *16S* rRNA gene is highly variable for resolving taxonomic phylogenetic relationships of a wide range of mammalian species. Our findings are in agreement with Tillmar *et al.,* (2013) who conducted a phylogenetic analysis of 334 mammalian species using the 16S rRNA gene. The authors (Tillmar *et al.,* 2013) went on to describe that the strength of a marker to accurately describe or identify species, relies on its degree of separation between intraspecific variability and intraspecific diversity. The use of the *16S* rRNA gene in a meat species identification experiment is advantageous because, the 16S meta-genomic sequencing library preparation can be used prior to sequencing with an Illumina sequencing platform. Metagenomic studies are often carried out using the *16S* rRNA gene that contains variable regions that are among conserved regions (Amplicon, Clean-Up and Index, 2013). Therefore, this form of library preparation can be used for both species-specific and universal primers. More recently, Liu *et al.,* 2021 recommended the use of the 16S rRNA gene, as they observed no cross-reactivity of designed primer pairs. Furthermore, they noticed that the PCR they performed using the designed primer was simple, fast, sensitive, specific, and cost-effective. Finally, they noticed that the sensitivity, specificity, and accuracy in the primer pairs they

designed and that the primers were applicable in forensics to investigate blood spots or evidence belonging for human, sheep, goat, and cow (Liu *et al.,* 2021).

3.5.2. Maximum Likelihood phylogenetic analysis of combined genes.

We further went on and paired the *16S* rRNA gene with each of the other three genes in three separate tree combination, revealing some improvement in performance of the *12S* rRNA, *COX3* and *ATP6* genes. The *Capra hircus* species formed a monophyletic clade with the Ovis species in all the *16S* rRNA and *12S* rRNA gene combination, whereas this was not observed in the *12S* rRNA individual gene. Furthermore, in the *16S* rRNA and *12S* rRNA combined phylogenetic tree all the bat species (*Rhinolophidae*, *Phyllostomidae*, *Vespertilionidae* and *Hipposideridae*) formed a monophyletic clade, whereas in the *12S* rRNA individual tree the *Vespertilionidae* family was not grouped together with the rest of the bat species. Similarly, in the *16S* rRNA and *COX3* combined phylogenetic tree, all the bat species were grouped in one monophyletic clade. However, in the *COX3* individual tree the *Vespertilionidae* and *Phyllostomidae* families did not group with the rest of the bat species. Furthermore, in the *16S* rRNA and *COX3* combined tree, the *Halichoerus Grypus* (Seal) species of the *Phocidae* family did not group with the *Bubalus Bubalis* (Buffalo) species of the *Bovidae* family, as observed in the *COX3* individual tree. When the *ATP6* gene was combined with the *16S* rRNA gene all the shrew species (*Talpidae* and *Soricidae*) formed one monophyletic clade. Whereas, in the *ATP6* individual tree these two families did not group together. We also observed that when the *12S* rRNA, *COX3* and *ATP6* genes were combined with the *16S* rRNA genes, the *Aves* and *Reptalia* species formed a paraphyletic clade at the bottom of each tree, as opposed to amongst the mammals as observed in the individual trees of the *12S* rRNA, *COX3* and *ATP6* genes. We also observed more clades being formed when the *COX3* and *ATP6* genes were combined with the *16S* rRNA gene.

A combination of all four genes in one phylogenetic tree showed high bootstrap values of 100% in most of the species' groupings. A total of nine clades had the highest bootstrap values when all genes were combined in a phylogeny, followed by the *16S* rRNA and *ATP6* combination that had five clades with the highest bootstrap values. The *16S* rRNA and *12S* rRNA had two clades with the highest bootstrap values and this was similar to the *16S* rRNA and *COX3* combination. These results indicated that a multi-locus phylogenetic analysis of all four genes

would give the highest discriminatory potential as opposed to using an individual gene. Using all four genes is practical when running an analysis using sequences that have already been published on public domains. However, this may not be practical if you are considering a species identification molecular experiment in the lab using physical samples in the form of tissue, DNA or blood. Firstly, the use of all four genes in a multiplex experiment will be laborious and will result in pipetting errors, increased cross contamination and will be time consuming. Secondly, the use of all four genes will be more expensive as there would be a need to purchase all four genes and an increase in consumables used. Lastly, the computational power that will be required to analyse the sequenced data will be more expensive and time consuming as opposed to using just one gene.

3.6. Conclusion

Our experiment was successful in the phylogenetic analysis of mammalian species. The results demonstrated that the mitochondrial *16S rRNA*, *12S* rRNA, *COX3* and *ATP6* genes have the discriminatory power to identify mammalian species. However, the *16S* rRNA gene performed better than all the other 3 genes in terms of accurately discriminating interspecies and intraspecies and maintaining high bootstrap values throughout. The *12S* rRNA and *COX3* genes failed to group all bat species together in individual phylogenetic trees. Furthermore, the *COX3* gene grouped a seal species with a water buffalo. The *ATP6* gene could not group all shrew species together. All these discrepancies were resolved when the *12S* rRNA, *COX3* and *ATP6* genes were combined with the *16S* rRNA gene. Therefore, we confidently recommend the use of the *16S* rRNA gene as a universal marker in a meat species identification experiment. We recommend the *12S* rRNA, *COX3* and *ATP6* genes to be more suitable for use as species-specific markers. These genes are more suitable as species-specific markers, as they were not able to delineate some species correctly. Whereas the *16S* rRNA mitochondrial gene managed to group all related species and separate unrelated species correctly.

As Published in *Animals* **2022,** 12, 416.
https://www.mdpi.com/2076-2615/12/4/416

CHAPTER FOUR

4. A 16S Next Generation Sequencing Based Molecular and Bioinformatics Pipeline to Identify Processed Meat Products Contamination and Mislabeling

4.1. Abstract

Processed meat is a target in meat adulteration for economic gain. This study demonstrates a molecular and bioinformatics diagnostic pipeline that utilizes the mitochondrial 16S ribosomal RNA (rRNA) barcoding gene, to determine processed meat product mislabelling through Next Generation Sequencing. The pipeline developed utilized universal primers. A universal method does not require prior knowledge of the investigated species, allowing for identification of unknown species. Nine pure meat samples were collected and artificially mixed at different ratios, to verify the specificity and sensitivity of the pipeline. Nine pure meat samples were collected and artificially mixed at different ratios to verify the specificity and sensitivity of the pipeline. Processed meat products (n = 155), namely, minced meat (49), biltong (28), burger patties (35), and sausages (43), were collected across South Africa. Sequencing was performed using the Illumina MiSeq sequencing platform. Each sample had paired-end reads with a length of ± 300 bp. Quality control and filtering was performed using BBDuk (version 37.90a https://jgi.doe.gov/data-and-tools/bbtools/bb-tools-user-guide/bbduk-guide/). Each sample had an average of 134 000 reads aligned to the mitochondrial genomes using BBMap v37.90. The average fold coverage across a mitochondrial reference genome was used for further analysis. All species in the artificial DNA mixtures were detected. Processed meat samples had reads that mapped to the *Bos* (90% and above) genus, with traces of reads mapping to *Sus* and *Ovis* (2–5%) genus. Sausage samples showed the highest level of contamination with 46% of the samples having mixtures of beef, pork, or mutton in one sample. The pipeline demonstrated its specificity through identifying species with percentages as low as 0,1% across all samples This method can be used to authenticate meat products, investigate, and manage any form of mislabelling in the system.

Keywords: Processed meat; Adulteration; Mitochondrial *16S* rRNA gene; Next Generation Sequencing

4.2. Introduction

Meat species identification is a subject that has received special attention world-wide, mainly due to the increased incidence of fraudulent practices that have been reported (Cawthorn, Steinman and Hoffman, 2013; D'Amato *et al.*, 2013b; Walker, Burns and Burns, 2013; Cho, Dong and Cho, 2014). These reports have led to consumers demanding the accurate identification and labelling of meat products (Song *et al.*, 2019). Incidents of meat species substitution include contamination of a product with a cheaper-priced protein. For instance, replacing Grade A beef with rejected horse meat, replacing mutton with a lower grade of beef, or replacing mutton with pork (Walker, Burns and Burns, 2013; Cho, Dong and Cho, 2014). The addition of plant proteins, such as grain by-products or soyabeans, to meat products like beef patties and sausages has also been reported (Flores-Munguia, Bermudez-Almada and Vázquez-Moreno, 2000). Meat species substitution is common in processed meat products that are difficult to accurately identify morphologically once processed into value-added products. For in-stance, pork is intentionally added to beef products to reduce production costs (Doosti, Ghasemi Dehkordi and Rahimi, 2014). Once the two different meats are minced or ground it is difficult to identify them using the naked eye. Meat adulteration predominately occurs in ground meat products (Murugaiah *et al.*, 2009).

Consumers have a right to purchase meat products that are correctly labelled for reasons of health (allergies), religious belief, individual preference and ethics (Bottero and Dalmasso, 2011). Therefore, there is need for the accurate identification of meat species in processed meat products. Food labelling regulations require that ingredients in food products are accurately declared to consumers (Cawthorn, Steinman and Hoffman, 2013). The governing organizations in South Africa have issued new legislation to encourage clarity and the accurate explanation of food products, in response to consumer demand. These are the controls linked to Advertising and Labelling of Foodstuffs (R. 146/2010) comprised of a compulsory ingredient list on food labels (DoH, 2010) and the Consumer Protection Act (R. 147/2009), which prevents unfair

marketing and business practices and provides an improved standard of consumer information (DTI, 2009). The food regulation standard in Europe requires that meat products should be accurately labelled with information that includes the composition and percentage of ingredients included in the products (Commission, 2003).

Techniques for meat species identification need to be reliable, rapid, and cheap enough for routine applications. In the past, the identification of meat species has been conducted using protein-based methods, which entail different immunological, chromatographic, and electrophoretic methods (Koh *et al.*, 2011; Cho, Dong and Cho, 2014). However, the disadvantages of protein-based methods are that proteins are denatured by heat, salt and pressure, making protein-based methods unsuitable for the identification of species in seasoned, cured or dried meat, and meat patties (Alikord *et al.*, 2018; Xing *et al.*, 2019). Protein-based methods are also inaccurate in identifying species that have a close phylogenetic relationship due to cross-reactivity, for instance in poultry species (Koh *et al.*, 2011; Kesmen *et al.*, 2012; Cawthorn, Steinman and Hoffman, 2013; Tillmar *et al.*, 2013) deoxyribonucleic acid- (DNA) based methods are now preferred in place of protein-based methods, because DNA is more stable during heating and less likely to be disturbed during food processing (Koh *et al.*, 2011).

Mitochondrial DNA (mtDNA) is commonly used in the identification of meat species, as it can be extracted, undamaged, from cooked and processed meat products (Dai *et al.*, 2015; Kumar *et al.*, 2015). Mitochondrial DNA is commonly used in species identification, since mtDNA occurs in multiple copies (an average of 1 000 per cell), can withstand heat, salt, and pressure and can discriminate closely related species due to its high rate of evolution (Chen, Liu and Yao, 2010). Several mtDNA genes have been used in meat species identification, such as *cytochrome b*, *cytochrome c oxidase subunit 1* (*COI*), *NADH dehydrogenase subunit 2* and *5* (*ND2* and *ND5*), *ATPase 6* and *8*, mitochondrial *16S*, and *12S* ribosomal RNA (rRNA) genes (Di Pinto *et al.*, 2015). Mitochondrial DNA-based methods that have been used for species identification in the past are polymerase chain reaction (PCR) (Ballin, 2010), PCR-restriction fragment length polymorphism (PCR-RFLP) (Dalmasso *et al.*, 2004), species-specific PCR

(Kane and Hellberg, 2016; Drummond, Álvarez and Mullen, 2019), DNA hybridization (Tillmar *et al.*, 2013), multiplex PCR (Horreo *et al.*, 2013), and real-time PCR (Tillmar *et al.*, 2013; Wu, Zhong and Yang, 2018). The limitations of previous DNA-based methods are that species-specific PCR methods were used and, therefore, these studies targeted specific species as opposed to having a universal method that targets any species. Species-specific methods are advantageous when species in a sample are known, however, a universal method is a better approach for investigating multiple and unsuspected contaminations in meat products. Recently, there have been advances in DNA-based methods, specifically in DNA sequencing technologies. Next-Generation Sequencing (NGS) is a method that can generate sequenced data from degraded DNA, and one that can produce large amounts of sequenced data at a low cost and with minimum errors (Wadapurkar and Vyas, 2018). Furthermore, essentially no prior information of species is needed, making NGS technology a non-species-specific method (Wadapurkar and Vyas, 2018). With the NGS method, mtDNA barcoding genes can be sequenced using universal primers, without knowledge of which species are present and without targeting specific species. This, then, enables the identification of every species present in a sample, as against merely the suspected/hypothesised ones (Tillmar *et al.*, 2013). However, the database used in the identification of species needs to be comprehensive, such that it contains many species to enable accurate identification.

Meat species contamination has been reported in the South African meat sector (Cawthorn, Steinman and Hoffman, 2013; D'Amato *et al.*, 2013b; Tembe, Mukaratirwa and Zishiri, 2018). A routine universal diagnostic method that can be used by laboratories needs to be developed, as the methods used to date, in South Africa, have been species-specific. Meat producers can use this method to authenticate their products and gain consumers' confidence in the products they will purchase. Mitochondrion carries extra-chromosomal genetic material and contains high copy numbers as compared to single copy nuclear genes. Therefore, mitochondrial DNA is the preferred analytical tool in forensic, molecular, and zoological experiments. The objective of this study was, therefore, to develop a universal and robust diagnostic molecular and bioinformatics pipeline that can utilize the mitochondrial 16S ribosomal RNA (rRNA) barcoding gene to identify processed meat product mislabelling/contamination using NGS. Universal mitochondrial *16S* rRNA primers will be used in this study to identify different species, including those in mixed samples. Meat suppliers can possibly implement the current

method to authenticate their products, and the food industry may also use this method to reveal any form of mislabelling that may be present.

4.3. Materials and methods

4.3.1. Collection of pure meat samples to confirm the reliability of the species identification pipeline

Pure meat samples were collected as controls to verify the use of the 16S rRNA gene in the molecular and bioinformatics pipeline developed. This was done primarily to test the specificity, sensitivity, and ability of the pipeline to be used as a diagnostic method. Nine unprocessed pure meat samples from nine different species were collected from a local butchery in Pretoria (South Africa), transported in an icebox and stored at −20 °C. The pure meat samples were placed in separate plastics upon collection, transportation, and storage, to avoid any unintentional cross-contamination. These species were pig, cattle, sheep, chicken, turkey, goat, ostrich, duck, and kangaroo.

4.3.2. The collection of processed meat samples for species identification

Meat products were randomly collected from processing plants and retail outlets in the Gauteng and Free State provinces in South Africa for the species identification test. A total of 155 samples from the meat value chain were collected and analysed. Four different categories of processed meat products were collected for analysis, namely, minced meat (49), burger patties (35), biltong (28), and raw sausages (43). Some samples included information on which species they were produced from, and, of these, 22 were beef mince, 20 were beef patties, 17 were beef biltong, and 21 were beef sausages. All samples were transported in an ice box and stored at −20 °C.

4.3.3. DNA extraction

Genomic DNA from the pure meat samples used for the verification test was extracted manually from 40 mg of each meat sample. A Macherey–Nagel NucleoMag Tissue kit for

DNA purification from cells and tissue (Macherey–Nagel, Germany) was used for DNA extraction according to the Genomic DNA from Tissue user manual. The pure DNA was stored at −20 °C while awaiting further analysis. Thirteen two-species DNA mixtures of known species and composition were artificially mixed (Table 4.1). Two ratios were used for the DNA mixtures, 1:1 (50%:50%) and 0.9:0.1 (90%:10%). The artificially mixed samples were used to test the specificity of the 16S universal primers, by confirming the origin of the known species in the 1:1 ratio mixture. The ratio of 0.9:0.1 was used to test the sensitivity of the pipeline, using, as a metric thereof, the smallest amount of DNA that the pipeline could correctly identify at an affordable cost. Each mixed ration had three replicates. The concentration of the DNA of each pure meat sample used was normalized to two different concentrations, 25 ng/ul and 5ng/ul prior to running of the PCR. The differences in DNA concentration are explained in the PCR step. Genomic DNA for the species identification of the samples collected from the meat value chain was extracted from 300 mg of each processed meat sample, using a Hamilton Microlab Star automated liquid handler (Hamilton Inc.). A Macherey–Nagel NucleoMag Tissue kit for DNA purification from cells and tissue (Macherey–Nagel, Ger-many) was used for DNA extraction according to the Genomic DNA from Tissue user manual. The DNA concentration of the meat value chain samples was between 28–467 ng/ul prior to PCR testing. The quantification of DNA for all samples was checked using the Qubit® fluorescent dye method, and gel electrophoresis was used to assess the quality of the starting material. A ratio of A260/A280 was used to access the purity of all extracted DNA.

Table 4. 1: Ratios and species of artificial DNA mixtures

Mixture Ratio	Species 1: Species 2
1:1	pork ‖ beef, mutton ‖ chevon, chicken ‖ turkey, ostrich ‖ duck and beef ‖ kangaroo
9:1	pork ‖ beef, pork ‖ beef, mutton ‖ chevon, chevon ‖ mutton, chicken ‖ duck, duck ‖ chicken, ostrich ‖ duck and duck ‖ ostrich

4.3.4. PCR amplification of the mitochondrial *16S* rRNA gene

Polymerase chain reaction (PCR) for the mitochondrial 16S rRNA gene was per-formed using universal mammalian primers designed by (Tillmar *et al.*, 2013) and tailed with Nextera adapters (Table 4.2). Thermal cycling was performed in a Labnet MultigeneTM Gradient Thermal Cycler (Woodridge, USA) at a final volume of 50 ul. All the ratio mixtures had a normalized DNA concentration of 50 ng/ul in the PCR run. This was done so that if one ratio mixture did not identify the contained species, it would not be due to the DNA having a lower concentration. The PCR for the 50%:50% (1:1) ratio mixture contained 25 ul of 2X Hot start PCR mastermix, 5 ul of each forward and reverse primer (1 mM final concentration), 13 ul RNase-free water and 1 ul of 25 ng/ul DNA template of each species. This brought the total amount of DNA template for the 50%:50% ratio mixture to 2 ul and the concentration to 50 ng/ul. The PCR for the 90%:10% (0.9:0.1) ratio mixture contained 25 ul Kapa HiFi Hotstart Readymix (Roche, USA), 5 ul of each forward and reverse primer (1 mM final concentration), 12.2 ul RNase-free water, 1.8 ul of the 25 ng/ul DNA template for the species with a ratio of 90% and 1 ul of 5 ng/ul DNA template for the species with a ratio of 10%. This brought the total amount of DNA template for the 90%:10% ratio mixture to 2.8 ul and the concentration to 50 ng/ul. Just like the DNA extraction process, sterile tips and PCR tubes were not reused and the pipettes and work-bench area were disinfected with 70% ethanol between analyses.

The PCR for the samples from the meat value chain contained 25 ul of Kapa HiFi Hotstart Readymix (Roche, USA), 5 ul of each forward and reverse primer (1 mM final concentration), 13 ul RNase-free water, and 2 ul of DNA template. The PCR conditions for all samples were as follows: denaturation at 95 °C for 3 min, followed by 30 cycles of 90 °C for 20 sec, 65 °C for 30 sec, 72 °C for 30 sec, and finalization at 72 °C for 5 min. The PCR products for the mitochondrial 16S rRNA gene were 186 bp in length. The PCR products were viewed in 2% agarose gels in 1 X tris-acetate-EDTA (TAE) buffer at 90V for 45 min. The amplified products were visualized under ultra-violet light in a trans-illuminator. Purification of PCR products was performed using a Qiagen MiniElute® PCR purification kit (Qiagen, Germany) according to the manufacturer's protocol. Quantification of the purified samples was done using the Qubit® fluorescent dye method. A ratio of A260/A280 was used to access the purity of all extracted DNA. The purified products were stored at 4 °C prior to sequencing.

Table 4. 2: The oligodeoxynucleotide sequences of the universal primers for 16S rRNA gene amplification designed by Tillmar *et al.*, 2013 (the letters in small case are Nextera adapter tails).

16S Forward	5′tcgtcggcagcgtcagatgtgtataagagacagGACGAGAAGACCCTATTGGAGC 3′
16S Reverse	5′gtctcgtgggctcggagatgtgtataagagacagTCCGAGGTCRCCCCAACC 3′

4.3.5. Library preparation and Illumina MiSeq sequencing

Prior to sequencing, library preparation was performed using the 16S Me-ta-genomics Sequencing Library Preparation kit, according to the manufacturer's protocol (Illumina, Inc). Quality control of the sample library and quantification of the DNA library templates was performed. Quantification of DNA was done using Qubit® fluorescent dye method. The library size distribution was checked using a High Sensitivity DNA chip. Thereafter, the indexed libraries were normalized, pooled, and loaded onto an Illumina MiSeq reagent cartridge using MiSeq reagent kit v3 and 600 cycles. The paired end 2 × 300 bp sequencing was run on an Illumina MiSeq sequencer at 0.2 X coverage at the Biotechnology Platform, Agricultural Research Council, Onderstepoort, South Africa. The DNA from pure meat samples were each sequenced individually prior to artificially mixing the DNA, to confirm the origin of each meat type.

4.3.6. Bioinformatics and data analyses

Prior to species identification, quality control, adapter removal, decontamination, and error correction of the raw sequence data was done using BBDuk (version 37.90; https://jgi.doe.gov/data-and-tools/bbtools/bb-tools-user-guide/bbduk-guide/). All available mitochondrial genomes (10,788) were downloaded from the NCBI RefSeq database (https://ftp.ncbi.nlm.nih.gov/refseq/release/mitochondrion) accessed on 20 July 2020 [30]. Filtered reads were aligned to the complete mitochondrial genomes using BBMap v37.90 [31], (https://www.osti.gov/biblio/1241166-bbmap-fast-accurate-splice-aware-aligner) for species

identification and the average fold coverage across a mitochondrial reference genome was used for further analysis. The average fold of each sample was exported into Microsoft Excel to calculate the percentage of the number of reads that aligned to a reference, over how much of the reference was covered in a sample. The percentage average fold was used to determine samples that were contaminated and uncontaminated.

Statistical analysis was initially performed on the pure and artificially mixed samples using R software, to determine whether our pipeline would work in a controlled environment. The percentage deviation from the expected composition within the pure (100%) and artificially mixed (50%:50% and 90%:10%) samples was also determined, and these values were used to calculate the mean (using absolute values), median, standard deviation, and variance. Bar plots, plotting the percentage composition by species pre-sent in all samples, were constructed, showing samples that were contaminated and uncontaminated. A chi-square proportion test was used to determine whether there was a significant correlation between two categorical variables, i.e., contaminated, and uncontaminated meat samples. A *p-value* was determined because of the number of contaminated versus uncontaminated. Associations between contaminated versus un-contaminated meat were considered statistically significant only for *p-values* ≤ 0.05. Cramer's V test, which measures how strongly two categorical fields are associated, was also performed. The confidence interval (CI) was set at 95% and the number of samples observed (nobs) was also determined. The workflow of the molecular and bioinformatics pipeline is shown in Figure 4.1.

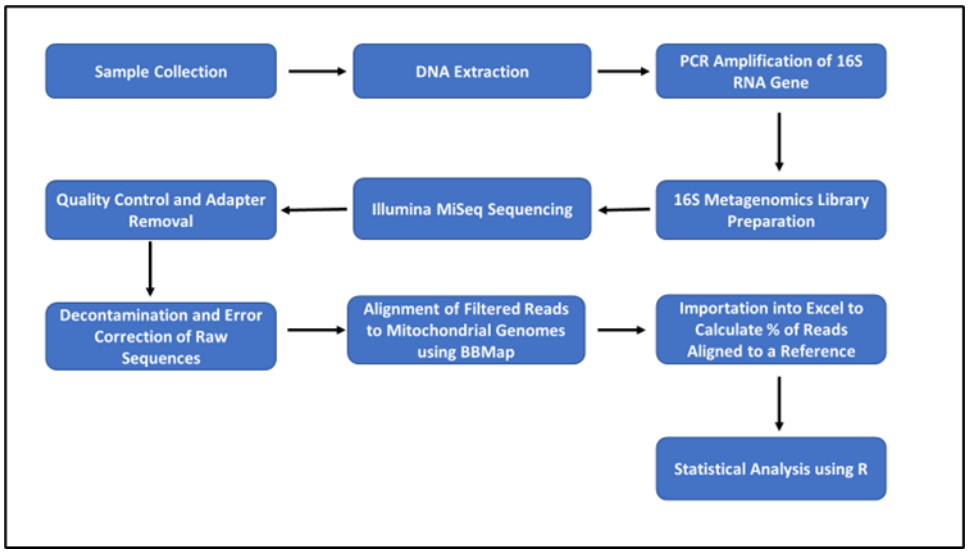

Figure 4. 1. Workflow for molecular and bioinformatics pipeline for species identification.

4.4. Results

4.4.1. Identification of meat species using pure DNA from known meat types

Paired-end reads, with a length of 300 bp, were sequenced using the Miseq sequencer. Each sample had an average of 156,863 reads, before quality filtering, and 134,230 reads after quality filtering. Nine pure meat samples with two replicates each were analysed, for a total of 18 analysed pure samples. The reads obtained mapped to the corresponding pure meat species, with a similarity of 98% and above for all meat types (Figure 2). Besides identifying the expected genera, traces of other meat species were observed (Appendix Table 4.1). The beef, mutton, and pork meat samples had reads with an average fold of 99% for the *Bos*, *Ovis*, and *Sus* genus, respectively. The chevon, chicken, and duck meat samples had reads with an average fold of 98% for the *Capra*, *Gallus*, and *Anas* genus, respectively. The turkey and kangaroo meat samples had reads with an average fold of 99% for the Meleagris, Struthio, and Macropus genus, respectively. One of the ostrich samples was contaminated with beef, as it showed a proportion with reads that mapped to the Bos genus (Figure 2). Based on these results from the controls, samples whose highest percentage average fold was less than 98% were considered contaminated.

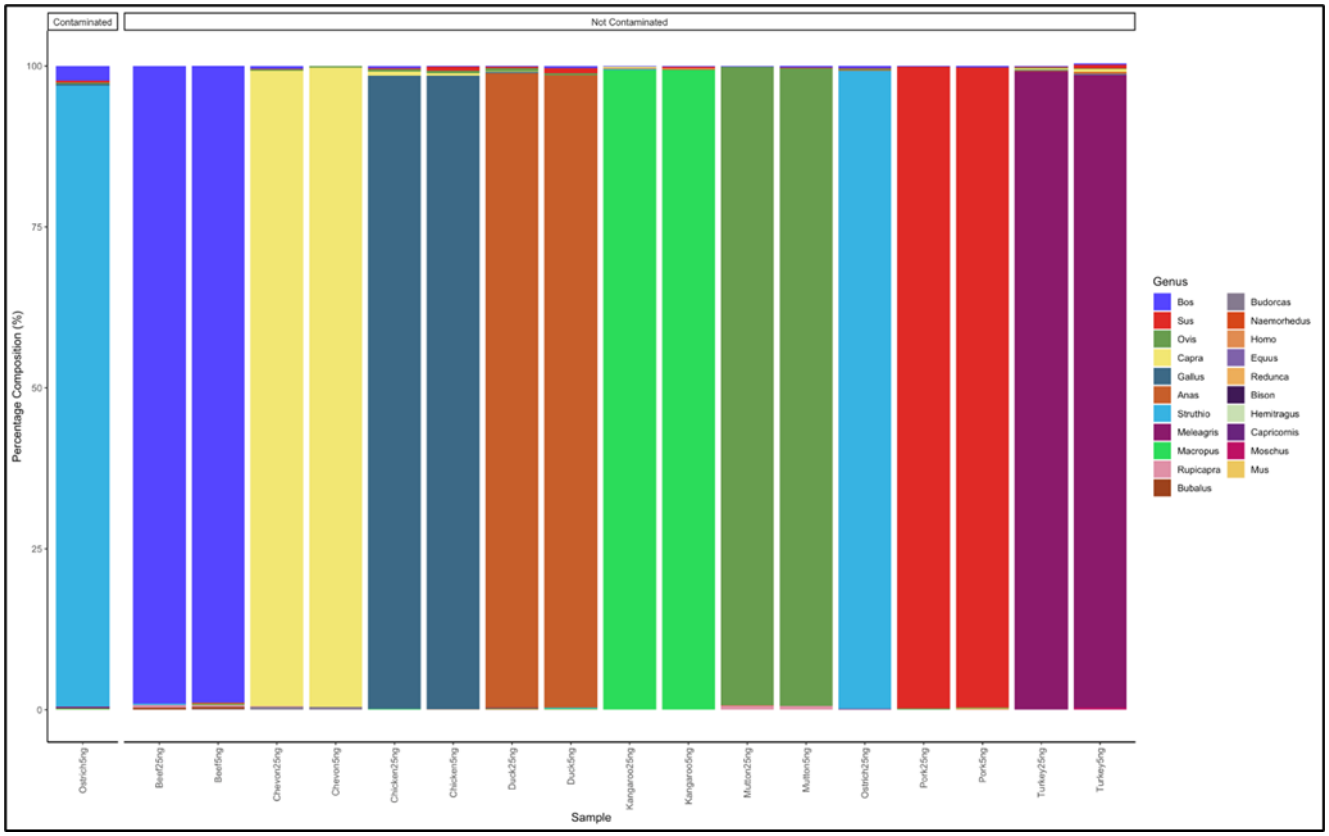

Figure 4. 2. Percentage average fold of pure DNA from known meat types.

4.4.2. Identification of meat species using pure DNA from known meat types artificially mixed at a ratio of 1:1

The pipeline identified all the meat types whose DNA were in the 1:1 ratio mixture. However, some positive and/or negative deviations were observed from the expected 50:50 percentages (Table S2). The beef (50%) and kangaroo (50%) DNA mixtures had reads with an average fold of 73% and 25% for the Bos (cattle) and Macropus (kangaroo) genus, respectively (Figure 3). The chevon (50%) and mutton (50%) DNA mixtures had reads with an average fold of 35% and 63% for the Capra (goat) and Ovis (sheep) genus, respectively. The chicken (50%) and turkey (50%) DNA mixtures had reads with an average fold of 48% and 50% for the Gallus (chicken) and Meleagris (turkey) genus, respectively.

The duck (50%) and ostrich (50%) DNA mixtures had reads with an average fold of 69% and 29% for the Anas (duck) and Struthio (ostrich) genus, respectively. The pork (50%) and beef (50%) DNA mixtures had reads with an average fold of 51% and 48% for the Sus and Bos

genus, respectively. The greatest deviations for the 1:1 (50% of each species) ratio were between beef (Bos): kangaroo (Macropus) where beef meat had an overestimation of 23% and kangaroo meat had an underestimation of 24%. The lowest deviations were between pork (Sus): beef (Bos), where pork had an overestimation of 1.5% and beef had an underestimation of 1,9% (Appendix Table 4.2).

4.4.3. Identification of pure DNA from known meat types artificially mixed at a ratio of 9:1

The pipeline showed that all the meat types whose DNAs were in the 9:1 ratio mixture were identified (Figure 4.3). Some positive and/or negative deviations were observed from the expected 90:10 percentage ratio (Table S3). The beef (90%) and pork (10%) DNA mixtures had reads with an average fold of 92% and 6% for the *Bos* (cattle) and *Sus* (pig) genus, respectively (Figure 3). The chicken (90%) and duck (10%) DNA mixtures, the reads had an average fold of 82% and 17% for the *Gallus* (chicken) and *Anas* (duck) genus, respectively. The duck (90%) and chicken (10%) DNA mixtures had reads with an average fold of 96% and 3% for the *Anas* (duck) and *Gallus* (chicken) genus, respectively. The duck (90%) and ostrich (10%) DNA mixtures had reads with an average fold of 86% and 13% for the *Anas* (duck) and *Struthio* (ostrich) genus, respectively. The goat (90%) and sheep (10%) DNA mixtures had reads with an average fold of 81% and 17% for the *Capra* (goat) and *Ovis* (sheep) genus, respectively. The ostrich (90%) and duck (10%) DNA mixtures had reads with an average fold of 81% and 18% for the *Struthio* (ostrich) and *Anas* (duck) genus, respectively. The pork (90%) and beef (10%) DNA mixtures had reads with an average fold of 91% and 8% for the *Sus* (pig) and *Bos* (cattle) genus, respectively. The mutton (90%) and chevon (10%) DNA mixtures had reads with an average fold of 92% and 6% for the *Ovis* (sheep) and *Capra* (goat) genus, respectively. The lowest deviations for the 9:1 (90% and 10%) of each species ratio were between pork (*Sus*): beef (*Bos*). Pork meat had an overestimation of 1.2% and beef meat had an underestimation of 1.5%. The greatest deviations were between ostrich (*Struthio*): duck (*Anas*). Ostrich meat had an underestimation of 8.8% and duck meat had an overestimation of 8.5% (Appendix Table 4.3).

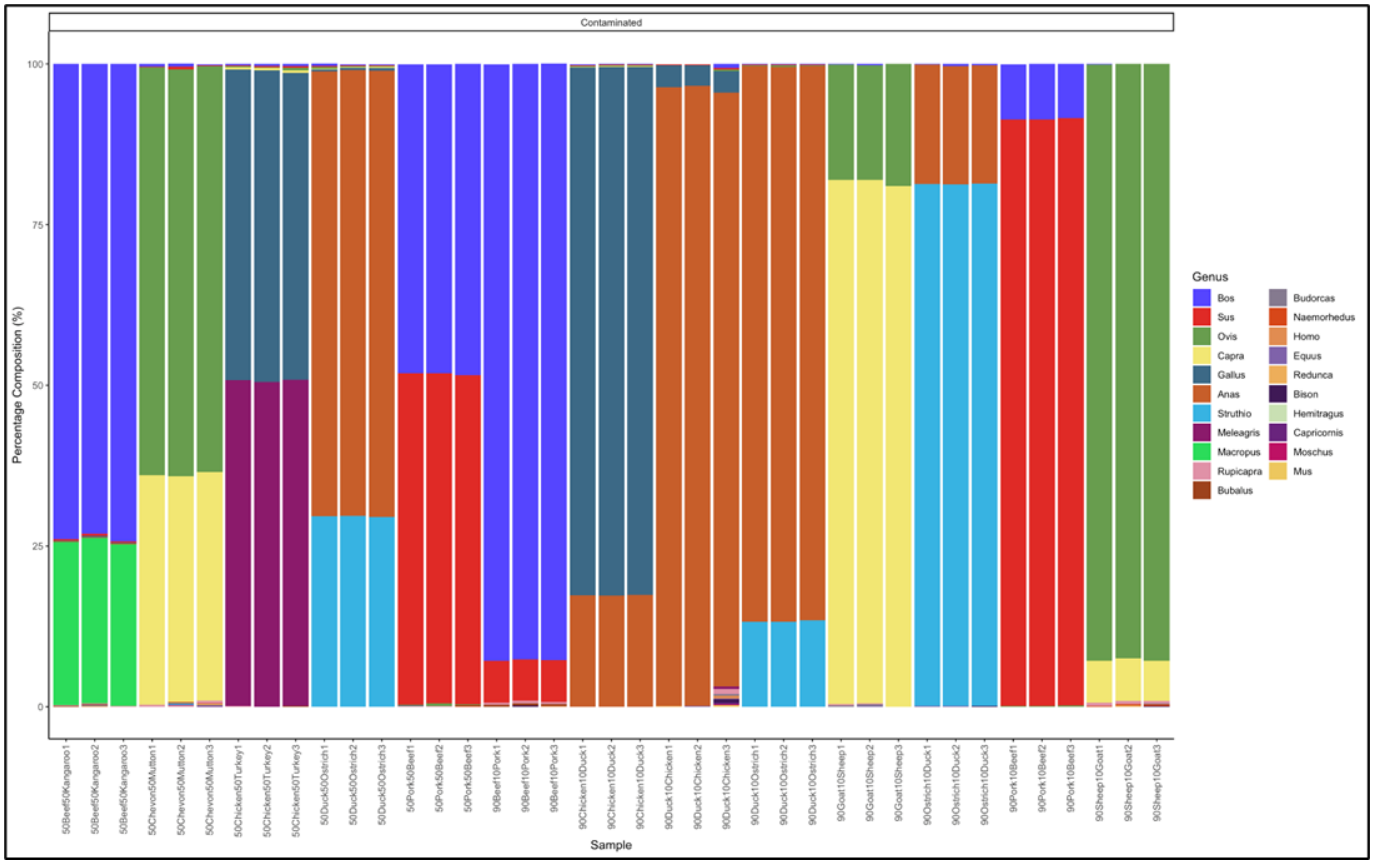

Figure 4. 3. Percentage of average fold of pure DNA from two known meat types artificially mixed at a ratio of 1:1 (50%:50%) and 9:1 (90%:10%).

4.4.4. Identification of species and species contamination of processed meat for which the meat types were not indicated on product labels

The percentage deviation (expected–observed) was determined and used to calculate the descriptive statistics (mean, median, standard deviation, variance, minimum value, and maximum value) of the pure and artificially mixed samples (Table 4.3). Measures of central tendency were described by the mean and median values, while measures of variability were described by standard deviation, variance, and minimum and maximum values. The range within the artificially mixed samples was 23.59% and 3.19% for the pure samples. The higher spread of values in the artificially mixed samples lead to a higher mean value, and that, in turn, resulted in a higher standard deviation and variance that was further from zero. In contrast, the smaller spread of values within the pure samples lead to a lower mean value that resulted in a lower standard and variance that was closer to zero**.**

Table 4. 3: Descriptive statistics of artificially mixed and pure samples.

Artificially Mixed % Deviation					
Min.	**Max.**	**Median**	**Mean**	**Standard Deviation**	**Variance**
1.0	24.59	9.08	13.76	6.98	48.66
Pure % Deviation					
Min.	**Max.**	**Median**	**Mean**	**Standard Deviation**	**Variance**
0.30	3.49	1.184	1.56	0.716	0.513

4.4.4.1. Biltong

The pipeline demonstrated that all the 11 biltong samples that had not specified from which species they were from were uncontaminated and were essentially beef (*Bos* genus) (Figure 4.4). The biltong samples, however, contained trace contaminants of other species (Appendix Table 4.4).

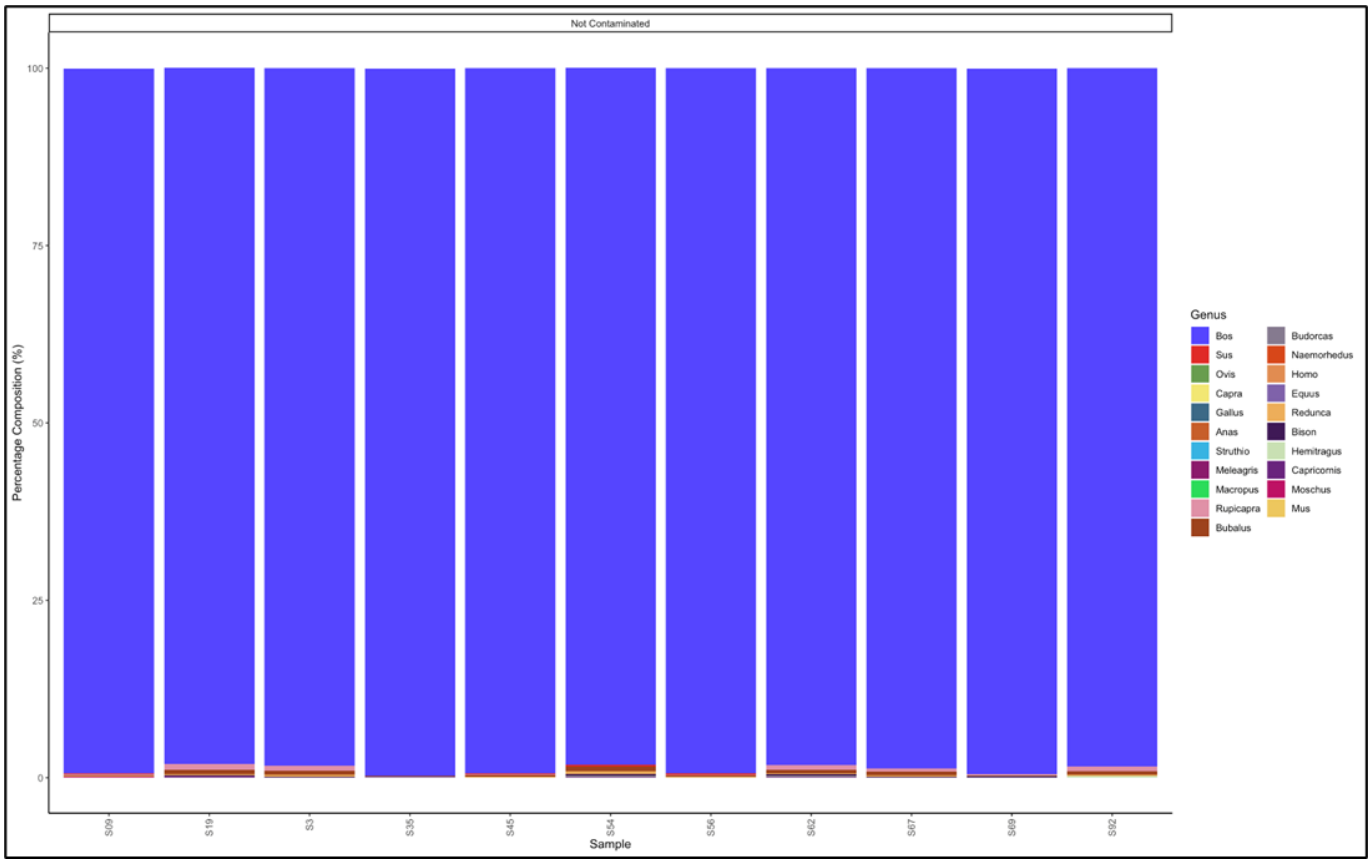

Figure 4. 4. Percentage composition of biltong samples not specified which species they are from.

The pipeline also demonstrated that 14 out of the 17 samples labelled as beef biltong were uncontaminated and were also essentially from beef (*Bos* genus) (Figure 4.5). The contaminated samples, descending from most contaminated, were predominantly composed as follows: Sample 16: *Bos* (cattle) (57.5) and *Sus* (pig) (37.4); Sample S: *Bos* (cattle) (93.6) and *Ovis* (sheep) (4.9); and Sample 117: *Bos* (cattle) (97.9) and *Bubalus* (Buffalo) (0.7). The major contaminants of the labelled beef biltong products were pork (*Sus*) and mutton (*Ovis*). Sample 16 contained beef (*Bos*) (57.5) but was contaminated with pork (*Sus*) (37.4) and mutton (*Ovis*) (4.6) (Appendix Table 4.5).

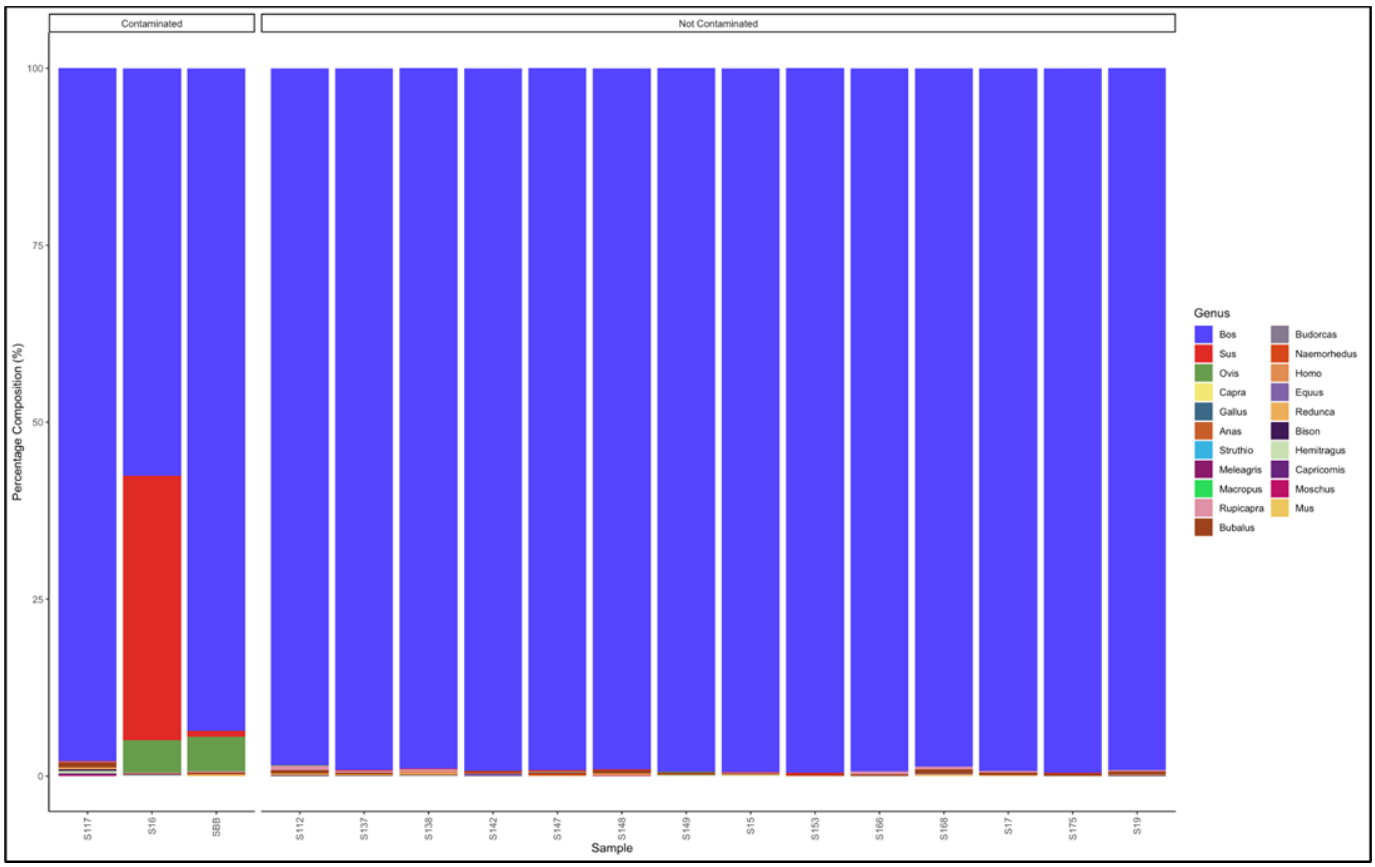

Figure 4. 5. Percentage composition of samples labelled as beef biltong.

4.4.4.2. Mince

The pipeline demonstrated that 23 out of 27 mince samples that had not specified from which species they originated were uncontaminated and were essentially from beef (*Bos* genus) (Figure 4.6). The five contaminated samples, descending from the most contaminated were predominantly composed as follows: Sample 65: *Bos* (cattle) (83,2) and *Ovis* (sheep) (16.4); Sample 78: *Bos* (cattle) (95,3) and *Sus* (pig) (3.8); Sample 4: *Sus* (pig) (97.7) and *Bos* (cattle) (2.2); and Sample 34: *Bos* (cattle) (97.7) and *Ovis* (sheep) (1,8). Sample 4 was evidently pork mince (predominantly *Sus* genus) contaminated with beef (*Bos* genus), while the rest were beef (*Bos* genus) mince contaminated with either pork (*Sus* genus) or mutton (*Ovis* genus) (Appendix Table 4.6).

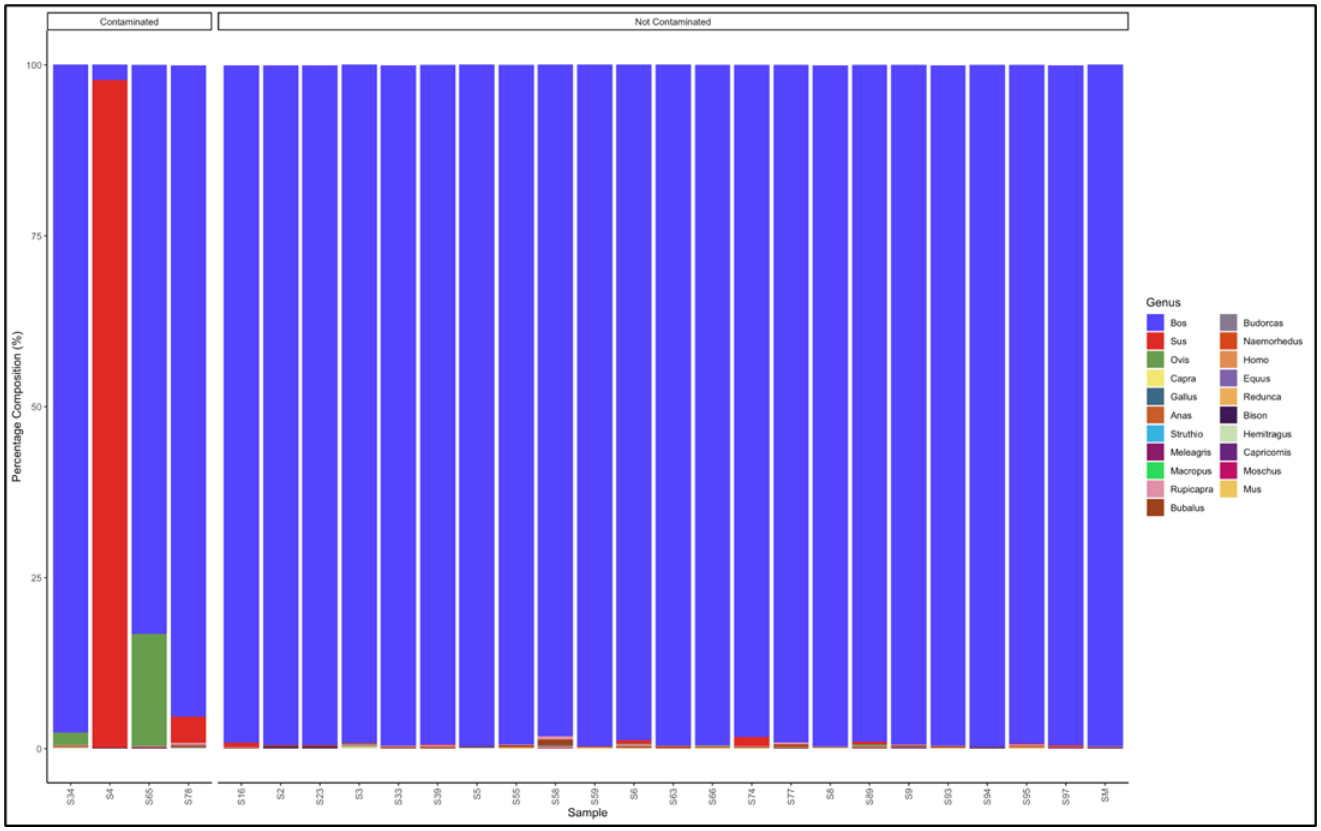

Figure 4. 6. Percentage composition of mince samples that had not specified which species they were from.

The pipeline also demonstrated that 20 out of the 22 samples labelled as beef mince were uncontaminated and were essentially from beef (*Bos* genus) (Figure 5.7). The contaminated descending from the most contaminated were predominantly composed as follows: Sample

183: *Bos* (cattle) (93.3) and *Sus* (pig) (6.1); and Sample 17: *Bos* (cattle) (97.0) and *Ovis* (sheep) (2.0). The major contaminants of the labelled beef mince products were pork (*Sus*) and mutton (*Ovis*). Two samples, S158 and S99 had traces (0.1% and 0.5%, respectively) of the *Homo* (human) genus (Appendix Table 4.7).

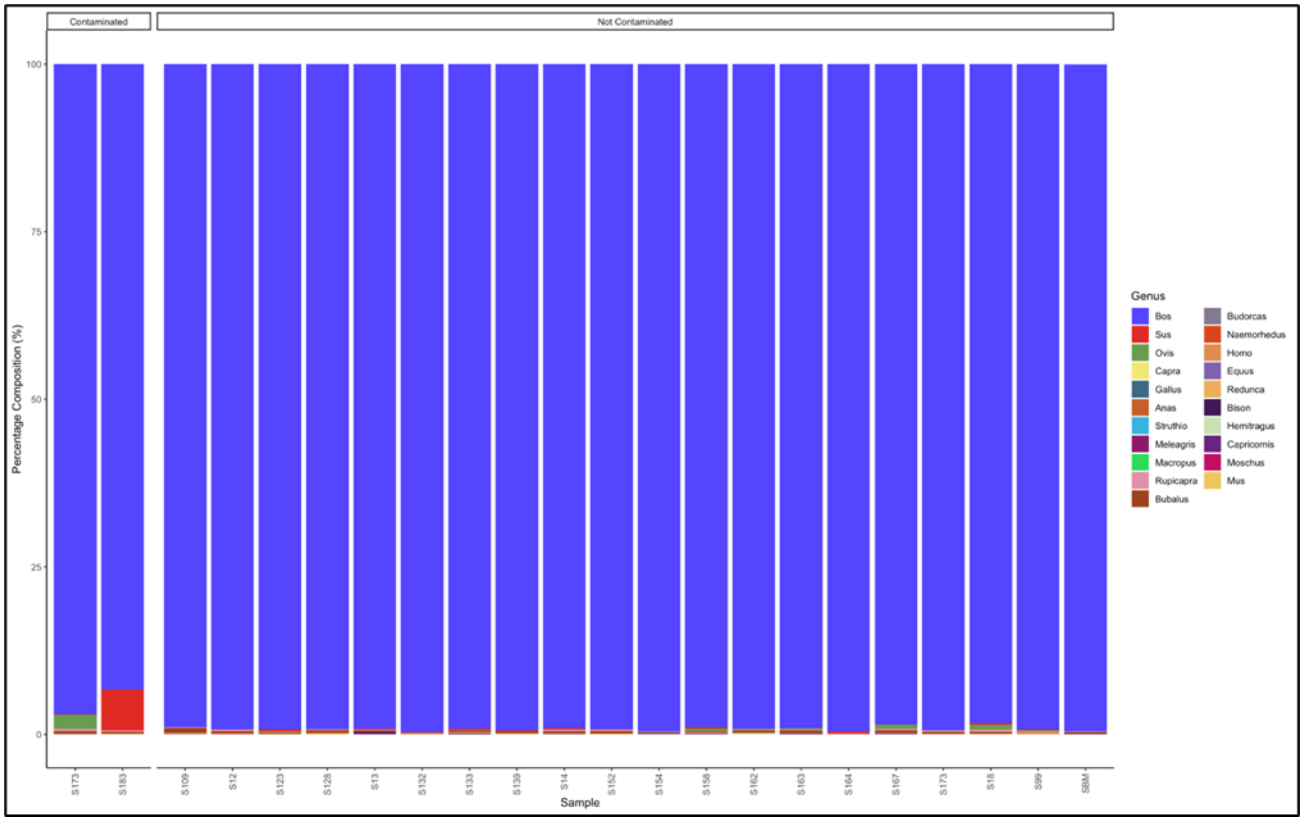

Figure 4. 7. Percentage composition of samples labeled as beef mince.

4.4.4.3. Patties

The pipeline demonstrated that 13 out of the 15 patty samples that had not specified which species they are from were uncontaminated and were essentially from beef (*Bos* genus) (Figure 4.8). The two contaminated samples, descending from the most contaminated, were predominantly composed as follows: Sample 15: *Sus* (pig) (59.6) and *Bos* (cattle) (40.0), and Sample 48: *Bos* (cattle) (91.8) and *Sus* (pig) (7.7). Sample 15 was a pork patty (*Sus* genus) contaminated with beef (B*os* genus), while sample 48 was a beef (*Bos* genus) patty contaminated with pork (*Sus* genus) (Appendix Table 4.8).

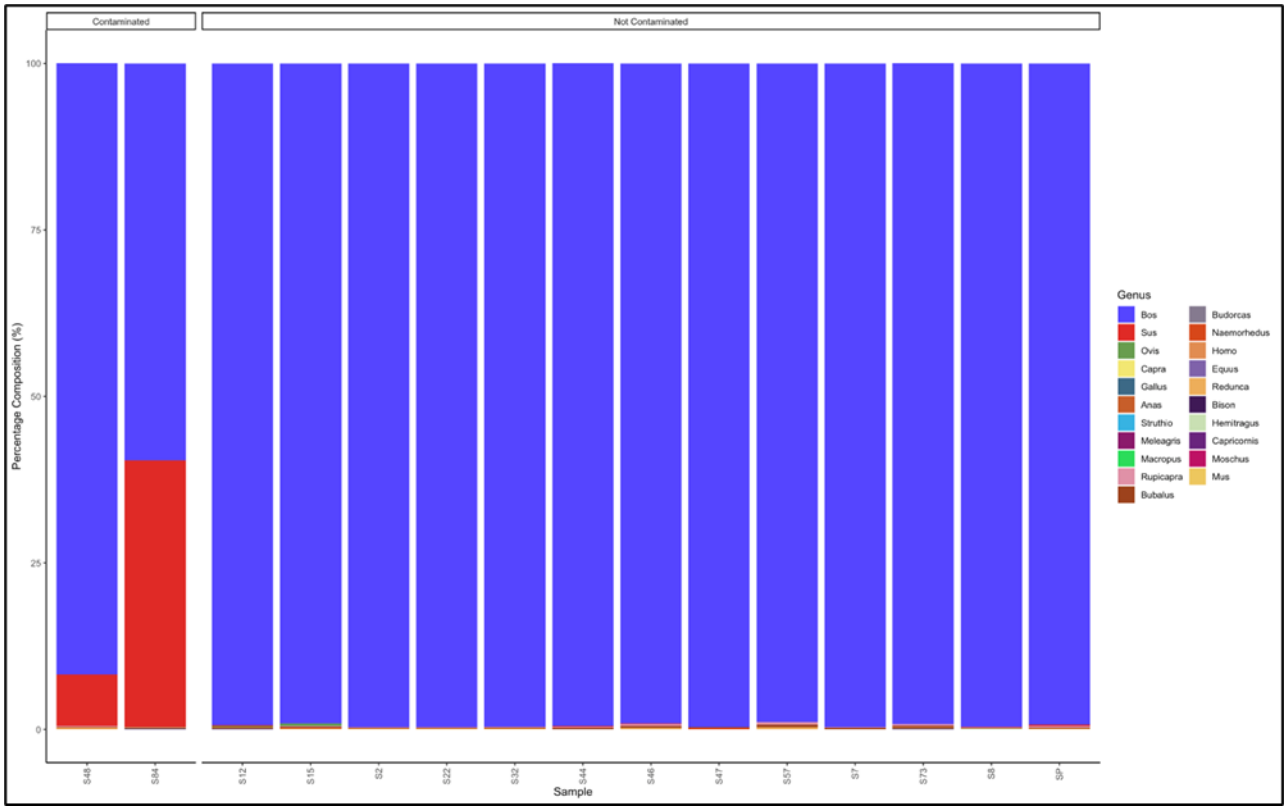

Figure 4. 8. Percentage composition of patty samples not specified which species they are from.

The pipeline also demonstrated that 13 out of 18 samples labelled as beef patty were uncontaminated and were essentially from beef (*Bos* genus) (Figure 4.9). The contaminated samples descending from the most contaminated were predominantly composed as follows: Sample 122: *Bos* (cattle) (64.5), *Ovis* (sheep) (34.3); Sample 112: *Bos* (cattle) (81.7) and *Ovis* (sheep) (17.7); Sample 179: *Bos* (cattle) (93.4) and *Ovis* (sheep) (6.2); Sample 113: *Bos* (cattle) (93.4) and *Ovis* (sheep) (5.6); and Sample 136: *Bos* (cattle) (94.9) and *Sus* (pig) (4.4). The major contaminants of the labelled beef patty products were mutton (*Ovis*) and pork (*Sus*). Sample 122 contained beef (*Bos*) (64.5) but was contaminated with mutton (*Ovis*) (34.3) (Appendix Table 4.9).

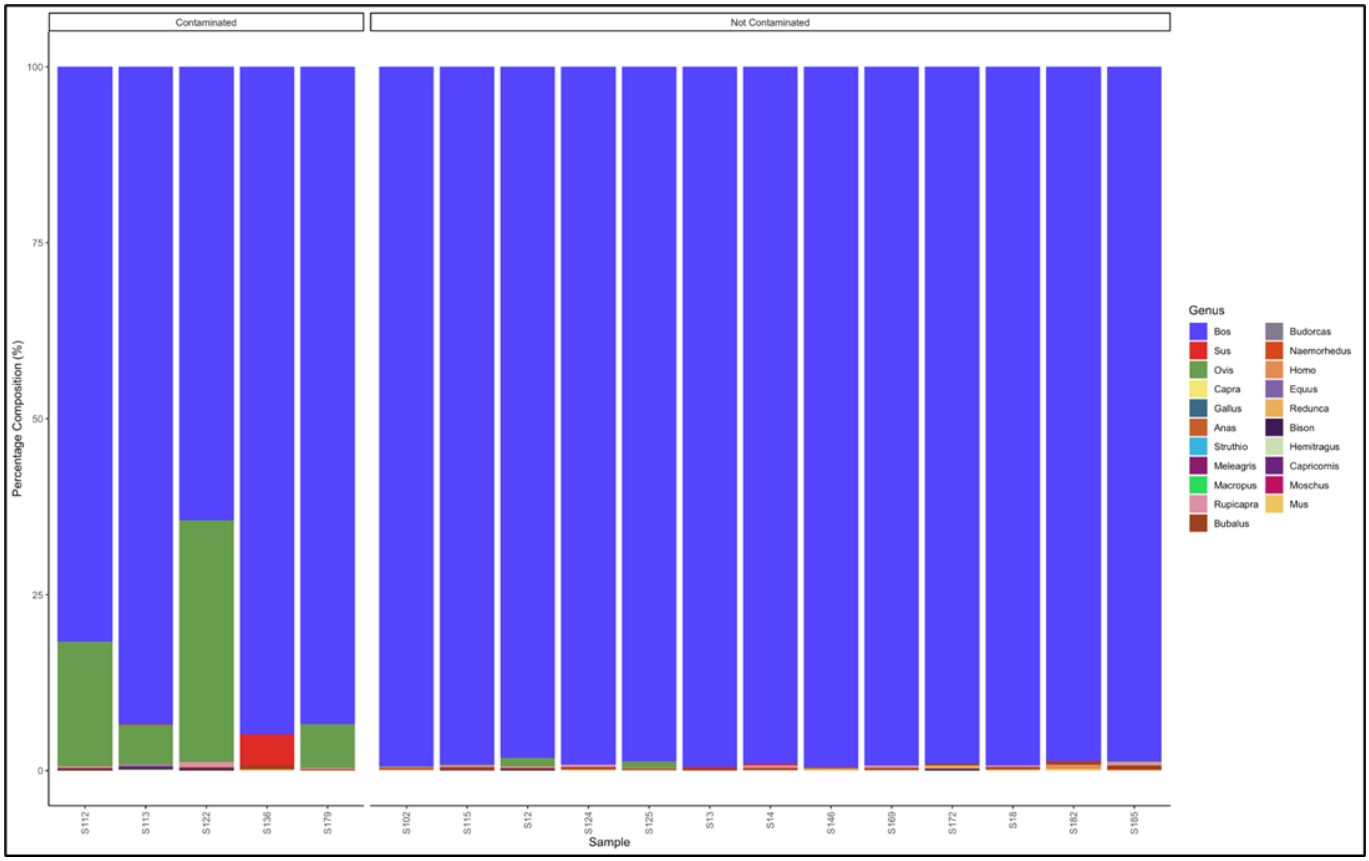

Figure 4. 9. Percentage composition of samples that were labelled as beef patty.

4.4.4.4. Sausages

The pipeline demonstrated that 14 out of the 21 sausage samples that had not specified which species they were from were uncontaminated and were essentially from beef (*Bos* genus) (Figure 4.10). The seven contaminated samples descending from the most contaminated were predominantly composed as follows: Sample 15: *Bos* (cattle) (37.1) *Sus* (pig) (38.5) and *Ovis* (sheep) (23,7); Sample 83: *Bos* (cattle) (74.7) and *Sus* (pig) (24.7); Sample 26: *Bos* (cattle) (91.0), *Ovis* (sheep) (5.2) and *Sus* (pig) (3.0); Sample 159: *Ovis* (sheep) (91.4) and *Bos* (cattle) (7.0); Sample 5: *Sus* (pig) (94,8) and *Bos* (cattle) (5,2); Sample 79: *Bos* (cattle) (95.9) and *Ovis* (sheep) (3.6); Sample 4: *Sus* (pig) (97.4) and *Bos* (cattle) (2.7); and Sample 68: *Bos* (cattle) (97.9), *Bubalus* (buffalo) (0.8) and *Rupicapra* (goat antelope) (0.6). Sample 15 was a mixed sausage made up of a substantial amount of beef (*Bos),* pork (*Sus*) and mutton (*Ovis*). Sample 159 was a mutton sausage contaminated with beef (*Bos* genus), while sample 4 and 5 were pork (*Sus* genus) sausages contaminated with beef (*Bos* genus (Appendix Table 4.10).

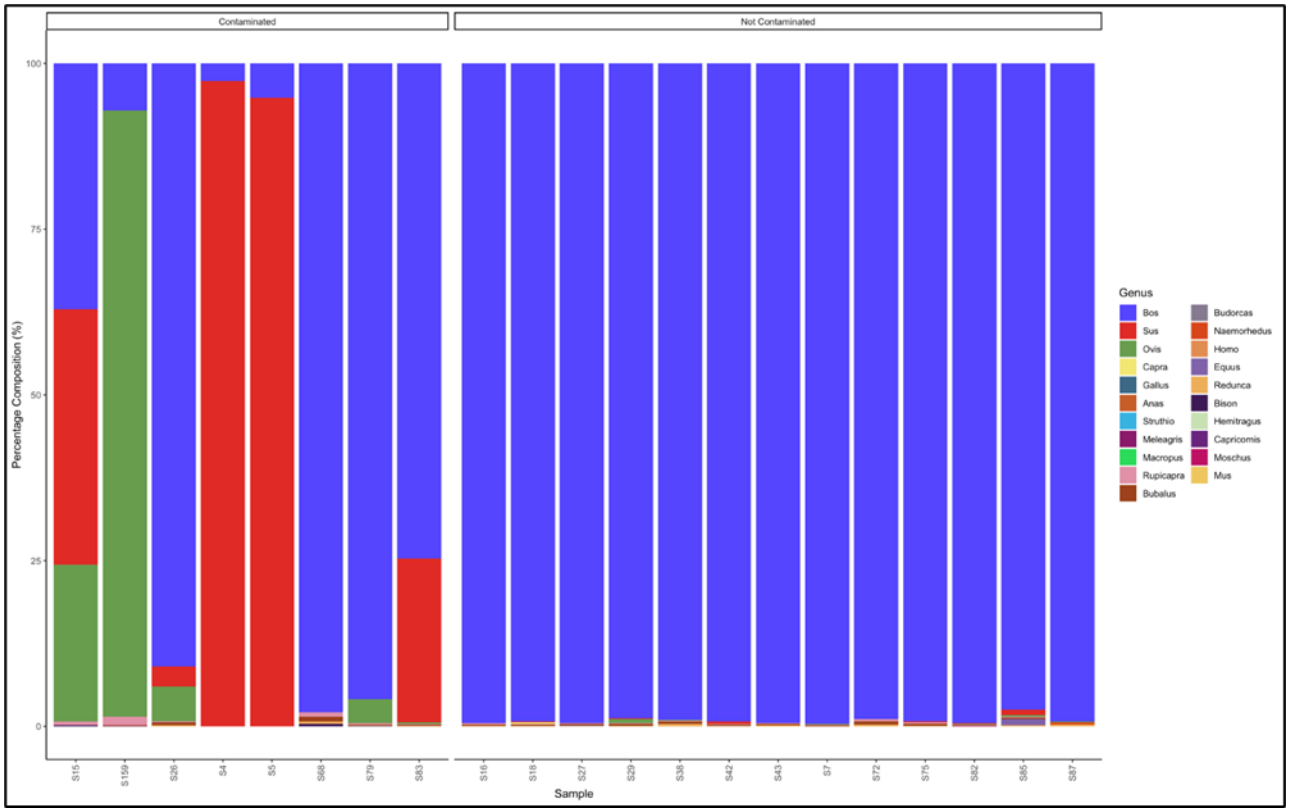

Figure 4. 10. Percentage composition for sausage samples not specified which species they are from.

The pipeline also demonstrated that 7 out of 21 (33%) samples labelled as beef sausage were uncontaminated and were essentially from beef (*Bos* genus) (Figure 4.11). The major contaminants of labelled beef sausage products were mutton (*Ovis*) and pork (*Sus*). Two samples, Sample 15: *Bos* (cattle) (18.1) and *Sus* (pig) (78.1) and Sample 135: *Bos* (cattle) (38.8) and *Sus* (pig) (60.1) can be considered as a mislabelled sample because the *Bos* (beef) genus represented a smaller percentage than the predominant *Sus* (pork) genus (Appendix Table 4.11).

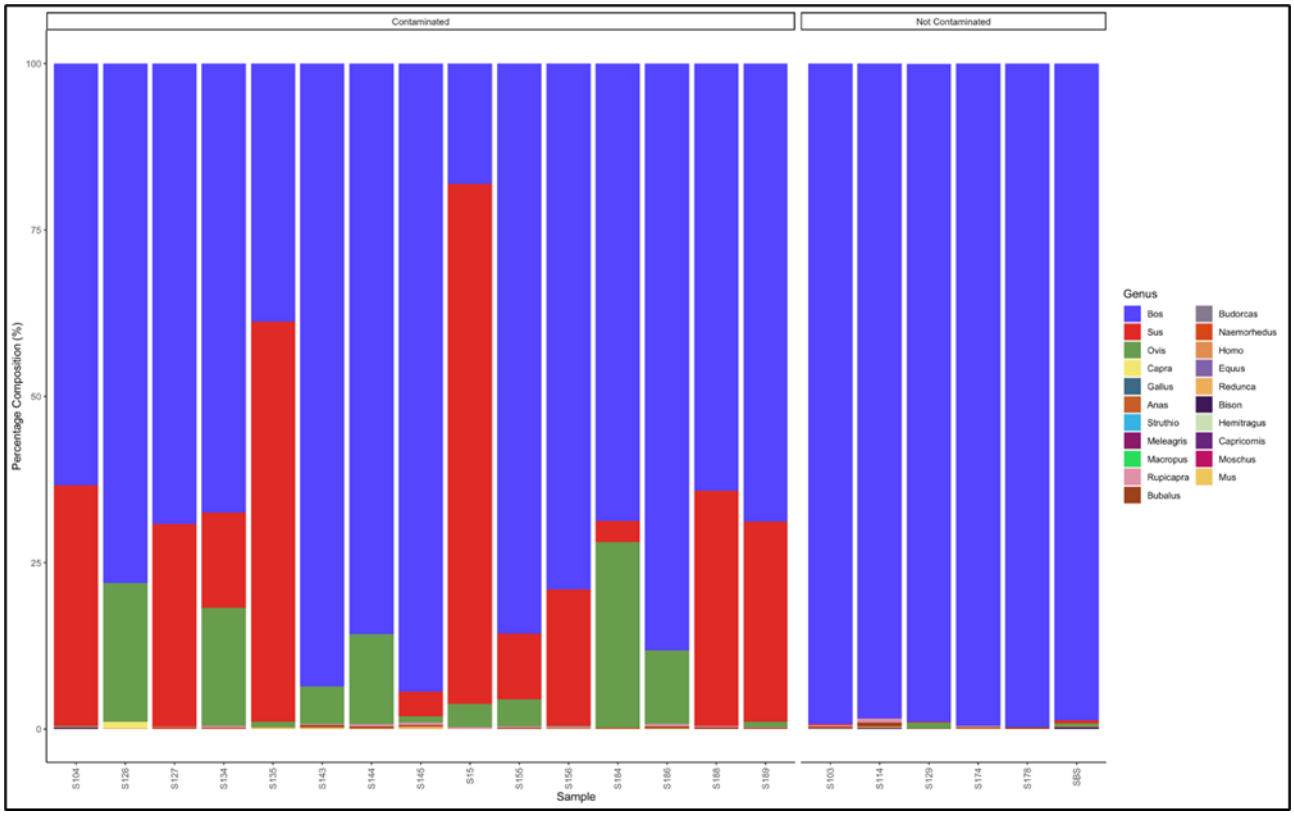

Figure 4. 11. Percentage composition of samples labelled as beef sausages.

4.4.5. Proportion Test of Two Categories (Contaminated vs. Not Contaminated) Using a Chi-Square Test

P-values of $p = 1.62 \times 10{-}4$, $p = 4.24 \times 10{-}10$ and $p = 1.95 \times 10{-}9$ were determined because of the chi-square test for the number of contaminated versus uncontaminated pure, artificially mixed and retail samples, respectively (Figure 4.12). There was 6% contamination in the pure samples and there was no statistically significant level of contamination. However, there was 100% and 26% contamination in the mixed and retail samples, respectively. The *p values* therefore indicate a significant level of contamination in the artificially mixed and retail samples. The overall *p value* for all three sample groups was $p = 1.85 \times 10{-}18$. Cramer's V association was 0.62, confidence interval (CI) was set at 95% (0.48, 0.75) and the number of samples observed (nobs) was 209 (Figure 4.12).

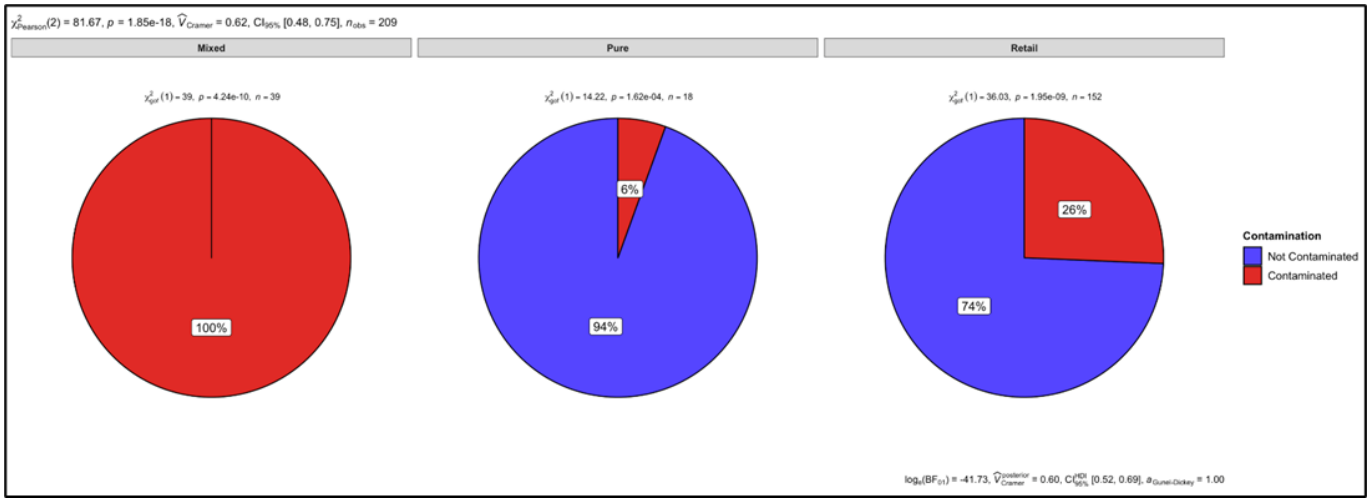

Figure 4. 12. Proportional test for mixed, pure, and retail samples.

The retail samples were then statistically analysed, according to the different meat types. There was no contamination observed in the biltong samples that had not specified which species they were from, and the p-value (p = 0.001) shows that there was no statistically significant level of contamination (Figure 13). There was contamination observed in the mince (15%), patties (13%), and sausage (38%) samples that had not specified which species they were from. The *p-values* for the mince (p = 2.56 × 10−4) and patty (p = 0.005) samples indicate that there was a statistically significant level of con-amination, however, there was no statistically significant level of contamination in the sausage (p = 0.275) samples (Figure 4.13). There was 18%, 9%, and 29% contamination in the beef biltong (p = 0.008), beef mince (p = 1.24 × 10−4), and beef sausage samples (p = 0.050), indicating a statistically significant level of observed contamination. However, regardless of finding 28% contamination in the beef patty samples, there was no statistically significant level of contamination (p = 0.059) (Figure 4.13). The overall p value for all the retail samples was p = 1.02 × 10−5. Cramer's V association was 0.48, confidence interval (CI) was set at 95% (0.21, 0.55) and the number of samples observed (nobs) were 152 (Figure 4.13).

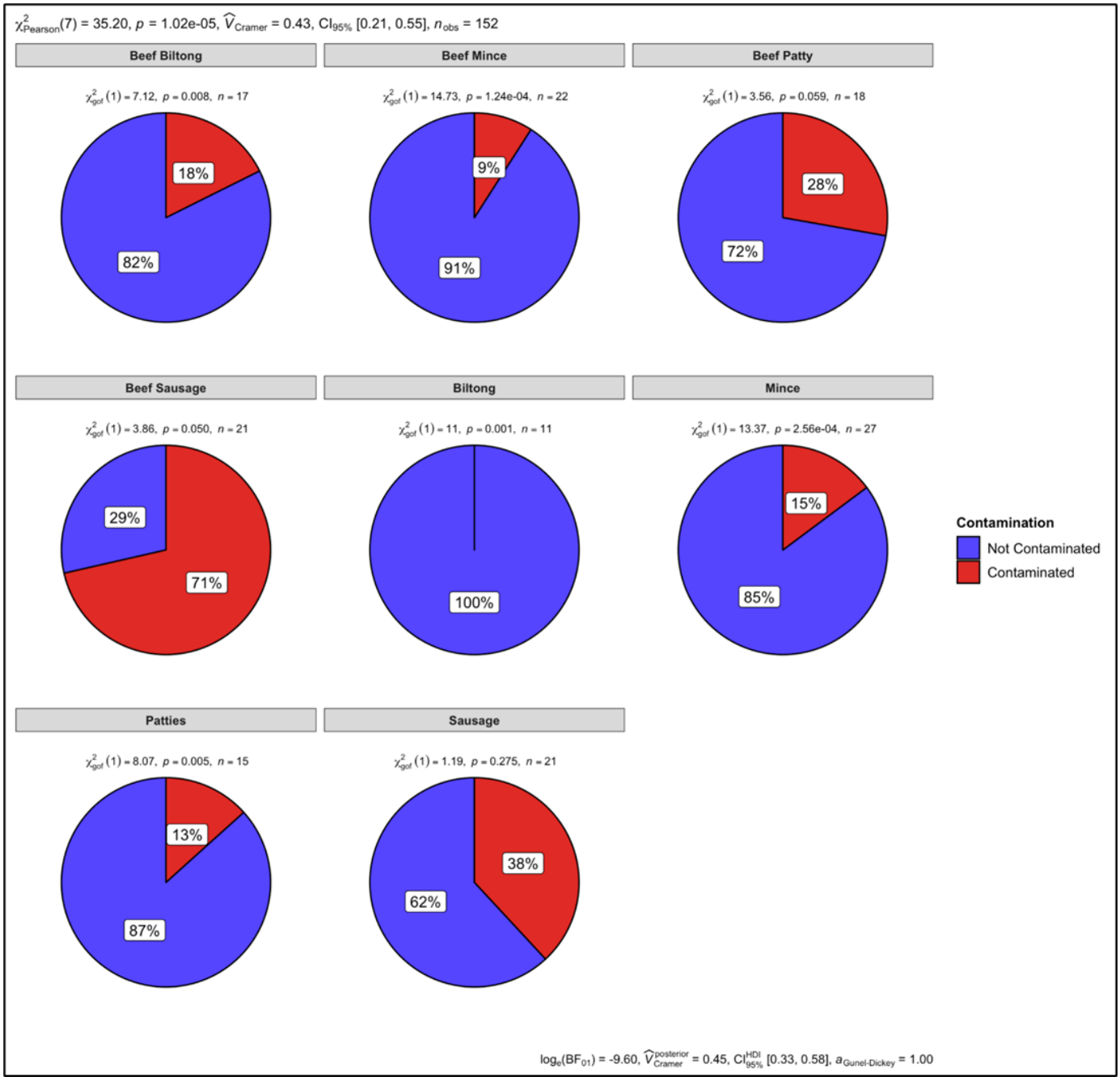

Figure 4. 13. Proportional test for biltong, mince, patty, and sausage samples.

4.5. Discussion

The main aim for this work was to develop a diagnostic pipeline for species identification in meat samples, including the identification of species in artificially mixed samples from different mammalian species. The mitochondrial 16S rRNA marker used in this study has

proven, in earlier studies, to have the power to detect individual species and even distinguish between closely related species (Tillmar *et al.*, 2013).

4.5.1. Pipeline for the identification of meat types using pure DNA from known meat samples

A pipeline, using the mitochondrial 16S rRNA gene and NGS, was established to initially identify known meat types from pure DNA that were not mixed with any other meat types. To achieve this, pure DNA from the respective meat types was used. The overall results demonstrated the ability of our protocol to identity pure DNA that is not mixed with other meat types, as all meat types from the known pure DNA were identified and mapped to the corresponding genera. The reads obtained mapped to the corresponding genus of each pure meat sample, with a similarity of ≤ 98% relative abundance for all meat types and with minor deviations of less than 2% relative abundance. One ostrich sample that was contaminated with beef DNA was a result of human error in the lab.

Contamination can either be intentional or unintentional. Intentional contamination occurs when deliberately adding a cheaper material to a product for economic gain. Unintentional contamination is the mistaken introduction of something into a product. This usually occurs through cross-contamination from the use of the same equipment amongst different products (Walker, Burns and Burns, 2013). In our experiment, the initial aim was to test the pipeline using 100% pure meat samples. There was no intention to assess contamination with other meat types. However, a contamination of less than 2% from other meat types was observed, which could be attributed to (i) trace amounts of other species having occupied abattoirs, butcheries, and retailers that slaughter, process, and sell multiple species' meat and use the same equipment for their processing, or (ii) a lack of maintenance of sequence databases could affect the stringency or sensitivity of species identification pipelines if new information is not added to a given database.

Overall, the initial verification test enabled us to determine some of the sensitivity thresholds of the pipeline. A threshold of 1% (w/w) for undeclared meat species in meat products was set by the Food Safety Authority (FSA) and Department for Environment Food and Rural Affairs (Defra) in Europe, after horse and pig DNA was identified in beef products (Zhang *et al.*, 2020).

Based on our analysis, a threshold of 2% is more practical when considering cross-contamination and database-stringency factors. However, a threshold of 2% carries implications for consumers who are mindful of their diet for religious reasons. The Jewish and Muslim communities prioritize the traceability and authenticity of the meat they consume, because they only consume meat from ritually slaughtered animals in accordance with their beliefs (Martuscelli *et al.*, 2020). Not declaring the meat species composition in a product violates their rights as consumers.

4.5.2. Identification of meat types using pure DNA from known meat samples artificially mixed at a ratio of 1:1 and 9:1

Having established the workflow and thresholds, the pipeline was further used to identify meat types from artificially mixed DNA at ratios of either 1:1 or 9:1. The aim of this part of the experiment was to simulate the conditions of retail market and deter-mine whether the pipeline could identify species in mixed DNA. The 1:1 ratio simulated retailers that intentionally contaminate meat products and do not try to hide it. This type of contamination mainly occurs for economic gain, by intentionally adding a cheaper product to the primary product (Cawthorn, Steinman and Hoffman, 2013; Doosti, Ghasemi Dehkordi and Rahimi, 2014; Tembe, Mukaratirwa and Zishiri, 2018). The pipeline demonstrated that the DNA of all meat types in the 1:1 ratio mixture were identified. However, some deviations were observed from the expected 50:50 percentages, and minute traces of species not included in some ratio mixtures. The major deviations were observed amongst the mutton: chevon, duck: ostrich and beef: kangaroo ratio mixtures. The deviation in the mutton: chevon ratio mixture may have been due to sheep and goats having similarities between their genomes, since they originate from the same family, Bovidae, and sub-family *Caprinae*. Previous research has demonstrated that sheep and goats evolved from the same ancestor *Rupicaprids* (goat antelopes) in the Pleistocene era (Sun *et al.*, 2004; Xing *et al.*, 2019). This may have resulted in an over or underestimation in the mutton: chevon ratio mixture. Similarly, ducks and ostriches have similarities in their genomes as they are both Aves species. Previous comparative cytogenic work has suggested that there is a preserved sequence homology between the Z and W chromosomes in ducks and ostriches, since recombination was suppressed (Nanda *et al.*, 2008). Furthermore, research has also demonstrated that the ostrich IgM isotype has a 66% and 63.1% sequence identity with the Cα and Cμ genes of the duck, respectively (Nanda *et al.*, 2008). The overestimation of beef

meat and underestimation of kangaroo meat in the beef: kangaroo ratio mixture may have been due to the cattle genome being sequenced more than the kangaroo genome. The sequencing of the bovine genome was initiated in 2002 (Gibbs *et al.*, 2002) and has continued, with several other works since published (Elsik, Tellam and Worley, 2009; Zimin *et al.*, 2009; Liao *et al.*, 2013; Weldenegodguad *et al.*, 2019; Rosen *et al.*, 2020). The kangaroo genome, on the other hand, was only first sequenced in 2011 (Murchison and Adams, 2011; Renfree *et al.*, 2011), even though the benefits of sequencing the kangaroo genome where initially discussed and published in 2003 (Wakefield and Graves, 2003). Further research on sequencing the kangaroo genome has been published (Deakin, 2013; Nilsson *et al.*, 2018), but there is a clear indication from the number of published articles that the kangaroo genome has been sequenced less than the cattle genome. Therefore, this may have resulted in an overestimation of cattle reads in the beef: kangaroo ratio mixture. The ever-decreasing cost of sequencing, coupled with increased efforts in sequencing non-conventional livestock species such as the kangaroo, will improve on the composition and quality of databases, which will also improve on the accuracy of the methods developed to date.

The 9:1 ratio simulated retailers that also intentionally contaminate meat products but try to hide it. This type of contamination occurs in situations where retailers intentionally add the fat or trimmings of certain meat species, such as pork, to improve the sensory value of some products (Bushnell, 2014). Just like the results of the 1:1 ratio mixture, the pipeline also managed to identify all the meat species whose DNA were in the 9:1 ratio, with some deviations from the expected 90:10 percentages. Descriptive statistics for the pure and artificially mixed samples were analysed. The standard deviation and variance indicated how close an observed value in a dataset is to the mean. A dataset with a smaller spread of values results in values closer to the mean, yielding smaller variance and standard deviation. In contrast, if a dataset has a wider spread of values, this results in values that are further from the mean, yielding a larger variance and standard deviation (Itoh *et al.*, 2010). The pure samples had lower standard deviations and variances that were closer to zero, meaning the values in the dataset had a smaller range and mean value. In contrast to the pure samples, the artificially mixed samples had higher standard deviation and variance values. This was brought about by a higher range within the dataset. The higher range of values may have been a result of the under and overestimation of expected percentages in the ratio mixtures. The use of mitochondrial genes in species identification has been used due to mitochondria having a mutation rate that is 10-

fold higher than that of nuclear genes, allowing for the discrimination of closely related species. Mitochondria is also abundant, with thousands of copies of DNA per cell, in comparison to nuclear genes that have single copies per cell (Yang *et al.*, 2014). However, the presence of several copies of mitochondrial DNA in a single cell can lead to either an underestimation (−70%) or overestimation (+160%) of species' DNA content (Flores-Munguia, Bermudez-Almada and Vázquez-Moreno, 2000). It has also been previously reported that there is a difference in binding efficiency of the universal primers for different species, resulting in a difference in the amplification efficiency and, therefore, leading to a large degree of error in quantitative analysis (Zhang *et al.*, 2020). The quantitative accuracy in meat species identification can be improved through the use of genes that have a single copy, the introduction of correction factors for primer amplification efficiency, designing degenerate primers, and controlling the number of amplification cycles and the amplification conditions (Zhang *et al.*, 2020).

4.5.3. Identification of retail processed meat products

The analyses of retail meat samples showed that beef was the main species found in most samples, since their reads predominantly mapped to the *Bos* genus. The samples that had indicated which species they were from on their product labels predominantly mapped to the *Bos* (cattle) genus (90% and above), confirming their origination from beef, as stated on the labels. There was, however, evidence of contamination and mislabelling of the pork (*Sus*) and/or mutton (*Ovis*) meat observed in most samples, but no mention of the presence of any other species on the labelling.

4.5.3.1. Biltong samples

All the unspecified biltong samples predominantly mapped to the *Bos* genus, however, there were minor traces of the *Sus* (pig) and *Ovis* (sheep) genera of less than 2% relative abundance, hence we concluded that there was no intentional contamination in the non-specified biltong samples. The beef biltong samples showed that pork had the highest percentage of contamination, with one of the samples having as high as 36% of the reads mapping to *Sus* genus. The contamination seemed intentional and for economic gain, as it is not practical to mistakenly add 36% of a different meat species, especially if it is a meat type that has been

previously reported to have a cheaper purchase price (Cawthorn, Steinman and Hoffman, 2013). Bottaro *et al.*, (2014) also reported of addition of low-valued meat and fat, such as pork, to high-valued meat, such as beef, as a form of intentional meat contamination for the purposes of economic gain.

4.5.3.2. Mince samples

Similarly, specified, and unspecified mince samples showed that the *Bos* (cattle) genus was predominant, with few samples contaminated with pork or mutton. There were a few contaminated mince samples that were contaminated with either mutton or pork. The percentages of pork found in the mince samples were between 2–3% relative abundance. These contamination percentages of may not necessarily have occurred intentionally for economic gain, since they were found at low percentages. Rather, the contamination may have been due to cross contamination from equipment not properly cleaned in operations that process multiple species (Tembe, Mukaratirwa and Zishiri, 2018). Similarly, in a South African study, Tembe, Mukaratirwa and Zishiri, (2018) concluded that the contamination of processed meat products was unintentional, and that the contamination may have been due to the use of the same equipment for processing different species. Regardless of the low contamination percentages in our study, the presence of pork has negative consequences to consumers who choose not to consume pork due to health reasons (Doosti, Ghasemi Dehkordi and Rahimi, 2014). The consumption of meat with undeclared allergens may cause an allergic reaction to certain consumers (Fajardo *et al.*, 2008; Maralit *et al.*, 2013; Doosti, Ghasemi Dehkordi and Rahimi, 2014). Previously, in the United States, allergy prevalence of 73%, 58%, and 41% to beef, pork and chicken, respectively, were reported in 57 patients suspected of being allergic to meat (Ayuso *et al.*, 1999).

4.5.3.3. Patty samples

The patty samples were also mainly from beef meat, with samples predominantly mapping to the *Bos* (cattle) genus. Pork was the main contaminate in the unspecified patty samples. One sample (Sample 48) had 7% pork in it, possibly a case of intentionally adding pork to a patty sample to improve its sensory and oxidative properties. The addition of pork meat or lard to

processed meat products has been previously reported (Martuscelli *et al.*, 2020), to improve the sensory properties and oxidative reactions of such processed meat products. Patty Sample 85 was intentionally mixed with two meat types, pork, and beef, as 59% of its read mapped to the *Sus* (pig) genus and 40% mapped to the *Bos* (cattle) genus. A case like this demonstrates intentional contamination for economic gain (unless if specified on the product label), since the production costs of pork are cheaper than those of beef (Cawthorn, Steinman and Hoffman, 2013), resulting in pork being cheaper to purchase. The major contaminant of labelled beef patty products was mutton (*Ovis*). This was unexpected, as mutton has a higher market price than beef. One of the reasons for substituting more expensive meat such as mutton for a cheaper meat such as beef may be due to the use of unmarketable trimmings from more expensive meat types (Doosti *et al.,* 2014). It is possible that intentional contamination with unmarketable mutton occurred in the contaminated beef patty samples, as one of the samples had as high as 34% of its reads map to the *Ovis* (mutton) genus.

4.5.3.4. Sausage samples

Our results demonstrated that pork and mutton were the main species that were contaminated in the sausage samples that had reads that predominately mapped to the *Bos* (beef) genus. However, there were samples of mutton and pork sausages contaminated with beef at percentages of 2–7% relative abundance. Some authors (Bottaro *et al.*, 2014; Di Pinto *et al.*, 2015) have indicated that the contamination of beef, in some meat products, may be from the addition of non-fat powdered milk to increase the overall yield and taste of the product. This may have been the case with the mutton and pork samples contaminated with beef. There was one sample (Sample 15) that was intentionally mixed with three meat types, beef, pork, and mutton, as the sample had 37%, 38%, and 23% of reads map to the *Bos* (cattle), *Sus* (pig), and *Ovis* (sheep) genus, respectively. This may have been a case of contamination for economic gain, because mutton and beef have a higher market price than pork (Tembe, Mukaratirwa and Zishiri, 2018).

The overall results of the specified and unspecified meat products indicated that pork was the main contaminate. Surowiec *et al.*, 2011 previously re-ported on undeclared pork and chicken in processed meat products such as burger patties and sausages and suggested that it could be from mechanically recovered meat (MRM), usually produced from pork and chicken carcasses.

According to Surowiec *et al.*, 2011, the addition of MRM, which is normally found in a paste-like form, represents a source of cheap protein in processed meat products, such as deli meats, burger patties, and sausages. This practice is, however, illegal in most countries, including South Africa (Cawthorn *et al.,* 2013). Furthermore, failure to declare the presence of other meat species in ingredients lists betrays consumer rights and has negative implications for consumers allergic to such contaminants and consumers whose religions observe dietary restrictions (Doosti *et a.,* 2014).

Similarly, in a South African study Cawthorn *et al.,* (2013), undeclared pork and mutton were found in minced meat, burger patties, and raw sausages labelled as beef, pork meat was the main undeclared meat type found in these meat samples. In another South African study on meat species' substitution, undeclared beef, pork, and lamb were found in commercially labelled wildlife meat products (D'Amato *et al.*, 2013b). More recently, another South African study revealed the presence of meat contamination in the province of KwaZulu-Natal (Tembe *et al.,* 2018). A high proportion of beef and mutton products were contaminated with pork and chicken. Undeclared species in the above-mentioned studies and in ours reflect that there is still a presence of meat adulteration in the South African meat market. Judging from the results we have observed from the retail samples, there seems to be intentional contamination for economic gain. There is need to improve product labelling to indicate every species within a meat product so that consumers can make informed decisions. Some major retailers in South Africa, such as Food Lover's Market, Pick and Pay, and Checkers, now mention the presence of multi-species on their meat product's labels. For example, a sausage sample, today, might be labelled as 70% beef, 20% pork, and 10% Water. This type of clarity in labelling assists consumers who prefer to avoid certain species for allergenic, religious, or ethical reasons.

Statistics indicate that beef has the highest gross value as compared with other meat species produced in South Africa, with an average of R 23.5 billion per annum (DAFF, 2018a). There was a slight decrease in cattle production between 2017–2018, due to farmers in South Africa not having enough cattle to slaughter. This led to an increase in beef market prices, as herds were replenished, which in turn decreased the consumption of beef and beef products. Consumers opted for cheaper alternatives, such as chicken and pork (DAFF, 2018a). The gross production of mutton in South Africa is an average of R 4.57 billion per annum (DAFF, 2018c). There was, however, a decline in sheep production from 2017 due to stock theft, which led to

an increase in demand and subsequent shortages in the supply of mutton (DAFF, 2017). This shortage in supply, coupled with high production costs, has led to high market prices of mutton in the South African meat market. Beef is more readily available; hence it is the main species processed into value-added products in South Africa. However, high production costs result in a higher purchasing price for beef. Mutton also has a higher purchasing price than the other meat types, mainly because it has higher production costs and is not readily available on the market. Chicken and pork have lower production costs in comparison with beef and mutton and this has led to them having a lower purchasing price and, therefore, being more frequently fraudulently added to higher-value products labelled as beef or mutton, for economic gain.

4.6. Conclusion

In conclusion, the chapter presents a universal diagnostic molecular method for the identification of meat species. The method used the mitochondrial *16S* rRNA gene, which has demonstrated its variability from the results of the phylogenetic analysis in Chapter 3. The verification experiment identified all species present in the known DNA mixtures, proving the accuracy of the pipeline in the species identification in the processed meat samples that were collected. Meat suppliers can possibly implement the current method to authenticate their products, and the food industry may also use this method to reveal any form of mislabeling that may be present within meat products.

CHAPTER FIVE

5. Development of SNP marker panel for discrimination and assignment of cattle breeds reared in South Africa using next generation sequencing technologies.

5.1. Abstract

Meat authentication is vital for consumer protection and especially important for beef products, as there is a rise in mislabeling of cheaper beef products with certified or branded beef breeds preferred by consumers. The aim of this study was to investigate the presence of breed specific SNPs within the mitochondrial DNA of cattle breeds that are reared in South Africa. A total of 42 animals from 13 cattle breeds, namely Afrikaner (4), Beefmaster (4), Boran (4) Charolais (2), Hereford (2), Nguni (2), Simbra (3), Bonsmara (4), Brahman (4), Drakensberger (4), Limousin (2), Santa (3) and Simmentaler (3) were used. Whole genome sequencing was performed on a single lane of the Illumina HiSeq 2500 (Illumina, San Diego, CA, United States) at 10X coverage. Quality check was performed using FASTQC and to avoid any mapping errors, whole genome sequences were mapped to the latest cattle genome ARS-UCD1.2 using BWA v0.7. Alignments were processed using GATK v3.3 and thereafter, variants were called across the entire mitochondria using HaplotypeCaller v3.3. All identified SNPs went through a hard filtering process to remove any low-quality SNPs. SNPs that passed quality control filtering were separated into homozygous and heterozygous SNPs and all homozygous SNPs were used to identify breed specific SNPs. A total of 8,3 billion reads were left after quality filtering, with an average of 198,3 million reads per sample. The reads had an average coverage of 8,8X, with the highest coverage in the Afrikaner breed (11,3X) and the lowest in the Limousin breed (7X). A total of 12 996 variants were identified after hard filtering and 12 633 of these were SNPs and 363 were Indels. The highest number of variants were identified in the Brahman breed (2 066) and the lowest in the Nguni breed (340). The highest number of homozygous SNPs were found in the Limousin breed (534) and the lowest in the Charolais breed (24). A total of 125 breed specific SNPs were identified in 12 of the 13 breeds. The Charolais breed had no breed specific SNPs. The highest number of breed specific SNPs were found in the Limousin breed (59) and the lowest in the Nguni breed (1). The *COX3* gene (22) contained the highest number of breed specific SNPs, followed by the *16S rRNA* gene (19). A total of 19 breed specific SNPs were shared amongst the breeds. The Limousin breed

(12) had the highest number of shared SNPs and the Nguni (1) and Simbra (1) had the lowest number of shared SNPs. This study provides the first insight of breed specific SNPs identified in the mitochondria of 13 cattle breeds reared in South Africa. The breed specific SNPs identified provide an understanding of the regions within mitochondrial genes that are unique in each breed. These regions include the *16S* rRNA and *COX3* mitochondrial regions and can be used in the authentication of beef meat in the meat industry.

Keywords: Beef meat, Authentication, Cattle breeds, Whole genome sequencing, SNPs, Mitochondria

5.2. Introduction

Beef meat is one of the most consumed meat types globally. Globally, pigs are the most consumed type of meat species, accounting for 36.3% of consumption, followed by poultry, cattle, sheep/goats accounting for 35.2%, 22.2% and 4.6% respectively (USDA, 2015). Meat adulteration involving the more expensive beef meat with cheaper meats has become a worldwide concern (Walker, Burns and Burns, 2013; Doosti, Ghasemi Dehkordi and Rahimi, 2014; Fang and Zhang, 2016; Zhang *et al.*, 2020), including South Africa (Cawthorn *et al.,* 2013; D'Amato *et al.*, 2013b; Tembe *et al.,* 2018; Chaora *et al.*, 2022). The replacement of high value beef products with lower value products for economic gain has become common in the meat industry more especially in the beef industry (Cho *et al.,* 2014; Doosti *et al.,*2014; Yang *et al.*, 2014; Lin *et al.,* 2019). Meat from dairy breeds has generally been considered of low eating quality, when compared to beef breeds (Surányi *et al.*, 2021). Consumers are increasingly getting concerned about the breed of the beef meat they purchase, and these reasons depend on the cost, religion preference, nutritional value and even pleasure (Surányi *et al.*, 2021). The beef industry worldwide has in the past introduced branding of beef products from certain cattle breeds. These branded beef products can be found in retail stores or restaurants. The branding may be dependent on how the cattle was fed, purity of the breed or the purity of the hide colour. For instance the Angus breed was certified as *Certified Angus Breed* in 1978 by the USDA Agricultural marketing Services, and became the first branded beef in the United States (Bass, 2016). The Angus branding depends on carcass specifications and live animal phenotypic identification (Bass, 2016). In South Africa the Wagyu breed was certified as Certified South African Wagyu Beed (CSAWB), that is fully traceable with no

hormones and growth stimulants allowed (Coleman, 2017). Apart from the Wagyu breed, South Africa has also introduced breed specific beef meat to the local market from breeds such as Angus, Afrikaner and Bonsmara (Brits, 2017; Lombard *et al.,* 2017).Therefore, there is a need to develop an accurate and reliable method for tracing and identifying cattle breeds in beef meat products in order to authenticate the purity of beef meat, safeguard consumer health, thereby satisfying consumer preference and eliminating meat fraud for economic gain.

Cattle breed traceability and assignment is of great importance to the beef industry, and it has been made possible through cutting-edge research that has previously been published (Dimauro *et al.*, 2013; Makina *et al.*, 2014; Zwane *et al.*, 2016; Lashmar *et al.*, 2021; Lashmar *et al.*, 2021). Several studies have demonstrated that the use of Single Nucleotide Polymorphisms (SNPs) for breed traceability and assignment is highly reliable (Negrini *et al.*, 2009; Dimauro *et al.*, 2013; Zwane *et al.*, 2016; Lashmar *et al.*, 2021). Genetic information of unknown individuals can be used to assign and allocate individual animals to a known population. The methods that can be used to select informative markers for breed discrimination and assignment of individuals to their respective populations have been previously published (Negrini *et al.*, 2009; Opara *et al.,* 2012; Zwane *et al.*, 2016). These methods include the use of high-density Single Nucleotide Polymorphism (SNP) assays, microsatellites and AFLP markers. Compared to microsatellites and AFLP markers, SNPs are available in large quantities within a genome (Lashmar *et al.*, 2021). A small number of SNPs can be used to reveal the genetic structure amongst breeds (Wilkinson *et al.*, 2011), and from there a small set of appropriate and informative SNPs can be used for breed assignment (Martínez-Camblor *et al.*, 2014). Previously, studies have been conducted on the use of SNPs to trace cattle breeds. In a study on European cattle breeds, Orrù *et al.*, (2009) tested 18 SNPs on their potential to classify individuals belonging to six different cattle breeds. Also in Europe, 90 SNPs were chosen to identify and trace four European beef products. They found that the percentage of accurate assignment ranged between 80 to 100 % (Dimauro *et al.*, 2013). In a Korean study, 90 SNPs were used to discriminate between Hanwoo Korean native cattle and Holstein cattle. The researchers reported a 100 % probability of discrimination (Cheong *et al.*, 2013).

High-density SNP assays like the BovineSNP50, GGP- 80K and bovine high-density contain a large number of SNPs and have been used to select informative SNPs for breed assignment (Matukumalli *et al.*, 2009; Zwane *et al.*, 2016). However, SNP genotyping assays are found to

be more informative in European breeds than indicine or indigenous African breeds (Zwane *et al.*, 2016). For example, studies that have used the BovineSNP50 chip in indigenous South African breeds, have demonstrated lower minor allele frequencies and linkage disequilibrium, compared to European taurine breeds (Edea *et al.*, 2013; Makina *et al.*, 2014). A study by Zwane *et al.*, (2016) using the BovineSNP50 to select signatures of selection, demonstrated that South African indigenous breeds had lower number of breed-specific informative markers. South African indigenous cattle alongside European breeds that were earlier introduced to the country, play a vital role in beef production within the country (Scholtz, 2010). High throughput sequencing has become popular in areas of genomics, and this has been made possible by Next Generation Sequencing (NGS) technologies. Through NGS a large number of DNA molecules can be rapidly and cost effectively sequenced at the same time, producing large amounts of data, that can be used to infer the genomic makeup of species being analysed (Le Roex *et al.*, 2012; Mullen *et al.*, 2012). The cost of whole genome sequencing (WGS) has decreased over the years, due to the introduction of advanced and cost effective sequencing technologies (Czech *et al.*, 2018). Whole genome sequencing allows the generation of large amounts of data from individuals and breeds. The use of WGS allows for the identification of breed specific SNPs and can also be used to construct breed-specific reference genomes. Previously, a study conducted by Czech *et al.*, (2018) used WGS from 936 bulls to identify breed-specific SNPs in Angus, Brown Swiss, Hereford, Fleckvieh, Jersey, Limousin and Simmental breeds. These SNPs were in turn used to construct breed-specific reference genomes. Approximately, 10,4 million SNPs were discovered by Choi *et al.*, (2014) in Korean Hanwoo, Jeju Heugu and Holstein cattle, of which 54% of them were novel SNPs. Population genetics studies have focused on the identification of mutations within mitochondrial DNA (mtDNA). It has also been found that mutations within the mitochondrial demonstrate useful genetic information, that can be used for tracing the origin of breeds and identifying individual animals (Chung, 2013). However, SNP identification within cattle mitochondria has been limited to the non-coding D-Loop region in the past, as it is believed that SNPs in the coding regions may not fully explain genetic diversity (Chung, 2013). The use of whole mtDNA sequences can be used to extrapolate useful information to identify SNPs that can be used to assign individuals to cattle breeds.

To our knowledge, there is no published study that has been conducted to identify breed-specific SNPs that can be used for beef meat authentication in the South African beef industry.

It has also been previously predicted that the sequencing of indigenous South African breeds, can lead to the discovery of new breed-specific SNPs that can be used to develop custom made SNP chips for the cattle population in South Africa (Zwane *et al.*, 2019). Therefore, the aim of this study was to investigate the presence of breed-specific SNPs using the entire mitochondrial genome of 13 European and Indigenous cattle breeds reared in South Africa for use in breed assignment and traceability.

5.3. Materials and Methods

5.3.1. Ethics Approval

The study was approved by the University of South Africa CAES Animal Research Ethics Review Committee (Reference number: 2016/CAEC/030).

5.3.2. Sample collection and DNA extraction

Cattle hair samples were randomly collected from a total of 42 animals from 13 breeds, including Afrikaner (4), Beefmaster (4), Boran (4) Charolais (2), Hereford (2), Nguni (2), Simbra (3), Bonsmara (4), Brahman (4), Drakensberger (4), Limousin (2), Santa (3) and Simmentaler (3). Genomic DNA was extracted from hair samples using a Hamilton Microlab Star automated liquid handler (Hamilton Inc.). A Macherey–Nagel NucleoMag Tissue kit for DNA purification from cells and tissue (Macherey–Nagel, Germany) was used for DNA extraction according to the Genomic DNA from Tissue user manual. The DNA concentration of the cattle breed samples was between 28–467 ng/ul prior to sequencing. The quantification of DNA for all samples was checked using the Qubit® fluorescent dye method, and gel electrophoresis was used to assess the quality of the starting material. A ratio of A260/A280 was used to access the purity of all extracted DNA.

5.3.3. Library preparation and Illumina sequencing

Library preparation and sequencing was performed at the Agricultural Research Council Biotechnology Platform (ARC – BTP) in Pretoria, South Africa. Each individual animal was pooled according to breed type using 170 ng of DNA per animal. Library preparation was performed using the TruSeq DNA sample preparation kit v2 (Illumina, San Diego, CA, United

States), using 1ug of genomic DNA according to the manufacturer's instructions. DNA was fragmented using a Covaris E220 sonicator (350bp), followed by the ligation of adaptors and 12 cycles of polymerase chain reaction (PCR). Clusters were generated on the flow cell using the automated cBot Cluster Generation System (Illumina, San Diego, CA, United States). Whole genome sequencing was performed on a single lane of the Illumina HiSeq 2500 (Illumina, San Diego, CA, United States) at 10X coverage. The reads were paired end with a read length of 125bp. The sequencing images were analysed using Bcl2fastq v2.0 (Illumina) to generate raw fastq files.

5.3.4. Alignment of sequences

Raw sequenced reads were checked for base quality using FastQC (Andrews, 2017) and thereafter, adapter trimming was performed using Trimmomatic v0.33 (Bolger, Lohse and Usadel, 2014) to remove Nextera Transposase sequence adapters and low quality reads. Reads were trimmed if bases had an average Phred quality score that did not exceed 20. After trimming, pairs of DNA sequences for which each read exceeded 35bp were retained for further analysis. To avoid any mapping errors by mapping the mitochondrial reads only, whole genome sequenced reads were aligned to the ARS-UCD1.2 reference genome using BWA v0.7 (Li and Durbin, 2009) using default parameters. The BWA software is a package that is used to map lowly divergent sequences against a large reference genome (Li and Durbin, 2009). The alignments were converted to BAM format using SAMTools v1.2 (Li *et al.*, 2009).

5.3.5. Variant calling and annotation

The Genome Analysis Toolkit (GATK) v3.3 was used to process the alignments according to GATK Best Practices Pipeline using the genomic variant call format (GVCF) workflow for downstream SNP and indel calling (Van der Auwera *et al.*, 2013). RealignerTargetCreator was used to identify poorly mapped regions in the alignments and these alignments were realigned around indels using GATK and IndelRealigner. Duplicates were marked using Picard MarkDuplicates v.1.135 (http://broadinstitute.github.io/picard) and base quality score were recalibrated using GATK BaseRecalibrator, which resulted in a final BAM file for each sample. Variant calling on the BAM files was done using the complete mitochondria genome of each sample. Variants were called across the entire mitochondria using HaplotypeCaller v3.3 (Van

der Auwera *et al.*, 2013) to create a multi-variant VCF file. Genotypes were called for each breed with a minimum genotype quality of 20. To avoid false discovery of variants hard filtering of identified variants within the mitochondria was performed using the following criteria: Phred scaled polymorphism probability (QUAL) < 30.0, variant confidence normalized by depth (QD) < 2.0, mapping quality (MQ) < 40.0, strand bias (FS) > 60.0, HaplotypeScore > 13.0, MQRankSum < −12.5, and ReadPosRank-Sum < −8.0 (Choi *et al.*, 2015). All SNPs that passed this criterion were then used to identify breed specific SNPs within the breeds. All of the filtered SNPs were also annotated and assigned to functional categories using snpEff (Cingolani *et al.*, 2012) and the *Bos taurus* reference genome ARS-UCD1.2.

5.3.6. Identifying breed-specific SNPs and annotation

All the identified SNPs that passed the hard filtering process were separated into homozygous (fixed) genotypes called in all individuals within a breed and heterozygous (variable) genotypes in each breed. To identify breed specific, the first step was to remain with individuals that were the same. This meant removal of SNPs with genotypes that had different alleles (heterozygous) and retaining of those that had the same allele pairs (homozygous) from the dataset. The resultant dataset, therefore, contained only homozygous SNPs that were used for further analyses. The second step was to develop an in-house script that was used to identify breed specific SNPs within the mitochondria of each of the 13 breeds. If a breed had a different genotype from other breeds within a certain position on the mitochondria, then that breed was considered to have a breed specific SNP at that position.

5.4. Results

5.4.1. Sequencing and mapping

A total of 42 animals from 13 different breeds, namely, Afrikaner (4), Beefmaster (4), Boran (4) Charolais (2), Hereford (2), Nguni (2), Simbra (3), Bonsmara (4), Brahman (4), Drakensberger (4), Limousin (2), Santa (3) and Simmentaler (3), were sequenced at 10X coverage using the Illumina HiSeq 2500 sequencer. Sequencing of the 13 breeds obtained approximately 11,7 billion raw paired-end sequences with an average of 280,4 million reads per sample (Appendix 5.1). After quality filtering and trimming, a total of 8,3 billion reads

(Appendix 5.1) passed quality control, with an average of 198,3 million reads per sample (Table 5.1).

The highest number of raw reads were found in Boran breed (347,3 million) and the lowest in the Limousin breed (220,3 million). The highest number of clean reads were found in the Brahman breed (224,7 million) and the lowest in the Limousin breed (159,9 million) (Table 5.1). The highest percentage of properly paired reads were found in the Charolais breed (98%) and the lowest were found in the Limousin breed (84,5%), with an average of 95,3% across all breeds (Table 5.1). Using BWA (Li and Durbin, 2009), clean reads were mapped to the latest bovine reference genome ARS-UCD1.2 (Li and Durbin, 2009), with an average mapping of 97,1% per sample (Table 5.1). The reads covered an average of 8,8X of the genome across all samples, with the highest coverage in the Afrikaner breed (11,3X) and the lowest in the Limousin breed (7X) (Table 5.1).

5.4.2. Variant Detection

A total of 13 539 variants were identified in the 13 breeds before hard filtering, where 13 169 of these were SNPs and 370 were Indels. After quality filtration a total 12 996 remained, where 12 633 SNPs and 363 Indels remained, making it 96% of variants remaining after hard filtration (Table 5.2). The greatest number of total variants were identified in the Brahman breed (2 066) and the lowest in the Nguni breed (340). The Brahman breed had the highest number of SNPs (1 998) and Indels (68) after hard filtering. The Nguni breed had the lowest number of SNPs (352) and Indels (8) after hard filtering (Table 5.2). The identified SNPs were also classified as homozygous and heterozygous SNPs (Table 5.2). The highest number of homozygous SNPs were found in the Limousin breed (534) and the lowest in the Charolais breed (24). The highest number of heterozygous SNPs were found in the Brahman breed (1 872) and the lowest in the Nguni breed (289) (Table 5.2).

Table 5. 1: Sequencing and mapping results of 13 cattle breeds

Breed	Number of Animals	Average Raw Reads	Average Clean Reads	Properly Paired %	Average Mapping %	Average Coverage

Afrikaner	4	263 325 626	211 766 075	95,6	99,1	11,3X
Beefmaster	4	258 096 189	187 850 308	95,1	96,8	8,1X
Bonsmara	4	272 435 566	196 376 670	94,2	95,7	8,6X
Boran	4	347349385,5	215987290,5	96,6	99,1	9,4X
Brahman	4	308 893 570	224 706 892	94,3	96,7	9,8X
Charolais	2	266 812 323	190 760 763	98	99,1	8,3X
Drakensberger	4	282 821 197	211 937 466	97,0	9,2	9,2X
Hereford	2	282 821 197	211 937 466	97,5	99,0	9,2X
Limousin	2	220 380 040	159 945 042	84,5	85,5	7,0X
Nguni	3	279 501 077	192 384 591	97,7	99,6	8,3X
Santa	3	280 107 813	164 603 367	97,8	99,5	7,1X
Simbra	3	257 429 735	197 109 919	93	94,6	8,6X
Simmentaler	3	252 159 763	213 478 978	97	99	9,3X
Means		274 779 499	198 372 679	95,3	97,1	8,8X

Table 5. 2: Summary of SNPs and Indels identified in 13 cattle breeds

		SNPs				Indels	
Breed	**Total No. of Variants**	**Total No. of SNPs**	**No. of SNPs after Filtration**	**No. of Homozygous SNPs after Filtration**	**No. of Heterozygous SNPs after Filtration**	**Total No. of Indels**	**No. of Indels after Filtration**
Afrikaner	682	660	647	76	571	22	22
Beefmaster	1 451	1 406	1 349	69	1 280	45	45
Bonsmara	1 531	1 499	1 442	151	1 291	32	32
Boran	1 512	1 463	1388	58	1 330	49	45

Brahman	2 175	2 105	1 998	126	1 872	70	68
Charolais	820	800	765	24	741	20	20
Drakensberger	742	723	699	80	619	19	19
Hereford	443	432	416	89	327	11	11
Limousin	934	915	885	534	351	19	18
Nguni	374	366	352	46	289	8	8
Santa	1 302	1 268	1 217	120	1 084	34	34
Simbra	1 085	1 058	1 018	48	970	27	27
Simmentaler	488	474	457	104	353	14	14
Total	**13 539**	**13 169**	**12 633**	**1 525**	**11 078**	**370**	**363**

Table 5. 3: Number of breed specific SNPs per mitochondria gene per breed

MT Gene	Position	Afr[1]	Bee[2]	Bon[3]	Bor[4]	Bra[5]	Dra[6]	Her[7]	Lim[8]	Ngu[9]	San[10]	Simb[11]	Sim[12]	Total SNPs in MT Gene
Control Region	0 - 768	-	-	-	-	-	-	-	3	-	-	-	0	**3**
12S rRNA	837 - 1801	-	-	2	-	-	-	-	-		-	-	1	**3**
16S rRNA	1875 - 3628	1	-	-	1	-	-	-	14	-	-	-	3	**19**
ND1	3704 - 4678	1	-	-	-	-	-	-	-	-	-	-	-	**1**
tRNA^{Met}	4824 - 4893	1	-	-	-	-	-	-	-	-	-	-	-	**1**
ND2	4894 - 5940	-	-	1	-	-	-	-	3	-	-	-	1	**5**
tRNA^{Trp}	5941 - 6012	-	-	-	-	-	-	-	1	-	-	-	-	**1**
tRNA^{Ala}	6013 - 6082	-	-	-	-	-	-	-	4	-	-	-	-	**4**
COX1	6329 - 7879	1	-	-	-	-	-	-	2	-	-	-	1	**4**
tRNA^{Ser}	7880 - 7950	-	-	-	-	-	-	-	2	-	-	-	-	**2**
tRNA^{Asp}	7955 - 8021	-	-	-	1	-	-	-	-	-	-	1	-	**2**
ATP8	8803 - 8969	-	-	-	-	-	-	-	2	-	-	-	-	**2**
ATP6	8960 - 9643	-	-	-	-	-	-	-	-	-	-	-	8	**8**
COX3	9643 - 10428	1	1	-	-	2	-	1	15	-	2	-	-	**22**
ND4	11209 - 12597	-	2	3	-	4	-	1	2	-	-	-	-	**12**

ND5	12804 - 14651	1	1	2	-	-	6	1	11	1	5	1	-	**29**
CTYB	15252 - 16392	1	1	1	-	1	1	1	-	-	1	-	-	**7**
Total		**7**	**5**	**7**	**2**	**7**	**7**	**4**	**59**	**1**	**8**	**2**	**13**	**125**

[1] – Afrikaner, [2] – Beefmaster, [3] – Bonsmara, [4] – Boran, [5] – Brahman, [6] – Drakensberger, [7] – Hereford, [8] – Limousin, [9] – Nguni, [10] – Santa, [11] – Simbra and [12] – Simmentaler.

5.4.3. Breed Specific SNPs identification

A total number of 125 breed specific SNPs were identified in the mitochondria of 12 breeds of the 13 breeds (Table 5.3). There were no breed specific SNPs that were identified in the mitochondria of the Charolais breed. The highest number of breed specific SNPs were found in the Limousin breed (59) and the lowest in the Nguni breed, which only had one breed specific SNP (Table 5.3). The highest number of breed specific SNPs were found in the *ND5* mitochondrial gene with 29 breed specific SNPs, followed by the *COX3* mitochondrial gene with 22 breed specific SNPs, and the 16*S rRNA* gene with 19 breed specific SNPs (Table 5.3). The lowest number of breed specific SNPs were found in the *ND1*, $tRNA^{Trp}$ *and* $tRNA^{Met}$ mitochondrial genes, each with one breed specific SNP in the Limousin and Afrikaner breeds, respectively (Table 5.3). A total of 19 SNPs were shared in more than one breed (Table 5.4). Most of the shared SNPs were found in the Limousin breed (12 positions) and the lowest in the Nguni (1 position) and Simbra (1 position) breeds. There were 8 breed specific SNPs that were a combination of two breeds, 10 breed specific SNPs that were a combination of three breeds and 1 breed specific SNP that was a combination of five breeds (Table 5.4). One position (10 149) had a biallelic SNP in both the Brahman and Limousin breeds (Table 5.4). The highest number of shared SNPs within a gene were found in the *ND5* gene, with 8 shared SNPs (Table 5.4). The lowest number of shared SNPs within a gene were found in the *16S rRNA* gene, with 1 shared SNP in the Boran and Afrikaner breeds (Table 5.4).

Table 5. 4: Shared SNPs

SNP Position	Breeds Sharing	No. of Breeds Sharing	MT Gene
3270	Boran and Afrikaner	2	*16S rRNA*
10113	Limousin, Hereford, and Santa	3	*COX3*
10149*	Brahman and Limousin	2	*COX3*
10232	Brahman and Limousin	2	*COX3*
10347	Afrikaner and Santa	2	*COX3*
11720	Beefmaster, Bonsmara and Hereford	3	*ND4*
11721	Bonsmara, Beefmaster and Hereford	3	*ND4*
11695	Brahman, Limousin, and Beefmaster	3	*ND4*
11697	Brahman and Limousin	2	*ND4*
13128	Limousin and Bonsmara	2	*ND5*
13201	Beefmaster, Drakensberger, Afrikaner, Nguni and Simbra	5	*ND5*
13626	Limousin, Drakensberger and Santa	3	*ND5*
13632	Limousin, Drakensberger and Santa	3	*ND5*
13633	Limousin, Drakensberger and Santa	3	*ND5*
13637	Limousin, Drakensberger and Santa	3	*ND5*
13638	Limousin, Drakensberger and Santa	3	*ND5*
14010	Limousin and Hereford	2	*ND5*
15814	Bonsmara and Hereford	2	*CYTB*
16148	Beefmaster, Drakensberger and Afrikaner	3	*CYTB*

*Position 10149 had a biallelic SNP for the Brahman and Limousin breeds

5.5. Discussion

5.5.1. The importance of identification of breed specific SNPs in cattle breeds

The identification of breed specific SNPs in cattle breeds is of vital importance to the meat industry, as identified markers can be used to authenticate beef products. Consumers have the right to choose beef breeds that they consume, therefore this has led to breed specific branding of beef breeds such as Angus, Wagyu, Bonsmara, Hanwoo (Cheong *et al.*, 2013; Bass, 2016; Brits, 2017; Coleman, 2017; Lombard *et al.,* 2017), just to mention a few. The certification is based on how the cattle are fed, purity of the breed or even purity of the hide. The Hanwoo breed is a Korean branded breed that is well known for its marbling and carcass quality and is sold at a higher price than other imported beef and domestic Holstein beef in Korea (Cheong *et al.*, 2013). The Angus breed is originally an American cattle breed, and its certification of beef meat is based on carcass specifications that must be met in addition to the phenotypic appearance. These include marbling, maturity, hot carcass weight, ribeye size, fat thickness, muscling thickness, free to capillary raptures, color of the carcass and neck hump size (Bass, 2016). The Wagyu breed is originally a Japanese breed that has gained popularity in the beef industry worldwide, with the largest production being found in Japan (1,64 million) followed by Australia (390 000) (Erasmus, 2018). In South Africa a Wagyu breed certification production was launched in 2018, by the Wagyu Society of South Africa (Waygu SA) (Erasmus, 2018). The certification of Wagyu in South Africa is based on the genetics of the animal and marbling score. Certified Wagyu animals need to have been bred by a fullblood or pure-bred Wagyu SA registered sire. Furthermore, the meat from the purebred animals should have a marbling score of four and above (Erasmus, 2018; Steinberg, 2021). Another beef breed that is branded in South Africa is the Bonsmara breed. In 2016 the Sernick Group launched South Africa's first certified Bonsmara beef brand called Certified Sernick Bonsmara Beef, which gave consumers satisfaction when it came to purchasing 100% authentic South African Bonsmara beef (Brits, 2017).

Some consumers prefer selecting meat from a specific breed based on its brand certification. In Korea, the Hanwoo beef breed is preferred over imported breeds such as the Holstein dairy breed, due to its superior marbling and carcass quality, resulting in the Hanwoo breed being

priced higher. Cases of mislabeling of the Hanwoo breed in the Korean meat market have occurred, henceforth prompting Cheong *et al.*, 2013 to develop 90 SNPs to discriminate between the Hanwoo breed and other breeds including the Holstein breed. In Europe some food products are recognized by the name of their region leading to some products having a Protected Geographical Indication certification (PGI). The PGI certification gives consumers assurance of the conditions and procedures undergone to protect the regional name of certain products. For instance, in Spain veal is produced in the province of Navarre and is certified as PGI-Certified *Ternera de Navarre* (Beriain *et al.*, 2021). In the USA beef may not necessarily have a PGI certification, however, the Department of Agriculture in the USA has certified beef programs and one of them is the Certified Angus Beef based on the Angus breed (Beriain *et al.*, 2021).

5.5.2. The use of the cattle mitochondria to identify breed specific SNPs

The current study was conducted to identify breed-specific SNPs in the mitochondria of 13 cattle breeds reared in South Africa. These included Afrikaner, Beefmaster, Boran, Charolais, Hereford, Nguni, Simbra, Bonsmara, Brahman, Drakensberger, Limousin, Santa and Simmentaler. The whole genome of cattle breeds sequenced in our study were sequenced at a sequencing depth of 10X for identification of breed-specific SNPs in the mitochondria. The mitochondria genome is 16,5kb in size (Chung, 2013), therefore, we believe that a sequencing depth of 10X was sufficient to identify breed-specific SNPs within the cattle breed mitochondria. Previous studies (Fernandes Júnior *et al.*, 2020) have indicated that a whole genome sequencing depth of 10X is sufficient to identify SNPs of high quality. In addition, the 1000 bull genomes project requires animals sequenced at a minimum depth of 10X (http://www.1000bullgenomes.com/). This, furthermore, gave us confidence that we would manage to identify high quality and informative SNPs in the mitochondria of the cattle breeds. The use of whole genome sequences to search for SNPs within cattle breeds is more time consuming and requires a lot of computational power, which ends up being very expensive. Our aim was to develop a pipeline to identify breed-specific SNPs in the mitochondria of cattle breeds, so these SNPs can be used to authenticate branded beef products within the South African beef industry. Identification of SNPs within the mitochondria of the breeds in our study, required less computational power and was less time consuming.

Previous studies that have conducted experiments of SNP discovery in cattle breeds reared in South Africa, have either looked at SNPs found in the entire cattle genome or used genotyped data. (Zwane *et al.*, 2016, 2019; Lashmar *et al.*, 2021). The whole genome sequences in our study were aligned to the latest cattle genome – ARS-UCD 1.2. A study conducted by Rosen *et al.*, (2020) comparing the previous cattle genome – UMD3.1.1 with ARS-UCD 1.2 indicated that ARS-UCD 1.2 showed more continuity accuracy and completeness. Subsequently, more recent whole genome sequencing studies have also used the latest reference cattle genome for alignment (Júnior *et al.*, 2020; Maiorano *et al.*, 2022). To avoid any false alignments, we aligned the entire genomes of the cattle breeds to the reference genome as opposed to aligning only the mitochondria genome of each breed. Mapping to the ARS-UCD 1.2 gave an average mapping percentage of 97% and average coverage of 8.8X. Our average coverage was lower than previous studies (Choi *et al.*, 2015; Zwane *et al.*, 2019; Fernandes Júnior *et al.*, 2020) that had an average of 10,7X, 21X and 18,5X, respectively. However, we were confident that we would be able to identify breed specific SNPs in the mitochondria of the cattle breeds in our study due to the small size of the cattle mitochondria genome. In the past, studies have used mitochondrial DNA to identify genetic variations, because the mitochondrial region have mutations that are five times higher than other genetic material (Mannen *et al.*, 2004; Hsieh *et al.*, 2006). According to Chung, 2013, the mitochondrial genome can be used to trace back the origin of breeds as well as identifying the breed belonging to individual animals. Earlier studies (Troy *et al.*, 2001; Hsieh *et al.*, 2006) that have used mitochondria DNA to search for genetic variants have mainly focused on the *D-Loop* gene, because of the gene's high mutations. It was believed that SNPs identified in the coding region may not be able to clearly elucidate genetic diversity in genetic variations of low frequencies. However, Chung, 2013 recommended the use of the entire mitochondrial region to study the genetic diversity in breeds and also identify individuals within a particular breed. To our knowledge this is the first study in South Africa to identify breed specific SNPs in the mitochondria of cattle breeds. Previous South African studies (Zwane *et al.*, 2019) that have looked at SNP identification in sequenced data, have focused on looking at the whole genome as opposed to just the mitochondria.

5.5.3. Variant detection and annotation

The total number of SNPs that were identified in the mitochondria of 13 breeds in our study were 12 633 after hard filtering. The Brahman breed had the highest number of SNPs followed by the Bonsmara breed and the lowest number of SNPs were found in the Hereford breed. The highest number of SNPs identified in the Brahman breed was expected, as the Brahman breed is a *Bos taurus* breed closely related to the reference genome that was a Hereford (*Bos taurus*) breed. However, to our surprise the Hereford breed individuals used in our study had the lowest number of identified SNPs, even though the reference genome used was from the Hereford breed. The total number of SNPs identified in our study were much higher than those identified by Chung, 2013, who discovered only 742 SNPs after using first generation sequencing on amplified products of whole mitochondria DNA. Our results indicate that NGS can identify a large number of informative SNPs within cattle mitochondrial sequences.

5.5.4. Breed specific SNPs identified

Breed specific SNPs are SNPs that are polymorphic in a single breed, and one of the allele is fixed in other breeds (Pant *et al.*, 2012; Mengistie *et al.*, 2022). Breed specific SNPs that have different alleles fixed in different breeds have a high discriminatory power (Pant *et al.*, 2012), as compared to SNPs that are not specific within a particular breed. Identifying SNPs in the cattle mitochondria genomes led us to further investigate the presence of breed specific SNPs in the 13 cattle breeds within our study. These SNPs can be used to authenticate beef meat products, especially those sold as certified or branded beef products. The authenticity of meat products in global markets has become important due to fraudulent practices that have been reported across the meat industry (Tembe *et al.,* 2018; Omran *et al.*, 2019; Hossain *et al.*, 2021). In order to identify breed specific SNPs, we needed to use SNPs that are the same, therefore, the SNPs used would need to carry the same alleles. Therefore, we selected only homozygous SNPs that were found in each animal after hard filtering. Homozygous SNPs carry the same allele pairs, whereas heterozygous SNPs carry different allele pairs. A total of 125 breed specific homozygous SNPs were identified in 12 of the 13 breeds in our study. The Charolais breed had the lowest number of homozygous SNPs, which could have resulted in not finding any breed specific SNPs in the mitochondria. There is not much evidence of published data on the use of sequenced data to detect the presence of breed specific SNPs in the mitochondria of the Charolais breed. However, there have been reports of Charolais breed specific SNPs being identified using genotypic data. Mengistie *et al.*, 2022 genotyped cattle breeds using the 80K

SNP Bead Cheap and identified 8 903 breed specific SNPs in 37 Charolais breed cattle. The origins of the Charolais breed are not clearly known, however, it is believed to have originated from France around the 800s – 900s. The Charolais breed is exotic to South Africa and has a market share of 4.2% (Bisschoff and Lotriet, 2013).

The highest number of breed specific SNPs in our study were found in the Limousin breed and the lowest were found in the Nguni breed. The Limousin breed had the highest number of homozygous SNPs; therefore, it was expected to find the highest number of breed specific SNPs in this breed. The highest number of breed specific SNPs were found in the *ND5* (29) gene, followed by the *COX3* (22) gene and *16S* rRNA (19) genes. The breeds that had breed specific SNPs in the *ND5* region included the Afrikaner, Beefmaster, Bonsmara, Drankensberger, Hereford, Limousin, Nguni, Santa and Simmentaler breeds. Looking that the *COX3* gene, the breeds that contained breed specific SNPs within that region include Afrikaner, Beefmaster, Brahman, Hereford, Limousin, and Santa. In the *16S* rRNA gene the breeds that had breed specific SNPs within that region were Afrikaner, Boran, Limousin and Simmentaler. The Limousin breed had the highest number of breed specific SNPs in each of the three mitochondrial genes that contained the highest number of breed specific SNPs. We expected the *COX3* and *16S rRNA* genes to be some of the genes that contained higher numbers of breed specific SNPs within our study, as they are amongst some of the mitochondrial genes that are believed to contain a higher number of SNPs (Zhang *et al.*, 2020). Zhang *et al.*, 2020 suggested that there is a high level of breed identification when using complete mitochondrial sequencing. Furthermore, both the *COX3* and *16S rRNA* have been used in several meat species identification studies that include beef products (D'Amato *et al.*, 2013b; Tillmar *et al.*, 2013; Liu *et al.*, 2021; Spychaj *et al.*, 2021). In Chapter 3 of this study, we conducted a multi-locus phylogenetic analysis to find out which mitochondrial gene amongst *16S* rRNA, *12S* rRNA, *COX3* and *ATP6* genes had the highest discriminatory potential. We observed that that the *16S* rRNA gene had the highest discriminatory potential. In Chapter 4 of this study (Chaora *et al.*, 2022) used the *16S* rRNA gene to identify meat species in pure and processed meat samples. The *16S* rRNA gene managed to identify species in pure, artificially mixed, and processed meat samples. Most of the processed meat samples were of beef origin. If we look at the previous results from Chapter 3 and 4 of the current study and the breed-specific SNPs results we observed in this current experiment, we can be confident that the breed-specific SNPs identified in the *16S* rRNA gene can be used to identify the cattle breeds.

Apart from the breed-specific SNPs identified in the *16S* rRNA gene in this study, we can also use the breed-specific SNPs identified in the *COX3* and *ND5* gene to identify the respective cattle breeds that were observed in these genes. Studies that have looked at the use of the *ND5* in cattle breed identification, have discriminated Korean or Japanese cattle breeds from imported breeds. Yoon *et al.,* 2008 managed to discriminate Korean Hanwoo beef from imported beef using the ND5 gene. Sasazaki *et al.,* 2004 discriminated the Japanese Black beef from imported Australian beef using PCR-RFLP. More recently, Kawaguchi *et al.,* 2018 conducted a study that discriminated Japanese Wagyu beef from Australian Wagyu beef using the *ND5* gene. We also identified breed specific SNPs in the *COX1* (4) and *cyt b* genes (7). The breeds that had breed-specific SNPs within the *COX1* gene include Afrikaner, Limousin and Simmentaler. Looking at the *cyt b* gene, the breeds that contained breed-specific SNPs within the gene were Afrikaner, Beefmaster, Bonsmara, Brahman, Drakensberger, Hereford and Santa. The *COX1* and *cyt b* have previously been used in identification of meat adulteration (Cawthorn, Steinman and Hoffman, 2013; Tembe, Mukaratirwa and Zishiri, 2018; Pan *et al.*, 2020; Zhang *et al.*, 2020). The adulteration reported included beef products being mislabeled. Therefore, the breed-specific SNPs observed in the *COX1* and *cyt b* in our study genes can be used to identify the above-mentioned cattle breeds in a meat identification study. Other studies that looked at breed specific SNPs have looked at fewer breeds, but larger numbers of individuals per breed. Czech et al., 2018 looked breed specific SNPs in whole genome sequences of 936 bulls from the Angus, Jersey, Simmentaler, Limousin, Hereford, Brown Swiss and Fleckvieh breeds. The highest breed specific SNPs were found in the Jersey breed (130 070) and the lowest in the Simmentaler breed (197). In an earlier study, Chung, 2013 looked at the presence of SNPs from sequenced amplified mitochondrial PCR products of 40 Korean Native cattle and 113 GenBank sequences from Angus, Japanese Black, Holstein breeds. The study managed to identify two unique nucleotide mitochondrial positions (2536 and 9682) that can distinguish the Japanese Black from other breeds. However, no significant SNPs were observed for the other breeds in the study.

A total of 19 breed specific SNPs were shared in more than one breed from our dataset. The highest number of shared SNPs were found in the Limousin breed, which was expected since the Limousin breed had the highest number of homozygous SNPs. The *ND5* gene had the highest number of breed specific SNPs that were shared, and the lowest number of shared SNPs

were found in the *16S* rRNA gene. Therefore, positions in the *16S* rRNA gene that carry breed specific SNPs for cattle breeds can be used for cattle breed identification, without the risk of errors from SNPs that are shared in more than one breed. We further believe the *ND5* gene would not be the best marker of choice given that it had a considerable number of SNP positions that were shared amongst breeds. The number of shared breed-specific SNPs identified in our study are almost like those identified in the study by Chung, 2013, who found 29 shared SNPs in the mitochondria of Korean Native cattle, Angus, Japanese Black, and Holstein breeds. Contrast to our study, Czech et al., 2018 found 445 breed specific SNPs shared in Angus, Jersey, Simmentaler, Limousin, Hereford, Brown Swiss and Fleckvieh breeds from whole genome sequences. The common breed with shared SNPs in this study was the Simmentaler breed. Weldenegodguad et al., 2019, did not identify breed specific SNPs in their study, however, they discovered 6,2 million shared SNPs in 15 whole genome sequences of Eastern Fincattle, Western Fincattle and Yakutian cattle breeds. In a South African study conducted in indigenous Afrikaner, Drakensberger and Nguni breeds, a total of 10,2 million SNPS were shared in the whole genome of 90 animals. If we were to conduct a network analysis of the breed-specific SNPs we identified within each gene, we believe we would observe an interaction of the breeds within the genes that contained breed-specific SNPs. Furthermore, we would also observe an interaction amongst breeds that shared SNPs and an indication of which SNPs were fixed in a particular breed. We, therefore, recommend network analyses of the breed-specific SNPs identified as a future study, to see the interaction between breeds and the shared and fixed breeds.

5.6. Conclusion

Our investigated breed specific and shared SNPs in the mitochondria of 13 cattle breeds reared in South Africa. The SNPs detected in the study indicate that mitochondria sequences carry genetic tools that can be used to discriminate cattle breeds. Identification of breed specific SNPs in 12 out of the 13 breeds in our study indicated to us that our pipeline and criteria used to identify breed specific SNPs can be used within the breeds in our study and can also be implemented in other breeds. The breed specific SNPs identified provide an understanding of the regions within mitochondrial genes that are unique in each breed. These regions can be used in the authentication of beef meat in the meat industry. Furthermore, the pipeline developed can be implemented to identify breed specific SNPs in other cattle breeds that were

not included in this study. Most of the breed specific SNPs identified were found in the *ND5*, *COX3* and *16S rRNA* mitochondrial genes. These genes have been used in meat species identification including beef meat types. Therefore, positions with breed specific SNPs identified in the above-mentioned genes can be used to develop genetic assays to authenticate branded and breed certified beef meat sold within the meat industry worldwide. The use of the mitochondria genome as opposed to the whole cattle genome was less time consuming and used less computational power for variant detection. For that reason, we believe that mitochondria sequences can be used to identify quality and informative variants in cattle.

CHAPTER SIX

6. General Discussion and Conclusions

6.1. Introduction

This chapter consolidates all the significant and most important outcomes of the thesis and brings to light how the thesis contributes to meat species identification. The horsemeat scandal that occurred in the United States of America in 2013 (Walker, Burns and Burns, 2013), where horsemeat was found in beef products, has led to food fraud and meat species substitution taking precedence in the meat industry. Thereafter, various studies were published indicating that meat fraud not only occurred in Europe, but also in South Africa (Cawthorn, Steinman and Hoffman, 2013; D'Amato *et al.*, 2013a; Tembe, Mukaratirwa and Zishiri, 2018), in Italy (De Battisti *et al.*, 2014; Di Pinto *et al.*, 2015), in China (Cai *et al.*, 2017) and in the United States (Kane and Hellberg, 2016), just mention a few countries. Consumers need to be protected from food fraud not only due to economic gain, but also because it breaks consumer trust and poses health, religious and ethical risks. The meat industry requires methods that can be used to prove the authenticity of products and declared meat species. Deoxyribonucleic acid (DNA) methods are the common methods that have been used in meat species identification in the past, and these include polymerase chain reaction (PCR), restriction fragment length PCR (PFLP), quantitative PCR (qPCR) and droplet digital PCR (ddPCR) (Kumar *et al.*, 2015; Alikord *et al.*, 2018). Unfortunately, these methods are mainly species-specific and not ideal for meat products containing multiple species.

Next Generation Sequencing (NGS) is a technology that can be used to sequence large numbers of unknown and mixed meat species simultaneously in a single run. The use of NGS is becoming extremely popular in food testing studies that include seafood (Giusti *et al.*, 2017), herbs and species (Barbosa *et al.*, 2019) and meat species (Xing *et al.*, 2019; Cottenet *et al.*, 2020; Liu *et al.*, 2021). These studies have demonstrated the food fraud still occurs in the food industry. A more recent study has demonstrated that meat mislabeling is still present in South Africa (Chaora et al., 2022). Therefore, there is need for reliable molecular and bioinformatics tools to be developed for accurate data interpretation. Next Generation Sequencing is believed to be an advanced tool for food authenticity that will become a refence point in the future

(Haynes *et al.*, 2019; Cottenet *et al.*, 2020). The main aim for conducting this study was to evaluate genomics and bioinformatics pipelines that will facilitate utilization of genetic markers and NGS technology to identify and discriminate meat species in the South African meat industry.

6.2. Justification of the three experimental chapters

In the first experimental chapter (Chapter 3) the main objective was to conduct a multi-locus phylogenetic analysis of mammalian species to determine the discrimination power of the *16S rRNA*, *12S rRNA*, *ATP6* and *COX3* mitochondrial genes. Mitochondrial genes are commonly used in meat species identification, since mtDNA contains multiple copies, can withstand harsh conditions, such as heat and pressure, and mtDNA is also capable of discriminating closely related species due to its high rate of revolution (Chen, Liu and Yao, 2010; Kumar *et al.*, 2015). Mitochondrial genes also contain phylogenetic information that differentiate species at an intraspecies and interspecies level (Zhang *et al.*, 2020). It is important to identify the ideal mitochondrial gene that identifies and accurately group closely related species, as well as separate species that are not closely related. Once the ideal gene is identified it can then be used in meat species identification. The phylogenetic analysis included 263 species that comprised of 76 taxonomic families. A phylogenetic analysis was initially performed on each individual gene to see which gene would perform the best. Thereafter, the genes were combined in a phylogenetic analysis to ascertain whether that will improve the performance of the least performing genes.

All four mitochondrial genes managed to separate distantly related species and group closely related species with a common ancestry, however some genes showed some errors in the groupings. We observed 23 clades in the *16S* rRNA and *12S* rRNA individual trees, 21 clades in the *COX3* individual tree and 20 clades in the *ATP6* individual tree. We observed a monophyletic clade formed by the *Bovidae* family (cattle, sheep, goats, antelopes, gazellas, buffalos and Bisons) in all four individual gene phylogeny trees. Within the *16S* rRNA and *ATP6* trees, we observed strong bootstrap values of 100% within the *Bos* species (cattle, aurochs and banteng). However, the *12S* rRNA and *COX3* genes had lower bootstrap values of 99% and 76 – 100%, respectively for the *Bos* species. Furthermore, in the *COX3* individual phylogenetic tree the *Halichoerus grypus* (Grey Seal) species from the *Phocidae* family

grouped with the *Bubalus bubalis* (Water Buffalo) with 100% bootstrap. This indicated the *COX3* gene incorrectly grouped these species. Therefore, the *COX3* gene may not be an ideal gene to use as a universal marker for meat species identification. The antelope species (*Antilope cervicapra* and *Antidorcas marsupialis*), although closed related to the *Bos* species were well separated with a bootstrap value of 100% in the *16S* rRNA, *12S* rRNA and *COX3* genes and a lower value of 99% in the *ATP6* genes. Previous studies (D'Amato et al., 2013a) have reported on game meat being mislabelled as beef meat. Therefore, it is important to use an ideal gene that will be able to differentiate game meat from beef meat in species identification. The *Capra* species (goat) formed a monophyletic with the *Ovis* species (sheep) in the *16S* rRNA, *12S* rRNA and *COX3* genes, with bootstrap values of 100%, 85% and 84%, respectively. However, the *Capra hircus* species did not group together with the *Ovis* species in the *ATP6* tree, despite their close relation. Indicating failure of the *ATP6* gene to accurately group certain closely related species. Recent studies (Soman *et al.*, 2020) have demonstrated that mutton (sheep) meat is often mislabelled as chevon (goat) meat for economic gain, since mutton meat costs more that chevon meat. Therefore, a gene that can accurately discriminate chevon meat from mutton meat is needed for the identification of these two meat types. The *Suidae* (Warthogs and Pigs) formed a monophyletic clade separate from the *Bos* species in all the four individual phylogenetic trees. This indicated that all four genes can be used in meat species identification of samples containing beef and pork meat. Adulteration of beef meat products with pork meat has been reported (Cawthorn *et al.,* 2013; Ha *et al.*, 2017; Tembe *et al.,*2018; Chaora *et al.*, 2022). This is usually done for economic gain, since pork is cheaper than beef. We observed that the bat (*Rhinolophidae, Phyllostomidae, Vespertilionidae and Hipposideridae*) species formed a monophyletic clade in the *16S* rRNA and *ATP6* individual genes. However, the *12S* rRNA and *COX3* genes, failed to group all the bat species in one clade in their respective phylogenetic trees. This further indicates that the *COX3* and *12S* rRNA genes may not be ideal to use as a universal marker, but more for species-specific studies.

Looking further in Chapter 3, the *Suidae* family (pig, warthog and wild boar) grouped together and formed a monophyletic clade with *Pecari tajacu* (Collard Peccary) in all four genes. These species are of the same order *Artiodactyla*. The *Cricetidae* (vole) and *Muridae* (mouse and rat) families formed well defined monophyletic clades in the *16S* rRNA, *12S* rRNA and *COX3* genes, although the *16S* rRNA gene had the highest bootstrap values of 100%. The bootstrap values in the *12S* rRNA and *COX3* genes were 99% and 61%, respectively. However, there

was no definitive clade formed between the *Cricetidae* and *Muridae* families in the *ATP6* gene. The *Talpidae* and *Soricidae* (Moles) species formed a monophyletic in the *16S* rRNA, *12S* rRNA and *COX3* gene individual trees. However, in the *ATP6* gene individual tree, the *Talpidae* species did not group together with the *Soricidae* species. These two scenarios suggest that the *ATP6* gene may not be an ideal universal marker in meat species identification. Given these results, we recommended the *16S* rRNA gene as the most ideal gene to use as a universal marker for meat species identification. The *16S* rRNA gene had higher bootstrap values in all identified clades and accurately grouped closely related species and separated species that we not closely related. We paired the *12S* rRNA, *COX3* and *ATP6* genes with the *16S* rRNA gene to see if there would be an improvement in performance. When we paired the *16S* rRNA gene with each other the other genes, there were some improvements in the bootstrap values that had been lower in other individual genes, as well as an improvement in the accuracy of some groupings. For instance, when the *COX3* and *12S* rRNA gene were each paired with the *16S* rRNA, there was an increase to 100% in bootstrap values in the *Bos* species grouping. We observed that in the *16S* rRNA and *COX3* combined phylogenetic tree the *Halichoerus grypus* (Grey Seal) species did not group with the with the *Bubalus bubalis* (Water Buffalo), but instead grouped with the other seal species. All the rat species (*Rhinolophidae, Phyllostomidae, Vespertilionidae and Hipposideridae*) that did not manage to group together in the *12S* rRNA and *COX3* individual phylogenetic trees, managed to group together all the species together when they were paired with the *16S* rRNA gene. When the *ATP6* gene was combined with the *16S* gene, the *Capra* family (goat) grouped with the *Ovis* family (sheep). Whereas when a phylogeny analysis was performed on the *ATP6* gene alone, these two families did not group together. Furthermore, when the *ATP6* tree was combined with the *16S* rRNA gene, all the mole (*Talpidae* and *Soricidae*) species managed to group together, as opposed to the *ATP6* individual tree. We combined all four mitochondrial genes in a phylogeny and there was a definite improvement in bootstrap values in the same clades that were formed in the *16S* rRNA individual tree. Most clades had a bootstrap value of 100%. However, a multiplex meat species identification experiment containing four genes will not be cost effective. These results were evidence of the strong discriminatory potential of the *16S* rRNA gene. Overall, the ideal marker for meat species identification should be one that is variable enough to identify ang group closely related species and accurately separate species that are not closely related. From the results we observed in Chapter 3, we recommended the *16S* rRNA gene to be used in our next experimental chapter involving meat species identification of processed meat samples.

Our phylogeny results indicated that the *16S* rRNA gene had higher discriminatory potential when compared to *12S* rRNA, *COX3* and *ATP6* genes. However, we had no knowledge of whether the *16S* rRNA gene would be an ideal universal marker for meat species identification and this is the question that Chapter 4 answered.

In experimental Chapter 4, the main objective was to develop a universal and robust diagnostic molecular and bioinformatics pipeline that can utilize the mitochondrial 16S ribosomal RNA (rRNA) barcoding gene to identify processed meat product mislabelling/contamination using NGS. The pipeline involved DNA extraction, *16S* rRNA PCR amplification, NGS sequencing of the PCR amplicons and bioinformatics analysis of the sequences. To determine whether the *16S* rRNA mitochondrial gene was an ideal marker for meat species identification, we initially used artificially mixed pure meat to test the sensitivity and specificity of the molecular and bioinformatics pipeline we developed and to see whether the pipeline would be able to identify meat species that are in mixed samples. The ratio mixtures used were 1 (50%) :1 (50%) and 9 (90%): 1 (10%). Looking at the artificially mixed samples, we observed that the pipeline managed to identify all the species that were included in 1:1 ratio mixture. These ratio mixtures included Pork (50%) : Beef (50%), Mutton (50%) : Chevon (50%), Chicken (50%) : Turkey (50%), Ostrich (50%) : Duck (50%) and Beef (50%) : Kangaroo (50%) meat types. Similarly, the meat species in the 9 :1 ratio were also identified. The ratio mixtures included Pork (90%) : Beef (10%), Beef (90%) : Pork (10%), Mutton (90%) : Chevon (10%), Chevon (90%) : Mutton (10%), Chicken (90%) : Duck (10%), Duck (90%) : Chicken (10%), Ostrich (90%) : Duck (10%) and Duck (90%) : Ostrich (10%). After observing positive results from the artificially mixed samples, we collected processed meat samples across retail outlets to test whether the pipeline could identify meat species in environmental samples. The processed meat samples included biltong, mined meat, burger patties and raw sausages. These samples were either labelled which species they are from or not labelled. For the biltong meat, all samples that were not labelled which species they are from were predominantly from beef. However, three of the biltong samples that were labelled as beef, contained contamination (2 – 5%) from *Sus* (pork) and *Ovis* (lamb/mutton) species were observed. Similarly, three of the mince samples that were not labelled which species they are from were predominantly from beef and contained traces of *Sus* and *Ovis* (3 – 16%) species. However, one of the unlabelled mince samples was evidently from pork, as it predominantly contained *Sus* (97%) species with traces of *Bos* and *Ovis* (3%) species. For those samples labelled as beef mince, two of the samples

were contaminated and once again the contaminate were from the *Sus* and *Ovis* (2 – 6%) species. Two of the unlabelled patty samples had traces of contamination from *Sus* (7%) species. It was worrying to observe that the five samples labelled as beef patty had high contamination from the *Ovis* species, with percentages as high as 34%. Sausage samples had the highest number of contaminated samples, when compared to all the other processed meat samples. Seven samples that were not specified which species they are from where contaminated with Sus and Ovis species, with percentages ranging from 3 – 38%. A similar trend was observed in samples that were labelled as beef sausages, where we found contamination from *Sus* and *Ovis* species, with percentages as high as 78%. The results in Chapter four show that the *16S* rRNA gene can be used as a universal marker for meat species identification. These results also demonstrate that that the molecular and bioinformatics pipeline we developed can identify meat species in mixed samples with no prior knowledge of which species are contained in a particular sample. Judging from our results it is also evident that there is still mislabelling in the South African meat market, despite previous studies (Cawthorn, Steinman and Hoffman, 2013; D'Amato *et al.*, 2013a; Tembe, Mukaratirwa and Zishiri, 2018) conducted in South Africa. These reports should have been taken as a warning message by unscrupulous traders that want to gain economically, through mislabelling of processed meat products.

After observing that meat species mislabeling is still present in processed meat samples sold in South Africa and that most of the contamination was found in beef samples, we were interested in finding out whether the mislabeling of meat was also present in beef meat samples with specified or branded cattle breeds on their labels. Consumers nowadays have a preference to consuming branded or certified beef products from cattle breeds. The Angus breed was initially certified as *Certified Angus Breed* in the USA in 1978 (Bass, 2016). To our knowledge there are no published scholarly articles on certified Angus beef in South Africa. However, branded Angus beef meat products are sold in South Africa. The Sparta Beef company sells branded Angus and Wagyu beef products that are sold in retailers and restaurants around South Africa (https://www.sparta.co.za/, no date). The Wagyu breed was certified in South Africa as *Certified South African Wagyu Breed* that is fully traceable with no hormones and growth stimulants allowed (Coleman, 2017). Other breed specific beef products that are sold in South Africa are from the Afrikaner and Bonsmara breeds (Brits, 2017; Lombard, Van Zyl and Beelders, 2017). Before we could conduct an experiment on identification of cattle breeds on

specified or branded beef meat samples, we initially needed to develop a pipeline that can authenticate cattle breeds for traceability and assignment of the breeds. The main objective for experimental Chapter 5 was to investigate the presence of breed-specific SNPs using the entire mitochondria of 13 European and Indigenous cattle breeds reared in South Africa for use in breed assignment and traceability. The identified breed-specific SNPs can then be used to authenticate branded beef products sold in South Africa. Furthermore, the pipeline developed to identify breed-specific SNPs can be implemented for use in other breeds that were not included in our study. The cattle breeds used in our study were Afrikaner, Beefmaster, Boran, Charolais, Hereford, Nguni, Simbra, Bonsmara, Brahman, Drakensberger, Limousin, Santa and Simmentaler.

Whole genome sequencing was performed on the cattle breeds, however, only the mitochondrial genome was used to identify SNPs. We managed to identify 12 633 SNPS in the mitochondria of all breeds and of these 125 were breed specific SNPs. Unfortunately, we did not manage to identify breed-specific SNPs in the Charolais breed. The highest number of SNPs were observed in the exotic Brahman breed and the lowest in the indigenous Nguni breed. The Limousin breed had the highest breed-specific SNPs and once again the Nguni breed had the lowest. It was expected that the exotic breeds would have the highest number of SNPs, because the reference breed used for mapping of the sequences was sourced from the exotic Hereford breed. We were pleased to identify breed-specific SNPs in the mitochondria of all the breeds (except for Charolais) in our study, because this was evident that the pipeline we developed worked and can be implemented for us in other breeds. The identified breed-specific SNPs also gave us an insight on which mitochondrial genes contained breed specific SNPs. The *ND5* gene had the highest number of breed specific SNPs, followed by the *COX3* and *16S* rRNA gene with 22 and 19 breed-specific SNPs, respectively. The *ATP6* and *12S* rRNA genes that were part of the genes we analyzed in Chapter 3 also contained 8 and 3 breed specific SNPs, respectively. Twenty-nine of the identified breed-specific SNPs were shared amongst the breeds and *ND5* gene had the highest number, with 8 SNPs shared amongst breeds. The *COX3* gene had 3 shared breed-specific SNPs and the *16S* rRNA gene only had one shared breed-specific SNP. There were no shared breed-specific SNPs in the *12S* rRNA and *ATP6* genes. Shared breed specific SNPs are not ideal to use in cattle breed assignment and traceability since the breed-specific SNPs will be specific in more than one breed. We would recommend using the breed-specific SNPs identified in the *16S* rRNA gene to breed assignment

and traceability, as there would be less chances of errors in breed identification. Therefore, the *16S* rRNA gene is an ideal gene to use for identification of processed meat samples, as well as cattle breed identification. The *16S* rRNA gene performed the best in the phylogenetic analysis in Chapter 3, managed to identify meat species in artificially mixed and processed samples in Chapter 4 and finally, contained breed-specific SNPs in cattle breeds with minimal shared SNPs amongst breeds in Chapter 5. The mitochondrial positions within the 1*6S* rRNA gene that contained breed-specific SNPs can be used to design breed-specific primers for cattle that can be used in PCR, sequencing of the PCR amplicons and mapping of the sequences to a database of the desired breed. This information can be used to authenticate the particular breed in question.

6.3. General conclusion and recommendations

The current study demonstrated that NGS can be used as a tool in phylogeny analyses, meat species identification of processed meat samples, as well as identification of breed-specific SNPs in cattle breeds. The *16S* rRNA mitochondrial gene demonstrated exceptional discriminatory potential in mammalian species when compared to the *12S* rRNA, *COX3* and *ATP6* mitochondrial genes and can therefore, be used as a universal marker. Using the *16S* rRNA gene we developed a molecular and bioinformatics pipeline that managed to identify meat species in artificially mixed pure meat samples and processed meat samples in the South African market. From that part of the study, it was evident that meat species mislabeling is still present in the South African market. We further went on to identify breed specific SNPs in twelve cattle breeds that are reared in South Africa. Most of the breed specific SNPs were found in the *ND5*, *COX3* and *16S* rRNA genes. These SNPs can be used to authenticate cattle breeds.

6.4. Study Limitations and future studies

Although we identified some cases of meat species mislabeling/substitution in some of the processed meat samples we analyzed in Chapter 4, one of the limitations we faced was that some meat samples were not specified which species they are from. Therefore, we could not determine if a sample was mislabeled or not when we identified a mixture of species in those

samples. For instance, one of the unlabeled patty samples contained 59% *Sus* species and 40% *Bos* species. Since the patty sample was not labeled which species it is from, there is no way of determining whether this sample was intentionally mislabeled or substituted. We also experienced some percentage deviations from the ratio mixtures of the artificially mixed pure meat samples we analyzed in Chapter 4. For instance, in the 9 (90%) :1 (10%) ratio mixture of Ostrich: Duck, Ostrich reads with an average fold of 81,2% for the *Struthio* genus and Duck had reads with an average fold of 18,5% for the *Anas* genus. The presence of several copies of mitochondrial DNA in a single cell can lead to either an underestimation (−70%) or overestimation (+160%) of species' DNA content (Flores-Munguia *et al.,* 2000). Mitochondrial genes are ideal for species identification, especially in species that are closely related due to mitochondria having a mutation rate that is 10-fold higher than that of nuclear genes. However, in studies where you would like to quantify the percentage of species present in a meat sample, the use of mitochondrial genes may not be ideal. In such instances it would be better to use nuclear genes. We were pleased to identify breed-specific SNPs in the mitochondria of 12 breeds that are reared in South Africa (Chapter 5), however we were not able to include other cattle breeds that are sold as branded breeds in South Africa, such as the Angus and Wagyu breeds. Identifying breed-specific SNPs in these breeds will be beneficial in developing a pipeline to authenticate all branded breeds sold in South Africa, that include Afrikaner, Bonsmara, Angus and Wagyu breeds. The number of animals that we used for the identification of breed-specific SNPs was a total of 42, with the number of animals per breed ranging from 2 – 4. A larger population size could have resulted in identification of more breed-specific SNPs and possibly identifying breed specific SNPs in the Charolais breed.

Some of the future studies that could be considered in meat species identification of processed meat is the use of nuclear genes, to quantify the amount of DNA present in the sample. Previous studies (Zhang *et al.*, 2020) have indicated that there is a difference in binding efficiency of the universal primers for different species, resulting in a difference in the amplification efficiency and, therefore, leading to a large degree of error in quantitative analysis. The breed specific SNPs in cattle can be used to design primers that can identify cattle breeds through PCR. Furthermore, these results can be validated through sequencing of the PCR amplicons and mapping the sequences to a known database of the breed. The breed-specific SNPs identified in the indigenous breeds of South Africa (Bonsmara, Afrikaner, Nguni and Drakensberger) can also be used to develop custom made SNP chips for the cattle population in South Africa

7. REFERENCES

Ali, M. E. *et al.* (2018) 'Multiplex polymerase chain reaction-restriction fragment length polymorphism assay discriminates of rabbit, rat and squirrel meat in frankfurter products', *Food Control*, 84, pp. 148–158.

Alikord, M. *et al.* (2018) 'Species identification and animal authentication in meat products: a review', *Journal of Food Measurement and Characterization*, 12(1), pp. 145–155.

Alkan, C. *et al.* (2009) 'Personalized copy number and segmental duplication maps using next-generation sequencing', *Nature genetics*, 41(10), p. 1061.

Amaral, J. S. *et al.* (2014) 'Authentication of a traditional game meat sausage (Alheira) by species-specific PCR assays to detect hare, rabbit, red deer, pork and cow meats', *Food Research International*, 60, pp. 140–145.

Andrews, S. (2017) 'FastQC: a quality control tool for high throughput sequence data. 2010'.

Van der Auwera, G. A. *et al.* (2013) 'From FastQ data to high-confidence variant calls: the genome analysis toolkit best practices pipeline', *Current protocols in bioinformatics*, 43(1), pp. 10–11.

Ayuso, R. *et al.* (1999) 'IgE antibody response to vertebrate meat proteins including tropomyosin', *Annals of Allergy, Asthma & Immunology*, 83(5), pp. 399–405.

Ballin, N. Z. (2010) 'Authentication of meat and meat products', *Meat science*, 86(3), pp. 577–587.

Barbosa, C. *et al.* (2019) 'Study on commercial spice and herb products using next-generation sequencing (NGS)', *Journal of AOAC International*, 102(2), pp. 369–375.

Bartlett, S. E. and Davidson, W. S. (1991) 'Identification of Thunnus tuna species by the

polymerase chain reaction and direct sequence analysis of their mitochondrial cytochrome b genes', *Canadian Journal of Fisheries and Aquatic Sciences*, 48(2), pp. 309–317.

Bass, P. D. (2016) 'The scientific basis of the Certified Angus Beef brand carcass specifications'.

De Battisti, C. *et al.* (2014) 'Pyrosequencing as a tool for rapid fish species identification and commercial fraud detection', *Journal of Agricultural and Food Chemistry*, 62(1), pp. 198–205. doi: 10.1021/jf403545m.

Baum, D. (2008) 'Reading a phylogenetic tree: the meaning of monophyletic groups', *Nature Education*, 1(1), p. 190.

Beriain, M. J. *et al.* (2021) 'Physicochemical and Sensory Assessments in Spain and United States of PGI-Certified Ternera de Navarra vs. Certified Angus Beef', *Foods*, 10(7), p. 1474.

Bisschoff, C. and Lotriet, R. (2013) 'The Drakensberger as competitive breed of cattle in the South African beef industry', in *19th International Farm Management Congress*, pp. 39–49.

Böhme, K. *et al.* (2019) 'Review of recent DNA-based methods for main food-authentication topics', *Journal of Agricultural and Food Chemistry*, 67(14), pp. 3854–3864.

Bolger, A. M., Lohse, M. and Usadel, B. (2014) 'Trimmomatic: a flexible trimmer for Illumina sequence data', *Bioinformatics*, 30(15), pp. 2114–2120.

Bottaro, M. *et al.* (2014) 'Detection of mislabeling in packaged chicken sausages by PCR', *Albanian Journal of Agricultural Sciences*, p. 455.

Bottero, M. T. and Dalmasso, A. (2011) 'Animal species identification in food products: Evolution of biomolecular methods', *The Veterinary Journal*, 190(1), pp. 34–38.

Brits, M. (2017) 'Good, better, branded beef', *Dié Rooi Ras*, 1(1), pp. 63–65.

Bushnell, B. (2014) *BBMap: a fast, accurate, splice-aware aligner*. Lawrence Berkeley

National Lab.(LBNL), Berkeley, CA (United States).

Cai, Y. *et al.* (2017) 'Detection and quantification of beef and pork materials in meat products by duplex droplet digital PCR', *PLoS ONE*, 12(8), pp. 1–12. doi: 10.1371/journal.pone.0181949.

Carrera, E. *et al.* (2000) 'Identification of smoked Atlantic salmon (Salmo salar) and rainbow trout (Oncorhynchus mykiss) using PCR-restriction fragment length polymorphism of the p53 gene', *Journal of AOAC International*, 83(2), pp. 341–346.

Cawthorn, D. M., Steinman, H. A. and Hoffman, L. C. (2013) 'A high incidence of species substitution and mislabelling detected in meat products sold in South Africa', *Food Control*, 32(2), pp. 440–449. doi: 10.1016/j.foodcont.2013.01.008.

Cermakova, E. *et al.* (2023) 'Identification of Fish Species and Targeted Genetic Modifications Based on DNA Analysis: State of the Art', *Foods*, 12(1), p. 228.

Chaora, N. S. *et al.* (2022) 'A 16S Next Generation Sequencing Based Molecular and Bioinformatics Pipeline to Identify Processed Meat Products Contamination and Mislabelling', *Animals*, 12(4), p. 416.

Chen, S.-Y., Liu, Y.-P. and Yao, Y.-G. (2010) 'Species authentication of commercial beef jerky based on PCR-RFLP analysis of the mitochondrial 12S rRNA gene', *Journal of Genetics and Genomics*, 37(11), pp. 763–769.

Cheong, H. S. *et al.* (2013) 'Development of discrimination SNP markers for Hanwoo (Korean native cattle)', *Meat science*, 94(3), pp. 355–359.

Cho, A.-R., Dong, H.-J. and Cho, S. (2014) 'Meat species identification using loop-mediated isothermal amplification assay targeting species-specific mitochondrial DNA', *Korean Journal for Food Science of Animal Resources*, 34(6), p. 799.

Choi, J.-W. *et al.* (2014) 'Whole-genome analyses of Korean native and Holstein cattle breeds

by massively parallel sequencing', *PloS one*, 9(7), p. e101127.

Choi, J.-W. *et al.* (2015) 'Whole-genome resequencing analysis of Hanwoo and Yanbian cattle to identify genome-wide SNPs and signatures of selection', *Molecules and cells*, 38(5), pp. 466–473.

Chung, H. (2013) 'Phylogenetic analysis and characterization of mitochondrial DNA for Korean native cattle'.

Cingolani, P. *et al.* (2012) 'A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3', *Fly*, 6(2), pp. 80–92.

Coleman, A. (2017) 'Wagyu breed sets out to revolutionise SA beef industry', *Farmer's Weekly*, 2017(17048), pp. 34–35.

Cottenet, G. *et al.* (2020) 'Evaluation and application of a next generation sequencing approach for meat species identification', *Food Control*, 110, p. 107003.

Czech, B. *et al.* (2018) 'Identification and annotation of breed-specific single nucleotide polymorphisms in Bos taurus genomes', *PLoS One*, 13(6), p. e0198419.

D'Amato, M. E. *et al.* (2013a) 'Where is the game? Wild meat products authentication in South Africa: a case study', *Investigative genetics*, 4(1), p. 6.

D'Amato, M. E. *et al.* (2013b) 'Where is the game? Wild meat products authentication in South Africa: A case study', *Investigative Genetics*, 4(1), p. 1. doi: 10.1186/2041-2223-4-6.

DAFF (2017) *The Profile of the South African Mutton Value Chain*. Pretoria, South Africa.
DAFF (2018a) *A Profile of the South African Beef Market Value Chain*. Pretoria, South Africa.

DAFF (2018b) *A Profile of the South African Ostrich Market Value Chain*. Available at: https://www.nda.agric.za/doaDev/sideMenu/Marketing/Annual Publications/Commodity

Profiles/Ostrich Market Value Chain Profile 2018.pdf.

DAFF (2018c) *A Profile of the South African Ostrich Market Value Chain*. Pretoria, South Africa.

DAFF (2019) *A profile of the South African Beef Market Value Chain*. Pretoria, South Africa.

DAFF (2020a) *A Profile of the South Africa Chevon Market Value Chain*. Pretoria, South Africa.

DAFF (2020b) *A profile of the South African broiler market value chain*. Pretoria, South Africa.

DAFF (2020c) *A Profile of the South African Mutton Market Value Chain*. Pretoria, South Africa. Available at: https://www.daff.gov.za.

DAFF (2020d) *A profile of the South African Pork Market Value Chain*. Pretoria, South Africa. Available at: https://www.daff.gov.za.

DAFF (2020e) *The Broiler Industry. Broilers and Eggs*. Available at: https://www.daff.gov.za/daffweb3/Portals/0/General Publications/Agricultural Marketing Extension Training Paper No.9 Broilers and Eggs.pdf.

Dai, Z. *et al.* (2015) 'Species authentication of common meat based on PCR analysis of the mitochondrial COI gene', *Applied biochemistry and biotechnology*, 176(6), pp. 1770–1780.

Dalmasso, A. *et al.* (2004) 'A multiplex PCR assay for the identification of animal species in feedstuffs', *Molecular and cellular probes*, 18(2), pp. 81–87.

Dalrrd (2021) *Abattoir list all provinces*. Available at: http://www.dalrrd.gov.za.

Deakin, J. E. (2013) 'Genome Sequence of an Australian Kangaroo, Macropus eugenii', *eLS*.

Dimauro, C. *et al.* (2013) 'Use of the canonical discriminant analysis to select SNP markers

for bovine breed assignment and traceability purposes', *Animal genetics*, 44(4), pp. 377–382.

DoH (2010) 'Foodstuffs, Cosmetics and Disinfectants Act, 1972 (Act 54 of 1972), regulations relating to the labelling and advertising of foodstuffs (R. 146/2010)', *South Africa: Government Printing Offices.*

Doosti, A., Ghasemi Dehkordi, P. and Rahimi, E. (2014) 'Molecular assay to fraud identification of meat products', *Journal of Food Science and Technology*, 51(1), pp. 148–152. doi: 10.1007/s13197-011-0456-3.

Drummond, L., Álvarez, C. and Mullen, A. M. (2019) 'Proteins Recovery From Meat Processing Coproducts', *Sustainable Meat Production and Processing*, pp. 69–83. doi: 10.1016/b978-0-12-814874-7.00004-3.

DTI (2009) 'Consumer Protection Act (Act no. 68 of 2008) (R. 467/2009)', *South Africa: Government Printing Offices.*

Dupuis, Julius; Roe, Amanda; and Sperling, F. (no date) *Multi-locus species delimitation in closely related animals and fungi: one marker is not enough | Enhanced Reader*.

Edea, Z. *et al.* (2013) 'Genetic diversity, population structure and relationships in indigenous cattle populations of Ethiopia and Korean Hanwoo breeds using SNP markers', *Frontiers in Genetics*, 4, p. 35.

Elsik, C. G., Tellam, R. L. and Worley, K. C. (2009) 'The genome sequence of taurine cattle: a window to ruminant biology and evolution', *Science*, 324(5926), pp. 522–528.

Erasmus, D. (2018) 'A bullish outlook for Wagyu in SA', *Farmer's Weekly*, 2018(18027), pp. 44–46.

Fajardo, V. *et al.* (2008) 'Differentiation of European wild boar (Sus scrofa scrofa) and domestic swine (Sus scrofa domestica) meats by PCR analysis targeting the mitochondrial D-loop and the nuclear melanocortin receptor 1 (MC1R) genes', *Meat Science*, 78(3), pp. 314–

322.

Fang, X. and Zhang, C. (2016) ‘Detection of adulterated murine components in meat products by TaqMan© real-time PCR’, *Food chemistry*, 192, pp. 485–490.

Farag, M. R. *et al.* (2015) ‘Identification of different animal species in meat and meat products: trends and advances’, *Adv. Anim. Vet. Sci*, 3(6), pp. 334–346.

Fernandes Júnior, G. A. *et al.* (2020) ‘Whole-genome sequencing provides new insights into genetic mechanisms of tropical adaptation in Nellore (Bos primigenius indicus)’, *Scientific reports*, 10(1), pp. 1–7.

Floren, C. *et al.* (2015) ‘Species identification and quantification in meat and meat products using droplet digital PCR (ddPCR)’, *Food Chemistry*, 173, pp. 1054–1058. doi: 10.1016/j.foodchem.2014.10.138.

Flores-Munguia, M. E., Bermudez-Almada, M. C. and Vázquez-Moreno, L. (2000) ‘A research note: Detection of adulteration in processed traditional meat products’, *Journal of Muscle Foods*, 11(4), pp. 319–325.

Folmer, O. *et al.* (no date) ‘DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates Mol Mar Biol Biotechnol. 1994; 3: 294–9’. PMID.

Gibbs, R. *et al.* (2002) ‘Bovine genomic sequencing initiative’.

Giromini, C. and Givens, D. I. (2022) ‘Benefits and Risks Associated with Meat Consumption during Key Life Processes and in Relation to the Risk of Chronic Diseases’, *Foods*, 11(14), p. 2063.

Giusti, A. *et al.* (2017) ‘Seafood identification in multispecies products: assessment of 16SrRNA, cytb, and COI Universal Primers’ efficiency as a preliminary analytical step for setting up metabarcoding next-generation sequencing techniques’, *Journal of agricultural and*

food chemistry, 65(13), pp. 2902–2912.

Global Data (2021) *South Africa Meat Market Analysis and Forecasts, 2015-2025 – Analyzing Product Categories and Segments, Distribution Channel, Competitive Landscape, Packaging and Consumer Segmentation*. Available at: https://www.globaldata.com/store/report/south-africa-meat-market-analysis/ (Accessed: 11 January 2023).

Gorski, I. *et al.* (2016) 'Nyama Choma culture: Implications of increased red meat and alcohol consumption in East Africa', *Journal of Sustainable Development*, 9(6), p. 96.

Gray, C. L., Van Zyl, A. and Strauss, L. (2016) '"Midnight anaphylaxis" to red meat in patients with alpha-gal sensitisation: a recent discovery in the food allergy world and a case report from South Africa: guest review', *Current Allergy & Clinical Immunology*, 29(2), pp. 102–104.

Gupta, N. and Verma, V. K. (2019) 'Next-generation sequencing and its application: empowering in public health beyond reality', *Microbial Technology for the Welfare of Society*, pp. 313–341.

Ha, J. *et al.* (2017a) 'Identification of pork adulteration in processed meat products using the developed mitochondrial DNA-based primers', *Korean Journal for Food Science of Animal Resources*, 37(3), pp. 464–468. doi: 10.5851/kosfa.2017.37.3.464.

Ha, J. *et al.* (2017b) 'Identification of pork adulteration in processed meat products using the developed mitochondrial DNA-based primers', *Korean Journal for Food Science of Animal Resources*, 37(3), p. 464.

Han, F. *et al.* (2020) 'Detection of beef adulterated with pork using a low-cost electronic nose based on colorimetric sensors', *Foods*, 9(2). doi: 10.3390/foods9020193.

Haynes, E. *et al.* (2019) 'The future of NGS (Next Generation Sequencing) analysis in testing food authenticity', *Food Control*, 101, pp. 134–143.

He, C. *et al.* (2022) 'Detection and Quantification of Adulterated Beef and Mutton Products by

Multiplex Droplet Digital PCR', *Foods*, 11(19), p. 3034.

He, H. *et al.* (2018) 'Molecular Authentication of Meats from Three Terrestrial Birds Based on Pcr-Rflp Analysis of the Mitochondrial 12S rRNA Gene', *Brazilian Journal of Poultry Science*, 20, pp. 651–656.

Horreo, J. L. *et al.* (2013) 'Universal primers for species authentication of animal foodstuff in a single polymerase chain reaction', *Journal of the Science of Food and Agriculture*, 93(2), pp. 354–361.

Hossain, A. *et al.* (2021) 'Detection of species adulteration in meat products and Mozzarella-type cheeses using duplex PCR of mitochondrial cyt b gene: A food safety concern in Bangladesh', *Food Chemistry: Molecular Sciences*, 2, p. 100017.

Hsieh, H. M. *et al.* (2006) 'Species identification of Kachuga tecta using the cytochrome b gene', *Journal of Forensic Sciences*, 51(1), pp. 52–56. doi: 10.1111/j.1556-4029.2005.00004.x.

https://www.sparta.co.za/ (no date) *Sparta The Beef Co.* Available at: https://www.sparta.co.za/our-brands/ (Accessed: 9 January 2023).

Islam, A. *et al.* (2021) 'Meat origin differentiation by polymerase chain reaction-restriction fragment length polymorphism', *International Journal of Food Properties*, 24(1), pp. 1022–1033.

Itoh, A. *et al.* (2010) 'Comparison of statistical tests for habitat associations in tropical forests: a case study of sympatric dipterocarp trees in a Bornean forest', *Forest ecology and management*, 259(3), pp. 323–332.

Javonillo, R. *et al.* (2010) 'Relationships among major lineages of characid fishes (Teleostei: Ostariophysi: Characiformes), based on molecular sequence data', *Molecular Phylogenetics and Evolution*, 54(2), pp. 498–511.

Kane, D. E. and Hellberg, R. S. (2016) 'Identification of species in ground meat products sold on the US commercial market using DNA-based methods', *Food Control*, 59, pp. 158–163.

Kesmen, Z. *et al.* (2012) 'Detection of chicken and turkey meat in meat mixtures by using Real-Time PCR Assays', *Journal of Food Science*, 77(2), pp. C167–C173.

Klinbunga, S. *et al.* (2003) 'Genetic diversity and molecular markers of the tropical abalone (Haliotis asinina) in Thailand', *marine biotechnology*, 5(5), pp. 505–517.

Koh, B.-R.-D. *et al.* (2011) 'Development of species-specific multiplex PCR assays of mitochondrial 12S rRNA and 16S rRNA for the identification of animal species', *Korean Journal of Veterinary Service*, 34(4), pp. 417–428.

Kumar, A. *et al.* (2015) 'Identification of species origin of meat and meat products on the DNA basis: a review', *Critical reviews in food science and nutrition*, 55(10), pp. 1340–1351.

Lashmar, Simon Frederick *et al.* (2021) 'A Within-and Across-country Assessment of the Genomic Diversity and Autozygosity of South African and Eswatini Nguni Cattle'.

Lashmar, Simon F *et al.* (2021) 'Assessing single-nucleotide polymorphism selection methods for the development of a low-density panel optimized for imputation in South African Drakensberger beef cattle', *Journal of Animal Science*, 99(7), p. skab118.

Li, H. *et al.* (2009) 'The Sequence Alignment/Map format and SAMtools', *BIOINFORMATICS APPLICATIONS NOTE*, 25(16), pp. 2078–2079. doi: 10.1093/bioinformatics/btp352.

Li, H. and Durbin, R. (2009) 'Fast and accurate short read alignment with Burrows–Wheeler transform', *bioinformatics*, 25(14), pp. 1754–1760.

Li, Y. *et al.* (2020) 'Comparative review and the recent progress in detection technologies of meat product adulteration', *Comprehensive Reviews in Food Science and Food Safety*, (May). doi: 10.1111/1541-4337.12579.

Liao, X. *et al.* (2013) 'Whole genome sequencing of Gir cattle for identifying polymorphisms and loci under selection', *Genome*, 56(10), pp. 592–598.

Lim, D. K. *et al.* (2017) 'Combination of mass spectrometry-based targeted lipidomics and supervised machine learning algorithms in detecting adulterated admixtures of white rice', *Food Research International*, 100, pp. 814–821.

Lin, C.-C., Tang, P.-C. and Chiang, H.-I. (2019) 'Development of RAPD-PCR assay for identifying Holstein, Angus, and Taiwan Yellow Cattle for meat adulteration detection', *Food Science and Biotechnology*, 28(6), pp. 1769–1777.

Liu, X. *et al.* (2021) 'Application of Next-Generation Sequencing Technology Based on Single Gene Locus in Species Identification of Mixed Meat Products', *Journal of Food Quality*, 2021.

Lombard, W. A., Van Zyl, J. H. and Beelders, T. R. (2017) 'Testing red meat consumer preferences', *FarmBiz*, 3(10), pp. 8–11.

Lu, D. *et al.* (2021) 'Prospects and challenges of using electrochemical immunosensors as an alternative detection method for SARS-CoV-2 wastewater-based epidemiology', *Science of The Total Environment*, 777, p. 146239.

Luo, A. *et al.* (2011) 'Potential efficacy of mitochondrial genes for animal DNA barcoding: a case study using eutherian mammals', *BMC genomics*, 12(1), pp. 1–13.

Maiorano, A. M. *et al.* (2022) 'Signatures of selection in Nelore cattle revealed by whole-genome sequencing data', *Genomics*, 114(2), p. 110304.

Makina, S. O. *et al.* (2014) 'Genetic diversity and population structure among six cattle breeds in South Africa using a whole genome SNP panel', *Frontiers in genetics*, 5, p. 333.

Mandli, J. *et al.* (2018) 'Enzyme immunoassay (ELISA/immunosensor) for a sensitive detection of pork adulteration in meat', *Food chemistry*, 255, pp. 380–389.

Mane, B. G. *et al.* (2013) 'Sequence analysis of mitochondrial 16S rRNA gene to identify meat species', *Journal of applied animal research*, 41(1), pp. 77–81.

Mannen, H. *et al.* (2004) 'Independent mitochondrial origin and historical genetic differentiation in North Eastern Asian cattle', *Molecular phylogenetics and evolution*, 32(2), pp. 539–544.

Maralit, B. A. *et al.* (2013) 'Detection of mislabeled commercial fishery by-products in the Philippines using DNA barcodes and its implications to food traceability and safety', *Food Control*, 33(1), pp. 119–125.

Martínez-Camblor, P. *et al.* (2014) 'Standard statistical tools for the breed allocation problem', *Journal of Applied Statistics*, 41(8), pp. 1848–1856.

Martuscelli, M. *et al.* (2020) 'Safety, Quality and Analytical Authentication of ḥalāl Meat Products, with Particular Emphasis on Salami: A Review', *Foods*, 9(8), p. 1111.

Matukumalli, L. K. *et al.* (2009) 'Development and characterization of a high density SNP genotyping assay for cattle', *PloS one*, 4(4), p. e5350.

Mengistie, D. *et al.* (2022) 'Detection and Evaluation of Breed-specific SNPs and Minor allele frequency in Ethiopian Indigenous and European Beef Cattle Breeds'.

Mohd-Hafidz, M. M. *et al.* (2020) 'Detection of pork in processed meat products by species-specific PCR for halal verification: food fraud cases in Hat Yai, Thailand', *Food research*, 4(S1), pp. 244–249.

Montowska, M. *et al.* (2015) 'Authentication of processed meat products by peptidomic analysis using rapid ambient mass spectrometry', *Food chemistry*, 187, pp. 297–304.

Mullen, M. P. *et al.* (2012) 'Polymorphism discovery and allele frequency estimation using high-throughput DNA sequencing of target-enriched pooled DNA samples', *BMC genomics*, 13(1), pp. 1–12.

Murchison, E. P. and Adams, D. J. (2011) 'Sequencing Skippy: the genome sequence of an Australian kangaroo, Macropus eugenii', *Genome Biology*, 12(8), pp. 1–3.

Murcott, M. J. (2021) 'Innovative regulation of meat consumption in South Africa: An environmental rights perspective', *Potchefstroom Electronic Law Journal/Potchefstroomse Elektroniese Regsblad*, 24(1).

Murugaiah, C. *et al.* (2009) 'Meat species identification and Halal authentication analysis using mitochondrial DNA', *Meat science*, 83(1), pp. 57–61.

Nakyinsige, K., Man, Y. B. C. and Sazili, A. Q. (2012) 'Halal authenticity issues in meat and meat products', *Meat science*, 91(3), pp. 207–214.

Nanda, I. *et al.* (2008) 'Synteny conservation of the Z chromosome in 14 avian species (11 families) supports a role for Z dosage in avian sex determination', *Cytogenetic and genome research*, 122(2), pp. 150–156.

Negrini, R. *et al.* (2009) 'Assessing SNP markers for assigning individuals to cattle populations', *Animal Genetics*, 40(1), pp. 18–26. doi: 10.1111/j.1365-2052.2008.01800.x.

Nicolas, V. *et al.* (2012) 'Assessment of three mitochondrial genes (16S, Cytb, CO1) for identifying species in the Praomyini tribe (Rodentia: Muridae)', *PLoS ONE*, 7(5), pp. 1–11. doi: 10.1371/journal.pone.0036586.

Nilsson, M. A. *et al.* (2018) 'Speciation generates mosaic genomes in kangaroos', *Genome Biology and Evolution*, 10(1), pp. 33–44.

Njaramba, Jane Kagure *et al.* (2021) 'Detection of Species Substitution in the Meat Value Chain by High-Resolution Melting Analysis of Mitochondrial PCR Products', *Foods*, 10(12), p. 3090.

Njaramba, Jane K *et al.* (2021) 'Species substitution in the meat value chain by high-resolution

melt analysis of mitochondrial PCR products', *bioRxiv*.

Omran, G. A. *et al.* (2019) 'Species DNA-based identification for detection of processed meat adulteration: is there a role of human short tandem repeats (STRs)?', *Egyptian Journal of Forensic Sciences*, 9(1), pp. 1–8.

Opara, A., Razpet, A. and Logar, B. (2012) 'Breed assignment test of Slovenian cattle breeds using microsatellites', *Livestock production as a technological and social challenge*, pp. 167–170.

Orrù, L. *et al.* (2009) 'Characterization of a SNPs panel for meat traceability in six cattle breeds', *Food Control*, 20(9), pp. 856–860.

Ouso, D. O. *et al.* (2020) 'Three-gene PCR and high-resolution melting analysis for differentiating vertebrate species mitochondrial DNA for biodiversity research and complementing forensic surveillance', *Scientific reports*, 10(1), pp. 1–13.

Pan, Y. *et al.* (2020) 'Combining a COI Mini-Barcode with Next-Generation Sequencing for Animal Origin Ingredients Identification in Processed Meat Product', *Journal of Food Quality*, 2020. doi: 10.1155/2020/2907670.

Pant, S. D. *et al.* (2012) 'Use of breed-specific single nucleotide polymorphisms to discriminate between Holstein and Jersey dairy cattle breeds', *Animal biotechnology*, 23(1), pp. 1–10.

Pawar, H. N., Agrawal, R. K. and Brah, G. S. (2013) 'Expression, purification and characterization of recombinant Heat Shock Protein 70 (HSP70) from sheep and goat species.', *International Journal of Current Microbiology and Applied Sciences*, 2(11), pp. 440–452.

Di Pinto, A. *et al.* (2015) 'Occurrence of mislabeling in meat products using DNA-based assay', *Journal of Food Science and Technology*, 52(4), pp. 2479–2484. doi: 10.1007/s13197-014-1552-y.

Qin, P. *et al.* (2019) 'A sensitive multiplex PCR protocol for simultaneous detection of chicken,

duck, and pork in beef samples', *Journal of food science and technology*, 56(3), pp. 1266–1274.

Quinto, C. A., Tinoco, R. and Hellberg, R. S. (2016) 'DNA barcoding reveals mislabeling of game meat species on the US commercial market', *Food Control*, 59, pp. 386–392.

Rafudeen, A. (2013) 'The Orion Cold Storage Saga: Debating 'Halaal'in South Africa', *Alternation Journal*, (11), pp. 134–162.

Raharjo, T. J. *et al.* (2019) 'Forgery Detection Beef with Mice Meat (Mus musculus) in Meatballs Using Real-Time Polymerase Chain Reaction (Real-Time PCR) Primer Specific for a Target Mitochondrial DNA ND-1 Gene', *Indonesian Journal of Chemistry*, 19(1), pp. 89–96.

Rasmussen Hellberg, R. S. and Morrissey, M. T. (2011) 'Advances in DNA-based techniques for the detection of seafood species substitution on the commercial market', *Journal of Laboratory Automation*, 16(4), pp. 308–321. doi: 10.1016/j.jala.2010.07.004.

Renfree, M. B. *et al.* (2011) 'Genome sequence of an Australian kangaroo, Macropus eugenii, provides insight into the evolution of mammalian reproduction and development', *Genome biology*, 12(8), pp. 1–26.

Le Roex, N. *et al.* (2012) 'Novel SNP discovery in African buffalo, Syncerus caffer, using high-throughput sequencing', *PloS one*, 7(11), p. e48792.

Rosen, B. D. *et al.* (2020) 'De novo assembly of the cattle reference genome with single-molecule sequencing', *Gigascience*, 9(3), p. giaa021.

Rout, M. ; P. S. R. A. (2018) 'Methods for detection of adulteration in meat', (May), pp. 1–7. doi: 10.22271/ed.book10a01.

Rubinoff, D. and Holland, B. S. (2005) 'Between two extremes: mitochondrial DNA is neither the panacea nor the nemesis of phylogenetic and taxonomic inference', *Systematic biology*, 54(6), pp. 952–961.

SA Stud Book (2016) *SA Stud Book Annual Report 2016*. Available at: http://studbook.co.za/images/photos/Annual_Report_2016_a.pdf.

Scholtz, M. M. (2010) *Beef breeding in South Africa.* Agricultural Research Council.

Sen, Shuvro, Antara, N. and Sen, Shusmita (2021) 'Factors influencing consumers' to take ready-made frozen food', *Current Psychology*, 40(6), pp. 2634–2643.

Shahbendeh, P. (2019) *Global Meat Industry - Statistics and Facts*, *https://www.statista.com/topics/4880/global-meat-industry/*.

Shan, L. C. *et al.* (2016) 'Consumer views on "healthier" processed meat', *British Food Journal*.

Shehata, H. R. *et al.* (2017) 'Droplet digital polymerase chain reaction (ddPCR) assays integrated with an internal control for quantification of bovine, porcine, chicken and turkey species in food and feed', *PLoS One*, 12(8), p. e0182872.

Singh, V. P. and Neelam, S. (2011) 'Meat species specifications to ensure the quality of meat: a review', *Int J Meat Sci*, 1(1), pp. 15–26.

Soman, M. *et al.* (2020) 'Detecting mislabelling in meat products using PCR–FINS', *Journal of food science and technology*, 57(11), pp. 4286–4292.

Sommerville, M., Khumalo, L., Kamuti, T. and Brooks, S. (2021) 'Game On! Understanding the Emerging Game Meat Value Chain in South Africa'. Available at: chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://www.issuelab.org/resources/38715/38715.pdf.

Song, Q. *et al.* (2019) 'Monitoring of sausage products sold in Sichuan Province, China: A first comprehensive report on meat species' authenticity determination', *Scientific Reports*, 9(1), pp. 1–9.

South African Market Insights (2022) *Cost of different types of meat in South Africa over the last couple of years*. Available at: http://www.southafricanmi.com/the-cost-of-a-meat-prices-in-south-africa-8july2022.html (Accessed: 11 January 2023).

Spychaj, A. *et al.* (2021) 'A Practical Approach to Identifying Processed White Meat of Guinea Fowl, Rabbit, and Selected Fish Species Using End-Point PCR', *International Journal of Food Science*, 2021.

Steinberg, A. (2021) 'Extra care goes a long way in Wagyu breeding', *Farmer's Weekly*, 2021(21011), pp. 44–46.

Sun, W. *et al.* (2004) 'Genetic differentiation between sheep and goats based on microsatellite DNA', *Asian-australasian journal of animal sciences*, 17(5), pp. 583–587.

Surányi, J. *et al.* (2021) 'Electronic Tongue as a Correlative Technique for Modeling Cattle Meat Quality and Classification of Breeds', *Foods*, 10(10), p. 2283.

Surowiec, I. *et al.* (2011) 'Metabolomic approach for the detection of mechanically recovered meat in food products', *Food Chemistry*, 125(4), pp. 1468–1475.

Suryawan, G. Y., Suardana, I. W. and Wandia, I. N. (2020) 'Sensitivity of polymerase chain reaction in the detection of rat meat adulteration of beef meatballs in Indonesia', *Veterinary World*, 13(5), p. 905.

Tamura, K. *et al.* (2007) 'MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0', *Molecular biology and evolution*, 24(8), pp. 1596–1599.

Tanabe, S. *et al.* (2007) 'A real-time quantitative PCR detection method for pork, chicken, beef, mutton, and horseflesh in foods', *Bioscience, Biotechnology and Biochemistry*, 71(12), pp. 3131–3135. doi: 10.1271/bbb.70683.

Taylor, W. A. *et al.* (2020) 'Jobs, game meat and profits: the benefits of wildlife ranching on

marginal lands in South Africa', *Biological Conservation*, 245, p. 108561.

Tembe, D., Mukaratirwa, S. and Zishiri, O. T. (2018) 'Undeclared Meat Species in Processed Meat Products from Retail Franchises in the Durban Metropole, KwaZulu-Natal Province, South Africa, Using Species-specific DNA Primers', *Food Protection Trends*, 38(6), pp. 440–449.

The Commission of the European Communities (2003) 'COMMISSION DIRECTIVE 2003/126/EC of 23 December 2003 on the analytical method for the determination of constituents of animal origin for the official control of feedingstuffs', 2003(807), pp. 78–84.

Tillmar, A. O. *et al.* (2013) 'A universal method for species identification of mammals utilizing next generation sequencing for the analysis of DNA mixtures', *PLoS ONE*, 8(12), pp. 1–9. doi: 10.1371/journal.pone.0083761.

Troy, C. S. *et al.* (2001) 'Genetic evidence for Near-Eastern origins of European cattle', *Nature*, 410(6832), pp. 1088–1091.

Unajak, S. *et al.* (2011) 'Identification of species (meat and blood samples)using nested-PCR analysis of mitochondrial DNA', *African Journal of Biotechnology*, 10(29), pp. 5670–5676. doi: 10.5897/AJB10.2432.

USDA (2015) *The South African meat market*, *United States Department of Agriculture*.

Wadapurkar, R. M. and Vyas, R. (2018) 'Computational analysis of next generation sequencing data and its applications in clinical oncology', *Informatics in Medicine Unlocked*, 11, pp. 75–82.

Wakefield, M. J. and Graves, J. A. M. (2003) 'The kangaroo genome: leaps and bounds in comparative genomics', *EMBO reports*, 4(2), pp. 143–147.

Walker, M. J., Burns, M. and Burns, D. T. (2013) 'Horse Meat in Beef Products- Species Substitution 2013', *Journal of the Association of Public Analysts (Online)*, 41(November), pp.

67–106.

Wang, F. *et al.* (2020) ‘Identification of chicken-derived ingredients as adulterants using loop-mediated isothermal amplification’, *Journal of Food Protection*, 83(7), pp. 1175–1180.

Wang, L., Hang, X. and Geng, R. (2018) ‘Molecular detection of adulteration in commercial buffalo meat products by multiplex PCR assay’, *Food Science and Technology*, 39, pp. 344–348.

Wang, W. *et al.* (2018) ‘Spectral detection techniques for non-destructively monitoring the quality, safety, and classification of fresh red meat’, *Food Analytical Methods*, 11(10), pp. 2707–2730.

Weldenegodguad, M. *et al.* (2019) ‘Whole-genome sequencing of three native cattle breeds originating from the northernmost cattle farming regions’, *Frontiers in genetics*, 9, p. 728.

Wiener, M. (2011) ‘Questionable meat products in the spotlight’, *Eye Witness News 10*.

Wilkinson, S. *et al.* (2011) ‘Evaluation of approaches for identifying population informative markers from high density SNP chips’, *BMC genetics*, 12(1), pp. 1–14.

Wu, T., Zhong, N. and Yang, L. (2018) ‘Identification of adulterated and non-adulterated Norwegian salmon using FTIR and an improved PLS-DA method’, *Food analytical methods*, 11(5), pp. 1501–1509.

Xing, R.-R. *et al.* (2019) ‘Application of next generation sequencing for species identification in meat and poultry products: A DNA metabarcoding approach’, *Food control*, 101, pp. 173–179.

Xu, H. *et al.* (2022) ‘Visual Detection of Chicken Adulteration Based on a Lateral Flow Strip-PCR Strategy’, *Foods*, 11(15), p. 2351.

Xu, R. *et al.* (2018) ‘Multiplex TaqMan locked nucleic acid real-time PCR for the differential

identification of various meat and meat products', *Meat science*, 137, pp. 41–46.

Yang, L. *et al.* (2014) 'Species identification through mitochondrial rRNA genetic analysis', *Scientific Reports*, 4, pp. 1–11. doi: 10.1038/srep04089.

Yang, L. *et al.* (2018) 'Rapid Identification of Pork Adulterated in the Beef and Mutton by Infrared Spectroscopy', *Journal of Spectroscopy*, 2018. doi: 10.1155/2018/2413874.

Zhang, Y. *et al.* (2020) 'Simultaneous identification of animal-derived components in meats using high-throughput sequencing in combination with a custom-built mitochondrial genome database', *Scientific Reports*, 10(1), pp. 1–11.

Zimin, A. V *et al.* (2009) 'A whole-genome assembly of the domestic cow, Bos taurus', *Genome biology*, 10(4), pp. 1–10.

Zion Market Research (2017) *Processed Meat Market*, *https://www.zionmarketresearch.com/news/processed-meat-market*.

Zvereva, E. A. *et al.* (2015) 'Enzyme immunoassay and proteomic characterization of troponin I as a marker of mammalian muscle compounds in raw meat and some meat products', *Meat science*, 105, pp. 46–52.

Zwane, A. A. *et al.* (2016) 'Genome-wide identification of breed-informative single-nucleotide polymorphisms in three South African indigenous cattle breeds', *South African Journal of Animal Science*, 46(3), pp. 302–312.
Zwane, A. A. *et al.* (2019) 'Genome-wide SNP discovery in indigenous cattle breeds of South Africa', *Frontiers in Genetics*, 10, p. 273.

APPENDIX A

CHAPTER THREE

Appendix A3. 1: Species in the phylogeny analysis

Scientific Name	**Common Name**	***Family***
distoechurus pennatus	Feather Tailed Possum	*Acrobatidae*
anas platyrhynchos	Mallard Duck	*Anatidae*
balaena mysticetus	Bowhead Whale	*Balaenidae*
balaenoptera acutorostrata	Mink Whale	*Balaenopteridae*
balaenoptera edeni	Brydes Whale	*Balaenopteridae*
balaenoptera musculus	Blue Whale	*Balaenopteridae*
balaenoptera omurai	Omuras Whale	*Balaenopteridae*
heterocephalus glaber	Naked Mole Rat	*Bathyergidae*
aepyceros melampus	Impala	*Bovidae*
ammotragus lervia	Barbary Sheep	*Bovidae*
antidorcas marsupialis	Springbok	*Bovidae*
antilope cervicapra	Blackbuck	*Bovidae*
bison bison	American Bison	*Bovidae*
bison bonasus	European Bison	*Bovidae*
bos indicus	Zebu Cattle	*Bovidae*
bos javanicus	Banteng	*Bovidae*
bos primigenius	Aurochs	*Bovidae*
bos taurus	Cattle	*Bovidae*
bubalus bubalis	Buffalo	*Bovidae*
budorcas taxicolor	Takin	*Bovidae*
capra hircus	Goat	*Bovidae*
capricornis crispus	Japanese Serow	*Bovidae*
capricornis swinihoei	Taiwan Serow	*Bovidae*
kobus leche	Lechwe	*Bovidae*
naemorhedus caudatus	Long Tailed Goral	*Bovidae*
oryx dammah	Scimitar Horned Oryx	*Bovidae*
oryx gazella	Gemsbok	*Bovidae*
ovis ammon	Argali	*Bovidae*
ovis aries	Sheep	*Bovidae*
ovis canadensis	Bighorn sheep	*Bovidae*
pantholops hodgsonii	Tibetan Antelope	*Bovidae*
pseudois schaeferi	Dwarf Blue Sheep	*Bovidae*
tragelaphus imberbis	Lesser Kudu	*Bovidae*
camelus bactrianus	Bactrian Camel	*Camelidae*

camelus dromedarius	Dromedary	*Camelidae*
camelus ferus	Wild Bactrian Camel	*Camelidae*
lama glama	Guanaco	*Camelidae*
lama guanicoe	Guanaco	*Camelidae*
vicugna vicugna	Vicuna	*Camelidae*
canis latrans	Coyote	*Canidae*
canis lupus	Tibetan Wolf	*Canidae*
nyctereutes procyonoides	Racoon Dog	*Canidae*
cavia porcellus	Guinea Pig	*Caviidae*
cebus albifrons	Fronted Capuchin	*Cebidae*
saimiri boliviensis	Bolivian Squirrel Monkey	*Cebidae*
saimiri sciureus	Common Squirrel Monkey	*Cebidae*
chlorocebus aethiops	Grivet	*Cercopithecidae*
chlorocebus pygerythrus	Vervet Monkey	*Cercopithecidae*
chlorocebus sabaeus	Green Monkey	*Cercopithecidae*
chlorocebus tantalus	Tantalus Monkey	*Cercopithecidae*
macaca sylvanus	Barbary Marcaque	*Cercopithecidae*
macaca thibetana	Tibetan Marcaque	*Cercopithecidae*
nasalis larvatus	Proboscis Monkey	*Cercopithecidae*
papio hamadryas	Hamadryas Baboon	*Cercopithecidae*
pygathrix nemaeus	Red Shanked Duoc	*Cercopithecidae*
pygathrix nigripes	Black Shanked Duoc	*Cercopithecidae*
rhinopithecus avunculus	Tonkin Snub nosed Monkey	*Cercopithecidae*
tarsius bancanus	Horsefields Tarsier	*Cercopithecidae*
cervus albirostris	Thorolds Deer	*Cervidae*
cervus elaphus	Red Deer	*Cervidae*
cervus nippon	Ussiri Sika Deer	*Cervidae*
elaphodus cephalophus	Tufted Deer	*Cervidae*
hydropotes inermis	Water Deer	*Cervidae*
moschus berezovskii	Chinese Forest Musk Deer	*Cervidae*
moschus moschiferus	Siberian Musk Deer	*Cervidae*
muntiacus reevesi	Reeves Muntjac	*Cervidae*
muntiacus vuquangensis	Giant Muntjac	*Cervidae*
odocoileus virginianus	White Tailed Deer	*Cervidae*
rangifer tarandus	Reindeer	*Cervidae*
rucervus eldii	Elds Deer	*Cervidae*
chrysochloris asiatica	Cape Golden Mole	*Chrysochloridae*
eremitalpa granti	Grands Golden Mole	*Chrysochloridae*
cricetulus griseus	Chinese Hamster	*Cricetidae*
mesocricetus auratus	Goldern Hamster	*Cricetidae*
microtus fortis	Reed Vole	*Cricetidae*
microtus kikuchii	Taiwan Vole	*Cricetidae*

microtus levis	Southern Vole	*Cricetidae*
myodes regulus	Korean Red Necked Vole	*Cricetidae*
neodon irene	Irene's Scrub Vole	*Cricetidae*
proedromys liangshanensis	Vole	*Cricetidae*
crocodylus niloticus	Nile Crocodile	*Crocodylidae*
galeopterus variegatus	Sunda Flying Lemur	*Cynocephalidae*
dasyurus hallucatus	Northern Qoull	*Dasyuridae*
sarcophilus harrisii	Tasmanian Devil	*Dasyuridae*
daubentonia madagascariensis	Aye Aye	*Daubentoniidae*
delphinus capensis	Long Beaked Common Dolphin	*Delphinidae*
eothenomys chinensis	Chinese White Dolphin	*Delphinidae*
grampus griseus	Rissos Dolphin	*Delphinidae*
inia geoffrensis	Amazon River Dolphin	*Delphinidae*
platanista minor	Indus River Dolphin	*Delphinidae*
pontoporia blainvillei	LaPlata Dolphin	*Delphinidae*
sminthopsis crassicaudata	Fat Tailed Dunnart	*Delphinidae*
sousa chinensis	Fat Tailed Dunnart	*Delphinidae*
stenella attenuata	Pantropical Spotted Dolphin	*Delphinidae*
stenella coeruleoalba	Stripped Dolphin	*Delphinidae*
tursiops aduncus	Common Bottlenose Dolphin	*Delphinidae*
tursiops truncatus	Indo Pacific Bottlenose Dolphin	*Delphinidae*
jaculus jaculus	Lesser Egyptian Jeboa	*Dipodidae*
dugong dugon	Dugong	*Dugongidae*
elephas maximus	Asian Elephant	*Elephantidae*
mammuthus columbi	Columbian Mammoth	*Elephantidae*
mammuthus primigenius	Woolly Mammoth	*Elephantidae*
equus africanus	Donkey	*Equidae*
equus caballus	Horse	*Equidae*
equus hemionus	Onanger	*Equidae*
echinosorex gymnura	Moonrat	*Erinaceidae*
erinaceus europaeus	European Hare	*Erinaceidae*
hemiechinus auritus	Long Eared Hedgehog	*Erinaceidae*
hylomys suillus	Short Tailed Gymnure	*Erinaceidae*
eschrichtius robustus	Gray Whale	*Eschrichtiidae*
felis catus	Cat	*Felidae*
lynx rufus	Bobcat	*Felidae*
neofelis nebulosa	Clouded Leopard	*Felidae*
panthera leo	Asiatic Lion	*Felidae*
panthera pardus	Leopard	*Felidae*
panthera tigris	Tiger	*Felidae*

prionailurus bengalensis	Amur Leopard Cat	*Felidae*
puma concolor	Cougar	*Felidae*
galago senegalensis	Senegal Bushbaby	*Galagidae*
otolemur crassicaudatus	Thick Tailed Bushbaby	*Galagidae*
hipposideros armiger	Great Roundleaf Bat	*Hipposideridae*
gorilla gorilla	Western Gorilla	*Hominidae*
homo sapiens	Human	*Hominidae*
pan paniscus	Bonobo	*Hominidae*
pan troglodytes	Common Chimpanzee	*Hominidae*
pongo abelii	Sumatran Orangutan	*Hominidae*
pongo pygmaeus	Bornean Orangutan	*Hominidae*
hylobates agilis	Agile Gibbon	*Hylobatidae*
hylobates lar	Lar Gibbon	*Hylobatidae*
nomascus gabriellae	Red Cheeked Gibbon	*Hylobatidae*
nomascus siki	Southern White Cheeked Gibbon	*Hylobatidae*
symphalangus syndactylus	Saimang	*Hylobatidae*
propithecus coquereli	Coquerels Sifaka	*Indriidae*
eulemur fulvus	Brown Lemur	*Lemuridae*
eulemur macaco	Black Lemur	*Lemuridae*
eulemur mongoz	Mongoose Lemur	*Lemuridae*
lemur catta	Ring Tailed Lemur	*Lemuridae*
lepilemur hubbardorum	Hubbards Sportive Lemur	*Lepilemuridae*
lepus capensis	Brown Hare	*Leporidae*
lepus europaeus	European Hare	*Leporidae*
oryctolagus cuniculus	European Rabbit	*Leporidae*
lipotes vexillifer	Baiji	*Lipotidae*
loris tardigradus	Slender Loris	*Lorisidae*
nycticebus coucang	Sunda Slow Loris	*Lorisidae*
perodicticus potto	Potto	*Lorisidae*
presbytis melalophos	Sumatran Surili	*Lorisidae*
varecia variegata	Black White Ruffles Lemur	*Lumuridae*
lagorchestes hirsutus	Rufous Hare Wallaby	*Macropodidae*
lagostrophus fasciatus	Banded Hare Wallaby	*Macropodidae*
elephantulus elephantulus	Elephant Shrew	*Macroscelididae*
macroscelides proboscideus	Round Ear Elephant Shrew	*Macroscelididae*
mammut americanum	American mastodon	*Mammutidae*
manis pentadactyla	Chinese Pangolin	*Manidae*
manis tetradactyla	Chinese Pangolin	*Manidae*
caperea marginata	Pygmy Right Whale	*Meobalaenidae*
kogia breviceps	Pygmy Sperm Whale	*Meobalaenidae*
apodemus agrarius	Striped field Mouse	*Muridae*
apodemus chejuensis	Jeju Striped Field Mouse	*Muridae*

apodemus chevrieri	Chevrieri Field Mouse	*Muridae*
apodemus peninsulae	Korean Filed Mouse	*Muridae*
glis glis	Edible Dormouse	*Muridae*
leggadina lakedownensis	Lakeland Down Mouse	*Muridae*
mus musculus	House Mouse	*Muridae*
mus terricolor	Earth Colored Mouse	*Muridae*
pseudomys chapmani	Pebble Mound Mouse	*Muridae*
rattus exulans	Polynesian Rat	*Muridae*
rattus fuscipes	Bush Rat	*Muridae*
rattus leucopus	Mottle Tailed Rat	*Muridae*
rattus lutreolus	Australian Swamp Rat	*Muridae*
rattus norvegicus	Brown Rat	*Muridae*
rattus praetor	Large New Guinea Spiny Rat	*Muridae*
rattus rattus	Black Rat	*Muridae*
rattus sordidus	Dusky Field Rat	*Muridae*
rattus tunneyi	Tunneys Rat	*Muridae*
rattus villosissimus	Long Haired Rat	*Muridae*
enhydra lutris	Sea Otter	*Mustelidae*
lutra lutra	European Otter	*Mustelidae*
martes flavigula	Yellow Throated Marten	*Mustelidae*
martes melampus	Japanese Marten	*Mustelidae*
martes zibellina	Sable	*Mustelidae*
meles meles	European Badger	*Mustelidae*
myrmecobius fasciatus	Numbat	*Myrmecobiidae*
tamandua tetradactyla	Long Tailed Pangolin	*Myrmecophagidae*
nannospalax ehrenbergi	Middleeast Blind Mole Rat	*Ochotonidae*
ochotona collaris	Collard Pika	*Ochotonidae*
ochotona curzoniae	Plateau Pika	*Ochotonidae*
odobenus rosmarus	Walrus	*Odobenidae*
ornithorrhynchus anatinus	Platypus	*Ornithorhynchidae*
orycteropus afer	Aardvark	*Orycteropodidae*
arctocephalus forsteri	New Zealand Fur Seal	*Otariidae*
arctocephalus pusillus	Brown Fur Seal	*Otariidae*
eumetopias jubatus	Stellar Sea Lion	*Otariidae*
neophoca cinerea	Australian Sea Lion	*Otariidae*
echymipera rufescens	Long Nosed Echymipera	*Peramelidae*
isoodon macrourus	Northern Brown Bandicoot	*Peramelidae*
dactylopsila trivirgata	Stripped Possum	*Petauridae*
phalanger vestitus	Steins Cuscus	*Phalangeridae*
trichosurus vulpecula	Common Bushtail Bossum	*Phalangeridae*
gallus gallus	Chicken	*Phasianidae*
meleagris gallopavo	Turkey	*Phasianidae*
cystophora cristata	Hooded Seal	*Phocidae*

erignathus barbatus	Bearded Seal	*Phocidae*
halichoerus grypus	Grey Seal	*Phocidae*
lobodon carcinophaga	Crabeater Seal	*Phocidae*
phoca fasciata	Ribbon Seal	*Phocidae*
phoca groenlandica	Harp Seal	*Phocidae*
phoca largha	Spotted Seal	*Phocidae*
pusa caspica	Caspican Seal	*Phocidae*
muntiacus crinifrons	Hairy Fronted Muntjac	*Phocoenidae*
phocoena phocoena	Harbour Porpoise	*Phocoenidae*
artibeus jamaicensis	Jamaican Fruit Bat	*Phyllostomidae*
physeter catodon	Sperm Whale	*Physeteridae*
dendrohyrax dorsalis	Western Tree Hyrax	*Procaviidae*
procavia capensis	Rock Hyrax	*Procaviidae*
procyon lotor	Raccoon	*Procyonidae*
pteropus dasymallus	Ryukyu Flying Fox	*Pteropodidae*
dicerorhinus sumatrensis	Sumatran Rhinoceros	*Rhinocerotidae*
diceros bicornis	Black Rhinoceros	*Rhinocerotidae*
rhinoceros sondaicus	Javan Rhinoceros	*Rhinocerotidae*
rhinolophus ferrumequinum	Korean Greater Horseshoe Bat	*Rhinolophidae*
rhinolophus formosae	Formosan Wolly Horseshoe Bat	*Rhinolophidae*
rhinolophus luctus	Woolly Horsehoe Bat	*Rhinolophidae*
rhinolophus pumilus	Least Horseshoe Bat	*Rhinolophidae*
rhinopithecus bieti	Black Snubnosed Monkey	*Rhinolophidae*
rhinopithecus brelichi	Gray Snubnosed Monkey	*Rhinolophidae*
rhinopithecus roxellana	Goldern Snubnosed Monkey	*Rhinolophidae*
rhinopithecus strykeri	Burmese Snubnosed Monkey	*Rhinolophidae*
sciurus vulgaris	Red Squirrel	*Sciuridae*
episoriculus fumidus	Taiwanese Brown Toothes Serow	*Soricidae*
eospalax baileyi	Plateau Zokor	*Spalacidae*
eospalax rothschildi	Rothschilds Zokor	*Spalacidae*
struthio camelus	Ostrich	*Struthionidae*
phacochoerus africanus	Common Warthog	*Suidae*
sus scrofa	Wild Boar	*Suidae*
sus scrofa domesticus	Domestic pig	*Suidae*
sus scrofa taiwanensis	Taiwan pig	*Suidae*
tachyglossus aculeatus	Short Beaked Echidna	*Tachyglossidae*
galemys pyrenaicus	Pyrenean Desmen	*Talpidae*
mogera wogura	Japanese Mole	*Talpidae*
pipistrellus abramus	Japanese House bat	*Talpidae*
talpa europaea	European Mole	*Talpidae*
urotrichus talpoides	Japanese Shrew Mole	*Talpidae*

tarsius syrichta	Phillipine Tarsier	*Tarsiidae*
pecari tajacu	Collard Peccary	*Tayassuidae*
echinops telfairi	Lesser Hedgehog Tenrec	*Tenrecidae*
thryonomys swinderianus	Greater Cane Rat	*Thryonomyidae*
thylacinus cynocephalus	Thylacine	*Thylacinidae*
trichechus manatus	West Indian Manatee	*Trichechidae*
tupaia pupaia	Nothern Tree Shrew	*Tupaiidae*
ailuropoda melanoleuca	Giant Panda	*Ursidae*
ailurus fulgens	Red Panda	*Ursidae*
arctodus simus	Short Faced Bear	*Ursidae*
helarctos malayanus	Sun Bear	*Ursidae*
melursus ursinus	Sloth Bear	*Ursidae*
tremarctos ornatus	Spectacled Bear	*Ursidae*
ursus americanus	American Black Bear	*Ursidae*
ursus arctos	Brown Bear	*Ursidae*
ursus maritimus	Polar Bear	*Ursidae*
ursus spelaeus	Cave Bear	*Ursidae*
ursus thibetanus	Asian Black Bear	*Ursidae*
lasiurus borealis	Red Rat	*Vespertilionidae*
plecotus auritus	Brown Big eared bat	*Vespertilionidae*
vombatus ursinus	Common Wombat	*Vombatidae*

Appendix A3. 2: *16S* rRNA Sequences
16S Multiple Seq Analysis

Appendix A3. 3: *12S* rRNA Sequences
12S Multiple Seq Analysis

Appendix A3. 4: *COX3* Sequences
COX3 Multiple Seq Analysis

Appendix A3. 5: *ATP6* Sequences

ATP6 Multiple Seq Analysis

Appendix A3. 6: *16S* rRNA Phylogenetic Tree
16S rRNA Phylogenetic Tree

Appendix A3. 7: *12S* rRNA Phylogenetic Tree
12S rRNA Phylogenetic Tree

Appendix A3. 8: *COX3* Phylogenetic Tree

COX3 Phylogenetic Tree

Appendix A3. 9: *ATP6* Phylogenetic Tree

ATP6 Phylogenetic Tree

Appendix A3. 10: *16S* rRNA and *12S* rRNA Combined Phylogenetic Tree

16S rRNA and *12S* rRNA Combined Phylogenetic Tree

Appendix A3. 11: *16S* rRNA and *COX3* Combined Phylogenetic Tree

16S rRNA and *COX3* Combined Phylogenetic Tree

Appendix A3. 12: *16S* rRNA and *ATP6* Combined Phylogenetic Tree

16S rRNA and *ATP6* Combined Phylogenetic Tree

Appendix A3. 13: *16S* rRNA, *12S* rRNA, *COX3* and *ATP6* Combined Phylogenetic Tree

16S rRNA, *12S* rRNA, *COX3* and *ATP6* Combined Phylogenetic Tree

CHAPTER FOUR

Appendix A4. 1: Percentage average fold of pure DNA from known meat types

Known Meat type (100%)	Genus (% Of Average Fold)	Percentage deviation
Beef	*Bos* (Cattle) (99), *Bubalus* (Buffalo) (0,4), *Struthio* (Ostrich) (0,2) & *Rupicapra* (Goat antelope) (0,4)	1
Mutton	*Ovis* (Sheep) (99,1), *Bos* (Cattle) (0,1), *Rupicapra* (Goat Antelope) (0,6), *Naemorhedus* (Goral) (0,1) & *Anas* (Duck) (0,1)	0.9
Pork	*Sus* (Pig) (99,7), *Ovis* (Sheep) (0,2), & *Bos* (Cattle) (0,1)	0.3
Chevon	*Capra* (Goat) (98,7), *Ovis* (Sheep) (0,2), *Budorcas* (Gnu Goat), *Bos* (Cattle) (0,4), *Rupicapra* (Goat Antelope) (0,2), & *Sus* (Pig) (0,1)	1.3
Chicken	*Gallus* (Chicken) (98,5), *Bos* (Beef) (0,3), *Sus* (Pig) (0,2), *Capra* (Goat) (0,6), & *Ovis* (Sheep) (0,3)	1.4
Duck	*Anas* (Duck) (98,5), *Capra* (Goat) (0,1), *Ovis* (Sheep) (0,5), *Gallus* (Chicken) (0,2), *Bos* (Beef) (0,1), *Meleagris* (Turkey) (0,3), & *Sus* (Pig) (0,2)	1.5
Turkey	*Meleagris* (Turkey) (99,1), *Sus* (Pig) (0,1), *Anas* (Duck) (0,2), *Capra* (Goat) (0,2), *Ovis* (Sheep) (0,1), *Gallus* (Chicken) (0,1), & *Bos* (Beef) (0,1)	0.9
Ostrich	*Struthio* (Ostrich) (99,3), *Bos* (Beef) (0,3), *Meleagris* (Turkey) (0,1), *Sus* (Pig) (0,1), *Anas* (Duck) (0,1), & *Ovis* (Sheep) (0,1)	0.7
Kangaroo	*Macropus* (Kangaroo) (99,4), *Capra* (Goat) (0,2), *Struthio* (Ostrich) (0,2), *Bos* (Cattle) (0,1), & *Anas* (Duck) (0,1)	0.6

Appendix A4. 2: Percentage of average fold of pure DNA from two known meat types artificially mixed at a ratio of 1:1

Mixture Ratio	Meat Type (%)	Genus (% Of Average fold)	Percentage Difference with Respect to Input Ratio
1:1	Pork (50%) Beef (50%)	*Sus* (Pig) (51,5) *Bos* (Cattle) (48,1) *Ovis* (Sheep) (0,1) *Bubalus* (Buffalo) (0,1) *Rupicapra* (Goat Antelope) (0,1)	1,5 -1,9
1:1	Chevon (50%)	*Capra* (Goat) (35,7)	-14,3

	Mutton (50%)	*Ovis* (Sheep) (63,4) *Budorcas* (Gnu Goat) (0,1) *Bos* (Cattle) (0,4) *Rupicapra* (Goat Antelope) (0,3) *Sus* (Pig) (0,1)	13,4
1:1	Chicken (50%) Turkey (50%)	*Gallus* (Chicken) (45,3) *Meleagris* (Turkey) (50,8) *Bos* (Cattle) (0,2) *Sus* (Pig) (0,1) *Capra* (Goat) (0,3) *Ovis* (Sheep) (0,2)	-4,7 0,8
1:1	Duck (50%) Ostrich (50%)	*Anas* (Duck) (69,2) *Struthio* (Ostrich) (29,9) *Sus* (Pig) (0,2) *Capra* (Goat) (0,1) *Ovis* (Sheep) (0,2) *Gallus* (Chicken) (0,1) *Bos* (Cattle) (0,3)	19,2 -20,1
1:1	Beef (50%) Kangaroo (50%)	*Bos* (Cattle) (73,9) *Macropus* (Kangaroo) (25,1) *Bubalus* (Buffalo) (0,2) *Gallus* (Chicken) (0,1) *Rupicapra* (Goat Antelope) (0,2) *Sus* (Pig) (0,2) *Anas* (Duck) (0,1)	23,9 24,9

Appendix A4. 3: Percentage average folds of pure DNA from known meat samples artificially mixed at a ratio of 9:1

Mixture Ratio	Meat Type (%)	Genus (% Of Average Fold)	% Difference with Respect to Input Ratio
9:1	Pork (90%) Beef (10%)	*Sus* (Pig) (91,3) *Bos* (Cattle) (8,5) *Ovis* (Sheep) (0,1)	1,3 -1,5
9:1	Beef (90%) Pork (10%)	*Bos* (Cattle) (92,7) *Sus* (Pig) (6,6) *Bubalus* (Buffalo) (0,3) *Rupicapra* (Goat Antelope) (0,3)	2,7 -3,4
9:1	Chevon (90%) Mutton (10%)	*Capra* (Goat) (81,5) *Ovis* (Sheep) (17,4) *Budorcas* (Gnu Goat) (0,3) *Bos* (Cattle) (0,1) *Rupicapra* (Goat Antelope) (0,2)	-8,5 7,4
	Mutton (90%)	*Ovis* (Sheep) (92,7)	2,7

	Chevon (10%)	*Capra* (Goat) (6,4) *Rupicapra* (Goat Antelope) (0,5) *Naemorhedus* (Goral) (0,1)	-3,6
9:1	Chicken (90%) Duck (10%)	*Gallus* (Chicken) (82,1) *Anas* (Duck) (17,5) *Capra* (Goat) (0,1) *Ovis* (Sheep) (0,1) *Sus* (Pig) (0,1) *Bos* (Cattle) (0,1)	-7,9 7,5
9:1	Duck (90%) Chicken (10%)	*Anas* (Duck) (96,3) *Gallus* (Chicken) (3,4) *Sus* (Pig) (0,1)	6,3 -3,6
9:1	Ostrich (90%) Duck (10%)	*Struthio* (Ostrich) (81,2) *Anas* (Duck) (18,5) *Bos* (Cattle) (0,1)	-8,8 8,5
9:1	Duck (90%) Ostrich (10%)	*Anas* (Duck) (86,4) *Struthio* (Ostrich) (13,2) *Ovis* (Sheep) (0,1) *Sus* (Pig) (0,1) *Bos* (Cattle) (0,1)	-3,6 3,2

Appendix A4. 4: Percentage average fold of DNA from retail biltong for which the meat types were not indicated on product labels (N=11)

Sample Number	Genus (% Of Average fold)
S19	*Bos* (Cattle) (98,1), *Bubalus* (Buffalo) (0,6), *Rupicapra* (Goat Antelope) (0,8), *Redunca* (Reedbuck) (0,2) and *Bison* (Bison) (0,1)
S35	*Bos* (Cattle) (99,7), & *Bubalus* (Buffalo) (0,1)
S3	*Bos* (Cattle) (98,3), *Bubalus* (Buffalo) (0,7), *Rupicapra* (Goat Antelope) (0,7), *Redunca* (Reedbuck) (0,2) and *Bison* (Bison) (0,1)
S45	*Bos* (Cattle) (99,4), *Bubalus* (Buffalo) (0,2), *Rupicapra* (Goat Antelope) (0,1), *Sus* (Pig) (0,1) and *Redunca* (Reedbuck) (0,1)
S54	*Bos* (Cattle) (98,2), *Bubalus* (Buffalo) (0,7), *Redunca* (Reedbuck) (0,3), *Sus* (Pig) (0,2), *Hemitragus* (Tahr) (0,2), *Capricornis* (Serow) (0,1) and *Bison* (Bison) (0,1)
S56	*Bos* (Cattle) (99,4), *Bubalus* (Buffalo), (0,2), *Rupicapra* (Goat Antelope) (0,1), *Sus* (Pig) (0,1) and *Redunca* (Reedbuck) (0,1)
S62	*Bos* (Cattle) (98,2), *Rupicapra* (Goat Antelope) (0,6), *Bubalus* (Buffalo) (0,5), *Redunca* (Reedbuck) (0,2), *Capricornis* (Serow) (0,1), *Bison* (Bison) (0,1), and *Hemitragus* (Tahr) (0,1)
S67	*Bos* (cattle) (98,7), *Bubalus* (Buffalo) (0,5), *Rupicapra* (Goat Antelope) (0,4), *Redunca* (Reedbuck) (0,2) and *Bison* (Bison) (0,1)
S69	*Bos* (Cattle) (99,4), *Bubalus* (Buffalo) (0,2), *Bison* (Bison) (0,1), *Rupicapra* (Goat Antelope) (0,2) and *Redunca* (Reedbuck) (0,1),
S92	*Bos* (Cattle) (xx) (98,4), *Rupicapra* (Goat Antelope) (0,7), *Bubalus* (Buffalo) (0,5) and *Redunca* (Reedbuck) (0,2)
S09	*Bos* (Cattle) (99,2), *Bubalus* (Buffalo) (0,3), *Rupicapra* (Goat Antelope) (0,2) *Sus* (Pig) (0,1) and *Redunca* (Reedbuck) (0,1)

Appendix A4. 5: Percentage average fold of DNA from retail meat products labelled as beef biltong (N=17)

Sample Number	Genus (% Of Average Fold)
S©	*Bos* (Cattle) (**93,6**), *Ovis* (Sheep) (4,9), *Sus* (Pig) (0,8), *Bubalus* (Buffalo) (0,2) *Bison* (Bison) (0,1), *Rupicapra* (Goat Antelope) (0,1) and *Redunca* (Reedbuck) (0,1)
112	*Bos* (Cattle) (98,4), *Bubalus* (Buffalo) (0,6), *Rupicapra* (Goat Antelope) (0,6), *Redunca* (Reedbuck) (0,2), *Bison* (Bison) (0,1) and *Ovis* (Sheep) (0,1)
117©	*Bos* (Cattle) (**97,9**), *Bubalus* (Buffalo) (0,7), *Hemitragus* (Tahr) (0,3), *Capricornis* (Serow) (0,2), *Redunca* (Reedbuck) (0,2), *Moschus* (Musk-deer) (0,1), *Sus* (Pig) (0,1), *Bison* (Bison) (0,1) and *Rupicapra* (Goat Antelope) (0,1)
137	*Bos* (Cattle) (99,0), *Bubalus* (Buffalo) (0,4), *Rupicapra* (Goat Antelope) (0,2), *Redunca* (Reedbuck) (0,2), *Sus* (Pig) (0,1) and *Bison* (Bison) (0,1)
138	*Bos* (Cattle) (99,0), *Bubalus* (Buffalo) (0,4), *Rupicapra* (Goat Antelope) (0,3), *Redunca* (Reedbuck) (0,2), *Sus* (Pig) (0,1) and *Bison* (Bison) (0,1)
142	*Bos* (Cattle) (99,5), *Bubalus* (Buffalo) (0,2), *Bison* (Bison) (0,1) and *Redunca* (Reedbuck) (0,1)
147	*Bos* (Cattle) (99,2), *Sus* (Pig) (0,2), *Bubalus* (Buffalo) (0,2), *Ovis* (Sheep) (0,1), *Rupicapra* (Goat Antelope) (0,1) and *Redunca* (Reedbuck) (0,1)
148	*Bos* (Cattle) (99,0), *Bubalus* (Buffalo) (0,5), *Redunca* (Reedbuck) (0,2), *Sus* (Pig) (0,1) and *Hemitragus* (Tahr) (0,1)
149	*Bos* (Cattle) (99,4), *Bubalus* (Buffalo) (0,3), *Ovis* (Sheep) (0,1) and *Redunca* (Reedbuck) (0,1)
15	*Bos* (Cattle) (99,4), *Sus* (Pig) (0,1), *Bubalus* (Buffalo) (0,1), *Rupicapra* (Goat Antelope) (0,1) and *Redunca* (Reedbuck) (0,1)
153	*Bos* (Cattle) (99,6), *Sus* (Pig) (0,2) and *Bubalus* (Buffalo) (0,1)
16©	*Bos* (Cattle) (**57,5**), *Sus* (Pig) (37,4), *Ovis* (Sheep) (4,6), *Bubalus* (Buffalo) (0,1) and *Rupicapra* (Goat Antelope) (0,1)
166	*Bos* (Cattle) (99,3), *Bubalus* (Buffalo) (0,2), *Rupicapra* (Goat Antelope) (0,2) and *Redunca* (Reedbuck) (0,1)
168	*Bos* (Cattle) (98,6), *Bubalus* (Buffalo) (0,6), *Rupicapra* (Goat Antelope) (0,4), *Redunca* (Reedbuck) (0,2) and *Bison* (Bison) (0,1)
17	*Bos* (Cattle) (99,2), *Bubalus* (Buffalo) (0,4), *Rupicapra* (Goat Antelope) (0,2) and *Redunca* (Reedbuck) (0,1)
175	*Bos* (Cattle) (99,5), *Bubalus* (Buffalo) (0,3), *Sus* (Pig) (0,1) and *Redunca* (Reedbuck) (0,1)
19	*Bos* (Cattle) (99,1), *Bubalus* (Buffalo) (0,4), *Rupicapra* (Goat Antelope) (0,2), *Bison* (Bison) (0,1) and *Redunca* (Reedbuck) (0,1)

Appendix A4. 6: Percentage average folds of DNA from retail minced meat for which the meat types were not indicated on product labels (N=27)

Sample Number	Genus (% Of Average fold)
S	*Bos* (Cattle) (99,7), *Bubalus* (Buffalo) (0,1) and *Rupicapra* (Goat Antelope) (0,1)
16	*Bos* (Cattle) (99,0), *Sus* (Pig) (0,6), *Bubalus* (Buffalo) (0,1) and *Rupicapra* (Goat Antelope) (0,1)
2	*Bos* (Cattle) (99,4), *Bubalus* (Buffalo) (0,3), *Bison* (Bison) (0,1) and *Redunca* (Reedbuck) (0,1)

23	*Bos* (Cattle) (99,4), *Bubalus* (Buffalo) (0,3), *Bison* (Bison) (0,1) and *Redunca* (Reedbuck) (0,1)
3	*Bos* (Cattle) (99,2), *Rupicapra* (Goat Antelope) (0,2), *Sus* (Pig) (0,1), *Ovis* (Sheep) (0,1) and *Redunca* (Reedbuck) (0,1)
33	*Bos* (Cattle) (99,5), *Bubalus* (Buffalo) (0,2), *Rupicapra* (Goat Antelope) (0,2) and *Redunca* (Reedbuck) (0,1)
34©	*Bos* (Cattle) (**97,7**), *Ovis* (Sheep) (1,8), *Bubalus* (Buffalo) (0,2), *Rupicapra* (Goat Antelope) (0,1) and *Redunca* (Reedbuck) (0,1)
39	*Bos* (Cattle) (99,3), *Rupicapra* (0,3), *Bubalus* (0,2) and *Sus* (Pig) (0,1)
4©	*Sus* (Pig) (**97,7**) and *Bos* (Cattle) (2,2)
5	*Bos* (Cattle) (99,7) and *Bubalus* (Buffalo) (0,1)
55	*Bos* (Cattle) (99,4), *Bubalus* (Buffalo) (0,2), *Rupicapra* (Goat Antelope) (0,1) and *Redunca* (Reedbuck) (0,1)
58	*Bos* (Cattle) (98,2), *Bubalus* (Buffalo) (0,8), *Rupicapra* (Goat Antelope) (0,5) and *Redunca* (Reedbuck) (0,3)
59	*Bos* (Cattle) (99,7) and *Bubalus* (Buffalo) (0,1)
6	*Bos* (Cattle) (98,8), *Sus* (Pig) (0,5), *Bubalus* (Buffalo) (0,2), *Rupicapra* (Goat Antelope) (0,3), *Ovis* (Sheep) (0,1) and *Redunca* (Reedbuck) (0,1)
63	*Bos* (Cattle) (99,6), *Bubalus* (Buffalo) (0,1) and *Rupicapra* (Goat Antelope) (0,1)
65©	*Bos* (Cattle) (**83,2**), *Ovis* (Sheep) (16,4), *Rupicapra* (0,2) and *Bubalus* (Buffalo) (0,1)
66	*Bos* (Cattle) (99,3), *Bubalus* (Buffalo) (0,2), *Rupicapra* (Goat Antelope) (0,1), *Ovis* (Sheep) (0,1) and *Redunca* (Reedbuck) (0,1)
74	*Bos* (Cattle) (98,3), *Sus* (Pig) (1,3), *Bubalus* (Buffalo) (0,2), *Rupicapra* (Goat Antelope) (0,1) and *Redunca* (Reedbuck) (0,1)
77	*Bos* (Cattle) (99,0), *Bubalus* (Buffalo) (0,3), *Rupicapra* (Goat Antelope) (0,3), *Bison* (Bison) (0,1) and *Redunca* (Reedbuck) (0,1)
78©	*Bos* (Cattle) (**95,3**), *Sus* (Pig) (3,8), *Bubalus* (Buffalo) (0,2), (Goat Antelope) (0,3), *Bison* (Bison) (0,1), *Rupicapra* and *Redunca* (Reedbuck) (0,1)
8	*Bos* (Cattle) (99,6), *Bubalus* (Buffalo) (0,1) and *Rupicapra* (Goat Antelope) (0,1)
89	*Bos* (Cattle) (98,8*), Ovis* (Sheep) (0,4), *Sus* (Pig) (0,2), *Bubalus* (Buffalo) (0,2), *Rupicapra* (Goat Antelope) (0,2) and *Redunca* (Reedbuck) (0,1)
9	*Bos* (Cattle) (99,4), *Bubalus* (Buffalo) (0,2), *Rupicapra* (Goat Antelope) (0,2), *Sus* (Pig) (0,1) and *Redunca* (Reedbuck) (0,1)
93	*Bos* (Cattle) (99,5), *Bubalus* (Buffalo) (0,2), *Rupicapra* (Goat Antelope) (0,1) and *Redunca* (Reedbuck) (0,1)
94	*Bos* (Cattle) (99,8) and *Bubalus* (Buffalo) (0,1)
95	*Bos* (Cattle) (99,3), *Bubalus* (Buffalo) (0,2), *Bison* (Bison) (0,1), *Rupicapra* (Goat Antelope) (0,2) and *Redunca* (Reedbuck) (0,1)
97	*Bos* (Cattle) (99,5), *Bubalus* (Buffalo) (0,1), *Rupicapra* (Goat Antelope) (0,2), *Sus* (Pig) (0,1) and *Redunca* (Reedbuck) (0,1)

Appendix A4. 7: Percentage average fold of DNA from retail meat products labelled as beef mince

Sample Number	Genus (% Of Average Fold)
S	*Bos* (Cattle) (99,5), *Bubalus* (Buffalo) (0,2), *Rupicapra* (Goat Antelope) (0,1) and *Redunca* Reedbuck) (0,1)
109	*Bos* (Cattle) (98,9), *Bubalus* (Buffalo) (0,7), *Redunca* (Reedbuck) (0,3) and *Bison* (Bison) (0,1)
12	*Bos* (Cattle) (99,3), *Bubalus* (Buffalo) (0,3), *Rupicapra* (Goat Antelope) (0,2) and *Redunca* (Reedbuck) (0,1)

123	*Bos* (Cattle) (99,3), *Bubalus* (Buffalo) (0,2), *Sus* (Pig) (0,2), *Ovis* (Sheep) (0,1), *Rupicapra* (Goat Antelope) (0,1) and *Redunca* (Reedbuck) (0,1)
128	*Bos* (Cattle) (99,2), *Bubalus* (Buffalo) (0,3), *Rupicapra* (Goat Antelope) (0,2) and *Redunca* (Reedbuck) (0,1)
13	*Bos* (Cattle) (99,2), *Bubalus* (Buffalo) (0,2), *Sus* (Pig) (0,1), *Bison* (Bison) (0,1) and *Rupicapra* (Goat Antelope) (0,1)
132	*Bos* (Cattle) (99,6), *Sus* (Pig) (0,1), *Bubalus* (Buffalo) (0,1), *Rupicapra* (Goat Antelope) (0,1) and *Redunca* (Reedbuck) (0,1)
133	*Bos* (Cattle) (99,1), *Bubalus* (Buffalo) (0,2), *Sus* (Pig) (0,2), *Rupicapra* (Goat Antelope) (0,1), *Bison* (Bison) (0,1), *Redunca* (Reedbuck) (0,1)
139	*Bos* (Cattle) (99,4), *Bubalus* (Buffalo) (0,2) *Sus* (Pig) (0,1), *Rupicapra* (Goat Antelope) (0,1) and *Redunca* (Reedbuck) (0,1)
14	*Bos* (Cattle) (99,0), *Sus*0,3 *Bubalus*0,3 *Rupicapra*0,2 *Redunca*0,1
152	*Bos* (Cattle) (99,2), *Bubalus* (Buffalo) (0,3), *Rupicapra* (Goat Antelope) (0,2), *Sus* (Pig) (0,1) and *Redunca* (Reedbuck) (0,1)
154	*Bos* (Cattle) (99,5), *Bubalus* (Buffalo) (0,1) and *Rupicapra* (Goat Antelope) (0,1)
158	*Bos* (Cattle) (98,9), *Ovis* (Sheep) (0,4), *Sus* (Pig) (0,2), *Homo* (Human) (0,1), *Bubalus* (Buffalo) (0,1) and *Redunca* (Reedbuck) (0,1)
162	*Bos* (Cattle) (99,1), *Bubalus* (Buffalo) (0,3), *Ovis* (Sheep) (0,2), *Bison* (Bison) (0,1), *Rupicapra* (Goat Antelope) (0,1) *Redunca* (Reedbuck) (0,1)
163	*Bos* (Cattle) (99,1), *Bubalus* (Buffalo) (0,3), *Ovis* (Sheep) (0,2), *Bison* (Bison) (0,1), *Rupicapra* (Goat Antelope) (0,1) and *Redunca* (Reedbuck) (0,1)
164	*Bos* (Cattle) (99,5), *Sus* (Pig) (0,3), *Bubalus* (Buffalo) (0,1) and *Rupicapra* (Goat Antelope) (0,1)
167	*Bos* (Cattle) (98,5), *Ovis* (Sheep) (0,6), *Bubalus* (Buffalo) (0,3), *Rupicapra* (Goat Antelope) (0,3) and *Redunca* (Reedbuck) (0,1)
17©	*Bos* (Cattle) (**97,0**), *Ovis* (Sheep) (2,0), *Bubalus* (Buffalo) (0,3), *Rupicapra* (Goat Antelope) (0,3), *Sus* (Pig) (0,1), *Bison* (Bison) (0,1) and *Redunca* (Reedbuck) (0,1)
173	*Bos* (Cattle) (99,4), *Bubalus* (Buffalo) (0,3), *Rupicapra* (Goat Antelope) (0,1) and *Redunca* (Reedbuck) (0,1)
18	*Bos* (Cattle) (98,3), *Ovis* (Sheep) (0,7), *Bubalus* (Buffalo) (0,3), *Sus* (Pig) (0,3), *Rupicapra* (Goat Antelope) (0,3) and *Redunca* Reedbuck (0,1)
183©	*Bos* (Cattle) (**93,3**), *Sus* (Pig) (6,1), *Bubalus* (Buffalo) (0,2), *Rupicapra* (Goat Antelope) (0,2) and *Redunca* (Reedbuck) (0,1)
99	*Bos* (Cattle) (99,1), *Homo* (Human) (0,5) and *Struthio* (Ostrich) (0,2)

Appendix A4. 8: Percentage average fold of DNA from retail patties for which the meat types were not indicated on product labels (N=15)

Sample Number	Genus (% Of Average Fold)
S	*Bos* (Cattle) (99,3), *Sus* (Pig) (0,2), *Bubalus* (Buffalo) (0,2), *Rupicapra* (Goat Antelope) (0,2) and *Redunca* (Reedbuck) (0,1)
12	*Bos* (Cattle) (99,4), *Bubalus* (Buffalo) (0,2), *Sus* (Pig) (0,1), *Ovis* (Sheep) (0,1) and Redunca (Reedbuck) (0,1)
15	*Bos* (Cattle) (99,1), *Ovis* (Sheep), (0,4), *Bubalus* (Buffalo) (0,2), *Rupicapra* (Goat Antelope) (0,1) and *Redunca* (Reedbuck) (0,1)
2	*Bos* (Cattle) (99,7), *Bubalus* (Buffalo) (0,1) *Rupicapra* (Goat Antelope) (0,1) and, *Redunca* (Reedbuck) (0,1)
22	*Bos* (Cattle) (99,7), *Bubalus* (Buffalo) (0,1) and *Rupicapra* (Goat Antelope) (0,1)
32	*Bos* (Cattle) (99,6), *Bubalus* (Buffalo) (0,2), *Rupicapra* (0,1) and *Redunca* (Reedbuck) (0,1)
44	*Bos* (Cattle) (99,5), *Bubalus* (Buffalo) (0,2) *Rupicapra* (Goat Antelope) (0, 2) and *Sus* (Pig) (0,1)

46	*Bos* (Cattle) (99,1), *Bubalus* (Buffalo) (0,3), *Rupicapra* (Goat Antelope) (0,3), *Sus* (Pig) (0,1) and *Redunca* (Reedbuck) (0,1).
47	*Bos* (Cattle) (99,7) and Bubalus (Buffalo) (0,1)
48©	*Bos* (Cattle) (**91,8**), *Sus* (Pig) (7,7), *Bubalus* (Buffalo) (0,1) *Bison* (Bison) (0,1), *Rupicapra* (Goat Antelope) (0,1) *Redunca* (Reedbuck) (0,1)
57	*Bos* (Cattle) (98,9), *Bubalus* (Buffalo) (0,4), *Rupicapra* (Goat Antelope) (0,4) and *Redunca* (Reedbuck) (0,1)
7	*Bos* (Cattle) (99,7), *Bubalus* (Buffalo) (0,1) and *Rupicapra* (Goat Antelope) (0,1)
73	*Bos* (Cattle) (99,2), *Bubalus* (Buffalo) (0,2) *Rupicapra* (Goat Antelope), (0,2) and *Redunca* (Reedbuck) (0,1)
8	*Bos* (Cattle) (99,6), *Sus* (Pig) (0,1) *Bubalus* (Buffalo) (0,1) and *Rupicapra* (Goat Antelope) (0,1)
84©	*Sus* (Pig) (**59,6**), *Bos* (Cattle) (40,0), *Bubalus* (Buffalo) (0,1), *Connochaetes* (Wildebeest) (0,1) and *Rupicapra* (Goat Antelope) (0,1)

Appendix A4. 9: Percentage average folds of DNA from retail meat products labelled as beef patty (N=18)

Sample Number	Genus (% Of Average Fold)
102	*Bos* (Cattle) (99,4), *Bubalus* (Buffalo) (0,2), *Ovis* (Sheep) (0,1), *Rupicapra* (Goat Antelope) (0,1) and *Redunca* (Reedbuck) (0,1)
113©	*Bos* (Cattle) (**93,4**), *Ovis* (Sheep) (5,6), *Sus* (Pig) (0,1), *Bubalus* (Buffalo) (0,1) and *Rupicapra* (Goat Antelope) (0,1)
115	*Bos* (Cattle) (99,1), *Bubalus* (Buffalo) (0,3), *Ovis* (Sheep) (0,2), *Bison* (Bison) (0,1), *Rupicapra* (Goat Antelope) (0,1) and *Redunca* (Reedbuck) (0,1)
112©	*Bos* (Cattle) (**81,7**), *Ovis* (Sheep) (17,7) *Rupicapra* (Goat Antelope) (0,2), *Bison* (Bison) (0,1) and *Bubalus* (Buffalo) (0,1)
12	*Bos* (Cattle) (98,2), *Ovis* (Sheep) (1,2), *Rupicapra* (Goat Antelope) (0,2), *Bubalus* (Buffalo) (0,1) and *Bison* (Bison) (0,1)
122©	*Bos* (Cattle) (**64,5**), *Ovis* (Sheep) (34,3), *Rupicapra* (Goat Antelope) (0,7) and *Bubalus* (Buffalo) (0,1)
124	*Bos* (Cattle) (99,2), *Bubalus* (Buffalo) (0,3), *Rupicapra* (Goat Antelope) (0,2) and *Redunca* (Reedbuck) (0,1)
125	*Bos* (Cattle) (98,7), *Ovis* (Sheep) (0,9), *Bubalus* (Buffalo) (0,2), *Rupicapra* (Goat Antelope) (0,1) and *Redunca* (Reedbuck) (0,1)
13	*Bos* (Cattle) (99,5), *Sus* (Pig) (0,2) and *Bubalus* (Buffalo) (0,1)
136©	*Bos* (Cattle) (**94,9**), *Sus* (Pig) (4,4), *Bubalus* (Buffalo) (0,2) and *Redunca* (Reedbuck) (0,1)
14	*Bos* (Cattle) (99,1), *Sus* (Pig) (0,2), *Bubalus* (Buffalo) (0,2), *Rupicapra* (Goat Antelope) (0,1) and *Redunca* (Reedbuck) (0,1)
146	*Bos* (Cattle) (99,5), *Sus* (Pig) (0,1), *Homo* (Huma) (0,1) and *Ovis* (Sheep) (0,1)
169	*Bos* (Cattle) (99,3), *Bubalus* (Buffalo) (0,3), *Rupicapra* (Goat Antelope) (0,1) and *Redunca* (Reedbuck) (0,1)
172	*Bos* (Cattle) (99,0), *Bubalus* (Buffalo) (0,4), *Rupicapra* (Goat Antelope) (0,2), *Redunca* (Reedbuck) (0,2) and *Bison* (Bison) (0,1)
179©	*Bos* (Cattle) (**93,4**), *Ovis* (Sheep) (6,2), *Bubalus* (Buffalo) (0,1) *Rupicapra* (Goat Antelope) (0,1)
18	*Bos* (Cattle) (99,3), *Bubalus* (Buffalo) (0,3), *Rupicapra* (Goat Antelope) (0,2) and *Redunca* (Reedbuck) (0,1)
182	*Bos* (Cattle) (98,7), *Homo* (Human) (0,6), *Mus* (Rat) (0,3), *Sus* (Pig) (0,1), *Bubalus* (Buffalo) (0,1) and *Redunca* (Reedbuck) (0,1)

185	*Bos* (Cattle) (98,7), *Bubalus* (Buffalo) (0,4), *Rupicapra* (Goat Antelope) (0,4), *Redunca* (Reedbuck) (0,2) and *Ovis* (Sheep) (0,1)

Appendix A4. 10: Percentage average folds of DNA from retail sausages for which the meat types were not indicated on product labels (N=21)

Sample Number	Genus (% Of Average Fold)
15©	*Bos* (Cattle) (**37,1**) *Sus* (Pig) (38,5), *Ovis* (Sheep) (23,7) and *Rupicapra* (Goat Antelope) (0,2)
159©	*Ovis* (Sheep) (**91,4**), *Bos* (Cattle) (7,0), *Bison* (Bison) (0,1), *Rupicapra* (Goat Antelope) (1,3) and *Naemorhedus* (Goral) (0,1)
16	*Bos* (Cattle) (99,5), *Bubalus* (Buffalo) (0,2), *Rupicapra* (Goat Antelope) (0,1) and *Redunca* (Reedbuck) (0,1)
18	*Bos* (Cattle) (99,3), *Bubalus* (Buffalo) (0,2), *Gallus* (Chicken) (0,2), *Sus* (Pig) (0,1), *Rupicapra* (Goat Antelope) (0,1) and *Redunca* (Reedbuck) (0,1)
26©	Bos (Cattle) (**91,0**), *Ovis* (Sheep) (5,2), *Sus* (Pig) (3,0), *Bubalus* (Buffalo) (0,3), *Rupicapra* (Goat Antelope) (0,2) and *Redunca* (Reedbuck) (0,1)
27	*Bos* (Cattle) (99,5), *Sus* (Pig) (0,1), *Bubalus* (Buffalo) (0,1) and *Rupicapra* (Goat Antelope) (0,1)
29	*Bos* (Cattle) (98,8), *Ovis* (Sheep) (0,6), *Bubalus* (Buffalo) (0,2), *Sus* (Pig) (0,1), *Rupicapra* (Goat Antelope) (0,1) and *Redunca* (Reedbuck) (0,1)
38	Bos (Cattle) (99,0), *Bubalus* (Buffalo) (0,3), *Rupicapra* (Goat Antelope) (0,2), *Ovis* (Sheep) (0,1) and *Redunca* (Reedbuck) (0,1)
4©	*Sus* (Pig) (**97,4**) and *Bos* (Cattle) (2,7)
42	*Bos* (Cattle) (99,3), *Sus* (Pig) (0,3), *Bubalus* (Buffalo) (0,2), *Rupicapra* (Goat Antelope) (0,1) and *Redunca* (Reedbuck) (0,1)
43	*Bos* (Cattle) (99,5), *Bubalus* (Buffalo) (0,2), *Rupicapra* (Goat Antelope) (0,1), *Redunca* (Reedbuck) (0,1) and *Hemitragus* (Tahr) (0,1)
5©	*Sus* (Pig) (**94,8**) and *Bos* (Cattle) (5,2)
68©	*Bos* (Cattle) (**97,9**), *Bubalus* (Buffalo) (0,8), *Rupicapra* (Goat Antelope) (0,6), *Redunca* (Reedbuck) (0,3) and *Bison* (Bison) (0,2)
7	*Bos* (Cattle) (99,6), *Bubalus* (Buffalo) (0,2), *Ovis* (Sheep) (0,1), *Bison* (Bison) (0,1), *Rupicapra* (Goat Antelope) (0,1) and *Redunca* (Reedbuck) (0,1)
72	*Bos* (Cattle) (98,9), *Bubalus* (Buffalo) (0,5), *Rupicapra* (Goat Antelope) (0,3) and *Redunca* (Reedbuck) (0,2)
75	*Bos* (Cattle) (99,2), *Bubalus* (Buffalo) (0,3), *Sus* (Pig) (0,2), *Rupicapra* (Goat Antelope) (0,2) and *Redunca* (Reedbuck) (0,1)
79©	*Bos* (Cattle) (**95,9**), *Ovis* (Sheep) (3,6), *Rupicapra* (Goat Antelope) (0,2), *Bubalus* (Buffalo) (0,1), *Bison* (Bison) (0,1) and *Redunca* (Reedbuck) (0,1)
82	*Bos* (Cattle) (99,5), *Sus* (Pig) (0,1), *Ovis* (Sheep) (0,1), *Bubalus* (Buffalo) (0,1) and *Rupicapra* (Goat Antelope) (0,1)
83©	*Bos* (Cattle) (**74,7**), *Sus* (Pig) (24,7), *Ovis* (Sheep) (0,2), *Bubalus* (Buffalo) (0,2), *Rupicapra* (Goat Antelope) (0,1) and *Redunca* (Reedbuck) (0,1)
85	*Bos* (Cattle) (97,5), *Equus* (Horse/Donkey) (0,9), *Sus* (Pig) (0,8), *Ovis* (Sheep) (0,2), *Bubalus* (Buffalo) (0,2), *Rupicapra* (Goat Antelope) (0,2) and *Redunca* (Reedbuck) (0,1)
87	*Bos* (Cattle) (99,3), *Rupicapra* (Goat Antelope) (0,2), *Sus* (Pig) (0,1), *Bubalus* (Buffalo) (0,1) and *Redunca* (Reedbuck) (0,1)

Appendix A4. 11: Percentage average folds of DNA from retail meat products labelled as beef sausage (N=21)

Sample Number	Genus (% Of Average Fold)
S	*Bos* (Cattle) (98,7), *Ovis* (Sheep) (0,5), *Sus* (Pig) (0,4) and *Homo* (Human) (0,1)
103	*Bos* (Cattle) (99,2), *Sus* (Pig) (0,2), *Bubalus* (Buffalo) (0,2), *Rupicapra* (Goat Antelope) (0,2), *Bison* (Bison) (0,1) and *Redunca* (Reedbuck) (0,1)
104©	*Bos* (Cattle) (**63,4**), *Sus* (Pig) (36,1), *Ovis* (Sheep) (0,1), *Bubalus* (Buffalo) (0,1) and *Rupicapra* (Goat Antelope) (0,1)
114	*Bos* (Cattle) (98,4), *Bubalus* (Buffalo) (0,6), *Rupicapra* (Goat Antelope) (0,6), *Redunca* (Reedbuck) (0,2) and *Bison* (Bison) (0,1)
126©	*Bos* (Cattle) (**78,1**), *Ovis* (Sheep) (20,8) and *Capra* (Goat) (1,0)
127©	*Bos* (Cattle) (**69,2**), *Sus* (Pig) (30,4), *Bubalus* (Buffalo) (0,1) and *Rupicapra* (Goat Antelope) (0,1)
129	*Bos* (Cattle) (98,8), *Ovis* (Sheep) (0,9), *Sus* (Pig) (0,1), *Bubalus* (Buffalo) (0,1) and *Rupicapra* (Goat Antelope) (0,1)
134©	*Bos* (Cattle) (**67,5**), *Ovis* (Sheep) (17,7), *Sus* (Pig) (14,3), *Rupicapra* (Goat Antelope) (0,3) and *Bubalus* (Buffalo) (0,1)
135©	*Bos* (Cattle) (**38,8**), *Sus* (Pig) (60,1), *Ovis* (Sheep) (0,8), *Homo* (Human) (0,2) and *Mus* (Rat) (0,2)
143©	*Bos* (Cattle) (**93,6**), *Ovis* (Sheep) (5,6), *Bubalus* (Buffalo) (0,2), *Rupicapra* (Goat Antelope) (0,2) and *Redunca* (Reedbuck) (0,1)
144©	*Bos* (Cattle) (**85,7**), *Ovis* (Sheep) (13,6), *Bubalus* (Buffalo) (0,2), *Rupicapra* (Goat Antelope) (0,2) and *Redunca* (Reedbuck) (0,1)
145©	*Bos* (Cattle) (**94,4**), *Sus* (Pig) (3,7), *Ovis* (Sheep) (0,9), *Bubalus* (Buffalo) (0,3), *Rupicapra* (Goat Antelope) (0,3) and *Redunca* (Reedbuck) (0,1)
15©	*Bos* (Cattle) (**18,1**), *Sus* (Pig) (78,1), *Ovis* (Sheep) (3,5) and *Rupicapra* (Goat Antelope) (0,2)
155©	*Bos* (Cattle) (**85,7**), *Sus* (Pig) (9,9), *Ovis* (Sheep) (4,0), *Bubalus* (Buffalo) (0,1) and *Rupicapra* (Goat Antelope) (0,1)
156©	*Bos* (Cattle) (**79,0**), *Sus* (Pig) (20,5), *Ovis* (Sheep) (0,1), *Bubalus* (Buffalo) (0,1), *Rupicapra* (Goat Antelope) (0,1) and *Redunca* (Reedbuck) (0,1)
174	*Bos* (Cattle) (99,5), *Ovis* (Sheep) (0,1), *Bubalus* (Buffalo) (0,1), *Rupicapra* (Goat Antelope) (0,1) and *Redunca* (Reedbuck) (0,1)
178	*Bos* (Cattle) (99,7) and *Bubalus* (Buffalo) (0,1)
184©	*Bos* (Cattle) (**68,7**), *Ovis* (Sheep) (27,9) and *Sus* (Pig) (3,2)
186	*Bos* (Cattle) (88,2), *Ovis* (Sheep) (11,0), *Bubalus* (Buffalo) (0,3), *Rupicapra* (Goat Antelope) (0,3) and *Redunca* (Reedbuck) (0,1)
188©	*Bos* (Cattle) (**64,2**), *Sus* (Pig) (35,3), *Bubalus* (Buffalo) (0,2) and *Rupicapra* (Goat Antelope) (0,1)
189©	*Bos* (Cattle) (**68,8**), *Sus* (Pig) (30,1), *Ovis* (Sheep) (0,8), *Bubalus* (Buffalo) (0,1) and *Rupicapra* (Goat Antelope) (0,1)

APPENNDIX B: Ethics Approval Letter

Ethics Approval Letter