Investigation of alternative molecular approaches for enhanced traceability of illegally traded Temminck's Ground Pangolins (*Smutsia temminckii*)

by

Sean Paul Heighton

Submitted in partial fulfilment of the requirements for the degree

Master of Science in Zoology

In the Department of Zoology and Entomology
Faculty of Natural and Agricultural Science

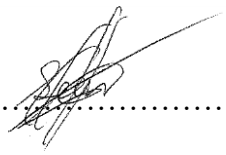
UNIVERSITY OF PRETORIA

December 2018

The financial assistance of the National Research Foundation (NRF) and Oppenheimer Memorial Trust (OMT) towards this research is hereby acknowledged. Opinions expressed, and conclusions arrived at are those of the author and are not necessarily to be attributed to the NRF or OMT

Declaration

I, **Sean Heighton** declare that the dissertation, which I hereby submit for the degree **Master of Science in Zoology** at the University of Pretoria, is my own work and has not previously been submitted by me for a degree at this or any other tertiary institution.

SIGNATURE: ……  ……

DATE: 12 December 2018

Investigation of alternative molecular approaches for enhanced traceability of illegally traded Temminck's Ground Pangolins (*Smutsia temminckii*)

Student: Sean Heighton [1]

Supervisors: Prof. Armanda D.S. Bastos[1], Dr. Darren W. Pietersen[1] and Dr. Grant Hall[1].

Department: [1]Mammal Research Institute, Department of Zoology and Entomology, University of Pretoria, Pretoria, 0002, South Africa

Degree: Master of Science in Zoology

Abstract:

Conservation regulatory bodies created to limit the overexploitation of wildlife are reliant on regulatory adherence. This over-reliance on regulation means that complementary approaches to assist with enforcement are often required. Wildlife forensics is one such approach that can be used to address the illegal trade in wildlife and pangolins in particular. With estimates from the legal trade being that over one million pangolins were trafficked in a ten-year period, it is clear that alternative approaches to curb illegal trade, which represents >98% of pangolin trade, are needed. Geographic traceability uses a range of markers and geo-referenced samples for tracking the origins of trafficked pangolins and their derivatives. In this study, the applicability of two molecular profiling approaches (stable isotopes and host microbes) were evaluated for their traceability potential, for three pangolin populations sampled from two sites in the Kalahari (xeric savanna) and one site in the Lowveld (mesic savanna).

The feasibility of carbon and nitrogen stable isotope traceability using pangolin scales was evaluated through mass spectrometry (Chapter 2). Results revealed that intra-scale and inter-individual variation are important considerations for traceability, whilst inter-scale

variation and scale harvesting methods (boiling or direct flaming of carcasses) are not. Given the high levels of intra-scale variation, development of a time and cost-effective sampling protocol suitable for forensic investigations was attempted. It was found that the use of duplicate samples of homogenised scales provides an averaged effect on scale variability. However, scale tips are comparatively isotopically enriched and should be taken into consideration during forensic analyses. Although preliminary results indicate that geographic separation based on isotopic ratios between Lowveld and Kalahari regions may be possible, geographic partitioning through use of stable isotopes could not be conclusively demonstrated due to the small sample size.

Baseline estimates of microbial diversity were established using two metagenomics approaches (cloning *versus* next generation sequencing (NGS)) targeting the 16S rRNA gene of bacteria and two taxonomic assignment methods (phylogeny testing *versus* a heuristic search tool; Chapter 3). These differed significantly with respect to microbial community composition with results suggesting that phylogeny testing in combination with NGS holds potential for pangolin traceability. Although whole microbial communities differed by 70.3% between the two Kalahari sites, constraints with respect to the number of localities sampled and the number of cloning colonies characterised are duly noted. The results also revealed that *Clostridium sordellii* and genospecies *C. novyi* sensu lato appear to be key site-specific taxa for presence/absence and topotype traceability methods, respectively.

Using a primer-set developed to target the phylogenetically-informative phospholipase C gene, the potential health risks and geographic traceability efficacy of the *C. novyi* sensu lato pathogenic genospecies group was investigated (Chapter 4). Although the prevalence was moderate at 17.65% and thus not feasible for traceability purposes, the potential public and animal health risks that the genospecies group poses, particularly for the illicit trade of pangolins, is of concern. Expanded studies using high throughput metagenomics and additional

stable isotopes hold promise for enhanced traceability of illegally traded pangolins, particularly when used together.

Keywords:

Geographic traceability, stable isotopes, microbial proxy tracing, pangolin, trafficking, regulatory bodies, *Smutsia temminckii*, *Clostridium novyi* sensu lato, forensics, public and agricultural health.

Acknowledgements

My utmost gratitude goes to my supervisors as without them this research would not have been possible. To Prof. Armanda Bastos, thank you so much for not only keeping this project afloat, but for being willing to help me in all of my research endeavours and crazy ideas outside of this project (I owe you so many articles in the coming months). I cannot thank you enough for making sure I had enough money to concentrate entirely on my M.Sc. and for always looking out for my best interests, even if it meant more work for you. To Dr Darren Pietersen, you have been invaluable to my understanding of pangolin conservation and have given me opportunities a young conservationist can only dream of. You have shown me what real hard work looks like and have been the best mentor I could have asked for, both within and outside of science. To Dr Grant Hall, thank you so much for putting in the hours to better my writing and interpretation skills. Even though this was frustrating at times, the lessons I have learnt are invaluable and I thank you for this. Thank you for welcoming me into your laboratory and for always being ready to teach me something related to stable isotopes with a whiteboard marker on the glass divider in front of the mass specs - how fitting. My sincerest thanks go to all three of you for working around the clock to submit this dissertation under a difficult deadline. I cannot thank you enough.

I would also like to thank Beryl Wilson from the McGregor Museum as well as Dr Jurie du Plessis and Dr Nico Avenant from the National Museum for going out of your way to provide the pangolin scale samples that proved to be so invaluable to this project. Thank you to Colin Rowles from Klaserie Private Nature Reserve as well as Gus van Dyk, Dylan Smith and Wendy Panaino from Tswalu Kalahari Reserve for your willingness to provide me with additional samples. A big thank you again to Darren for collecting the majority of these samples, without you creating a large sample size over the last decade, some of the techniques tested in this research would not have been possible. I would also like to acknowledge the staff

at the TADP BSL3 laboratory at Onderstepoort, the University of Pretoria's sequencing facility and the Department of Zoology and Entomology's molecular laboratory for providing a platform for research to be conducted at a high standard.

To my family and friends, thank you for being there to share my downfalls and triumphs. Thank you for understanding my absenteeism over the past few months and for lending your time and support. It has been a bumpy ride but keep buckled up, the PhD is on its way.

Last but not least, I would like to thank the Oppenheimer Memorial Trust and National Research Foundation (SARChI Chair M3B2 grant holder linked bursary) for giving me the opportunity to pursue this research on a full-time basis. A special thanks goes to Dr Duncan MacFadyen, Prof. Armanda Bastos and Prof. Jacek Banasiak for your invaluable contribution in attaining these bursaries.

Table of contents

List of figures

Chapter 1

Chapter 2

Chapter 3

Chapter 4

List of tables

Chapter 2

Chapter 3

Chapter 4

Chapter 1: Introduction

Rationale and background

There are currently ~7.3 billion human inhabitants that share the earth with a vast number and diversity of species, which includes an estimated 5 ± 3 million eukaryotes alone (Costello *et al.*, 2013). These species along with now extinct species have in some shape or form influenced the existence of our own species. However, through excessive human population growth, globalization and higher levels of disposable income, human beings are now the main drivers of biodiversity loss and have triggered what some consider to be the earth's sixth mass extinction event (Vitousek *et al*., 1997; Ceballos *et al*., 2010; Santos *et al*., 2011; Duckworth *et al*., 2012). This is underscored by extinction rates that are estimated to be 100–1 000 times higher than the pre-human background norm, and that are still increasing (Pimm *et al*., 1995; Hilton-Taylor *et al*., 2009). It has further been suggested that of all extant bird species, one in eight are at risk of extinction whilst the same can be said for approximately one-third of all amphibian species and nearly a quarter of all mammal species (Hilton-Taylor *et al*., 2009). As part of the Living Planet Index, McRae *et al*. (2017) noted that in the last 40 years there has been a 58% decline in vertebrate population abundance across 14 152 populations globally. This has placed added pressure on the survival of other species, the ecosystems we share and perhaps in due course, ourselves (Ceballos *et al*., 2010). One of the most significant environmentally adverse anthropogenic activities in recent times has been the international trade in wildlife, which, in addition to the effects of invasive species, causes ~30% of all global species threats (Lenzen *et al*., 2012).

International trade in wildlife

The term 'wildlife trade' extends to the sale and exchange of wild animal and plant resources by humans (Oldfield, 2003). Its annual market value based solely on import declarations is approximately US$ 300–332.5 billion with an estimated 5.5% yearly increase (Engler, 2008; Roe, 2008). This figure is predominantly due to timber and fishing industries which account for an estimated 80–90% of the total (Engler, 2008). However, this does not detract from the extent of trade in thousands of other species and their products that make up the remaining proportion.

Although the term 'wildlife trade' has an innate negative connotation in the minds of the public, its regulated structure combined with international pressure by governments, as well as international and regional trade regulatory bodies, in many cases prevents the overexploitation of wildlife. The best-known multilateral regulatory body of wildlife trade is the Convention on the International Trade in Endangered Species of Wild Fauna and Flora (CITES) which was established in 1975 in order to prevent the overexploitation of wildlife as a result of unsustainable levels of international trade (CITES, 2017a). This is a difficult undertaking considering the growing demand arising from the increased growth of the human population and the world economy (Warchol, 2004; Engler, 2008; Graham-Rowe, 2011). Regulatory bodies such as CITES were formulated by our growing consciousness of the overexploitation of wild animals and have since had some success stories (see: McNeill and Lichtenstein, 2003; Smith *et al*., 2008; Uscamaita and Bodmer, 2010; Coonan *et al*., 2014; Swaisgood *et al*., 2016). However, these successes are exceptions to the rule mainly due to the lack of enforcement by regulatory bodies and stakeholders, non-compliance to regulatory processes, lack of resources to devote to upkeep of regulatory agreements, lack of knowledge of species and market forces, political and emotional influence on regulatory decisions, and general faults with the design of the regulatory processes themselves (IUCN, 2000; Oldfield,

2003; Challender *et al*., 2015). These inherent flaws of regulatory conservation are compounded by the over-reliance on regulatory bodies, and may even lead to a growth in illegal trade due either to misconduct by the regulatory body or to the use of methods undertaken by governing bodies, industry or other stakeholders that aim to undermine regulatory processes of legal trade (Challender *et al*., 2015). Examples include renaming and mislabelling of seafood in order to circumvent fishing regulations (Jacquet and Pauly, 2008), the selling of illegally sourced products such as rosewood or sturgeon species in legal markets (UNODC, 2016), the increase in rhinoceros poaching due to increased market price after the introduction of more stringent trading regulations (Rivalan *et al*., 2007; Abensperg-Traun, 2009) and the increase in African Elephant (*Loxodonta africana*) poaching after the lifting of the ivory trade ban in the late 1990s (Hitch, 1998). The pangolin trade too has examples of methods that have been adopted to undermine regulatory processes. These include (i) the sale of pangolin products confiscated from the illegal trade at legal auctions as legal products by law enforcement authorities, (ii) the illegal sale of pangolin products to Traditional Chinese Medicine (TCM) retailers by authorized pharmaceutical companies that are supposed to provide patented pangolin products to designated TCM hospitals only, and (iii) the use of supposed pangolin breeding farms as a cover for the illegal trade (Xu *et al*., 2016; Challender and Waterman, 2017). The latter as a more recent approach whereby countries claim to house pangolins in these breeding farms, which under CITES law permits trade from the F2 generation onwards if the founding population was sourced legally (CITES, 2017b). However limited evidence of these breeding farms has been provided (Xu *et al*., 2016; Challender and Waterman, 2017). Pangolins are also notoriously difficult to breed in captivity (Hua *et al*., 2015), suggesting that individuals that are claimed to have originated from these breeding programmes are most likely being illegally sourced from the wild and traded under the guise of captive breeding. These examples highlight the problems that can arise when there is a fault with the design of the

regulatory process itself. It is evident that complementary conservation approaches should be considered in order to reduce the over-reliance and thus impact that regulatory flaws may have on species or ecosystem conservation.

Overexploitation through unsustainable levels of trade does not only adversely affect wildlife populations, but inherently has the potential to have an adverse, indirect effect on the ecosystems in which these populations reside (Cardinale *et al*., 2012). Growing evidence suggests that biodiversity loss has a major impact on ecosystems in general, particularly when the species affected play a major role in the physical structuring, productivity and biogeochemical cycling of ecosystems (Cardinale *et al*., 2012). It reduces the efficacy with which an ecosystem can recycle and decompose essential nutrients, produce biomass and capture resources - its so-called ecosystem functioning (Cardinale *et al*., 2012). An ecosystem functioning at reduced levels may have dire consequences to its wildlife, domestic animal and human inhabitants as ecosystem services are influenced by ecosystem health. These services are divided into four functioning groups by the Millennium Ecosystem Assessment (2005), *viz.* provisioning (fuel, water and food), support (soil formation and primary production), regulation (climate, disease and water) and culture (educational, aesthetic, recreational and spiritual).

Although much monetary wealth is generated in the short-term from the overexploitation of wildlife through trade, the effect it has on ecosystem functioning alone may exceed these monetary gains in the longer term (Millennium Ecosystem Assessment, 2005). It is estimated that in southern Africa alone, nature-based tourism accounts for half of the total tourism revenue, an estimated US$ 3.6 billion per annum (Scholes and Biggs, 2004). If this revenue is affected by species or population extinctions owing to the trade-related overexploitation of wildlife, it may have negative implications for the southern African economy. Similarly, when considering the potential for trade-related invasive species

introduction, which is suggested to be driven globally by international trade (Westphal *et al*., 2008), the associated costs are estimated to be US$ 120 billion annually for the United States of America alone (Pimentel e*t al*., 2005). Trade in animals in general also bear another cost related to public and agricultural health. The cost of disease transmission and spread due to trade of animals amounted to more than US$ 20 billion in direct and US$ 200 billion in indirect economic losses, between 2000–2010 (World Bank, 2010). The monkeypox epidemic in the United States of America (CDC, 2003), rinderpest epidemic in Africa (Daszak *et al*., 2000) and the Severe Acute Respiratory Syndrome (SARS) epidemic in China (Bell *et al*., 2004) are examples of disease epidemics that originated from animal trade and exemplify the need for well-defined and regulated trade. Although there are many examples of overexploitation under regulatory processes, a bigger problem arises when these regulatory processes are not enforced, and the exploitation of wildlife surpasses legal limits.

Illegal wildlife trade

Illegal wildlife trade forms part of trans-national crime, which includes drug and human trafficking, terrorism and smuggling of art or antiques (Warchol, 2004). The illegal wildlife trade is unregulated and estimated to have an annual market value of US$ 7–23 billion, with 132 144 seizures being recorded in 120 countries between 2004 and 2014 (Nellemann *et al*., 2016; UNODC, 2016). The large estimate range in market value reflects the inherent difficulties in quantifying this illicit trade due to its clandestine nature and lack of discrimination between what constitutes legal *versus* illegal trade (Duffy, 2016). These statistics are also only a fraction of the actual amount of trade occurring (Heinrich *et al*., 2016), giving a larger appreciation for the sheer level of illicit trade in wildlife. This level of trade is in many cases unsustainable as seen with the population decline of tigers (*Panthera tigris*) for ethnomedicinal purposes (Goodrich *et al*., 2015), great whales for food (Springer *et al*., 2003),

boas and pythons as fashion accessories (Pough *et al*., 1998), cycads for ornamental purposes (Warchol, 2004) as well as primate and bird species for use as pets (BirdLife International, 2000; Duarte-Quiroga and Estrada, 2003). Harvesting individuals from wild populations not only reduces genetic variation, but also alters population subdivisions or induces selective changes in a population, which in turn places added pressure on populations and may lead to an increased risk of extinction (Allendorf *et al*., 2008). This can, and has, progressed to the point of complete species or population extirpations partly due to illegal trade and poaching as in the case of a subspecies of Javan Rhinoceros (*Rhinoceros sondaicus annamiticus*) from Vietnam (Brook *et al*., 2012), the Miss Waldron's Red Colobus (*Piliocolobus badius waldronae*; Oates *et al*., 2000, although contested by Wagner, 2004), Spix's Macaw (*Cyanopsitta spixii*; Thomsen and Munn, 1988; Juniper and Yamashita, 1990) and the Western Black Rhinoceros (*Diceros bicornis longipes*; Emslie, 2011).

Although overexploitation occurs in the wildlife trade in general, the illicit trade in wildlife has far more wide-ranging consequences, as its unregulated nature results in a higher probability of complete species or population extirpation as seen with the aforementioned examples. This may increase the frequency with which ecosystem functioning and services are reduced. The chance of establishment of invasive species through illegal trade is also greater due to its unregulated nature (Maynard and Nowell, 2009). Illegal trade in animals also bears a greater risk of disease transmission and spread (Fèvre *et al*., 2006) as was the case with livestock trade from Somalia when the country's political instability resulted in increased unregulated trade of animals to other countries and thus the spread of a number of livestock-related diseases, with major economic implications for all trade partners (Food and Agriculture Organization, 2004).

The reality is that overexploitation is still occurring despite regulatory processes and international pressure. However, there is a reason why this trade is difficult to stop:

approximately 0.2–1 billion people are believed to use wildlife resources, whilst trade in wildlife is a major source of income for the poor across the globe (van Rijsoort, 2000; Roe, 2008). If regulations do not consider the cultures and circumstances of the people who live closest to the wild species or areas that they intend to conserve as well as the culture and circumstances of the people that create the demand for overexploitation of wildlife, these regulations are unlikely to lead to sustainable trade (Hutton and Webb, 2002; Oldfield, 2003). Unfortunately, this is a long-term goal as changing people's perceptions, cultural belief systems and circumstances is often seen as a transgenerational, long-term process. However, time is running out in many instances and thus more direct conservation approaches need to take priority in the interim.

The case of the pangolin

Pangolins are predominantly nocturnal mammals with an outer body covering of keratin scales (Speaman *et al*., 1967; Swart, 2013; Liu *et al*., 2016). They feed on ants and termites (myrmecophagous; Swart, 2013; Pietersen *et al*., 2016) and make up the monophyletic order Pholidota which diverged from its sister order Carnivora roughly 56.8–67.1 million years ago (Gaudin *et al*., 2009; du Toit *et al*., 2014; Choo e*t al*., 2016; Gaubert *et al*., 2017). The order is comprised of eight species from three genera, namely *Manis* (all four Asian species), *Phataginus* (two small, predominantly arboreal African species) and *Smutsia* (two large, terrestrial African species) which are found in tropical to inter-tropical zones of Africa and Asia (Figure 1.1; Gaudin *et al*., 2009; Swart, 2013; Gaubert *et al*., 2017). The Temminck's Ground Pangolin (*Smutsia temminckii*) has the largest distribution of the African pangolin species, ranging from south-eastern Chad and South Sudan southwards through East Africa to the northern regions of South Africa (Figure 1.1; Swart, 2013; Pietersen *et al*., 2014c). It is listed as globally Vulnerable on the IUCN Red List™ (Pietersen *et al*., 2014c), although a

reassessment of this species' conservation status incorporating data published subsequent to 2014 is warranted, and is likely to result in the up-listing of this species to Endangered (D.W. Pietersen, *pers. comm.*).

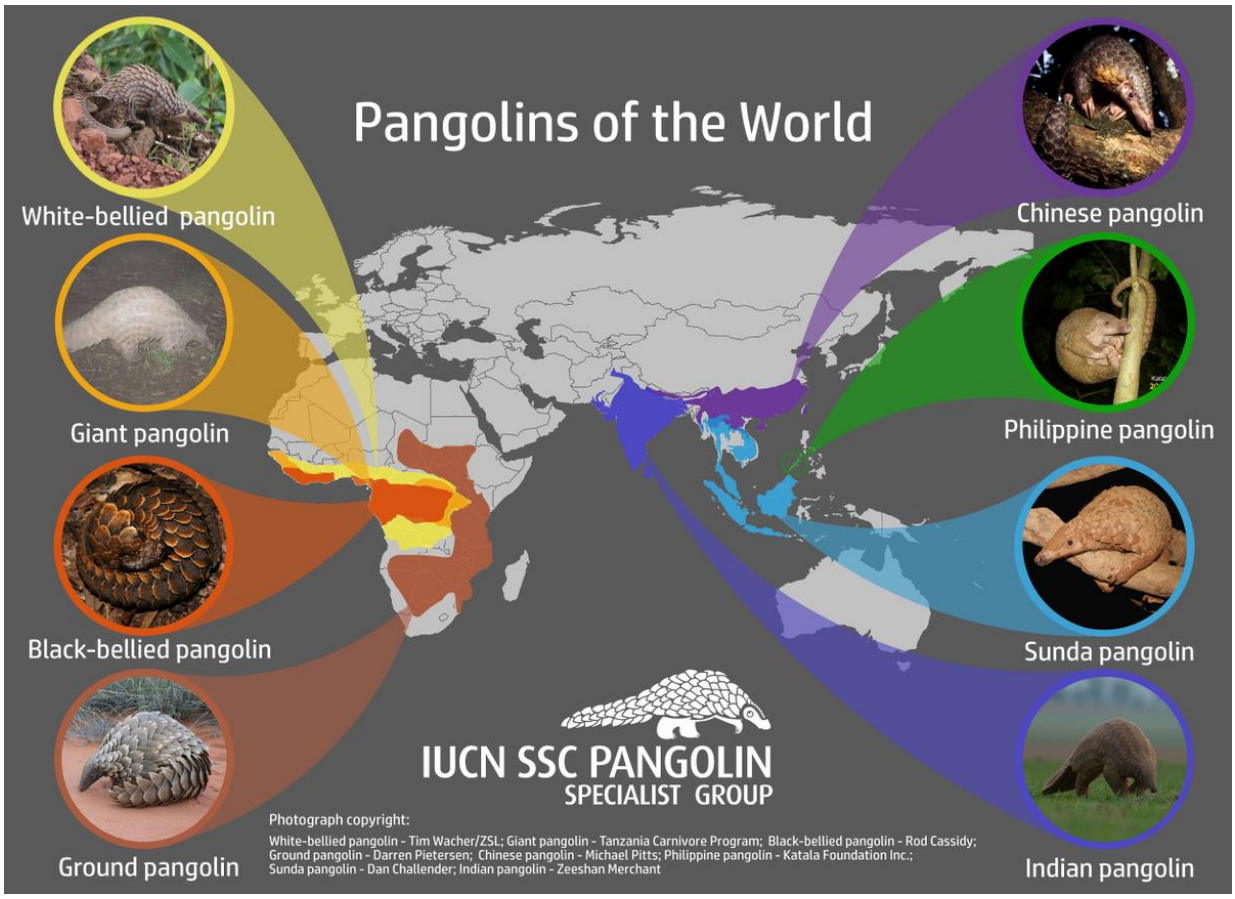


Figure 1.1: The geographic distribution of the eight extant pangolin species in Africa and Asia. The White-bellied (*Phataginus tricuspis*), Black-bellied (*P. tetradactyla*), Giant Ground (*Smutsia gigantea*) and Temminck's Ground (*S. temminckii*) Pangolins occur in Africa whilst the Chinese (*Manis pentadactyla*), Philippine (*M. culionensis*), Sunda (*M. javanica*) and Indian (*M. crassicaudata*) Pangolins occur in Asia. Image sourced from the IUCN Species Survival Commission Pangolin Specialist Group.

With the recent up-listing of all eight pangolin species from CITES Appendix II to Appendix I (CITES, 2017b), trade, both legal and illicit, has been recognised as a serious threat to these species' future survival. The illegal trade of pangolins between 2004 and 2014 was estimated to exceed one million individuals (Challender *et al.*, 2014) with pangolins constituting 5% of wildlife seizures documented in the World WISE database between 2005 and 2014, 2% more than that of rhinoceroses (UNODC, 2016). This has resulted in pangolins

being labelled the most heavily trafficked wild mammal in the world (Challender *et al*., 2014). The number of pangolins traded legally between 2007 and 2013 was 1 467, but the number seized from the illegal trade was 107 060 (UNODC, 2016), which essentially equates to a maximum 1.37% of the total pangolin trade being legal prior to the 2017 CITES uplisting. This indicates that conservation authorities have little control over the pangolin trade through regulation.

Recent research has suggested that Asian pangolin populations are dwindling, thus resulting in an increased demand for African pangolins and a spate in illegal trade of these populations (Challender and Hywood, 2012; Pietersen *et al*., 2014a,c; Heinrich *et al*., 2016; Pietersen *et al*., *in prep.*). Between 2012 and 2017 the four African pangolin species has seen a record high of up to 1.1 million individuals being traded illegally (Pietersen *et al*., *in prep*.) thus surpassing the initial figure of all eight pangolin species (Challender *et al*., 2014) in half the time. Between 2015 and 2016 there has been a 300% (6 300–18 923 kg) increase in the volume of scales traded illegally out of Africa (APWG, *unpubl. data*). Not only does this pose a risk for the continued survival of the four African pangolin species, but it may have a detrimental effect on the cultural significance for many African ethnic groups (Manwa and Ndamba, 2011; Soewu and Adekanola, 2011; Boakye *et al*., 2014, 2015). The integrity of African conservation authorities, impact on ecological processes (Cooney *et al*., 2015) and the risk of disease spread are also at stake (Smith *et al*., 2009b). In a preliminary assessment of microbial diversity, several bacterial species of epidemiological importance were found to be harboured by *S. temminckii*, some of which have zoonotic potential (Heighton, 2016). Furthermore, increased trade of stressed populations may increase the chance and rate of spread of diseases, potentially resulting in epidemics (Woodford, 2000; Godoy and Matushima, 2010).

Despite being the most heavily trafficked group of wild mammals, there are currently no mechanisms by which law enforcement or conservation authorities can accurately determine

the origin of trafficked pangolins. Using the origin of shipments to infer collecting sites is insufficient, as collectors and traders are known to smuggle pangolins and their derivatives across borders to exploit countries with more lax laws or law enforcement, or that have better infrastructure (especially harbours and airports) by which their consignments can be shipped (Heinrich *et al*., 2017; Mambeya *et al*., 2018; Pietersen *et al*., *in prep.*). This along with the limited control that regulatory bodies have on pangolin trade suggests that conservation initiatives with more direct outcomes should be investigated to complement the issues raised.

Geographic traceability through wildlife forensics

Of the various direct conservation initiatives available, wildlife forensics has received increased attention in the past few years, including at the recent Conference of the Parties to CITES (CoP17; Ogden and Linacre, 2015; Nellemann *et al*., 2016; TRAFFIC, 2016). Wildlife forensics is a discipline that incorporates analytical techniques in order to investigate wildlife crime, encapsulating everything from pathology and forensic entomology to ballistics of wildlife crime cases (Huffman and Wallace, 2012). Although there are many uses for forensic science, from determining the type and time of death to the determination of a species or individual that has been trafficked, the geographic traceability or assignment of an individual to origin is less refined but is becoming increasingly requested by investigators (Huffman and Wallace, 2012; Ogden and Linacre, 2015). Geographic assignment can be used to determine whether animals were legally or illegally sourced but can also be used to identify poaching hotspots, determine syndicate networks or even identify areas best-suited for repatriation of live individuals seized from trade (Fernandes and Caparroz, 2013; Ogden and Linacre, 2015).

Molecular forensics of traded individuals and derivatives is the most well-established and widely applied forensic technique owing to the customary precedent of accepting DNA evidence in court cases and the ubiquity of DNA forensic facilities (TRAFFIC, 2016). The

approach has been used in determining source populations of illegally traded products including cannabis (*Cannabis sativa*; Gilmore *et al*., 2007), Gopher Tortoises (*Gopherus polyphemus*; Schwartz and Karl, 2008), African Rhinoceroses (*Ceratotherium simum* and *Diceros bicornis*; Harper *et al*., 2013) and African Elephants (*Loxodonta africana*; Wasser *et al*., 2004). Research on pangolin traceability based on phylogeographic structuring of populations (du Toit, 2014; Gaubert *et al*., 2016) or identifying geographic variation in trade seizures (Zhang *et al*., 2015) has delivered varying levels of success. This is because the level of population assignment resolution and accuracy is based on the type and number of genetic markers used as well as the range and resolution of population samples obtained when developing phylogeographic traceability maps (Ogden and Linacre, 2015). The problem often faced with wildlife traceability of rare species is the undesirable effect that inbreeding and genetic drift have on the resolution of neutral genetic markers for traceability purposes (Arenas *et al*., 2017). This often leads to the need for markers that target geographically-linked traits which are not well characterized unless genome-wide studies have previously been conducted (Arenas *et al*., 2017). However, this sort of genomic research on wild animals is lagging compared to that of domestic animals, humans and diseases. This is attributed to less funding being available for wildlife research and the difficulty in obtaining good quality samples representative of the variation of entire populations (Arenas *et al*., 2017). It also explains why only two whole genomes of pangolins (both of Asian species) have been published to date (Choo *et al*., 2016).

DNA samples from wildlife forensic cases are often degraded or at low-copy which reduces the efficacy of obtaining sufficient DNA for molecular techniques (Arenas *et al*., 2017). Mitochondrial DNA can to some extent overcome this problem owing to its high copy number per cell as well as its circular genome structure which prevents degradation through exonucleases (Foran, 2006). However, relying solely on mitochondrial DNA dramatically

reduces the number of potentially informative markers available compared to that of an entire nuclear genome. This impacts the resolution, as observed in a phylogeographic study on *S. temminckii* from southern Africa which made use of three mitochondrial gene markers, *viz*. cytochrome *b*, cytochrome oxidase subunit 1 and the displacement loop/control region (du Toit, 2014). This technique was only able to separate populations from four southern African countries into two separate clades, namely a western clade comprising South Africa (Northern Cape Province) and Namibia, and an eastern clade inclusive of South Africa (Limpopo Province), Zimbabwe and Mozambique (du Toit, 2014).

Using microbial forensics as proxies for traceability, particularly in human cases, is a rapidly growing field which can overcome the aforementioned limitations of host forensics (Arenas *et al*., 2017). The advantage of microbial DNA is that it is present at high levels, with an average of 10 times more bacteria than human cells in a human body (Sender *et al*., 2016). This also introduces a wealth of additional traceability biomarkers which can complement host traceability efforts (Hampton-Marcell *et al*., 2017). Mammals (including pangolins) have significantly lower per base pair variation between individuals of the same species when compared to microbes, due to longer generation times, sexual replication and smaller populations, which in turn reduces the genome-wide mutation rates of a population (Drake *et al*., 1998, Baer *et al*., 2007). Although microbes usually replicate asexually resulting in static clonal lineages (which are nevertheless still more diverse than animal populations over a given time period), when sexual replication does occur, recombination or horizontal gene transfer is ubiquitous resulting in lineage mixing and the possible formation of new strains or taxa (Xu, 2006). This transfer of genetic material can go as far as whole plasmid transfers which can change the behaviour and pathogenicity of bacteria in a single generation (Skarin and Segerman, 2014). Furthermore, these changes can often be identified due to well characterised gene panels and the ability to sequence whole microbial genomes at lower cost due to their

genomes being significantly smaller than that of their host (Arenas *et al*., 2017), thus providing added variability and subsequent resolution to microbial traceability maps. The added benefit of microbial source tracing is that it is not limited to a single species but can also be applied to a community of species. Locality-specific microbial community signatures arise due to interactions which influence both the presence/absence of a species in a community, but also abundance in a particular host and environmental setting (Hampton-Marcell *et al*., 2017).

An alternative technique to geographic traceability is the use of stable isotope analysis which can provide accurate, high resolution, locality-specific data owing to its reflection of environmental-specific influences (West *et al*., 2006; Ben-David and Flaherty, 2012; Chesson *et al*., 2018). An advantage that stable isotope forensics has over DNA-based approaches is its robustness as samples can undergo extreme conditions (e.g. exposure to high temperatures) and can still be viable for analyses using certain isotopes (Harbeck *et al*., 2011). Old material can be used as in the case of archaeological stable isotope research (Bocherens, 2009), and the technique requires small quantities of a sample for analysis (< 600 µg; Viljoen *et al*., 2016). Stable isotope analyses may also have a faster turn-around time and are more cost effective than molecular techniques (Viljoen *et al*., 2016). There are a range of isotopes (e.g. $^{13}C/^{12}C$, $^{15}N/^{14}N$, $^{2}H/^{1}H$, $^{18}O/^{17}O/^{16}O$) that may be measured and can improve the geographic resolution as each one has a set of environmental factors that influence its signature (Rubenstein and Hobson, 2004; Viljoen *et al*., 2016; Chesson *et al*., 2018). Stable isotope analyses can be performed on samples that do not possess DNA which is vital for the tracing of some animal trade related samples such as hair or soil (Viljoen *et al*., 2016). The use of continuously growing biochemically inert samples (do not change in isotopic ratios over time) such as nails or hair may provide a chronological isotope record that may trace the movement of traded animals, helping to identify potential trade routes (West *et al*., 2006; Ben-David and Flaherty, 2012; Viljoen *et al*., 2016).

In light of the complementary factors associated with microbial and stable isotope forensics, along with research on pangolin (host) DNA traceability already underway by six research groups, research on stable isotope and microbial diversity techniques offer a means of complementing these ongoing pangolin traceability initiatives.

Project Aim and Objectives

The feasibility of using alternative molecular approaches for geographic traceability of *S. temminckii* will be assessed in order to:

1. Test potential sources of variation relating to the use of stable isotopes for geographic traceability of *S. temminckii* using scale samples, as well as provide preliminary traceability results for three South African study sites using carbon and nitrogen stable isotopes.
2. Test the effect that different taxonomic assignment methods and metagenomics approaches have on microbial community identification, which will facilitate the identification of key taxa that may be useful for traceability of *S. temminckii*.
3. Test the feasibility of using the microbial genospecies *Clostridium novyi* sensu lato as a proxy for the geographic traceability of *S. temminckii* between two South African study sites, and the zoonotic implications of this bacterial group in the pangolin trade.

Study sites

Scale samples were collected from 37 *S. temminckii* individuals at two study sites in the Kalahari (Northern Cape Province) and one site in the Lowveld (Limpopo Province; Figure 1.2). Spleen, lung and liver samples were collected from 34 deceased *S. temminckii* individuals from the same two sites in the Kalahari whilst additional samples were collected from a

confiscated *S. temminckii* individual that had succumbed to its injuries following its retrieval from the illegal trade in 2016. Individuals sampled, apart from a few specific individuals, were assigned a sample label (STEM - referring to the species *S. temminckii*), followed by a unique identifying number for sample identification (Appendix 1). Samples were collected under approval from the University of Pretoria's animal ethics committee (reference number: ECO18-17 with Amendment 1) and the Threatened or Protected Species permit (reference number: S02658) from the Department of Environmental Affairs (DEA).

The two Kalahari sites, namely Kalahari Oryx Private Game Farm and Tswalu Kalahari Reserve are situated approximately 100 km apart and form part of the more xeric, Kalahari savanna biome (Acocks, 1988; Rutherford and Westfall, 1994). Klaserie Private Nature Reserve is situated approximately 1 000 km north-east of the two Kalahari sites and forms part of the more mesic, Lowveld savanna biome (Acocks, 1988; Rutherford and Westfall, 1994).

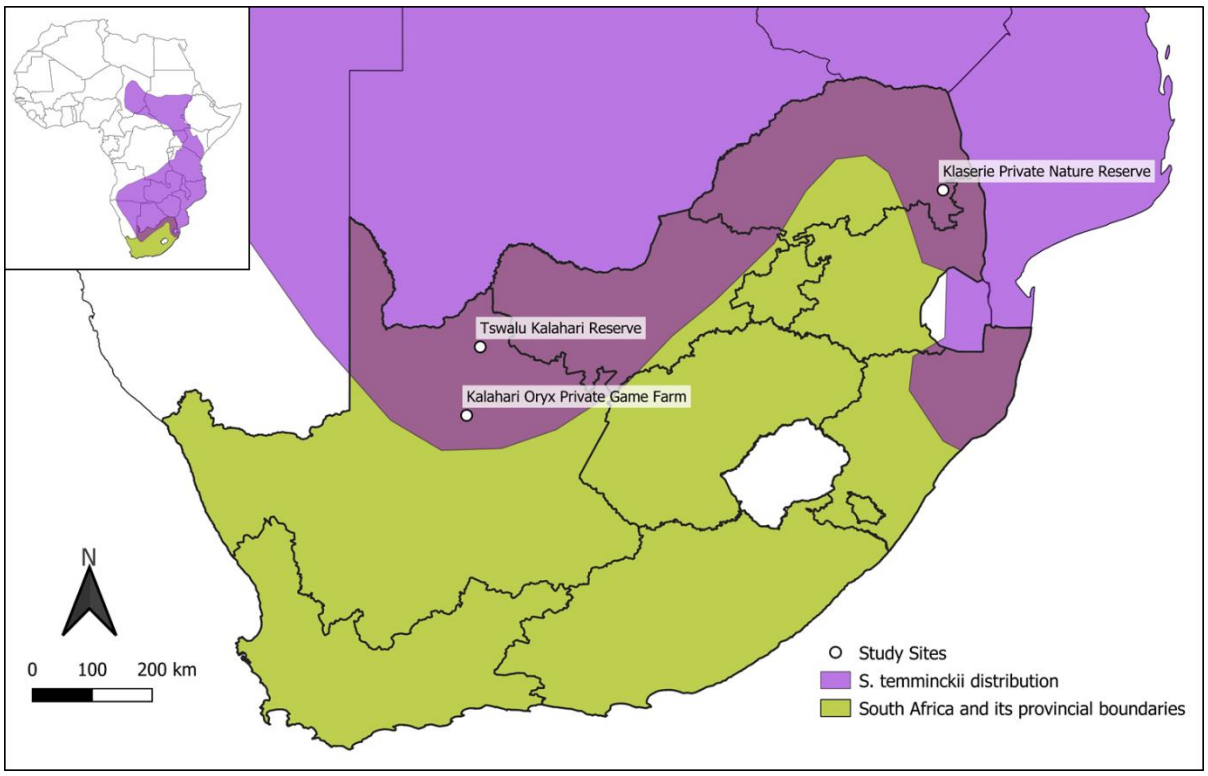


Figure 1.2: The three study sites in the Northern Cape and Limpopo Provinces of South Africa (green shaded area) and their relative positions within the known distribution of the Temminck's Ground Pangolin (*Smutsia temminckii*).

Chapter 2:

Isotopic variability of Temminck's Ground Pangolin (*Smutsia temminckii*) scales and their utility for geographic traceability

Introduction

The use of isotope ratio mass spectrometry and stable isotopes has increasingly been identified as a potential source of evidence for forensic cases (Gentile *et al.*, 2015; Dunn *et al.*, 2017) based on the method having large quantifiable ranges in values, being repeatable across laboratories, and in many cases being well understood in terms of the natural processes that lead to variability in isotope values (Cerling *et al.*, 2016). Such cases include money counterfeiting (Cerling *et al.*, 2016), food and beverage authentication (Gonzalvez *et al.*, 2009), plant-based drug trade (Booth *et al.*, 2010; Mallette *et al.*, 2016), bioterrorism threats (National Research Council, 2011) and even cases involving murder (Meier-Augenstein and Fraser, 2008) and war (Bartelink *et al.*, 2014). Stable isotope analyses have also been used in cases pertaining to the illegal trade of wildlife including distinguishing wild-caught from captive-bred animals (Serrano *et al.*, 2007; Kelly *et al.*, 2008; Retief *et al.*, 2014; Symes *et al.*, 2017; Alexander *et al.*, 2018), and determining the geographic origin and tracking of illegally poached and traded animals (Ehleringer *et al.*, 2008; Ben-David and Flaherty, 2012).

Background to stable isotope analyses

Isotopes are atoms of an element that have the same number of protons and electrons but differ in atomic mass due to differences in the number of neutrons (McKechnie, 2004; Sulzman, 2007). Stable isotopes are defined as isotopes of elements that are stable or non-radioactive, as in the case of $^{13}C/^{12}C$ and $^{15}N/^{14}N$ (Sulzman, 2007). Stable isotope analyses rely on the fact that these stable isotopic variants of an element, which are referred to as light or heavy isotopes, differ in relative abundance across various natural gradients due to different biological and physical processes (Hobson and Wassenaar, 1996; McKechnie, 2004). These gradients include altitude, latitude and longitude (Hobson and Wassenaar, 1996; McKechnie, 2004). These relative abundances can be measured as a ratio which can be used to characterize

specific food, water and habitat types across these gradients (Hobson and Wassenaar, 1996; McKechnie, 2004; West *et al.*, 2006). Changes in isotope ratios via biological and physical processes are known as fractionation when referring to a single reaction (i.e. evaporation, diffusion or an enzymatic reaction; Ben-David and Flaherty, 2012) and discrimination when referring to multiple reactions, of which some are not always known (i.e. enzymatic reactions; Cerling and Harris, 1999). Drivers of fractionation or discrimination, and thus isotopic variability, include differences in climate, food chain content and length, substrate composition, and tissue type (Bowen *et al.*, 2005). In order to determine isotopic variability across environmental gradients, the most suitable set of stable isotopes with regards to the question being investigated should be used with cognisance of their innate variability at the base of a food chain noted (Rubenstein and Hobson, 2004; Ben-David and Flaherty, 2012). To understand the variability of an animal's isotopic values, understanding sources of this variation in the terrestrial environmental (soil, water, air) and basal food chain (plant-based) levels of trophic ecology, is important. I shall use the term "background variability" when referring to terrestrial environmental and basal food chain variability. I shall also focus on carbon ($\delta^{13}C$) and nitrogen ($\delta^{15}N$) isotopes in order to give an indication of how their variability across a geographic scale can be used to separate geographically distinct pangolin populations.

Carbon and nitrogen background variability

Background variability in $\delta^{13}C$ values come from three main factors, namely the geographic location of atmospheric CO_2, climatic variability and variability in plant photosynthetic cycles (Ben-David and Flaherty, 2012; Chesson *et al.*, 2018b). Plant photosynthetic cycles also significantly affect the $\delta^{13}C$ values of a food chain with mainly grass species using Calvin photosynthetic cycle (C_3), woody species using the Hatch-Slack cycle (C_4) and succulent species using the Crassulacean acid metabolism (CAM) cycle (McCarron

and Knapp, 2001; Edwards and Ogburn, 2012). In southern Africa, both C_4 and C_3 grass species can be found, however C_4 grasses are by far the most prevalent across the landscape, particularly in drier and hotter areas such as the Kalahari (Vogel *et al.*, 1978; Ellis *et al.*, 1980). Plants utilizing the C_3 cycle range from -35‰ to -20‰, plants utilizing the C_4 cycle range from -18‰ to -10‰ and those using the CAM cycle range between C_3 and C_4 values due to the cycle following the C_3 pathway during the day and C_4 pathway at night (Dawson *et al.*, 2002; Marshall *et al.*, 2007; Tipple and Pagani, 2007). However, not all plant species are obligate regarding the photosynthetic cycle used as some CAM plants can be facultative in the photosynthetic cycle adopted (Marshall *et al.*, 2007). There is also currently no way to isotopically discriminate between C_4 and CAM plants using carbon (Marshall *et al.*, 2007). Utilizing additional isotopes such as deuterium (δD), however, does provide suitable means of isotopic distinction between these plant photosynthetic cycles (Sternberg *et al.*, 1984). In terms of climatic variability, natural gradients in temperature and partial pressure from the equator to the poles as well as low to high elevations result in $\delta^{13}C$ gradients with altitude and latitude, respectively (Morecroft and Woodward, 1990; Körner *et al.*, 1991; Kelly, 2000). The distribution of these plants is also important with C_4 plants known to decrease with increased altitude and latitude (Tieszen *et al.*, 1979), whilst differences in $\delta^{13}C$ values between mesic and xeric habitats due to water stress and light intensity are well recognised in C_3 plants (Farquhar *et al.*, 1989; Kelly, 2000; Diefendorf *et al.*, 2010; Kohn, 2010) but not as clearly in C_4 plants (Swap *et al.*, 2004). Importantly, carbon found in an animal's tissue is sourced through multiple macromolecular dietary sources including protein (amino acids), fatty acids, lipids and cholesterol (Koch, 2007). This can play a role in the resultant isotopic composition of an animal.

Background variability in $\delta^{15}N$ values is attributed to chemical and biotic variations in soil as well as the root depth of plants (Ben-David and Flaherty, 2012; Craine *et al.*, 2015;

Chesson *et al.*, 2018b). As with carbon, xeric and mesic environments are also distinguishable by nitrogen isotope ratios (Heaton *et al.*, 1986; Swap *et al*., 2004; Murphy and Bowman, 2009). Although the process driving it is unclear, nitrogen isotope ratios can also be used to differentiate between colder and warmer environments (Craine *et al.*, 2015). Prevailing climatic conditions affect the relative concentration of soil nitrogen, which in turn affects fractionation values during plant uptake, the level of microbial biomass (enriched ^{15}N), and the relative abundance of various forms of inorganic and organic nitrogen (Evans, 2001; Craine *et al.*, 2015). Unlike carbon, nitrogen in an animal's tissue is almost exclusively sourced through dietary protein (Koch, 2007).

Geographic traceability using stable isotopes

Geographic traceability through stable isotopes is a method in which investigators use isotopic variability across a geographic gradient or area to infer the origins of animals (Hobson, 2007; Chesson *et al.*, 2018a; Vander Zanden *et al.*, 2018a). This can be achieved by using one of two approaches. The first is to obtain geographically referenced isotopic values of basal items in a food chain (i.e. stable isotopes found in water, soil or plant material) and infer where an animal of unknown locality came from through these data (Bowen *et al.*, 2007), whilst the second is to obtain isotopic values of individuals of known locality and use these values as a reference to infer where other individuals of the same species originated from (Vogel *et al.*, 1990).

From the above terrestrial environmental variability in carbon and nitrogen isotopic ratios from different gradients (altitude, latitude and longitude), along with plant-based variability of photosynthetic pathways and nitrogen fixation, a mosaic of isotopic signatures across a geographic range are realised. This phenomenon has been defined as an isoscape, or isotopic landscapes, which was envisioned by West *et al.* (2006) as a way to create a range of

maps showing the distribution of isotopic variation across a landscape. These maps relate to the first approach of geographic traceability as they may provide a means to determine the source or origin of an object or organism across its geographic range as these background isotopic differences are incorporated into the food-web (West *et al.*, 2006; Chesson *et al.*, 2018b). For isoscapes to be used effectively, background isotope data must be determined for a designated region (Post, 2002; Newsome *et al.*, 2012). The fractionation factors, a factor by which fractionation of an isotope occurs, for the species under investigation and the respective tissues being analysed are also required, if not already known (Post, 2002; Newsome *et al.*, 2012). This can be problematic considering the number of assumptions and measurements needed to determine accurate fractionation factors of an organism's tissue. An example of this is of how δD and $\delta^{18}O$ values of an organism may change due to the evaporation of body water and exhalation of CO_2 through the skin, mouth and nasal cavity (Wong *et al.*, 1988). Excluding these factors during fractionation calculations may result in an overestimation of δD and underestimation of $\delta^{18}O$ values (Wong *et al.*, 1988). The same is true for $\delta^{13}C$ and $\delta^{15}N$ values whereby different volumes of protein and water intake can result in different fractionation values (Ambrose and DeNiro, 1986; Sealy *et al.*, 1987; Pearson *et al.*, 2003; Sponheimer *et al.*, 2003a; 2003b). These requirements make the initial phase of geographic traceability through isoscapes expensive, time consuming and inevitably less accurate. However, it has been successfully adopted in a number of studies including the mapping of δD and $\delta^{18}O$ from tap water as well as $^{87}Sr/^{86}Sr$ from bedrock and catchment water across the United States of America (Bowen *et al.*, 2005; Bataille and Bowen, 2012). It has also been adopted for the mapping of δD and $\delta^{18}O$ from precipitation across the entire globe and validated as a forensic application by tracing birds using their feathers (Bowen *et al.*, 2007).

One way to reduce variability in geographic traceability is to use the second approach, which is to measure isotopic ratios from a specific tissue type from individuals of known

geographic locations and use these data as a reference for deducing the geographic origin of individuals of unknown origin (Hobson, 2007; Vander Zanden *et al.*, 2018a). In a study by Vogel *et al.* (1990), bone and ivory samples of African Elephant (*Loxodonta africana*) populations across southern Africa were characterised using a combination of carbon, nitrogen and strontium isotope data. The isotopic resolution was high enough to distinguish populations within the Kruger National Park into three different groups, namely North, South-East and South-West (Vogel *et al.*, 1990). These results were mirrored in a similar study by van der Merwe *et al.* (1990), however these latter authors also employed $\delta^{13}C$ values to determine whether elephants mostly occupied forest or savanna habitats. These case studies were the first to investigate the use of stable isotopes for the traceability of wildlife in general, but also lead to a wealth of research on elephants specifically (see: Koch *et al.*, 1995; Cerling *et al.*, 1999; Ishibashi *et al.*, 1999; Cerling *et al.*, 2007; Codron *et al.*, 2012; Coutu *et al.*, 2016; Ziegler *et al.*, 2016). Research on Black (*Diceros bicornis*) and White (*Ceratotherium simum*) Rhinoceros horns followed suit with Hall-Martin *et al.* (1993) showing that not only were the species distinguishable isotopically, but different southern African populations were also isotopically distinct when using $\delta^{13}C$, $\delta^{15}N$, $^{87}Sr/^{86}Sr$ and $^{207}Pb/^{206}Pb$ isotope ratios. Cerling *et al.* (2018) were also able to separate rhino populations within Kenya using $\delta^{13}C$, $\delta^{15}N$ and $^{87}Sr/^{86}Sr$ isotope ratios of hair.

Traceability testing using stable isotopes

The aforementioned examples aimed to generate isotopic data to test whether isotopic analyses could be used as a forensic tool to trace the geographic origin and movement of traded specimens. In order for this to be viable, traceability estimates of an animal need to be an unambiguous reflection of the different landscapes through which an animal travels and derives its nutrition. This is heavily reliant on the size of an animal's home range and dispersal

capabilities, the types of isotopes used to reflect environmental gradients and the ability to predict spatial patterns over time (Koch, 2007; Coulton *et al.*, 2009; Chesson *et al.*, 2018b; Vander Zanden *et al.*, 2018a). The isotopic composition of an animal is also highly complex as it is a product of the background isotopic composition of the food chain in a specific area, the fractionation steps along the food chain leading to the dietary items it consumes as well as the organism's assimilation and excretion via physiological processes (West *et al.*, 2006; Hartman, 2011; Ben-David and Flaherty, 2012). This along with other sources of variability need to be accounted for, before stable isotope traceability can be accepted as a forensic tool.

Only a handful of the examples provided in the previous section investigated what the effect of individual variation within a population may have on traceability estimates (Koch *et al.*, 1995; Cerling *et al.*, 2007; Ziegler *et al.*, 2016). Although limited analyses were performed on data generated for elephants in Amboseli National Park, Koch *et al.* (1995) included the animals' gender, age at death and year of death in their analyses, and found that age and year of death were important factors in individual variability across the population when using δD, $\delta^{18}O$ and $^{87}Sr/^{86}Sr$ stable isotope ratios. Additionally, Ziegler *et al.* (2016) included statistics on the intra-site variance which accounted for 65% of the isotopic variability ($\delta^{13}C$, $\delta^{15}N$, $\delta^{18}O$, δD, and $\delta^{34}S$) across 507 samples from both African and Asian Elephant (*Elephas maximus*) populations worldwide. However, as with the elephants from Kenya (Cerling *et al.*, 2007), neither authors considered specific factors responsible for the observed variability. Potential sources of isotopic variation amongst individuals may include age-related growth rate, body mass, dietary preference, physiological status (nutrition and hydration) and assimilation efficiency (Hesslein *et al.*, 1993; Hobson *et al.*, 1993; Martínez del Rio and Wolf, 2005; Martínez del Rio and Carleton, 2012).

Additionally, isotopic ratios change as trophic levels change which is known as a trophic shift (Kelly, 2000; Ben-David and Flaherty, 2012). However, determining trophic shift

values are difficult as they are known to be highly variable, tissue specific, dependable on the species being analysed and diet specific (DeNiro and Epstein, 1978, 1981; McCutchan *et al*., 2003; Olive *et al.*, 2003; Barnes *et al.*, 2007; Kelly and Martínez del Rio, 2010; Ben-David and Flaherty, 2012), especially if the diet consists of ants (Tillberg *et al*., 2006).

An important aspect to understand and determine is the degree of isotopic variation between and within a tissue, which was only accounted for by a few these traceability studies (Koch *et al.*, 1995; Ziegler *et al.*, 2016; Cerling *et al.*, 2018), all of which looked at variability of bone, ivory, teeth and/or hair samples spanning a few months to the entire lifetime of an individual. With regards to sampling for geographic traceability, both inter- and intra-tissue variation are important factors to test as variation in tissue integration and turnover rates of stable isotopes have been documented (Hobson, 1999; Ben-David and Flaherty, 2012; Martínez del Rio and Carleton, 2012; Vander Zanden *et al*., 2015) and sampling protocols of tissues need to take this into account (Wassenaar and Hobson, 2006; Barnes *et al.*, 2008; Vander Zanden *et al.*, 2018a).

Isotopic signatures derived from dietary input in a certain area have a faster turnover rate in tissues such as blood and hence will lose that area's specific signature relatively quickly (within days) if an organism is taken out of its environment (Hobson and Clark, 1992). Tissues with comparatively slower turnover rates like bone, although dependent on type of bone, do not lose an area's specific signature as quickly as this may take years for an isotopic signature change (Hobson and Clark, 1992). These time periods are also dependent on an animal's metabolism and size. Some tissue types are biochemically inert after formation (feathers, hair, tusk, nails and scales) meaning that once isotopes have been incorporated, their isotopic ratios cannot change and do not have a turnover rate (Bowen *et al.*, 2005). These types of tissues make for better geographic origin determination due to reduced assumptions relating to turnover rates (Koch, 2007). However, these tissue types incorporate isotopes in a

chronological fashion which could lead to inaccurate estimations if they are variable across their growth patterns (dietary shifts, dispersal, age-related growth rates, *et cetera*) and are sampled incorrectly (Wassenaar and Hobson, 2006). Incorrect sampling would involve collecting point samples of parts of the biochemically inert tissue and excluding the rest of the sample or taking samples from different parts of the tissue in different cases. If the tissue is variable across its growth pattern, this could lead to incomparable samples between individuals and may only represent a part of the variability within the tissue (Wassenaar and Hobson, 2006). It is therefore integral for forensic applications, where analysing whole tissue samples may not always be possible, that the variability within a specific tissue is understood to determine the best practice of sample collection for traceability estimates (Wassenaar and Hobson, 2006; Vander Zanden *et al.*, 2018a).

This may be the case for the biochemically inert pangolin scale (Spearman, 1967) which shows isotopic variability across its length, with $\delta^{15}N$ values ranging from 6.6–9.4‰ and $\delta^{13}C$ values ranging from -15.6– -11.3‰ (Pietersen *et al.*, 2016). Apart from stable isotope integration and turn-over rates, tissues also vary in isotopic composition within the same organism due to different macromolecules (protein, fatty acids, lipids and cholesterol) incorporated into different tissues (Koch, 2007). Differences in $\delta^{13}C$ and $\delta^{15}N$ values between different tissues may be due to the level and types of essential (acquired directly from the diet) or non-essential (either directly acquired from the diet or generated within the body) macromolecules required for tissue growth and regeneration (Koch, 2007). This may also change over time in the same tissue due to changes in diets or bouts of physiological stress (Ambrose and DeNiro, 1986; Sealy *et al.*, 1987; Howland *et al.*, 2003; Fuller *et al.*, 2005; Mekota *et al.*, 2006; Podlesak and McWilliams, 2006). With this being an important consideration, choosing an easily accessible and relevant tissue type that is kept constant throughout the process whilst understanding the particular animal's dietary ecology is integral

to successful traceability tests (Vander Zanden *et al.*, 2018a). It is also integral to understand the sources of variability within a pangolin scale before it can be used as a possible sample for traceability.

Pangolin scales

Pangolin armour is in the form of keratinous scales that cover most of the body, apart from the ventral trunk, head and inner regions of the limbs (Spearman, 1967; Kingdon, 1971; Smithers, 1971). These scales make up an estimated 20–35% of the total weight, depending on the species (Kingdon, 1971; Pietersen, 2013) and are arranged in an imbricate manner (Figure 2.1a; Heath, 1992; Liu *et al.*, 2016a; Wang *et al.*, 2016). As a substantial proportion of the body is covered by scales, the choice of a particular scale for isotopic analyses may be important due to possible isotopic variation in scales from different locations on the body. Such isotopic discrepancies have been found in feathers from different body regions in birds (Smith and Dufty Jr, 2005).

The scale is comprised of alpha and beta keratin (Tong *et al.*, 1995) arranged in different lamella structures across three distinct macro layers varying in histological and histochemical characteristics (Figure 2.1b,c; Spearman, 1967; Liu *et al.*, 2016a). The dorsal plate, which comprises around 8% of the scale, is the furthest layer from the body and contains flattened keratinized cells with high levels of phospholipids and moderate levels of sulfhydryl bonds (Spearman, 1967; Liu *et al.*, 2016a; Wang *et al.*, 2016). The inner ventral plate layer, which is closest to the body, is only a few cells thick and consists of flattened keratinized cells with high levels of phospholipids (Spearman, 1967; Liu *et al.*, 2016a; Wang *et al.*, 2016). This layer has the majority of contact with the skin and contains high levels of disulphide bonds (Spearman, 1967). The layer between the dorsal and ventral plates (intermediate plate, 54% of the scale volume) differs from the aforementioned layers by containing circular keratinized cells and

high levels of disulphide bonds (Spearman, 1967; Liu *et al.*, 2016a; Wang *et al.*, 2016). These differences, together with each layer originating from a different region of the germinal epidermis (Spearman, 1967), suggests that there may be isotopic differences between the three layers and this possibility should be investigated when devising appropriate sampling protocols.

The dorsal surface of the scale has characteristic grooves that run longitudinally down the scale from the base to the tip (Figure 2.1b), although these disappear in very old individuals through consistent abrasion with the environment and surrounding scales (D.W. Pietersen, *pers. comm.*). These have been suggested to function as an anti-abrasive mechanism to soil by reducing surface contact and allowing grains of sand to furrow down grooves rather than be an indication of transverse growth patterns (Tong *et al.*, 2007). The reduction of nuclear staining as one moves away from the scale base and towards the free edge suggests that scale growth is longitudinal as reduction in nuclear staining is an indication of cell death (Spearman, 1967). This pattern of growth has also been observed in field studies (D.W. Pietersen, *pers. comm.*) and was therefore assumed as such for the purposes of this research.

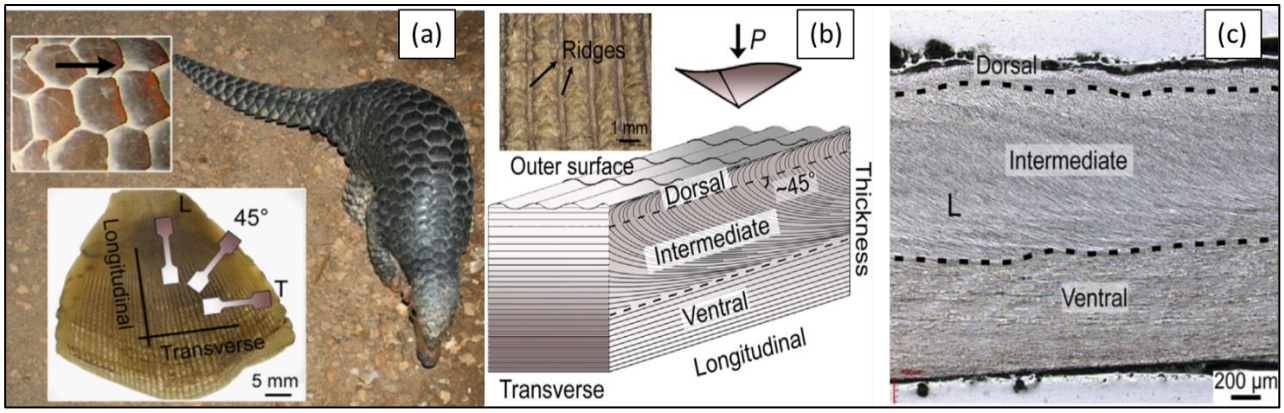


Figure 2.1: The scale arrangement and structure of a Chinese Pangolin (*Manis pentadactyla*) from the (a) macroscopic, (b) animated and (c) microscopic views. Image adapted from Liu *et al.* (2016).

Illegal trade of scales

Pangolin scales are the most frequently trafficked derivatives (Heinrich *et al.*, 2017) due to their perceived medicinal properties. Although Traditional Chinese Medicine pharmacopeia only prescribe Chinese Pangolin scales for the purposes of detoxifying and draining pus, relieving palsy, and stimulating lactation (Li, 1982; Qiu, 1985; Gao, 2012) there are approximately 70 patented traditional medicines that claim to cure anything from a simple gastro-intestinal problem to cardiac complications (Katuwal *et al.*, 2015; Xu *et al.*, 2016; Perera *et al.*, 2017). Scales are usually traded in bags, often by the ton, and are traded separately from live or skinned pangolin carcases which are sold in meat markets (Mwale *et al.*, 2016; Heinrich *et al.*, 2017). Scales are harvested from dead pangolins either through boiling the whole carcass or skin, or by placing them directly in the flames of a wood-burning fire (APWG *unpubl. data*). As these harvesting methods may affect the isotopic composition of the scales, this aspect was selected for further investigation in this study.

Studies on cremated bones have shown that temperatures above 200 °C significantly alter both $\delta^{13}C$ and $\delta^{15}N$ values (Schurr *et al.*, 2008; Harbeck *et al.*, 2011), however boiling bones for 1 hour and roasting them over coals for 2 hours had no effect on $\delta^{13}C$ and $\delta^{15}N$ values (DeNiro *et al.*, 1985). For bird feathers, temperatures above 300 °C altered δD values, however individual variation was the major driver of variance in the bird populations under study (Vander Zanden *et al.*, 2018b). There has been no research into the effects that direct flaming and boiling of other keratin-based biological materials, including pangolin scales, may have on their stable isotope values. This leaves a level of uncertainty regarding the impact of these treatments on traceability estimates. Due to scales being biochemically inert, the most heavily trafficked derivative, and being relatively easy to non-invasively sample from live individuals, the feasibility of isotopic traceability in pangolins will focus on scales.

Study aim

Before geographic traceability estimates can be derived with a high level of confidence, there are an array of factors of variation that should be tested first in order to quantify variability in $\delta^{13}C$ and $\delta^{15}N$ values (Barnes *et al.*, 2008; Ziegler *et al.*, 2016). These include:

1) Intra-scale variation: Test whether scale layer and point of sampling along the length of the scale (temporal) have a significant effect on $\delta^{13}C$ and $\delta^{15}N$ values.

2) Intra-individual variation: Test whether taking scale samples from different regions of the body will significantly affect $\delta^{13}C$ and $\delta^{15}N$ values.

3) Inter-individual variation: Define the level of variation between individuals of the same population as well as identify which variables of pangolin ecology and behaviour may drive this variability.

4) Variation between harvesting methods: Test whether boiling and direct flame harvesting methods have a significant effect on $\delta^{13}C$ and $\delta^{15}N$ values of scales.

Methods

Sample collection

Scale samples were collected from 37 Temminck's Ground Pangolin (*Smutsia temminckii*) individuals from South Africa (Appendix 1). Three whole scales were collected from each of 32 *S. temminckii* skins from collections held in the National Museum in Bloemfontein and the McGregor Museum in Kimberley. These skins originated from Kalahari Oryx Private Game Farm (hereinafter Kalahari Oryx) and Tswalu Kalahari Reserve (hereinafter Tswalu) situated in the Northern Cape Province of South Africa between 2009 and 2013 (Pietersen, 2013). Additional scale clippings were collected from the elongated elbow

scale (amounting to *ca.* 20% of the scale) from four live individuals at Tswalu in 2017. A bag of scales from an individual that died on an electrified game fence in the 1990's was collected from Klaserie Private Nature Reserve (hereinafter Klaserie) in Limpopo Province, South Africa (Figure 2.2).

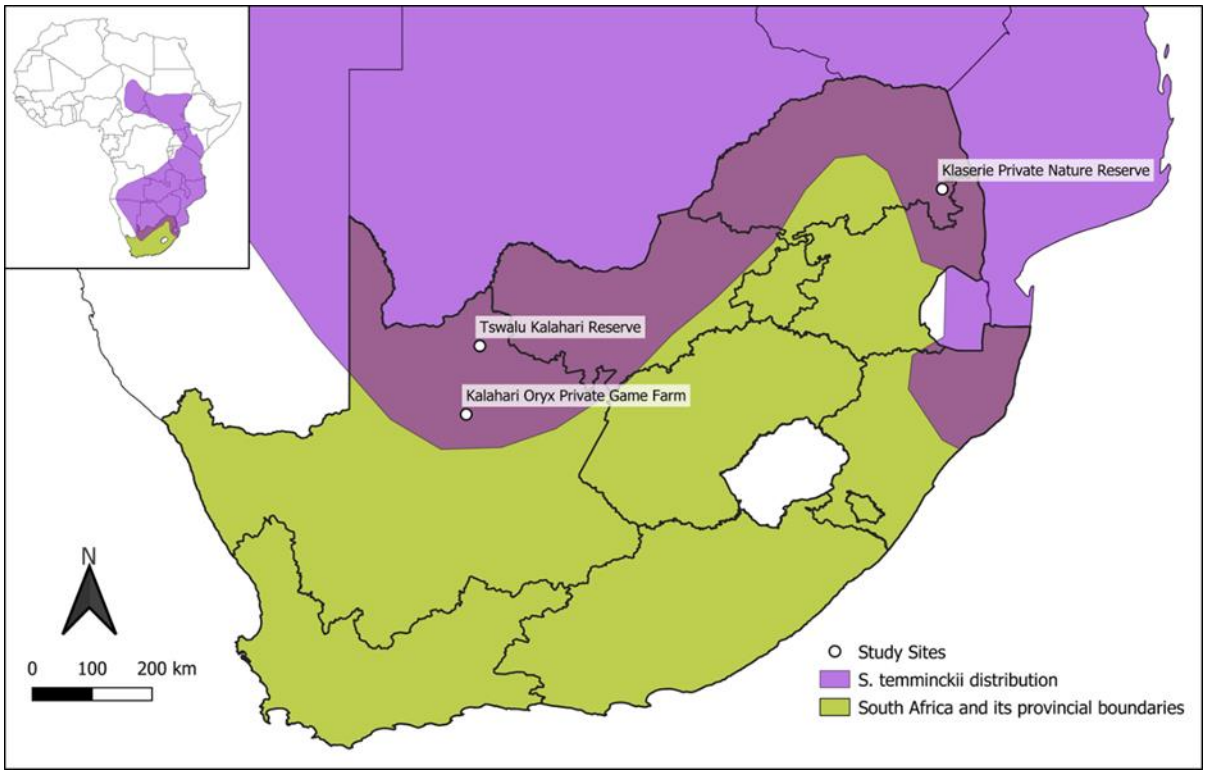


Figure 2.2: Map of the three study sites in the Northern Cape and Limpopo Provinces of South Africa (green shaded area) within the distribution of the Temminck's Ground Pangolin (*Smutsia temminckii*).

If museum study skins were intact three elongated elbow scales were sampled. In instances where the skins were not intact, the three largest and most rounded scales (situated at the lower back) were collected (hereinafter lower back scales). This sampling protocol was designed to negate possible isotopic discrepancies between different anatomical regions. Scales from six different anatomical regions were also collected from 13 of these skins.

Sample preparation

Scales were handled with nitrile gloves and equipment used was cleaned between each sample with 70% ethanol (EtOH) to avoid contamination. Soil particles that were attached to

scales were collected by scraping onto weighing paper using a scalpel, whilst skin was collected using forceps. These two sample types were placed in separate Eppendorf tubes for future traceability research. Scales were washed with a brush and distilled water at room temperature and dried with paper towels to remove any potential contaminants.

Experiment 1: Intra-scale variation

A lower back scale from seven Kalahari Oryx museum specimens (four males and three females), one Tswalu museum specimen and the Klaserie specimen (sexes unknown) were cut longitudinally in half using a hacksaw. The material produced during cutting was collected and stored in an Eppendorf tube as a homogenised scale sample. The cut edges were sanded with 600-grain sandpaper, followed by 1 000-grain sandpaper until they had an even, smooth surface for better visualisation of the three scale layers. Equidistant drillings were performed using a 0.1 mm taper pointed dental carbide burr (New Technology Instruments, Kahla, Germany) along the length of each scale from each of the three layers, namely; dorsal, ventral and intermediate (Figure 2.3). The resultant powder from each sample was collected and placed separately into labelled Eppendorf tubes.

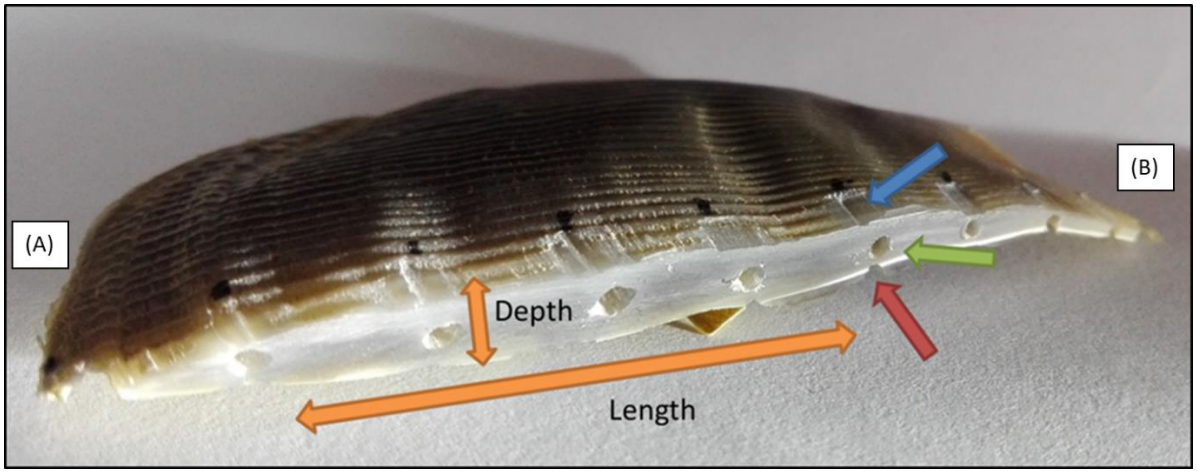


Figure 2.3: Lower back scale of a Temminck's Ground Pangolin (*Smutsia temminckii*) that has undergone micro-sampling across the length of the scale from the attached end (A) to the free edge (B) as well as across the depth of the scale from dorsal (blue arrow), to intermediate (green arrow) and ventral layers (red arrow).

Experiment 2: Intra-individual variation

Scales from six anatomical regions on the body of the Klaserie specimen, namely (1) the front limb (elbow joint), (2) flank near the belly, (3) lower back, (4) dorsal part of the tail, (5) side of the tail and (6) mid-ventral part of the tail were collected (Figure 2.4). In addition, duplicate scale samples from various anatomical regions for 13 of the 32 museum specimens from both Tswalu and Kalahari Oryx were included for better sample representation. These scales came from (1) the elbow, (2) flank near the belly, (3) lower back and (6) mid-ventral part of the tail (Figure 2.4). The same collection methodology for *Experiment 1* was used, whilst an additional sample was obtained from each scale tip (free-end) by drilling through all three layers and collecting the resultant powder.

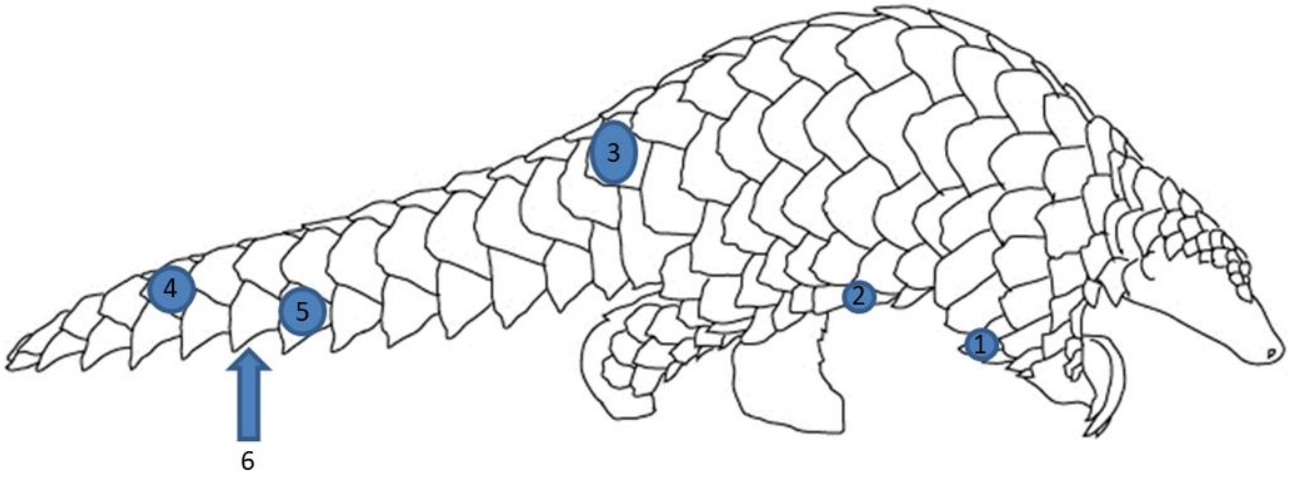


Figure 2.4: The six anatomical regions of the body of the Temminck's Ground Pangolin (*Smutsia temminckii*) where scale samples were collected.

Experiment 3: Inter-individual variation

Two homogenised scale samples and one tip sample were collected from each of the 33 *S. temminckii* specimens (museum and Klaserie samples) using the same methodology from *Experiment 2*. Additional scale tip samples were collected from four live individuals sampled at Tswalu (total of 37 scale tip samples).

Experiment 4: Variation between harvesting methods

The two remaining scales from each of the 32 museum specimens, along with samples from the Klaserie specimen, underwent a range of treatments to emulate the two main methods that poachers use to remove the scales from the pangolin's carcass, *viz.* boiling and direct exposure to fire. From anecdotal evidence, processing takes approximately 10–15 minutes during which time the whole carcass or skin is placed in a pot to boil or on direct flame (APWG, *unpubl. data*). One scale from each specimen was therefore boiled for 15 minutes whilst the other scale was placed in direct flame until slight browning around the periphery of the scale was observed (n=23; Figure 2.5a). A few scales were left for a longer period of time and the degree of charring was divided into three categories: slight charring around the edges (n=2; Figure 2.5b), major charring around the edges and inner ridge (n=2; Figure 2.5c) or charring over the entire scale (n=5; Figure 2.5d). A subjective observation index was used to classify the degree of charring because scales varied in size and in the amount of skin attached, thus time needed to brown. If these scales had been exposed to direct flames for the 10 to 15-minute processing time as thought to be used by poachers for whole carcasses, they would have been totally burnt. Single homogenised scale samples were collected from each scale, whilst additional micro-samples of charred and non-charred parts were collected from two specimens (STEM 10 and STEM 12) that were charred over the entire scale (as in Figure 2.5d).

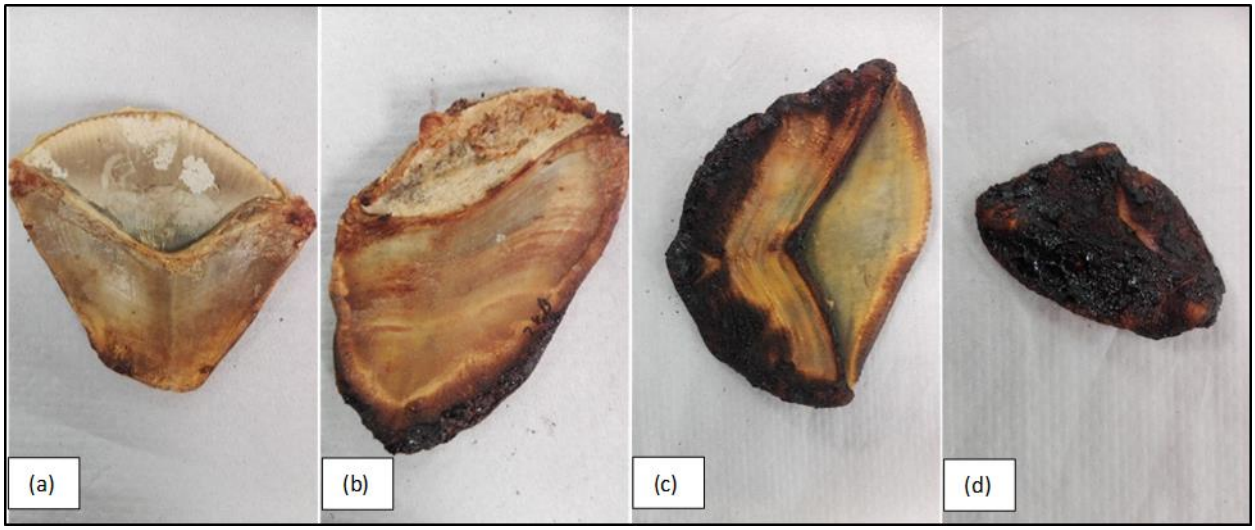


Figure 2.5: Levels of charring emulated with Temminck's Ground Pangolin (*Smutsia temminckii*) scales on an open wood fire. The majority of scales (a) underwent slight browning along the periphery, whilst others were taken off the flame after (b) slight charring along the periphery, (c) major charring along the periphery and inner ridge or (d) significant charring over the entire scale.

Isotopic analysis

For the first three experiments, duplicate aliquots of 0.5–0.6 mg from each homogenised sample were placed in tin capsules pre-cleaned in toluene, using sterile dissection forceps, whilst for samples from *Experiment 4*, only a single aliquot was weighed. Only single aliquots of 0.4–0.5 mg from each scale layer drilling from *Experiment 1* were weighed due to the small amounts of sample obtained.

Samples were combusted at 1020°C using an elemental analyser (Flash EA 1112 Series) coupled to a Delta V Plus stable light isotope ratio mass spectrometer via a ConFlo IV system (all equipment supplied by Thermo Fischer, Bremen, Germany), housed at the Core Stable Isotope Facility of the University of Pretoria. Two randomized duplicates were run after every ten samples to account for sample variation in measurements. To minimize instrument drift or memory effect, a blank sample and two laboratory running standards, namely Merck Gel ($\delta^{13}C$ = -20.26‰, $\delta^{15}N$ = 7.89‰, C% = 41.28, N% = 15.29) and DL-Valine ($\delta^{13}C$ = -10.57‰, $\delta^{15}N$ = -6.15‰, C% = 55.50, N% = 11.86) were run after every 11 unknown samples.

A total of ten standards on average were analyzed during each run. These standards were calibrated (2017) against international standards from the National Institute of Standards and Technology (NIST): NIST 1557b (bovine liver), NIST 2976 (mussel tissue) and NIST 1547 (peach leaves).

The precision of each stable light isotope ratio mass spectrometer run was determined using the DL-Valine standards (Standard deviation: δ^{15}N (‰) < 0.07; δ^{13}C (‰) < 0.05), whilst output data from each run was corrected using isotope values for the Merck Gel standards. Isotope ratios are expressed in delta notation using a per mille scale (‰), which are referenced to Vienna Pee-Dee Belemnite for carbon isotope values ($\delta^{13}C_{VPDB}$), and air for nitrogen isotope values ($\delta^{15}N_{Air}$), using the following equation:

$$\delta X(‰) = [(R_{sample}/R_{standard}) - 1]$$

where X= ^{15}N or ^{13}C and R represents ^{15}N/^{14}N or ^{13}C/^{12}C, respectively.

Statistical analyses

All statistical analyses were run in the Rstudio interface of R v3.5.1 (Rstudio Inc., Massachusetts, U.S.A), whilst assumptions related to these analyses were tested and upheld. Shapiro-Wilk and Levene's tests at an alpha value of 0.5 were used to determine normality and homogeneity of variance, respectively.

Intra-scale variation

A one-way Analysis of Variance (ANOVA) was performed to determine whether the scale layers (dorsal, ventral and intermediate plates) for all nine individuals (as well as the seven Kalahari Oryx individuals alone) differed significantly in their δ^{13}C values. A non-

parametric Kruskal-Wallis rank sum test was run to determine the same for δ^{15}N values due to a failed assumption of normality. A post-hoc Tukey multiple comparison analysis using the "multcomp" package (Hothorn *et al.*, 2008) was used to determine whether the scale layers were significantly different from each other for both δ^{13}C and δ^{15}N micro-sample values.

As the sex of the Tswalu and Klaserie individuals were unknown, only data from the seven Kalahari Oryx individuals were used to perform a F-statistics variance test, which was used to determine whether there was a significant difference between the sexes at the micro-sample level.

The data from *Experiment 3* were subjected to paired t-tests in order to determine whether the duplicates of the homogenised scale samples were significantly different to one another for δ^{13}C and δ^{15}N values. This test was designed to determine whether duplicate sampling of the homogenized scale was necessary. Paired t-tests were used to determine whether the averaged δ^{13}C and δ^{15}N values of homogenised scale samples (average of duplicate 1 and 2) and those of the tip samples were significantly different from one another. A linear regression model was created to test whether the pangolin's total length (as a proxy for age; Pietersen *et al.*, 2016) had a significant effect on the interaction between averaged δ^{13}C values of homogenised scale samples and their corresponding tip samples.

Intra-individual variation

One-way ANOVA's were performed to determine whether the location of a scale on the body had a significant effect on δ^{13}C values. Due to a normality assumption fail for δ^{15}N, a non-parametric Kruskal-Wallis rank sum test was performed. The averaged δ^{13}C and δ^{15}N values of homogenised scale samples and their duplicates were used, whilst separate tests investigating scale tip samples were included as the scale tip is the area of the scale that is generally collected from live individuals.

Inter-individual variation

Mixed effect models with specimen set as a random variable (duplicate samples of each individual) were performed to determine the effects that an individual's sex, length and year of sampling had on $\delta^{13}C$ and $\delta^{15}N$ values. Prior to this, data missing from the aforementioned factors were cleaned using the na.omit function reducing the dataset by six individuals (n = 26). The length of each individual was used as a proxy for its age, following Pietersen *et al.* (2016). Significance estimates were determined using the ANOVA function for Type II Wald chi-square tests.

Variation between harvesting methods

A one-way ANOVA was performed to determine whether the scale harvesting methods (control, boiling and direct flaming) differed significantly in their $\delta^{13}C$ and $\delta^{15}N$ values.

Results

A summary of all the statistical results are presented in tabular form for ease of reference (Table 2.1).

Table 2.1: Summary of statistical analyses performed for feasibility of Temminck's Ground Pangolin (*Smutsia temminckii*) traceability. Homogenised samples with a star (*) are when a single sample was collected from the homogenised material rather than two.

Variable tested	Samples used	Statistical test	Result
Intra-scale variation:			
Between layers for $\delta^{13}C$	Nine micro-sampled individuals	One-way ANOVA	Significant
Post-hoc for which layers differ ($\delta^{13}C$)	Nine micro-sampled individuals	Tukey multiple comparison	Dorsal significantly more enriched than ventral
Between layers for $\delta^{15}N$	Nine micro-sampled individuals	Kruskal Wallis	Not significant
Between sexes for $\delta^{13}C$ and $\delta^{15}N$	Kalahari Oryx micro-sampled individuals	F-statistics variance test	Males significantly more variant
Scale sampling protocol:			
Between duplicates of homogenised scale ($\delta^{13}C$, $\delta^{15}N$)	33 individuals (homogenised scale)	Paired t-test	Not significant for either $\delta^{13}C$ or $\delta^{15}N$
Between averaged duplicates and scale tip ($\delta^{13}C$, $\delta^{15}N$)	33 homogenised and tip samples	Paired t-test	Scale tips significantly more enriched for ^{13}C
Individual's length has an effect on above result for $\delta^{13}C$	33 homogenised and tip samples	Linear regression model	Not significant
Intra-individual variation:			
for $\delta^{13}C$	14 individuals (homogenised and tip samples from six anatomical regions)	One-way ANOVA	Not significant
for $\delta^{15}N$		Kruskal Wallis	Not significant
Inter-individual variation:			
Effect sex, body length, year had on $\delta^{13}C$ and $\delta^{15}N$	33 homogenised and tip samples	Mixed effects model	Not significant
Harvesting methods:			
Effect scale harvesting methods had on $\delta^{13}C$ and $\delta^{15}N$	33 homogenised* samples	One-way ANOVA	Not significant

Intra-scale variation

The variation in $\delta^{13}C$ and $\delta^{15}N$ values across the scale for each individual averaged 3.8‰ and 3.2‰, respectively (Table 2.2), but was as high as 5.6‰ for carbon (STEM 12 – Kalahari Oryx) and 5.6‰ for nitrogen (STEM 55 – Kalahari Oryx). STEM refers to the species specific (*S. temminckii*) sample identity label, with each number being unique to a specific individual in the study (see: Appendix 1).

Table 2.2: Sampling statistics for isotope ratios measured in Temminck's Ground Pangolins (*Smutsia temminckii*).

Parameters	**$\delta^{15}N$ values (‰)**		**$\delta^{13}C$ values (‰)**	
	Mean ± SD	Ave. Variation	Mean ± SD	Ave. Variation
Averaged values				
Micro-sampling sites	10.3 ± 0.2		-15.4 ± 0.5	
Layers	10.2 ± 0.2		-15.3 ± 0.4	
Kalahari Oryx Males (n = 4)	10.0 ± 1.1		-14.9 ± 1.1	
Kalahari Oryx Females (n = 3)	10.6 ± 0.8		-15.4 ± 1.0	
Intra-scale difference		3.2		3.8
Intra-individual difference	10.1 ± 0.4	2.9	-15.1 ± 0.6	2.3
Kalahari Oryx inter-individual difference (n = 31)	10.4 ± 1.4	7.0	-14.6 ± 1.2	5.9
Inter-individual difference for all sites (n= 33)	10.3 ± 1.5	7.0	-14.6 ± 1.2	5.9

Carbon and nitrogen delta values across the scales and layers of each of the nine individuals tested are shown in Figure 2.6.

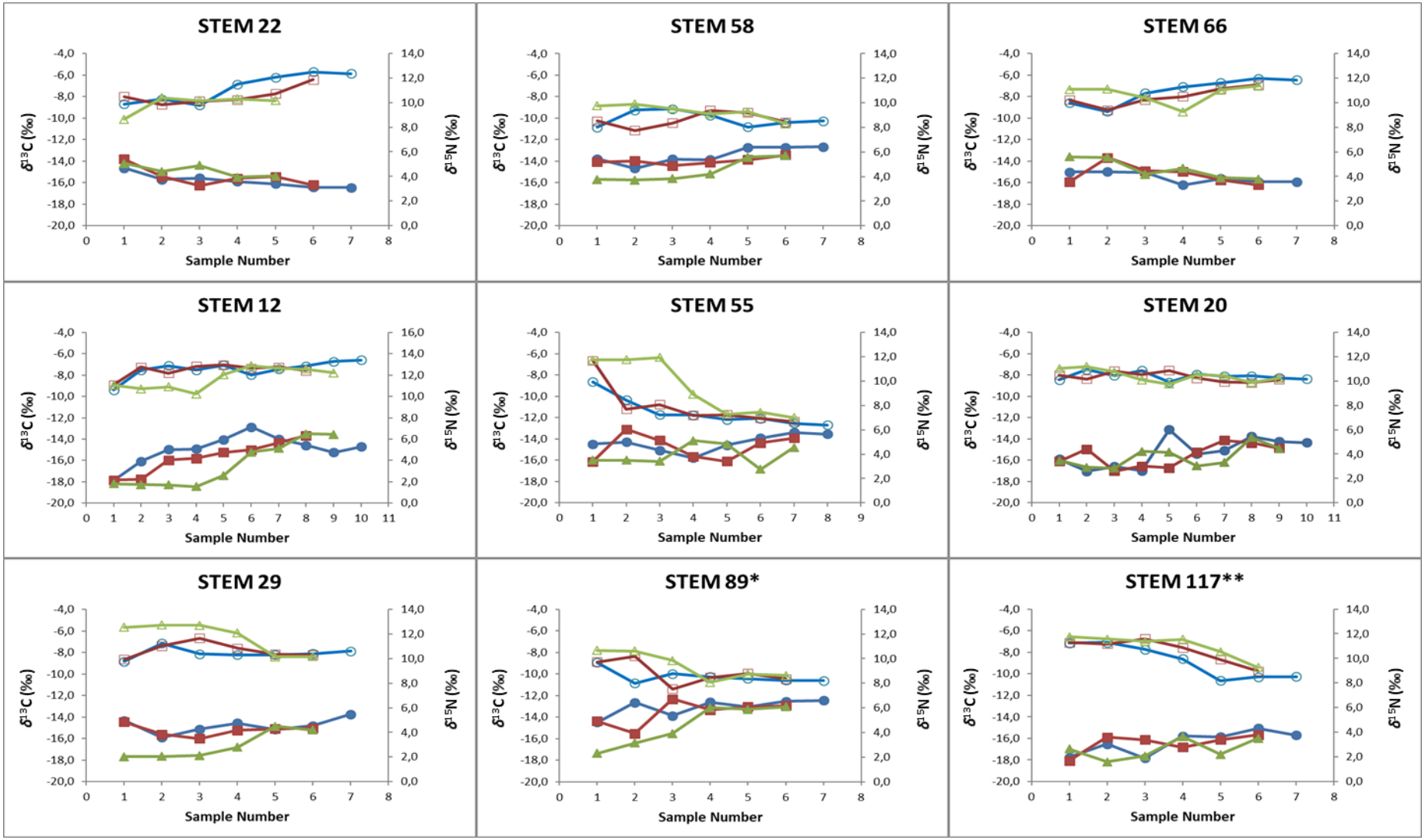


Figure 2.6: Isotope variation across the scales of nine Temminck's Ground Pangolin (*Smutsia temminckii*) specimens. The dorsal (blue), intermediate (red) and ventral (green) plate values are shown for $\delta^{13}C$ (solid symbols) and $\delta^{15}N$ (open symbols). Sample number one represents the fixed-edge. The specimen marked with an asterisk (*) is from Tswalu Kalahari Reserve, the one with two asterisks (**) is from Klaserie Private Nature Reserve and the remaining specimens are from Kalahari Oryx Private Game Farm.

The higher variability in carbon was further expressed across scale layers which differed significantly from each other for $\delta^{13}C$ (F = 4.803, p = 0.009) but not for $\delta^{15}N$ values (χ^2 = 2.008, p = 0.366; Figure 2.6). The dorsal (mean $\delta^{13}C$ = -14.9‰) and ventral plates (mean $\delta^{13}C$ = -15.7‰) were the only two layers that were significantly different from one another in $\delta^{13}C$ micro-sample values (p = 0.008) under the multiple comparisons test. A stepwise enrichment pattern from the ventral to dorsal layers is evident (Figure 2.7). These results were mirrored when testing only the seven Kalahari Oryx individuals.

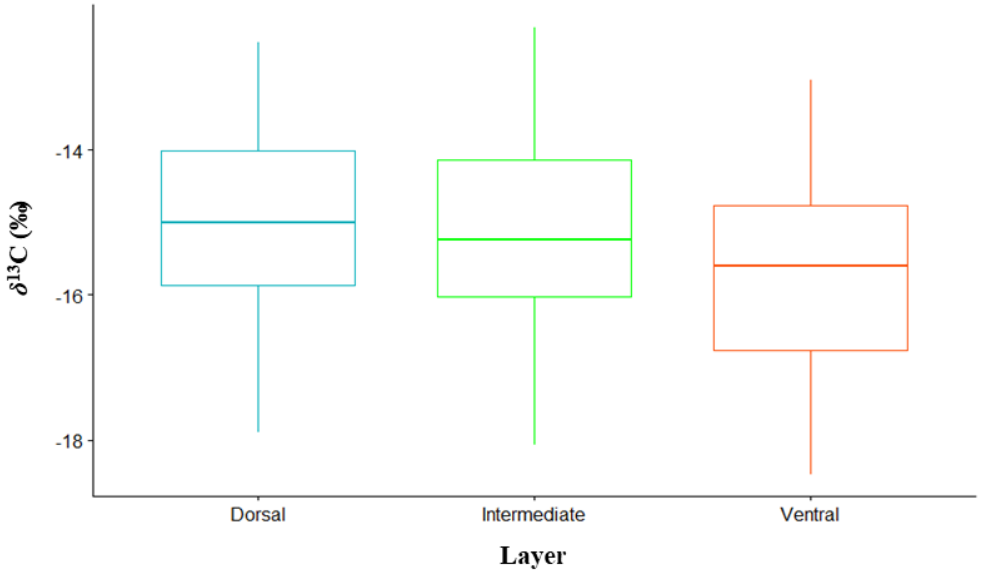


Figure 2.7: Variability in $\delta^{13}C$ values across the different scale layers for micro-samples of all nine Temminck's Ground Pangolin (*Smutsia temminckii*) specimens.

Kalahari Oryx males (n = 4) not only had significantly more inter-individual variation in both $\delta^{13}C$ (F = 1.718, p = 0.024) and $\delta^{15}N$ values (F = 6.274, p < 0.001) compared to females (n = 3; Figure 2.8), but they also had slightly more intra-scale variation (Table 2.2). This difference remained even after randomly excluding a single male from the analysis to balance the sample numbers. Nitrogen delta values also had a higher inter-individual variability than that of carbon (Figure 2.8).

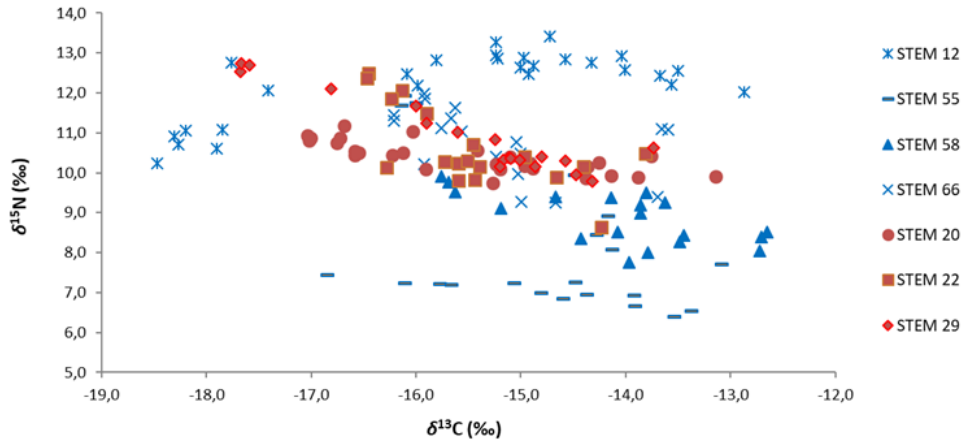


Figure 2.8: Variability in δ^{13}C and δ^{15}N values sampled across the scale growth axes of four male (blue) and three female (red) Temminck's Ground Pangolins (*Smutsia temminckii*) from Kalahari Oryx Private Game Farm.

No significant difference was found for either δ^{13}C (t = -1.728, p = 0.091) or δ^{15}N values (t = -0.684, p = 0.497) between homogenised scale sample duplicates (n = 33). The same was true between averaged values of the duplicate homogenised scale samples (duplicate 1 and 2 averaged) and the scale tips of those samples for δ^{15}N values (t = 0.974, p = 0.335). However, scale tips were significantly enriched in ^{13}C (t = -4.439, p < 0.001; Figure 2.9) compared to the homogenised samples. Due to this pattern in δ^{13}C values, a linear regression model was used to test whether length (as a proxy for age) had a significant effect on the interaction between the scale tip and the corresponding averaged homogenised sample duplicates, however this proved not to be significant (p > 0.05).

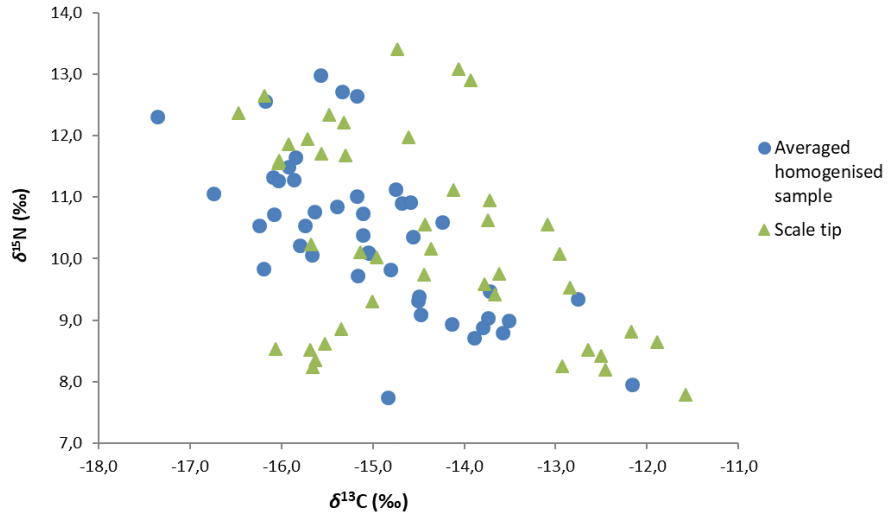


Figure 2.9: Scatterplot showing variation in $\delta^{13}C$ and $\delta^{15}N$ values for homogenised and scale tip samples of Temminck's Ground Pangolins (*Smutsia temminckii*).

Intra-individual variation

The anatomical position where scales were collected did not have a significant effect on $\delta^{13}C$ or $\delta^{15}N$ values ($p > 0.05$) of either averaged scale homogenates or scale tip samples. For the Klaserie sample, for which all six anatomical regions were represented, variation between the tip samples (encircled) and their corresponding homogenised scale samples was higher than that of the variability between scales from different body parts. Although one homogenised scale sample outlier from the limb was shown to have values closer to those of the scale tip average (Figure 2.10).

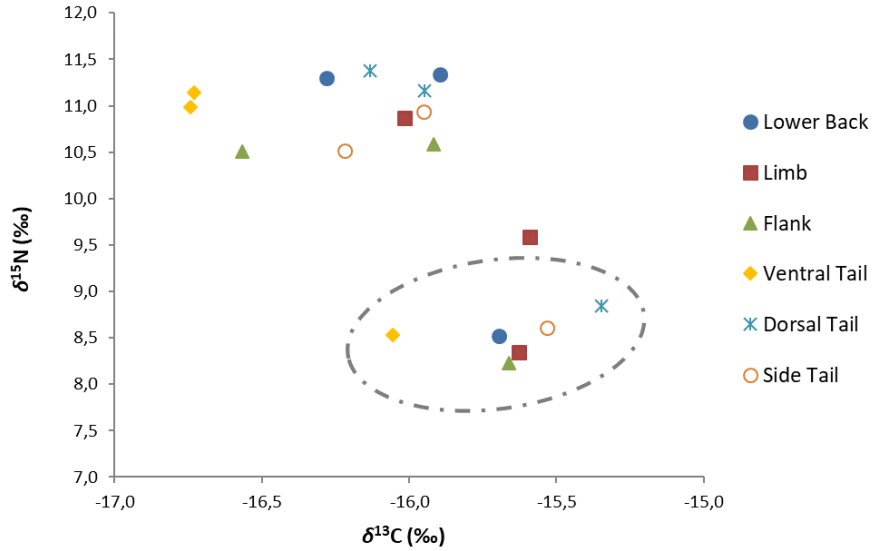


Figure 2.10: Scatterplot of $\delta^{13}C$ and $\delta^{15}N$ values of scales from different anatomical regions of a Temminck's Ground Pangolin (*Smutsia temminckii*) originating from Klaserie Private Nature Reserve. The encircled points are samples from the scale tip.

Inter-individual variation

The aforementioned results appear to indicate higher levels of variability between individuals than within scales or across anatomical regions (Figures 2.6 and 2.11).

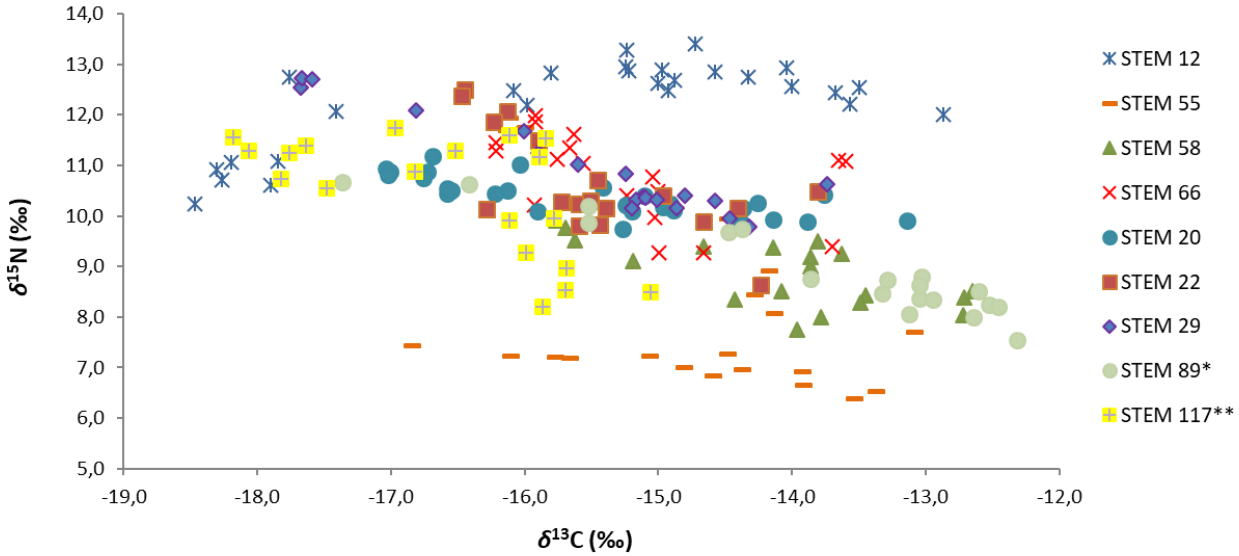


Figure 2.11: $\delta^{13}C$ and $\delta^{15}N$ values of the lower back scale from nine individuals that were micro-sampled (averaged across all three layers). All samples are from Kalahari Oryx Private Game Farm apart from STEM 89* (Tswalu Kalahari Reserve) and STEM 117** (Klaserie Private Nature Reserve).

This is supported by the high levels of residual variance in the statistical tests revolving around intra-scale and intra-individual variation, suggesting that the remaining source of variance may be due to differences between individuals and/or geographic areas. The former was shown to be the case in respective standard deviations of intra-scale, intra-individual and inter-individual comparisons for both $\delta^{13}C$ and $\delta^{15}N$ values (Table 2.2) along with scale tip samples of each individual across the three sites (n = 37; Figure 2.12). These values had a range of -17.5– -11.6‰ for $\delta^{13}C$ (difference of 5.9‰) and 6.4–13.4‰ for $\delta^{15}N$ (difference of 7.0‰) in the Kalahari Oryx samples alone as well as across all three study sites (Figure 2.12; Table 2.2). In order to determine the source of this variability, mixed effect models were performed, however sex, year of sampling and length of animal (as a proxy for age) had no significant effect on $\delta^{13}C$ or $\delta^{15}N$ values.

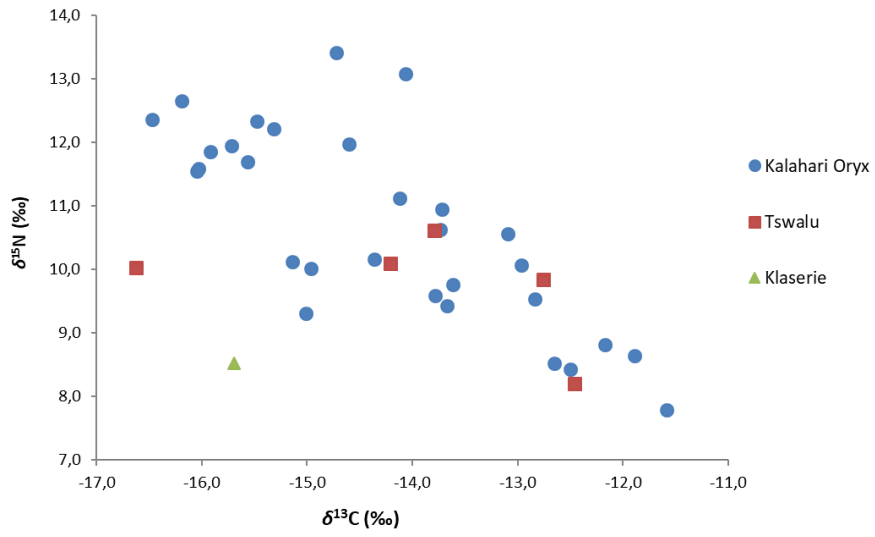


Figure 2.12: $\delta^{13}C$ and $\delta^{15}N$ values of scale tips from all 37 Temminck's Ground Pangolins (*Smutsia temminckii*) sampled from Kalahari Oryx Private Game Farm, Tswalu Kalahari Reserve and Klaserie Private Nature Reserve.

Variation between harvesting methods

The $\delta^{13}C$ (F = 0.18, p = 0.836) and $\delta^{15}N$ (F = 0.138, p = 0.871) values of scales harvested by different methods did not differ significantly (Figure 2.13).

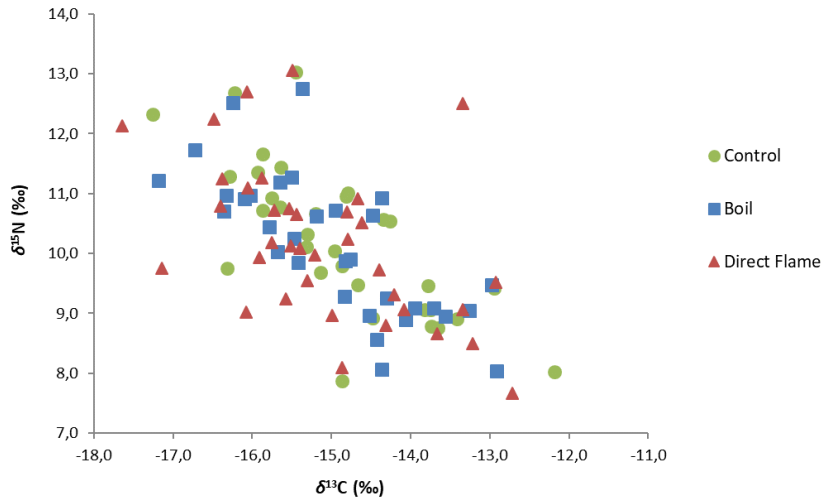


Figure 2.13: $\delta^{13}C$ and $\delta^{15}N$ values of different harvesting methods used during scale harvesting of Temminck's Ground Pangolins (*Smutsia temminckii*).

Discussion

Intra-scale variation

The results suggest a high level of intra-scale variation, particularly for carbon isotopic values ($\delta^{13}C$ = 3.8‰ and $\delta^{15}N$ = 3.2‰). The observed intra-scale $\delta^{13}C$ variation is higher than the observed variability within tusks of Kenyan elephants (Cerling *et al.*, 2007), but not that of Kruger National Park elephants (Codron *et al.*, 2012). Although tusks are biologically different from pangolin scales, this variability is nonetheless something to be cognisant of during scale sampling. The observed temporal isotopic variation as the scale grows may be due to seasonality (Cerling *et al.*, 2009; Codron *et al.*, 2012; Cerling *et al.*, 2018), prevailing climatic conditions (Ayliffe *et al.*, 1992; Hedges *et al.*, 2004), dietary shifts (Koch *et al.*, 1995; Codron *et al.*, 2012; Britton *et al.*, 2016), or an animal's physiological status over time (Ambrose and DeNiro, 1986; Sealy *et al.*, 1987; Mekota *et al.*, 2006).

Seasonality patterns across elephant hairs over a six year period not only showed a characteristic pattern of $\delta^{13}C$ values but that it was in direct co-variation with $\delta^{15}N$ values

(Cerling *et al.*, 2009). This co-variation was not found in pangolins. If keratin-based pangolin scales (myrmecophage) have a similar isotopic assimilation mechanism to that of elephant hair (herbivore), then it is unlikely that the intra-scale variability is attributable to seasonality (Koch *et al.*, 1995; Mekota *et al.*, 2006; Britton *et al.*, 2016). Regardless of this, pangolin scales have sophisticated anti-abrasive properties that minimizes wear (Tong *et al.*, 1995; 2000; 2007), suggesting a slow scale growth rate. This would mean that observing the effect of seasonality across a scale may require a higher resolution of sampling than what was carried out in this study, and thus is an unlikely factor of the observed variability. Due to scale growth rates being unknown, it is difficult to determine whether the variability observed was influenced by longer lasting prevailing climatic conditions (Hedges *et al.*, 2004). Dietary shifts, nutritional status and water stress may be responsible for the lack of a distinct pattern across scales and co-variation of patterns between individuals (Koch *et al.*, 1995; Mekota *et al.*, 2006; Britton *et al.*, 2016). Koch *et al.* (1995) showed that African Elephants in Amboseli National Park had significant levels of intra-sample variability for carbon only, as was also observed in this study. These authors attributed this variability to a shift in diet as the population had a higher grass intake due to disruption in migratory patterns within the park via the construction of fences (Koch *et al.*, 1995).

Although it is unlikely that dietary shifts occurred in these pangolin populations due to migratory disruptions, dietary shifts could have occurred by other means of migration or dispersion. Previous studies on the ecology of *S. temminckii* in both mesic and arid savanna habitats reported that juveniles established a home range within their natal home range for up to a year before dispersing to a new area to set up their own home range (Heath and Coulson, 1997; Pietersen *et al.*, 2014b). However, these newly-established home ranges were shown to be static over long periods of time (Heath and Coulson, 1997; Pietersen *et al.*, 2014b). The inverse relationship of carbon and nitrogen isotopes have not been observed in temporal dietary

shift studies of lab-reared broad whitefish (*Coregonus nasus*) muscle and liver, or African Elephant hair (Hesslein *et al.*, 1993; Cerling *et al.*, 2009) which provides additional evidence against this hypothesis. This inverse relationship may be attributed to bouts of physiological stress as ^{15}N becomes more enriched during nutrition and water stress (Ambrose and DeNiro, 1986; Sealy *et al.*, 1987; Hobson *et al.*, 1993; Fuller *et al.*, 2005; Mekota *et al.*, 2006) and ^{13}C becomes more depleted under times of protein stress (Mekota *et al.*, 2006). This sort of relationship was also observed in the majority of nine individuals that underwent scale micro-sampling in this study (Figure 2.6) as well as that of Pietersen *et al.* (2016). It is therefore possible that physiological stress may play an important role in isotopic variation across the length of scales than dietary niche shifts. This is corroborated by the arid environment of the Kalahari study sites (Pietersen, 2013), limited and often non-existent fat deposits observed during dissections of the individuals used during this study (D.W. Pietersen, *unpubl. data*), as well as evidence of physiological stress in body temperature fluctuations when prey availability is decreased (W. Panaino, *unpubl. data*). However, without measurable scale growth rates, it is difficult to determine the time period captured across the length of a scale and hypotheses can only be given. To determine growth rates, scales would require either repeated distance measurements across the scale at specific times or the use of Accelerator Mass Spectrometry (AMS) bomb-curve radiocarbon dating (Uno *et al.*, 2013).

The variability in $\delta^{13}C$ values between scale layers may be attributable to each layer originating from different regions of the germinal epidermis (Spearman, 1967) suggesting that the isotopic source and mechanism for growth (fractionation) may differ between layers. The source of carbon and nitrogen may differ in age as some may be derived from recent dietary intake whilst others may originate from stored sources such as fat. The change in relative percentages of organic bonds (phospholipids, sulfhydryl and disulphide) between different scale layers (Spearman, 1967) may also play a role. Why these results show a stepwise $\delta^{13}C$

enrichment pattern from ventral to dorsal scale layers is more difficult to determine as there is no apparent correlation with temporal growth variability.

A previous ecological study of *S. temminckii* in a mesic savanna biome (Lowveld) found males to have significantly larger home ranges than their female counterparts (Heath and Coulson, 1997). In contrast, a study in the xeric savanna biome (Kalahari) found males and females to have home ranges of comparable size that were also mutually overlapping, while the existence of roving individuals was recorded for the first time (Pietersen *et al.*, 2014b). These roving individuals are typically males who appear to wander without a home range for a significant period of time (seven years in males as opposed to 2–3 years for females; Pietersen *et al.*, 2014b). Although calculated from a limited number of individuals (seven in total), this may explain why males showed higher intra-scale δ^{13}C and δ^{15}N variability compared to females, as males have a higher chance of encountering a wider variety of prey species across a broader variety of microhabitats (Lindsey and Skinner, 2001; Davies *et al.*, 2003; Pietersen *et al.*, 2016; Leitner *et al.*, 2018). Gestation and lactation periods may also play a role in the observed difference between isotopic variability of males and females (Fuller *et al.*, 2004). Isotopic analyses of female human hair showed significantly different δ^{15}N values during gestation, however there was no significant effect on δ^{13}C values (Fuller *et al.*, 2004).

Regardless of the cause of intra-scale variability, it is clear that identification of a comparative and cost-effective sampling protocol to account for this variability both for seizures and for geographically referenced samples, is important. The most appropriate protocol for deceased individuals is to obtain a homogenised sample across the entire length (attached end to free edge tip) and depth (across all three layers) of each scale as well as a corresponding tip sample comprising all three layers. This bulk tissue analysis over micro-sampling for origin tracing has been stressed by Koch (2007). It provides averaged δ^{13}C and δ^{15}N values across the length and depth of the scale as well as allowing a comparison between

samples from deceased and live individuals, as only scale tip samples are feasible for live individuals. However, the amount of sample used for analysis (0.5–0.6 mg) taken from a large amount of homogenised scale may not represent the averaged effect and the necessity for duplicate samples of the homogenised material would need to be tested (Wassenaar and Hobson, 2006). This study showed that there was no significant difference between homogenised scale duplicate samples suggesting the material was well homogenised. However, the anomaly in the Klaserie individual where one homogenised duplicate was closer to tip isotopic ratios than that of the other duplicate, suggests that the collected material may have not been sufficiently homogenized and that using one subsample as a standard protocol may not be completely accurate. It is therefore suggested that duplicate, or even triplicate samples of homogenised scale samples be taken for forensic purposes, more especially from those localities for which sample sizes are low, so as to prevent signature anomalies (Ainur *et al.*, 2017).

The results show that tip samples were significantly more enriched in ^{13}C than the homogenised scale duplicates, leaving the possibility that these two sample types may not be comparable for traceability purposes. In order to determine the possibility of incorporating fractionation values for these purposes, it is necessary to understand the biological processes behind enrichment of scale tips. The scale tip is the oldest part of growth, and thus the part of growth when the animal was at its youngest across the length of the scale (Spearman, 1967, Heath, 1992). With this is mind, it may be that the ^{13}C enrichment of scale tips could be a result of age-associated changes in growth rates (Hesslein *et al.*, 1993; Hobson and Quirk, 2014; Webb *et al.*, 2016) as well as a possible effect of suckling mothers' milk (Fuller *et al.*, 2006; Tsutaya and Yoneda, 2015; Lübcker *et al.*, 2016). Isotopic assimilation processes of both $\delta^{13}C$ and $\delta^{15}N$ values vary between growing and non-growing life stages as well as at different rates of tissue growth (Hesslein *et al.*, 1993; Hobson and Quirk, 2014; Webb *et al.*, 2016). However,

animal length as a proxy of age had no significant effect on the interaction between scale tips and their respective homogenised scale samples for $\delta^{13}C$ values, arguing against this possibility for pangolin scales. If this was a factor, juveniles (shorter body length) would have a closer isotopic relationship between homogenised and scale tip samples as a higher percentage of the scale would be indicative of growth compared to adults with a lower percentage of the scale indicative of growth. A distinct depletion in both ^{13}C and ^{15}N was observed in Southern Elephant Seal (*Mirounga leonina*) pup whiskers when pups switched from suckling to independent foraging (Lübcker *et al.*, 2016). Human infant nails also showed a substantial enrichment of both ^{13}C and ^{15}N during the suckling period, with a ^{15}N enrichment factor shown to be larger than that of ^{13}C (Fuller *et al.*, 2006). The results do not show this pattern for $\delta^{15}N$ values in pangolin scales which are almost histologically undistinguishable from human nails (Spearman, 1967), suggesting an alternative explanation, particularly when considering that suckling infants will always be more enriched in ^{15}N due to the systematic diet-tissue enrichment pattern seen in $\delta^{15}N$ values (Kelly, 2000; Tsutaya and Yoneda, 2015).

An alternative hypothesis relates to the change in relative thickness of scale layers along the length of the pangolin scale (Liu *et al.*, 2016a). There is a marked increase in the relative thickness of the ^{13}C enriched dorsal layer compared to that of the more ^{13}C depleted intermediate and ventral layers from the base to the scale tip. This may be due to abrasion observed on the ventral side of the scale tip which may result in a more enriched ^{13}C scale tip (little or no ventral layer; Figure 2.14). Marked abrasion midway along the scale length is also observed on the dorsal side, which along with an observed increase in thickness of the ventral layer around the same area (the site where the scale attaches to the dermis), may result in a more ^{13}C depleted scale homogenate (Figure 2.14). The abrasion patterns are due to the imbricate arrangement of scales (Heath, 1992; Liu *et al.*, 2015; Wang *et al.*, 2016) and the resultant wearing down of one scale on the other.

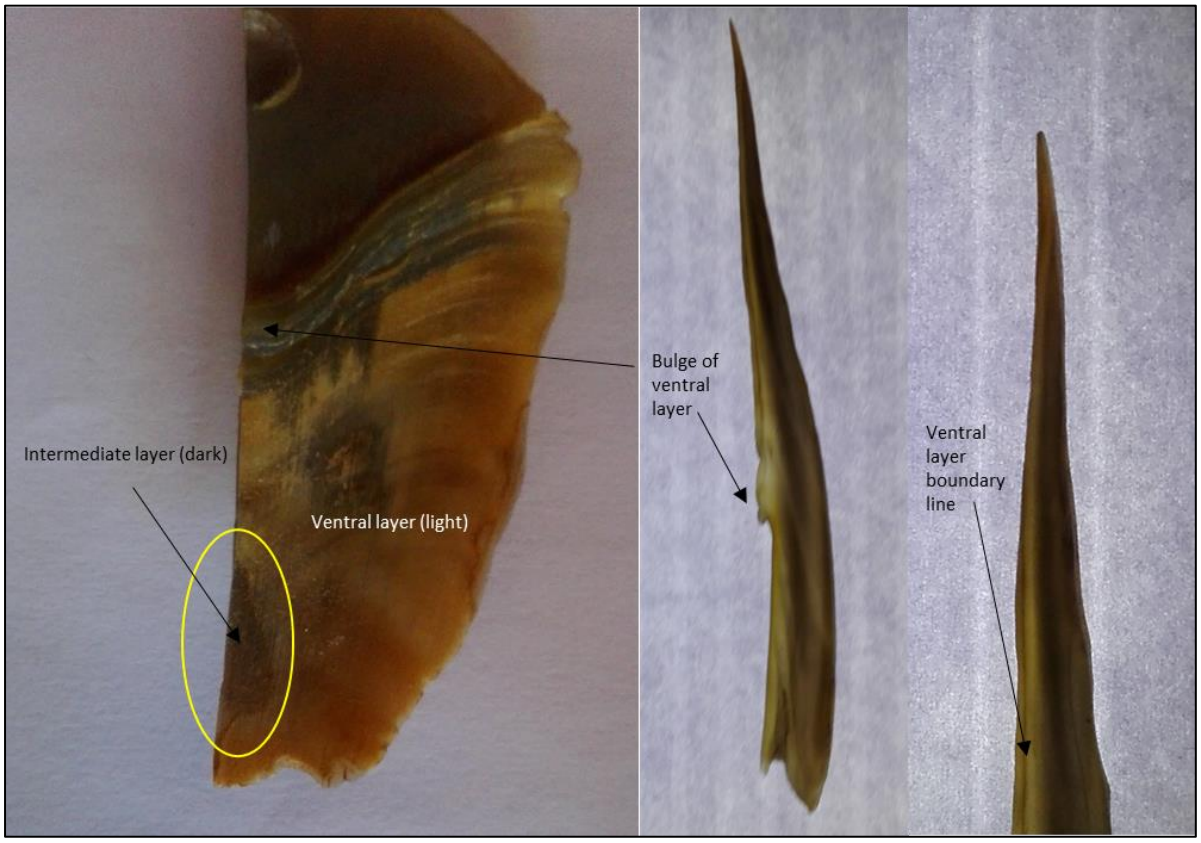


Figure 2.14: Lower back scale of a Temminck's Ground Pangolin (*Smutsia temminckii*) that has been cut in half and sanded in order to visualize the ventral layer bulge mid-way along the length of the scale (marking the site where the scale attaches to the dermis) and loss of the ventral plate layer at the tip due to abrasion.

A suggested minimum sampling protocol is the collection of homogenised scale material from deceased individuals to account for intra-scale variability, along with a corresponding scale tip sample to account for the observed enrichment between scale tips and homogenised samples. Although determining an enrichment factor between the scale tip and homogenised sample was considered, due to the possible effects of *in utero*, suckling and scale friction-related variability, it is unlikely that this will be viable (see: Figure 2.6). This means that geographic separation using only scale tips may not be accurate, and alternative sampling methods may be required. It is therefore suggested that a dorsal scale scraping taken along the length of the scale or small scale clippings taken along the side of the scale should be considered as alternative sampling methods. Both sample types can be ethically obtained from live individuals and once homogenised and validated would account for variability across the

scale. If only scale tip samples can be collected, homogenising these scale clippings may be the best way of reducing possible variation observed across the length of the scale as well as between homogenised and scale tip samples. Regardless of the protocol ultimately selected, it is suggested that material be submitted to forensic isotopic laboratories to ensure that samples can be prepared and analysed by trained personnel.

Intra-individual variation

Samples collected from six different anatomical regions of *S. temminckii* showed no significant difference in their isotopic composition, thus isotopic data from a scale from any region of the body will be representative of the individual's isotopic composition. This aligns with the suggestion that the scales of pangolins grow in a consistent manner across the entire body (Spearman, 1967; Heath, 1992). However, further isotopic analyses using a micro-sampling approach along the length of the scale for different body parts of an individual should be done to obtain an accurate and comparable temporal pattern between scales (see: Vander Zanden *et al.*, 2010). Although time consuming and costly, it is a necessary validation procedure required in order for evidence to be upheld in court.

Unlike bird feathers which grow and moult at different times for different anatomical regions, making specific feather sampling protocols integral for origin testing (Smith and Dufty Jr, 2005; Wassenaar and Hobson, 2006), pangolin scales are not shed nor do they grow at different times (Heath, 1992). This simplifies sampling protocols for pangolin scale seizures, as investigators will not have to identify scale samples from a predetermined anatomical region. This is important as scale seizures are often in the form of bags of scales that contain a number of anatomical regions, individuals and even pangolin species (Mwale *et al.*, 2016).

Inter-individual variation

The variability between individuals is significant, with nitrogen showing a higher level of variability ($\delta^{13}C$ = 5.9‰ and $\delta^{15}N$ = 7.0‰). The results concur with the findings of Pietersen *et al.* (2016) in $\delta^{15}N$ values having a higher inter-individual variability than $\delta^{13}C$ values, although $\delta^{15}N$ values in this study were almost two-fold higher. This variability may be attributed to age-related growth rates, body size or gender (Hesslein *et al.*, 1993; Martínez del Rio and Carleton, 2012; Hobson and Quirk, 2014), however none of these factors showed a significant effect on inter-individual $\delta^{13}C$ and $\delta^{15}N$ values both in this study and in Pietersen *et al.* (2016). Although water sources (below or above ground) may affect local nitrogen delta values, the Kalahari study sites (Tswalu and Kalahari Oryx) have no visible water bodies and rely on deep groundwater boreholes for water (Pietersen *per comm.*), suggesting that this may be an unlikely factor. Other sources of inter-individual variation such as diet preference, nutritional status, hydration and even individual variation in isotopic assimilation are factors that were not tested in this study (Ambrose and DeNiro, 1986; Sealy *et al.*, 1987; Hobson *et al.*, 1993; Martínez del Rio and Wolf, 2005; Barnes *et al.*, 2008; Newsome *et al.*, 2009). Research on Loggerhead Turtle (*Caretta caretta*) scutes suggested that individual variation in isotopic assimilation of food was an unlikely factor due to a large range in $\delta^{13}C$ and $\delta^{15}N$ values (Barnes *et al.*, 2008; Vander Zanden *et al.*, 2010). This may be the case for *S. temminckii*, however additional research is required before a conclusion can be reached. Although physiological status may be one of the more evidence-backed driving factors for intra-scale variability, it is unlikely that individuals were either in a constant state of stress or relief throughout a large percentage of the scale growth period to significantly affect the homogenised scale sample values to such an extent. However, with high levels of variability in carbon and nitrogen isotopic values found in both bird and mammal species alike due to different influences of diet such as protein deficit or excess (Pearson *et al.*, 2003; Sponheimer

et al., 2003a; 2003b), understanding pangolin nutritional ecology may provide better insights into this (Vander Zanden *et al.*, 2018a). Although the majority of the nine micro-sampled individuals showed a degree of overlap in their $\delta^{13}C$ and $\delta^{15}N$ values, two males (STEM 12 and STEM 55) from Kalahari Oryx had markedly different $\delta^{15}N$ range values (Figure 2.11). This suggests that there may be dietary specialisation occurring within the Kalahari Oryx population, particularly when isotopic variation in prey species utilised by *S. temminckii* at this site was 8.6‰ for carbon and 6.0‰ for nitrogen (Pietersen *et al.*, 2016). This is likely due to potential ant and termite prey species feeding on various plant, fungal and even animal tissues (McIlwee and Johnson, 1998). California Sea Otters (*Enhydra lutris nereis*) show a degree of dietary specialisation whereby 48% of the variability in $\delta^{13}C$ and $\delta^{15}N$ values at the inter-individual level can be attributed to diet (Newsome *et al.*, 2009). The same was found for Loggerhead Turtles but at much higher levels of variation than Sea Otters, which was correlated with a higher level of dietary specialisation (Vander Zanden *et al.*, 2010). This along with known variation of ant and termite diversity and abundance across a geographic scale in general (Lindsey and Skinner, 2001; Davies *et al.*, 2003; Leitner *et al.*, 2018), as well as in this study area (Pietersen *et al.*, 2016) suggests that inter-individual variability may be attributable to diet. However, a study on background $\delta^{15}N$ values (soil and plant) over a large Kalahari biome transect showed that local soil heterogeneity had a large effect on gross mineralization and nitrification rates, and thus $\delta^{15}N$ values (Aranibar *et al.*, 2008). This suggests that local soil variability may also influence the base of the food web of pangolins and thus will be reflected in the isotopic variability across a population. There are three major soil types identified in Kalahari Oryx (Pietersen, 2013), each most likely having its own $\delta^{15}N$ values (Aranibar *et al.*, 2008). The type of vegetation may also play a role with five different vegetation communities identified in Kalahari Oryx which were further separated into nine different vegetation types (Pietersen, 2013). Inter-individual variability therefore may not only be affected by an

individual's dietary preference but rather by the prey, vegetation and soil types present in an individual's home range. Furthermore, each individual's isotopic ratio appears to remain relatively static throughout the temporal scale compared to the ratios of other individuals, which fits with the suggested static home ranges of *S. temminckii* (Heath and Coulson, 1997; Pietersen *et al.*, 2014b). Further analyses on soil samples taken from these scales along with tracking and feeding data may provide additional insights into this hypothesis.

Geographic traceability

This study shows that the average $\delta^{13}C$ values (-14.6‰) of the Kalahari population of *S. temminckii* (Tswalu and Kalahari Oryx) fit within the range of C_4 (-18‰ to -10‰) plant carbon isotope values and not that of C_3 (-35‰ to -20‰) values, suggesting a higher abundance of these plant types (mainly grasses) at the base of its food chain (Marshall *et al.*, 2007). Hints of separation for $\delta^{13}C$ values between Lowveld (mesic) and Kalahari (xeric) savanna biomes are starting to show when comparing the Klaserie sample (average $\delta^{13}C$ values = -16.6‰) with that of the Kalahari Oryx and Tswalu samples. Factors that are known to influence plant $\delta^{13}C$ values between the two biomes are altitude, latitude, level of water stress and light intensity (Farquhar *et al.*, 1989; Morecroft and Woodward, 1990; Körner *et al.*, 1991; Kelly, 2000). However, the increased C_4 plant (grasses) prevalence in the Kalahari and increased C_3 plant (woody species) prevalence in the Lowveld suggests that the relative percentage of plants with different photosynthetic cycles may be a major driver in geographic partitioning between the two types of savanna (Tieszen *et al.*, 1979; Huntley, 1982). This fits well with the difference in termite $\delta^{13}C$ values of the mesic savanna dominated Kruger National Park (-20.1‰, which adjoins Klaserie; Sponheimer *et al.*, 2005) compared to that of the xeric savanna dominated Kalahari Oryx (-18.9‰; Pietersen *et al.*, 2016). Additional samples are required from the Klaserie region and the mesic savanna biome in general before conclusions can be made as

only one sample representing the mesic savanna biome was available. With additional samples, I expect to find a larger separation of these two savannas in $\delta^{15}N$ values as animals in more xeric environments generally have enriched ^{15}N values compared to that of more mesic environments due to more ^{15}N enriched soils and a higher level of ^{15}N depleted urea as a cost of reduced hydration (Ambrose and DeNiro, 1986; Schwarcz *et al.*, 1999; Hartman, 2011). However, from the data generated thus far, the inter-individual variability for the Kalahari Oryx site was higher than that of all three sites combined, and thus geographic traceability was not yet realised.

Variation between harvesting methods

There was no significant difference in isotopic values between scale harvesting methods, even after scales were completely charred. In a study on the cremation of cow tibia for forensic purposes, carbon isotopic values remained constant until reaching 200 °C (brown in colour with notable fissures) where-after they became significantly more depleted from the basal values as the temperature was increased (Harbeck *et al.*, 2011). Although $\delta^{15}N$ values of the tibia were not measured, it was noted that the heavier the isotope, the higher the temperature required before it would affect isotopic ratios (Harbeck *et al.*, 2011). This suggests that boiling should not affect pangolin scale $\delta^{13}C$ and $\delta^{15}N$ values. Exposure to direct flames would likely affect $\delta^{13}C$ values, considering the estimated temperature for a wood fire is around 320–400 °C at the flame tip and 900 °C in the continuous flame region (Babrauskas, 2006). However, flame temperature is highly stochastic possibly resulting in the length of time required to cause isotopic fractionation not being reached, whilst the thermodynamics of heat transfer from flames to pangolin scales may have reduced the actual temperature experienced by the scales (Cox and Chitty, 1982; Sullivan *et al.*, 2003; Babrauskas, 2006). Additionally, under laboratory observations, a significant part of the centre of the scales that underwent charring showed no

structural difference at a superficial level to that of the controls. This suggests that a significant part of the scale could have been below the necessary 200 °C and thus the carbon and nitrogen isotope ratios would not have been affected. These observations along with the stochastic nature of flaming scales on a wood fire and the time required to cause isotopic fractionation may explain the results found in this study. Future experiments involving additional isotopes should still be analysed under these harvesting method tests as they may be influenced differently. An example of such is the effect that boiling water may have on scale hydrogen and oxygen isotopic ratios due to possible movement of hydrogen and oxygen atoms between the boiled water and scales (von Holstein *et al*., 2018).

Study caveats

Although the sample size was large for a pangolin study, additional samples are required to increase confidence in the conclusions drawn. The sample set was also heavily biased toward the Kalahari Oryx site (84% of scale clippings and 94% of whole scale samples) with only one specimen (three whole scales) from Tswalu and Klaserie each. The Klaserie sample was also collected during the 1990's while all of the Kalahari samples were collected post-2009. Whilst the homogenised scale duplicate part of the sampling protocol showed potential for forensic purposes, use of a scale tip drilling did not, highlighting the need for a better sampling protocol for live individuals.

Conclusion and synthesis

Although the aim of this study was not to determine the ecological factors behind pangolin isotopic composition, this information may better guide wildlife forensic researchers in identifying best practices for accurate and reliable traceability estimates of pangolins (Wassenaar and Hobson, 2006; Barnes *et al.*, 2008). Based on these data, nutritional and water

stress (across scale length) and differences in origin of growth from the germinal epidermis (across scale depth) may be the major factors driving intra-scale isotopic variability, suggesting that homogenised bulk samples across the entire length and depth of the scale are the most suitable to characterise the geographic origin of individuals. There were no significant differences between duplicates of homogenised scale samples, however with the nature of forensic cases and the possible effects that one anomalous result may have on areas with limited sample sizes, duplicate or even triplicate sampling should be conducted. There was a significant enrichment in ^{13}C for tip samples when compared to homogenised duplicates, which may be due to relative thickness of scale layers differing across the scale length. Due to an enrichment factor between scale tips and whole scale homogenised samples not being viable, a more suitable method of scale sampling is required. The method would need to be able to account for intra-scale variation whilst concurrently being ethical for sampling of live individuals. The anatomical location of a scale had no effect on isotopic ratios, suggesting that investigators do not need to select specific scale types during seizure analyses. This study revealed that the highest variability was between individuals, specifically with regards to $\delta^{15}N$ values. This may be due to home range selection that can heavily influence $\delta^{15}N$ values in terms of differences in prey abundance as well as variation in soil and vegetation type across the landscape. Harvesting methods had no significant effect on isotopic ratios, suggesting that scale harvesting methods will not affect traceability estimates. These initial $\delta^{13}C$ and $\delta^{15}N$ isotopic data suggest that it may be possible to isotopically differentiate between the Lowveld and Kalahari regions, however analyses of additional Lowveld samples are required to confirm this. In order for this tool to be effective, higher resolution of geographic traceability needs to be realised by the inclusion of additional isotopes such as oxygen, hydrogen and sulphur.

Future work on traceability estimates should not only consider the factors of variation tested, but also consider the minimum requirements for method validation to reduce

discrepancies of forensic stable isotope analyses that may arise in a court of law (Dunn *et al.*, 2017). The feasibility of other methods of geographic traceability that complement this approach also need to be investigated, including the use of microbial profiles.

Chapter 3:

Microbial diversity in Temminck's Ground Pangolins (*Smutsia temminckii*): an evaluation of geographic traceability potential

Introduction

Microbial traceability encapsulates a variety of methods under the banners of genotypic (restriction-fragment length polymorphism typing, Sanger sequencing, next generation sequencing) or phenotypic (outer membrane protein profiles, antibody reactivity, serogrouping) identification in order to characterize distinct microbial communities and/or indicator organisms of a specific host or location (Scott *et al.*, 2002; Schürch and Siezen, 2010). The rationale behind microbial traceability through genotypic methods is the assumption that microbial communities are influenced by host or environmental-specific factors and can go as far as to influence the adaptation of a particular microbial species, with subsequent propagation to the remainder of the species population (Scott *et al.*, 2002; Fierer and Jackson, 2006; Gilbert *et al.*, 2016). These methods have been used in a variety of fields including the traceability of faecal contamination in environmental samples (Scott *et al.*, 2002; Sargeant *et al.*, 2011), pests in agriculture (Mazzaglia *et al.*, 2012), food-borne pathogens in public health cases (Abubakar *et al.*, 2007; Muellner *et al.*, 2013) and bioterrorism events (Keim *et al.*, 2004). They have also been used in tracing disease pandemics infecting domestic animal (Bastos *et al.*, 2003; Abolnik *et al.*, 2006), wildlife (Cleaveland *et al.*, 2000) and human populations (Gardy *et al.*, 2011; Omrani *et al.*, 2015; Wu *et al.*, 2015).

Molecular methods used for microbial traceability

The different fields that make use of microbial traceability use identical molecular techniques to determine source populations or locations. These range from simple methods such as determining which likely source populations are seropositive for a specific microbe just before an outbreak (Cleaveland *et al.*, 2000) to whole genome sequencing of the disease and subsequent network analysis of infected individuals (Gardy *et al.*, 2011). The majority of these cases are of a well-characterised species and thus authorities often have access to a large

amount of information relating to the characteristics of the species, including a list of known hosts or regions where the species may occur. This makes tracing the source of these outbreaks relatively easy compared to outbreaks that are from novel or understudied species, highly complex and finite cases (hospital cases), outbreaks that result from socio-economic changes rather than genetic ones, or outbreaks involving microbes that are omnipresent or environmental in origin (Beres *et al.*, 2010; Harris *et al.*, 2010; Schürch and Siezen, 2010; Gardy *et al.*, 2011; Snitkin *et al.*, 2012). When these types of outbreaks occur the use of traditional molecular methods such as restriction-fragment length polymorphism typing (RFLP), sequence repeat lengths (microsatellites, short tandem repeats), Sanger sequencing of housekeeping genes and even multi-locus sequence typing (MLST) often do not have the resolution required to trace their sources (Keim *et al.*, 2004; Schürch and Siezen, 2010; Rasko *et al.*, 2011). The advent of next generation sequencing technologies (NGS), particularly whole genome sequencing, single nucleotide polymorphisms (SNP's) and insertion deletion patterns (indels), has revolutionised microbial traceability due to the ability to scan the entire genome for changes between isolates and thus dramatically increase the resolution (Keim *et al.*, 2004; Schürch and Siezen, 2010; Rasko *et al.*, 2011). However, this has not meant the end of non-NGS based techniques. In fact, it has in many ways improved their resolution by being able to develop targeted approaches for specific taxa or strains, gene profiles, SNP's or even whole genomes using multi-locus sequence typing, Sanger sequencing and microarrays (Wu *et al.*, 2004; Loy *et al.*, 2005; He *et al.*, 2007). The latter is often seen as being in direct competition with NGS technologies for microbial ecology research (Roh *et al.*, 2010).

Microbial traceability as a proxy for host origins

One step further is using these technologies to identify microbial traceability estimates as proxies for estimating the origins of their hosts. This sort of traceability has a strong

anthropocentric focus which can be traced back to forensic cases investigating HIV transmission (Ou *et al.*, 1992; Schmedes *et al.*, 2016). However, traceability through microbial proxies are starting to make an appearance in the tracking of wildlife. The rationale behind this notion of microbial proxy tracing is that there is a direct link between the microbes of an animal or human and the environment in which they live as well as the population with which they interact (Sousa and Silva-Souza, 2001; Hewitt *et al.*, 2012).

There are different levels at which investigators can use microbial proxy tracing to determine animal geographic origins. At a microbial community level, Tatsadjieu *et al.* (2010) were able to demonstrate geographic separation, independent of the fish species tested, between three lakes in Cameroon (two being 100 km apart), based solely on a denaturing gradient gel electrophoresis (DGGE) analysis of bacterial taxa. This technique has also worked on other freshwater fish species in Asia (Montet *et al.*, 2012), some marine fish species (Smith *et al.*, 2009a; Higgins *et al.*, 2010) and seahorses (Cohen *et al.*, 2018).

The second level at which microbial proxy tracing can be used is through the identification of key microbial taxa that are site-specific and using their prevalence/absence as proxies for traceability. A study on Seabass (*Dicentrarchus labrax*) was able to attain a geographic resolution of <500 m by determining these key traceability-related genera when comparing bacterial communities between fish farms (Pimentel *et al.*, 2017). Although this level of resolution can be attributed to the fish farms acting as isolated populations, this method may still be viable for wild populations provided that there is sufficient separation between the populations in question. In a forensic investigation of a case of fraudulent geographic labelling involving the selling of supposedly foot and mouth disease (FMD) virus-free Cape Buffalo (*Syncerus caffer*) in South Africa, microbial analyses provided evidence for falsified documentation through origin tracing of microbes commonly found in *S. caffer* (Vosloo *et al.*, 2001). These authors initially used the site-specific presence/absence level approach to

determine that the buffalo most likely originated from the Kruger National Park (KNP) rather than the purported farm in a separate, disease-free province (Vosloo *et al.*, 2001). This was determined by the dual presence of *Theileria parva* (causative agent of corridor disease) and FMD virus, a phenomenon only seen in KNP buffalo at the time (Vosloo *et al.*, 2001).

The final level of microbial traceability is to identify geographically specific genotypes (topotypes) within a single microbial taxon. Vosloo *et al.* (2001) did this by comparing the sequence data of the isolated FMD virus from the buffalo in question to that of isolates from several southern African reserves. Phylogeny testing not only separated FMD geographically at a per reserve level, but managed to separate the KNP into northern and southern subregions, with the isolates of the buffalo in question aligning most closely with the northern subregion (Vosloo *et al.*, 2001). The five FMD genotypes have also been shown to have distinct geographic areas within the KNP for both Impala antelope (*Aepyceros melampus*) and Cape buffalo (Bastos *et al.*, 2000). A RFLP analysis of *Mycobacterium bovis*, causative agent of bovine tuberculosis (BTB) was used to investigate the most likely geographical origin of the BTB-infected cattle farm that was the source of spillover between cattle and buffalo in the KNP (Vosloo *et al.*, 2001). This research has led to the recent mapping of FMD topotypes, of the different serotypes, across the globe in order to determine livestock trade-related risk of transmission as well as to give insight into geographic origins of outbreaks through this trade (Di Nardo *et al.*, 2011). It also led to the reconstruction of host species' geographic movements over a 20-year period based purely on dated sampling events (Hall *et al.*, 2013). The ability to use all three levels of microbial traceability, *viz*. microbial communities, site-specific presence/absence keys and topographical variants of a single microbial species (topotypes), highlights the value of microbial approaches to host origin traceability.

Metagenomics as a tool for traceability

The aforementioned examples show the potential resolution and value of using microbial traceability as a proxy for host traceability. However, before this can be attempted it is important to establish baseline estimates of microbial species diversity for the specific host species across the geographic range in question. These data can be used directly for whole microbial community traceability but may also help guide investigators in determining which taxa will be key for topotyping or presence/absence-based traceability. The most widely used technique for establishing microbial community diversity is metagenomics which has evolved from culturing to PCR-cloning (polymerase chain reaction cloning) with subsequent Sanger sequencing to microarrays and high throughput next generation sequencing (Xu, 2006; Roh *et al.*, 2010; Zhou *et al.*, 2015). Microarrays are seen as a closed architecture system, because they can only identify taxa for which there are pre-designed target probes (Roh *et al.*, 2010; Zhou *et al.*, 2015). PCR-cloning works through the amplification of a single gene or gene region with universal primers, in combination with cloning of the amplified targets of multiple microbial species, with subsequent Sanger sequencing of individual clones. This means that PCR-cloning, along with NGS, is an open architecture system which presents a less biased approach to microbial community assessments whilst also permitting identification of new microbial species (Roh *et al.*, 2010; Zhou *et al.*, 2015). For both approaches, microbial discovery is only restricted by PCR amplification using universal primers prior to analyses (Roh *et al.*, 2010; Zhou *et al.*, 2015). Because these two approaches are open architecture systems, both were evaluated in this study.

PCR-cloning has the advantage of being highly reliable as researchers are in control of each step of the procedure. They can identify discrepancies in taxonomic assignment as each sequence is assigned directly through tools like BLAST (Basic Local Alignment Search Tool), they can use long sequence reads which reduces the chance for misclassifications and can

include more stringent taxonomic assignment tests such as phylogenetics, which allows for more resolution below the genus and species levels (Patel *et al.*, 2004; Kircher and Kelso, 2010; Samarajeewa *et al.*, 2015; Muñoz-Colmenero *et al.*, 2017). Additionally, error rates of Sanger sequencing are low and vary from 0.001% to just more than 1%, depending on the software used for read processing (Hoff, 2009). However, PCR-cloning is more expensive per sample, it has a long turnaround time and unless many colonies are sequenced it will only provide a fraction of the species that NGS identifies (Samarajeewa *et al.*, 2015; Muñoz-Colmenero *et al.*, 2017). This means that PCR-cloning approaches are often limited to presence/absence applications, due to the lower microbial diversity recovery potential (Samarajeewa *et al.*, 2015). NGS has the advantage of having short turnaround times as samples can be analysed simultaneously, it has a high throughput which allows for species abundance estimates, it can identify taxa occurring at low abundance and because it is increasingly being applied, cost-efficiency is steadily improving (Samarajeewa *et al.*, 2015; Muñoz-Colmenero *et al.*, 2017). The disadvantages of this technology are that sequence error rates are high and bioinformatics pipelines can be unreliable, often requiring manual revision and knowledge of the analyses (Muñoz-Colmenero *et al.*, 2017). The type of NGS platform used can also dramatically change the accuracy, read length, cost and turnaround time (Quail *et al.*, 2012). With both cloning and NGS technologies having different strengths and weaknesses as metagenomics approaches, it is important to establish which of the two approaches is best suited to identifying communities or key taxa that can be used as presence/absence or topotyping traceability estimates, so that a single metagenomics platform can be implemented.

Taxonomic assignment for traceability

Another consideration when applying metagenomics approaches is whether heuristic search tools such as BLAST are sufficient for determining species identity or whether

phylogenetic approaches should be given greater emphasis. BLAST is the most widely used heuristic approach for sequence identity mainly due to it being built into the National Center for Biotechnology Information (NCBI) software, which hosts one of the largest open-access nucleotide databases to date, GenBank (Jelić *et al.*, 2003; Mount, 2007). BLAST was developed as a faster alternative to dynamic programming algorithms which at the time needed excessive computational requirements and time to search (Altschul *et al.*, 1990). The problem relating to heuristic approaches is two-pronged: the first is the limitations of the heuristic approach itself whilst the second relates to the reference database used. When comparing BLAST assignment to distance matrices and phylogeny-testing based methods on simulated data, they all had a similar outcome unless taxa were absent from the database (Ross *et al.*, 2008). A study reviewing the databases used for 16S rRNA identification presented varied results, however curated databases that use sequences from bacterial strains of culture collections (i.e. MicroSeq and RIDOM) identified significantly more taxa correctly than did non-curated public databases such as GenBank or RDP-II (Woo *et al.*, 2008). One advantage of non-curated databases is that they are constantly updated and are larger than curated databases, meaning that if legitimate sequences can be identified, these databases can become advantageous.

Sources of variation in microbial communities

Regardless of the metagenomics method and downstream assignment analyses used, variation in the microbial communities between individuals and locations may stem from other sources including the type of organ used, date of sampling (seasonality and year), time lapse between host death and sampling and even cause of death (Carter *et al.*, 2015; Hyde *et al.*, 2015; Cohen *et al.*, 2018; Quaak *et al.*, 2018). Microbial community composition between different human body regions have been shown to differ, with intra-body region variation also

apparent particularly for the buccal region (Huse *et al.*, 2012; Quaak *et al.*, 2018). Additionally, temporal microbial community variation has been well documented in seahorses which showed more variation in microbial community between two consecutive sampling months at the same site than between geographically separated samples, even though these samples did separate on a geographic basis when compared over the same month (Cohen *et al.*, 2018). This temporal variation is important to consider for host samples being analysed, particularly carcasses. Temporal shifts in microbial communities of carcasses from aerobic to anaerobic as bloating progresses, a reduction in proteolytic bacteria during decay, increased prevalence of fly-related bacteria once flies can access the carcass as well as increased prevalence of soil-associated bacteria once the carcass has dehydrated and starts to skeletonize, may all affect traceability estimates (Howard *et al.*, 2010; Hyde *et al.*, 2013; Hyde *et al.*, 2015). A similar notion has been shown for seasonality, whereby carcasses differ in microbial profiles during decomposition in different seasons (Carter *et al.*, 2015). All of these factors are important sources of variation that need to be accounted for when conducting microbial traceability analyses as these may dramatically affect the outcomes.

Study aim

The aim of this study was to compare the impact of different taxonomic assignment methods and metagenomics approaches when establishing baseline estimates of bacterial diversity in *Smutsia temminckii*, and to evaluate the microbial proxy tracing potential for this species. This was achieved by:

1) Comparing universal bacterial 16S rRNA PCR-cloning and Sanger sequencing with Ion Torrent NGS.
2) Comparing a common heuristic search tool (BLAST) with phylogeny testing when evaluating microbial community diversity and taxon assignment.

3) Determining whether additional sources of variation (organ type, year of sampling, cause of death) significantly influence microbial communities of *S. temminckii*.
4) Comparing the bacterial community composition between Tswalu Kalahari Reserve and Kalahari Oryx Private Game Farm using PCR-cloning data.
5) Integrating the results in order to determine the best metagenomics approach for future traceability applications.

Methods

Sample collection

Lung and spleen samples were obtained from eight *S. temminckii* specimens, four from Tswalu Kalahari Reserve (hereinafter Tswalu) and four from Kalahari Oryx Private Game Farm (hereinafter Kalahari Oryx) as part of a previous ecological study (Appendix 1; Figure 3.1; Pietersen, 2013).

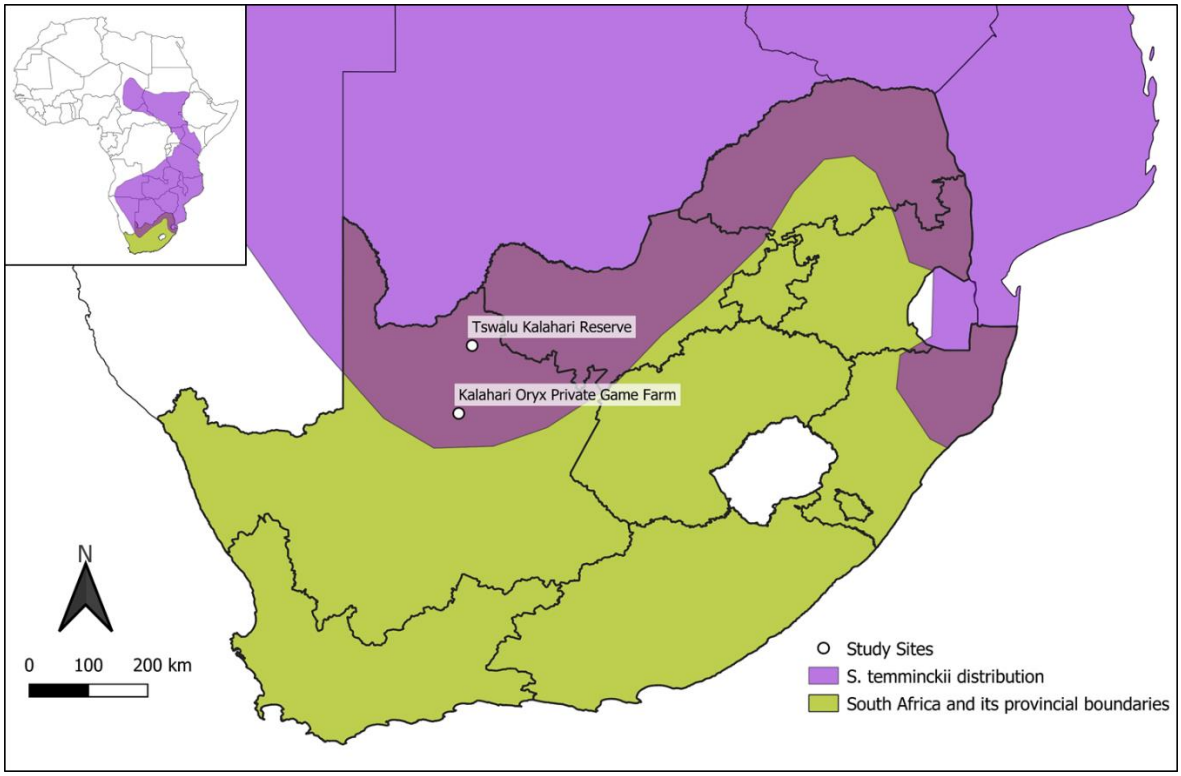


Figure 3.1: The two study sites in the Northern Cape Province of South Africa (green shaded area) and their relative positions within the known distribution of the Temminck's Ground Pangolin (*Smutsia temminckii*).

Specimens perished due to accidental electrocution on electrified game fencing as well as natural starvation and were collected within hours of death. Specimens were immediately dissected upon receipt or were individually sealed in sterile plastic bags and stored in a dedicated freezer at -20 °C until later dissection. Upon dissection, tissue samples of each animal were individually stored in 1.5 mL Eppendorf collection tubes containing absolute ethanol.

DNA Extractions

DNA extractions were conducted using a ZR Genomic DNA™ Tissue MiniPrep extraction kit (Zymo Research Corporation, California, U.S.A), following the manufacturers protocol for solid tissue. All extractions were performed within the Transboundary Animal Diseases Programme (TADP) Biosafety Level Three (BSL3) facility at the Onderstepoort Veterinary Institute, University of Pretoria, with section 20 approval (reference number: 12/11/1/1) from the South African Department of Agriculture, Fisheries and Forestry (DAFF).

Experiment 1: TOPO® TA Cloning of the 16S rRNA gene

All eight individuals were selected for the PCR-cloning-sequencing component, with DNA extracted from either lung or spleen, or both (Table 3.1) being used as template for the 16S rRNA amplification. Cloning was performed with the TOPO® TA Cloning® Kit for Sequencing (Thermo Fisher Scientific Inc.).

Table 3.1. Sampling details of the eight Temminck's Ground Pangolin (*Smutsia temminckii*) individuals that underwent cloning. Cloning reactions were conducted on both lung and spleen samples for individuals STEM 53 and STEM 70, with asterisks indicating the organ from which sequences underwent phylogeny testing.

Specimen	***Site***	***Organ***	***Year of sampling***	***Cause of death***	***No. of clones processed***
STEM 36	Kalahari Oryx	Spleen	2011	Electrocution	15
STEM 41	Kalahari Oryx	Lung	2011	Electrocution	11
STEM 53	Kalahari Oryx	Lung* and spleen	2011	Electrocution	63
STEM 70	Kalahari Oryx	Lung and spleen*	2012	Starvation	40
STEM 89	Tswalu	Lung	2012	Electrocution	11
STEM 115	Tswalu	Spleen	2015	Electrocution	12
STEM 116	Tswalu	Lung	2015	Electrocution	11
Pangolin 2	Tswalu	Spleen	2014	Starvation	11

Template DNA (3 µL) was added to a 37 µL Polymerase Chain Reaction (PCR) master mix containing DreamTaq™ DNA polymerase (Thermo Fisher Scientific Inc., Massachusetts, U.S.A) at a concentration of 1.5 units/40 µL, 10 X DreamTaq™ buffer (Thermo Fisher Scientific Inc.), 0.25 µM dNTP's (Thermo Fisher Scientific Inc.) and 0.4 µM of the broad-range, universal 16S rRNA primer set pA (Edwards *et al.*, 1989) and 1492R (Reysenbach *et al.*, 1992). Touchdown thermal cycling, with annealing at a final temperature of 56 °C was performed when amplifying the ~1.5 kbp amplicon. An extended elongation time was included during completion of thermal cycling in order to ensure that full length amplicons were 3' adenylated (Table 3.2).

Table 3.2: Touchdown thermal cycling conditions used for the broad-range, universal 16S rRNA gene amplification with primers 27F and 1492R, prior to TOPO® TA cloning.

Stage	*Denaturation*	*Annealing*	*Elongation*	*Cycles*
[1]	96°C; 10s			1
[2]	96°C; 12s	58°C; 30s	72°C; 1:50s	3
[3]	96°C; 12s	57°C; 25s	72°C; 1:50s	5
[4]	96°C; 12s	56°C; 20s	72°C; 1:45s	35
[5]			72°C; 20min	1

PCR products were visualised on 1.5% agarose gel under ultraviolet light with positive samples undergoing purification directly from the tube using a Roche High Pure PCR Product purification kit (Roche Diagnostics GmbH, Mannheim, Germany), following the manufacturers protocol. Purified samples were used for cloning, rather than using the PCR products directly, as suggested by the kit manufacturer. This decision was based on the higher transformation efficiency observed when the effects of a variety of components were initially evaluated, including (i) direct use of the PCR product, (ii) using PCR products with loading dye added, and (iii) using purified PCR products. Each of these three variables were evaluated using the control DNA provided with the kit, prior to attempting the 16S rRNA cloning experiments.

Selective plates were prepared prior to experiments by making up a 2.8% mixture of nutrient agar (LAB M, Neogen Corporation, Michigan, U.S.A), autoclaving it for 20 minutes at 121 °C and adding reconstituted kanamycin sulphate (Gibco, Thermo Fisher Scientific Inc.) to a final concentration of 50 µg/mL, once the agar had cooled to 50 °C. No additional reagents such as IPTG and X-gal were added as the TOPO® TA cloning kit does not use blue-white screening but rather relies on the pCR™4-TOPO® vector containing a lethal gene which is activated if no amplicon is inserted into the vector.

TOPO® TA cloning reaction mixtures were set up on a per-sample basis as follows: 2 µL purified product, 1 µL salt solution, 2 µL water and 1 µL of the pCR™4-TOPO® vector. Reaction mixtures were gently mixed and left to incubate for 5 minutes at room temperature prior to placement on ice to halt the cloning reaction. Following cloning inactivation, the One Shot® TOP10 chemically competent *Escherichia coli* cells were thawed on ice prior to adding 2 µL of the TOPO® TA cloning reaction mixtures of each sample to each competent cell vial. The resultant transformation mixture was left to incubate on ice for 5 minutes after which they were heat-shocked by placing the vials in a 42 °C water bath for 30 seconds prior to placing them directly on ice and adding 250 µL of room temperature Super Optimal broth with Catabolite repression (S.O.C.) medium. Vials with the transformed product were incubated at 37 °C with continuous shaking (200 RPM) for one hour. The product was aliquoted (40 µL) onto pre-warmed selective plates, spread using sterile plastic spreaders and incubated for 24 hours at 37 °C. All cloning reactions were conducted under a BSL2 laminar flow (Vivid Air, Pretoria, South Africa) in order to prevent contamination.

Resultant colonies were picked using sterile toothpicks and placed in 100 µL of phosphate-buffered saline (PBS) before undergoing a modified version of the Glucothiocyonate-silica method of Boom *et al.* (1990), with a final elution in 40 µL of 1 x Tris Elution buffer. A maximum of 15 colonies from each of the eight samples were randomly picked for extraction, apart from STEM 70 (40 colonies) and STEM 53 (63 colonies). More colonies were picked from the latter two samples due to these two samples being used for both the cloning and NGS components, and thus a larger number of colonies were required for comparative purposes. PCRs were run on DNA extracted from each of the colonies picked for evaluation (3 µL of extracted DNA) using a 37 µL master mix containing DreamTaq™ DNA polymerase (Thermo Fisher Scientific Inc.) at a concentration of 1.5 units/40 µL, 10 X DreamTaq™ buffer (Thermo Fisher Scientific Inc.), 0.25 µM dNTP's (Thermo Fisher

Scientific Inc.) and 0.4 µM of the cloning kit specific primer set. The primer set consisted of the M13 Forward and M13 Reverse primers which target vector sites flanking the inserted amplicon. PCR products were visualised under ultraviolet light on 1.5% agarose gel. Samples that showed positive irradiation bands of the expected ~1.5 kbp amplicon size were purified directly from the tube using the Roche High Pure PCR product purification kit (Roche Diagnostics GmbH), following the manufacturers protocol. Purified products were again visualised on 1.5% agarose gel for estimation of product concentration, prior to cycle sequencing with BigDye Terminator v3.1 cycle sequencing (Applied Biosystems, California, U.S.A), using the M13 forward primer. Cycle sequencing products underwent sodium acetate DNA precipitation before being sent to the University of Pretoria's core sequencing facility where all samples were denatured prior to sequencing on an ABI PRISM^{TM} 3100 Analyzer (Applied Biosystems, Foster City, California).

Sequences were visualized and edited using the Chromas package embedded in MEGA 7 (Kumar *et al.*, 2016) with taxonomic assignment of each sequence being achieved by determining the closest match in nucleotide identity and percentage coverage using the nucleotide BLAST search tool in NCBI (Altschul *et al.*, 1990). For phylogeny-based taxonomic assignment, reference datasets comprising of related sequences to BLAST assignments (including outgroup taxa) were downloaded from GenBank. As DNA amplicons can be inserted in either orientation during cloning, sequencing using the same primer (M13F) yielded either sense or antisense strands. In order to accommodate for this issue in a cost- and time-effective manner (forensic case requirements), genus-specific reference datasets corresponding to either sense or antisense sequence strands were generated with each undergoing independent analysis. These reference datasets were aligned using the ClustalW alignment tool embedded in MEGA 7 prior to undergoing p-distance Neighbour Joining analyses whereby nodal support was assessed by non-parametric bootstrap resampling (5 000 replicates). Samples were

assigned new taxon names reflecting their phylogenetic lineages under the strict taxonomic assignment method outlined in Meier *et al.* (2006) and Ross *et al.* (2008). These figures can be found in Appendices 2 and 3 for all genera identified during BLAST that required only one tree (inferred with either the sense or antisense sequence data) and Appendices 4–7 for all genera identified during BLAST that required both sense and antisense dataset analyses.

Experiment 2: Ion Torrent 16S™ Metagenomics (NGS)

For this component I used the Ion Torrent metagenomics platform as it has an error rate of 1.71% and has a quicker sequencing time (2 hours) compared to Illumina and PacBio platforms (Quail *et al.*, 2012). However, it is one of the more expensive NGS options to run at a per base pair basis (Quail *et al.*, 2012).

Extracted DNA from the spleens of two specimens, namely STEM 53 and STEM 70 (both originating from Kalahari Oryx), were tested under the following quality control measures: a DNA concentration above 10 ng/µL, and Nanodrop 2000 spectrophotometer (Thermo Fisher Scientific Inc.) readings above 1.6 for the 260/230 ratio and between 1.8–2.0 for 260/280. After passing the quality control phase, aliquots were sent to the University of Pretoria's core next generation sequencing facility for library preparation and sequencing. At the facility, 6 ng aliquots of each sample were split into two tubes (primer pools) and amplified over 27 cycles using the Ion Torrent 16S™ Metagenomics Kit (Thermo Fisher Scientific Inc., Massachusetts, U.S.A), whereby annealing times were extended to 1 minute instead of the manufacturer's guidelines of 30 seconds. The first tube of each sample contained a primer pool for variable regions two (V2), four (V4) and eight (V8) whilst the second contained a primer pool for variable regions three (V3), six to seven (V6–7) and nine (V9) of the 16S rRNA gene. The PCR products from each tube were purified using AMPure™ XPA kit (Thermo Fisher Scientific Inc.) with subsequent quality analysis and quantification of amplified DNA being

conducted using a Qubit 2.0 fluorometer (Invitrogen, Thermo Fisher Scientific Inc.) and qPCR (Ion Universal Library Quantitation Kit; Thermo Fisher Scientific Inc.), respectively. A genomic library was constructed for each sample through random fragmentation of amplified DNA with the Ion Plus Fragment Library Kit (Thermo Fisher Scientific Inc.), formation of sticky ends through smooth end polishing, ligation of Ion Xpress™ Barcode Adapters (Thermo Fisher Scientific Inc.) to the sticky ends, and subsequent amplification of DNA with these adapter primers. Amplicons larger than 450 bp were selected for template preparation using the Ion PGM™ Hi-Q™ OT2 Kit (Thermo Fisher Scientific Inc.), which involved ligation of DNA fragments to ion sphere particles (ISP's) and subsequent amplification of each fragment for each ISP using an emulsion PCR procedure (amplification of one fragment attached to its ISP per droplet). Ion sphere particles containing amplified fragments were washed over an Ion 318™ Chip with each ISP slotted into one well. Finally, an Ion PGM™ Hi-Q™ Sequencing Kit (Thermo Fisher Scientific Inc.) was used to sequence the amplified fragment for each ISP by washing over a single nucleotide at a time, whilst conductors in the well measured the addition of each nucleotide to the amplified fragment through cleavage and release of the H^+ ion (Rothberg *et al.*, 2011).

Analyses

The bioinformatic analyses of Ion Torrent metagenomics outputs were analysed using the 16S metagenomics pipeline in Ion Reporter™ Software v5.10.20 (Thermo Fisher Scientific Inc.). Sequencing reads were assigned a taxon identity by using a two-step approach where the reads were searched against the MicroSEQ™ ID database and reads that did not match this database subsequently searched against the GreenGenes database. Operational Taxonomic Units (OTUs) were assigned as follows: >97% similarity for genus identification and >99% for species identification. A minimum of ten sequence reads were needed before an OTU was

assigned to the output species dataset. In addition, unmapped reads, constituting a total of 13 sequences, were used in manual nucleotide BLAST searches against the GenBank database and subsequently added to the species dataset created from the bioinformatics outputs. QIIME™ software (Caporaso Lab, Northern Arizona University, USA) was used to run Shannon alpha diversity analyses on the microbial sequence data that were generated for both samples.

The baseline microbial community resulting from phylogeny testing of data obtained from the cloning component of the study was compared to that of the corresponding samples' BLAST results in a Permutational Multivariate Analysis of Variance Using Distance Matrices (PerMANOVA) analysis. A PerMANOVA was conducted on all BLAST and phylogeny tested cloning data using the "adonis2" package in order to test whether sampling site (Kalahari Oryx/Tswalu), organ type (lung/spleen), date of sampling and type of death (starvation/electrocution) had a significant effect on microbial species composition. The effect of site on microbial composition of phylogeny-tested cloning data was visualised using a cluster dendrogram and non-parametric Bray–Curtis multidimensional scaling (nMDS) analysis using the function metaMDS. Microbial communities that were identified using BLAST searches of cloned STEM 70 and STEM 53 spleen samples were compared to those of the corresponding lung samples, using cluster dendrograms (statistics were excluded due to the small sample size, n = 4). Species rarefaction curves were constructed for the species datasets that resulted from BLAST (cloning), phylogeny testing (cloning) as well as the Ion Torrent analyses. Shannon alpha diversity estimates were calculated for each cloned sample whilst the microbial communities inferred from the cloning and Ion Torrent NGS analyses were compared using a PerMANOVA analysis. The spleen as well as the combined spleen and lung microbial communities of STEM 70 and STEM 53 that were inferred from cloning were compared to the corresponding spleen microbial communities resulting from the Ion Torrent

metagenomics approach using cluster dendrograms (statistics were excluded due to the small sample size, n = 4). All statistical analyses were performed on species count data conducted under the "vegan" package in the Rstudio interface of R v3.5.1 (Rstudio Inc., Massachusetts, U.S.A) under test-specific assumptions.

Results

BLAST *versus* phylogeny testing of cloned data

The microbial community resulting from phylogeny testing was significantly different to that of the BLAST output on corresponding sample colonies (F = 1.827, p = 0.024), with the two identification methods only having four taxa in common (Figure 3.2).

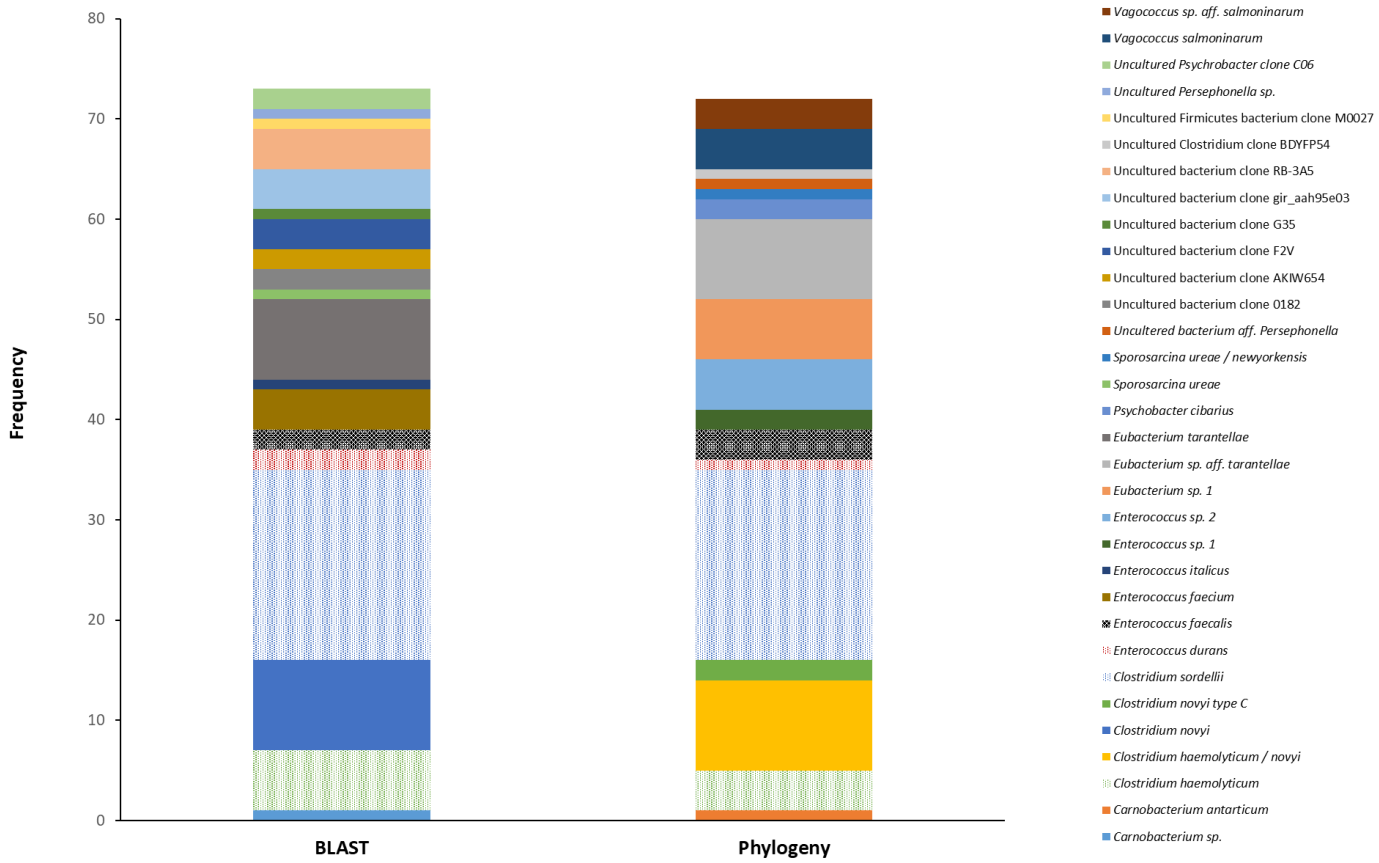


Figure 3.2: Difference in taxon identification and count between Basic Local Alignment Search Tool (BLAST) and phylogeny methods of taxonomic assignment for cloned samples of eight Temminck's Ground Pangolins (*Smutsia temminckii*). Patterned segments indicate taxa found in both methodologies.

Only one taxon, *Clostridium sordellii*, had the same prevalence (count) for both the BLAST and phylogeny testing methods, indicating shifts in taxon count data being more heavily influenced by the adopted methodology than taxon presence/absence data. Notable changes from BLAST to phylogenetic assignments include the amalgamation of some uncultured bacterial clones and shifts from uncultured clones to described *Enterococcus*, *Eubacterium*, *Psychobacter* and *Vagococcus* species (Figure 3.3; Appendix 8). Phylogeny testing was also able to identify a type strain, namely the *Clostridium novyi* type C strain.

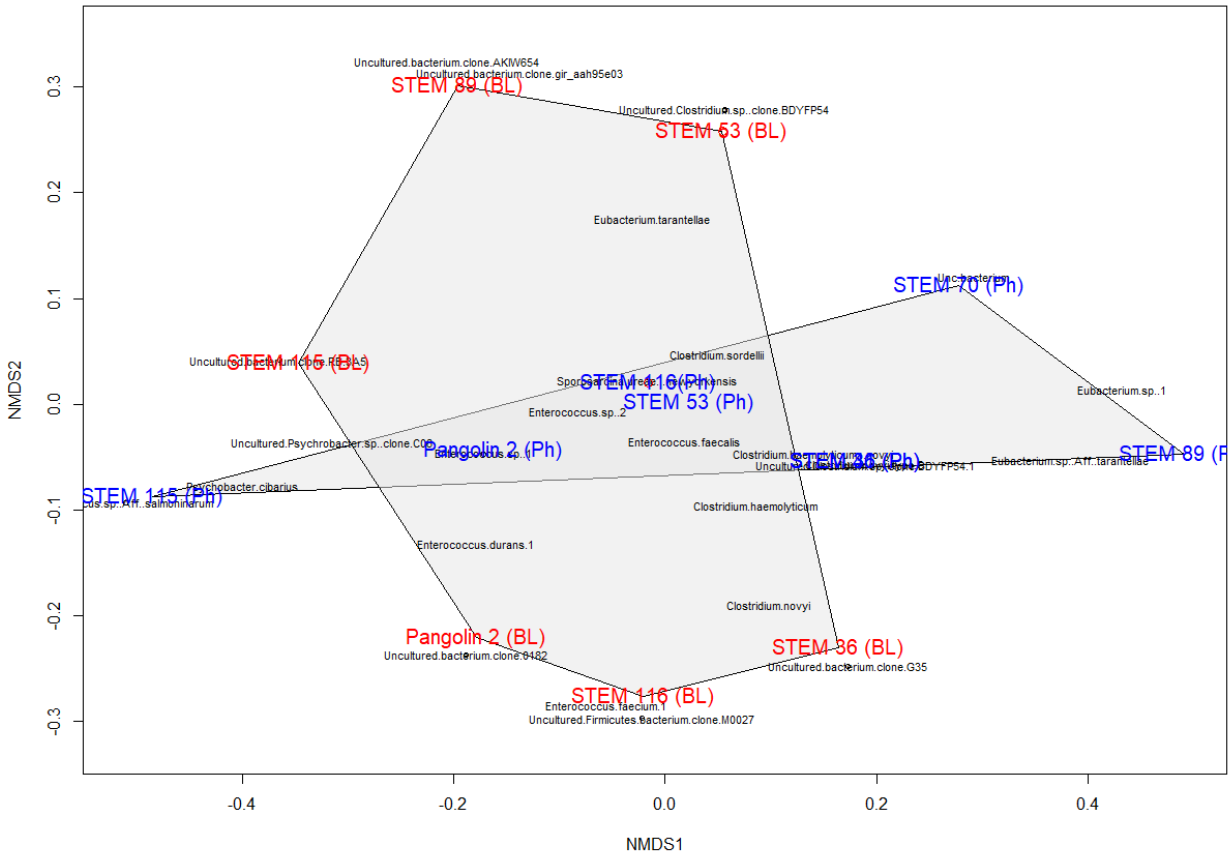


Figure 3.3: Non-parametric multidimensional scaling (nMDS) plot showing the similarity interactions between BLAST (red) and phylogeny tested (blue) assignment of taxa for cloned samples of eight Temminck's Ground Pangolins (*Smutsia temminckii*).

Noticeable cladal changes in phylogenetic trees were seen between sense and antisense related phylogenies of *Clostridium*, *Eubacterium* and *Enterococcus* (Appendices 4–6).

Geographic traceability of cloned data

The PerMANOVA showed no significant effect of geographic sampling site on microbial community composition ($p > 0.5$). This was further exemplified in the cluster dendrogram and nMDS plots, however the nMDS analysis illustrated community separation between sites for certain taxa (Figure 3.4). This prompted a per-taxon account of differences and similarities between the two sites, which yielded improved between-site separation (Appendix 9). From this it was observed that 12 out of 17 taxa differed between the two geographic sites giving a site difference of 70.59% in presence/absence data.

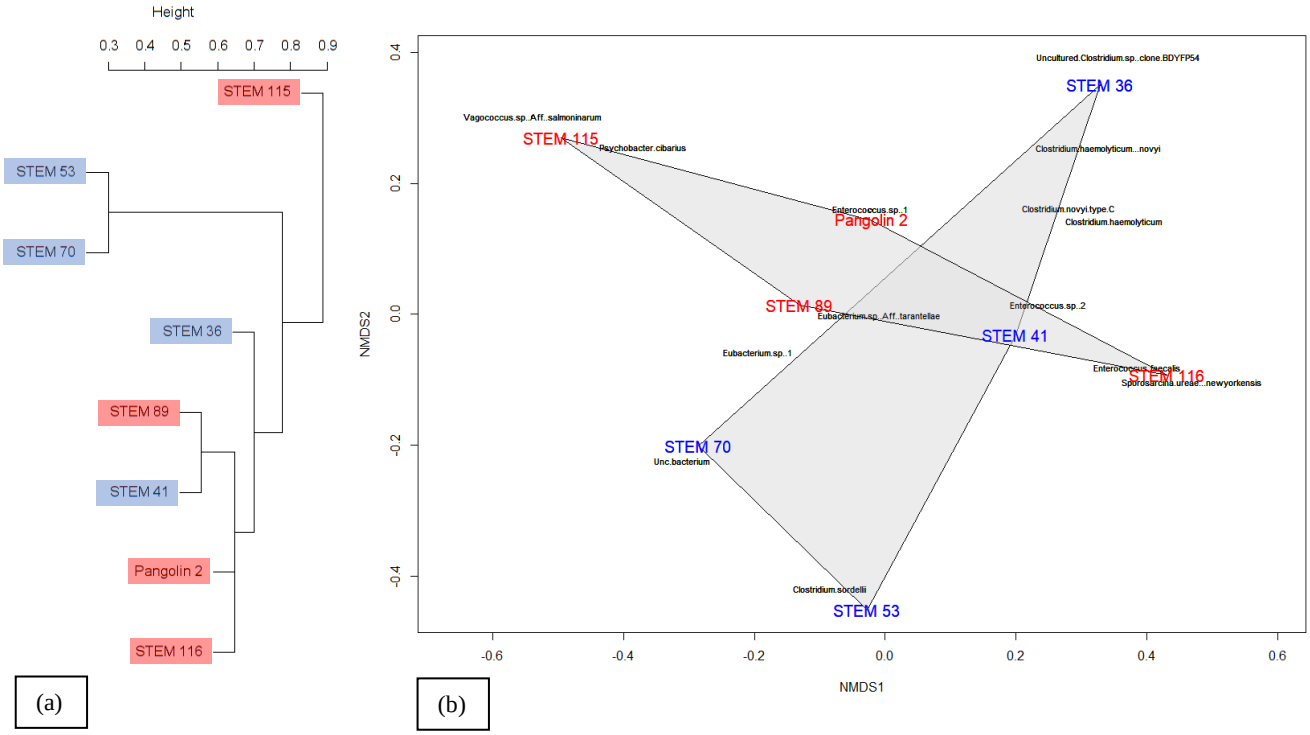


Figure 3.4: Phylogeny tested cloned microbial community composition comparisons between individuals from Kalahari Oryx Private Game Farm (blue) and Tswalu Kalahari Reserve (red) using (a) a cluster dendrogram and (b) non-parametric multidimensional scaling (nMDS) plot.

Additional variance factors of cloned data

The PerMANOVA showed no significant effect of organ type, sampling date or cause of death on microbial community composition ($p > 0.5$). In the case of organ type, this was further exemplified in the cluster dendrogram for STEM 70 and STEM 53 BLAST results with

microbial communities clustering by specimen for STEM 53 rather than organ type (Figure 3.5). However, according to the distance matrix from Figure 3.5, the microbial community composition of STEM 70 spleen is closer to STEM 53 (both organs) than STEM 70 lung. Differences in bacterial count data between the two organs for STEM 53 was only 7.02%, whilst it was 24.14% for STEM 70.

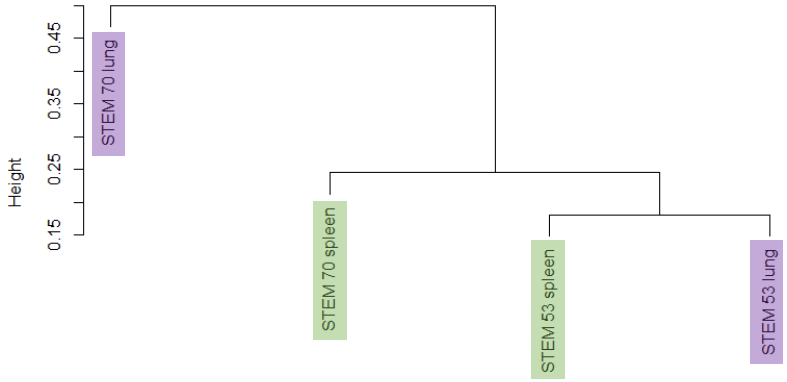


Figure 3.5: Cluster dendrogram comparing spleen (green) and lung (purple) microbial community composition for BLAST search results of cloned samples STEM 53 and STEM 70 of Temminck's Ground Pangolins (*Smutsia temminckii*).

Cloning *versus* NGS

The rarefaction curves of BLAST and phylogeny tested cloning data showed no sign of a plateau for each sample (Figure 3.6a). Additionally, the gradient for phylogeny tested samples is lower due to a lower number of assigned taxa as this method reduced the microbial taxa number by three through the amalgamation of different uncultured clones or identification of these clones as already identified taxa (Figure 3.6b and c). The NGS data reaches a plateau at 60 taxa due to the significantly higher number of sequences available for screening (Figure 3.6d). This aligns with the Shannon alpha diversity estimates whereby NGS data diversity (*H'* = 2.218–2.527) was higher than that obtained for the cloning component (*H'* = 0.973–2.164).

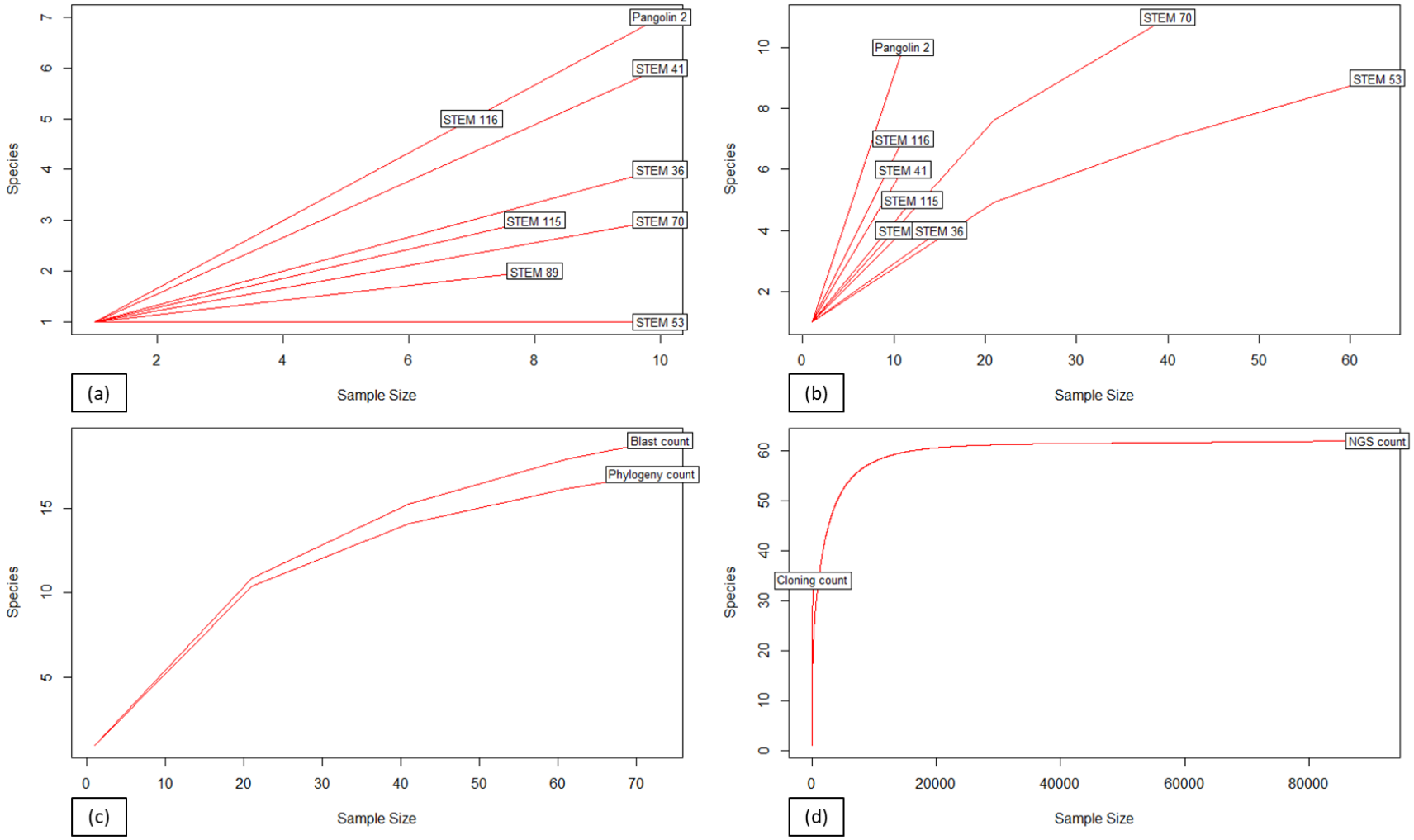


Figure 3.6: Rarefaction curves of the microbial communities of Temminck's Ground Pangolins (*Smutsia temminckii*) comparing species accumulation of (a) specimens that underwent phylogeny testing at a maximum of ten colonies each, (b) specimens that underwent BLAST searches at a maximum of 63 colonies, (c) the amalgamation of specimens for BLAST and corresponding phylogeny tested clones as well as (d) the amalgamation of all cloning data in comparison with Ion Torrent 16S Metagenomics (NGS).

The microbial community inferred from cloning differed significantly to that of the Ion Torrent NGS microbial community (F = 2.424, p = 0.009) with cloning results falling separately to that of NGS in the cluster dendrogram (Figure 3.7a). The same was shown when comparing microbial community composition of STEM 53 and STEM 70 cloning BLAST results (spleen alone as well as lung and spleen combined) with the metagenomics BLAST results of spleens from the same specimens (Figure 3.7b). *Clostridium sordellii* was the most abundant species for both methods when comparing only the spleen samples of STEM 70 and STEM 53, but was the only species to be recovered by both methods. This resulted in a species incongruency percentage between cloning and NGS of 98.48% for presence/absence data and 83.92% for count data.

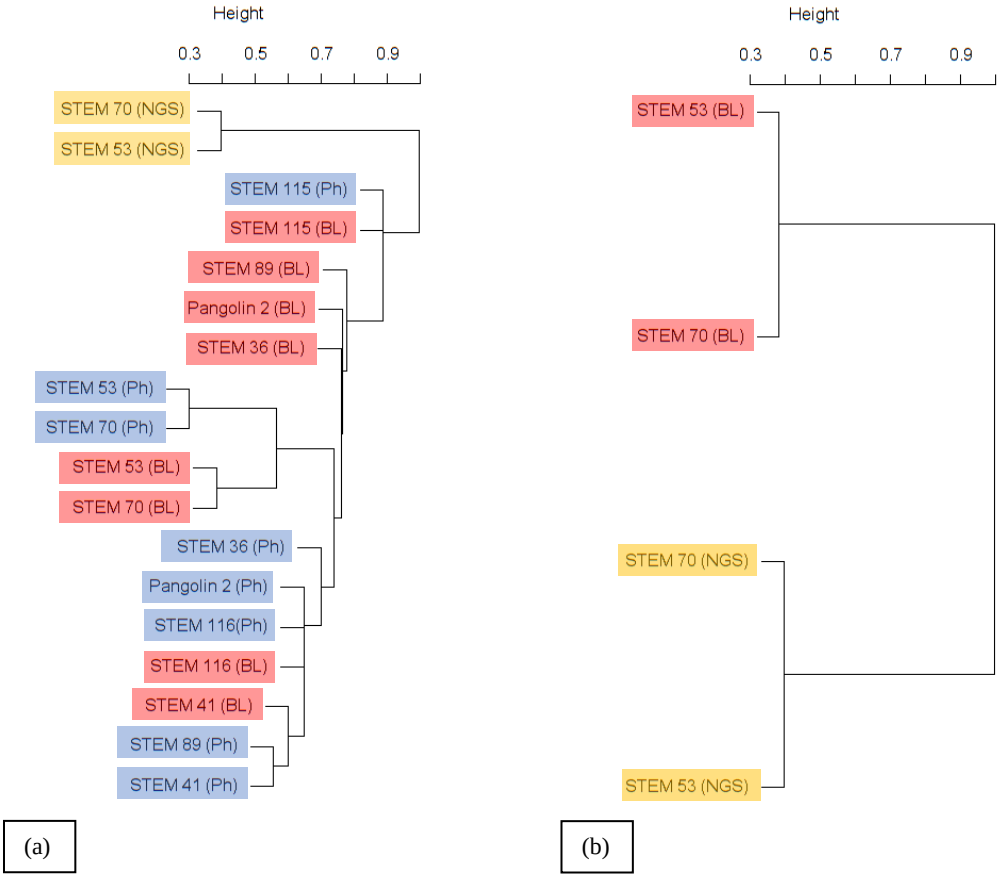


Figure 3.7: Cluster dendrograms of microbial community composition between (a) phylogeny-tested cloning data (blue), corresponding BLAST searches of cloning data (red) and Ion Torrent 16S Metagenomics (NGS) data (yellow) as well as between (b) STEM 53 and STEM 70 specific BLAST data of all colonies (red) and Ion Torrent 16S Metagenomics (NGS) data (yellow).

Discussion

BLAST versus phylogeny testing of cloned data

The results suggest that the method of taxonomic assignment is an important factor to consider when investigating microbial communities. Only four out of 32 identified taxa were congruent in presence/absence data between methods. This was exemplified when count data of each taxon was taken into consideration as only one taxon (*C. sordellii*) was found to be static between both methods. This is due to sequences being switched to closely related taxa depending on the type of assignment method used, in particular between taxa of the *Clostridium haemolyticum/novyi* group, the *Enteroccocus durans/faecalis/faecium* group as well as the *Eubacterium* group. Species in the *Clostridium haemolyticum/novyi* group are highly analogous to one another at the 16S rRNA gene region and are often described as being part of a genospecies group (Skarin *et al.*, 2011). Only a single base pair difference on the 16S rRNA gene separates *C. haemolyticum* and *C. novyi* type B from *C. novyi* type C whilst no difference is found between *C. haemolyticum* and *C. novyi* type B (Sasaki *et al.*, 2001; Skarin *et al.*, 2011). A similar phenomenon is observed for the *Enteroccocus* group with high levels (as high as 99.8%) of similarity between species of the group across 16S rRNA (Patel *et al.*, 1998). Taxon-assignment switching within the *Eubacterium* group is unlikely to be caused by similarity of described species considering *Eubacterium tarantellae* (the most closely related species to my sequences) is most closely related to species of the *Sarcina* genus (Vos *et al.*, 2011) and no *Sarcina* species were found in this study. This suggests that some sequences isolated in this study may represent novel taxa that are either closely related to, or a variant of, the *Eubacterium tarantellae* clade. This is likely considering that only three *Eubacterium tarantellae* strains have been uploaded to GenBank for the 16S rRNA gene region to date, and all strains isolated thus far have come from fish (Vos *et al.*, 2011). These discrepancies were not found in *C. sordellii*, perhaps due to the species being situated within a well-defined clade whereby the

closest relative in terms of 16S rRNA sequence homology is *Eubacterium tenue* at 99% (Vos *et al.*, 2011; Jyothsna *et al.*, 2016). It therefore becomes apparent that the level of species similarity when conducting BLAST searches is an important factor to consider in preventing false positives. It is also important to include count data rather than presence/absence data alone, when comparing taxonomic assignment methods.

The best method for taxonomic assignment could not be empirically determined in this study as determining if the species identified from the two methods are correct by using culture-based phenotypic methods as a reference was not attempted. Despite this shortcoming, the marked differences in community composition between the two methods suggests that a single method should be determined for future traceability research. In a study comparing phylogeny-based methodology to BLAST searches of the same sequences it was noted that phylogenetic techniques have a higher accuracy when species in question are present in databases (Ross *et al.*, 2008). However, when absent, a liberal phylogeny-based method of species determination had significantly higher false positive rates compared to BLAST searches whilst the strict phylogeny-based method performed the best (Ross *et al.*, 2008). This latter strict method was adopted in this study, which states that if a sequence is not nested within a monophyly of one species it is not assigned that species name (Meier *et al.*, 2006). Since many taxa were unidentified clones in this study it is suggested that the phylogeny testing method adopted may have had a lower false positive rate compared to BLAST searches (Ross *et al.*, 2008). Phylogeny testing may also out-perform BLAST searches from databases like GenBank due to the high number of misidentified/bad quality sequences in GenBank (Turenne *et al.*, 2001; Patel *et al.*, 2004; DeSantis *et al.*, 2006). This is because under phylogeny testing, reference sequences are viewed during construction of tree datasets (picking representative sequences of species and genera) as well as during alignment and trimming prior to analysis, which lends itself to the identification and subsequent removal of erroneous or incongruous sequences that

would otherwise be missed during BLAST searches. The method also facilitates an in-depth view of data at a per-taxon level and thus the identification of taxon-assignment errors or presence of duplicates. This includes instances where taxonomic changes have occurred in species, but existing database sequences have not been updated and thus sequences of the same taxon are treated as separate species. The assignments of *C. sordellii* and *Paeniclostridium sordelli* (Jyothsna *et al.*, 2016) or *Clostridium tarantellae* and *Eubacterium tarantellae* as separate species are examples of this issue in this study (Udey *et al.*, 1977; Lawson and Rainey, 2016). This is often related to non-curated databases such as GenBank having no mechanisms to correct for taxonomic name changes apart from author or third-party requests (Turenne *et al.*, 2001). Using curated databases may be a way around this, however incorrect assignment of *C. sordellii* and *P. sordelli* was found in the Ion Torrent Metagenomics output as well, which was only aligned with curated databases (MicroSeq and GreenGenes). This indicates that even though the latter databases are considered to be better, they are not completely free of imprecisions (Woo *et al.*, 2008). Phylogeny testing also seems to handle uncultured bacterial clones in a better manner. BLAST outputs suggested that sequences belonged to uncultured bacterial clones of unknown taxonomic providence, but these same sequences were assigned to a described taxon after strict phylogeny testing. Additionally, uncultured bacterial clones that appeared to represent distinct taxa in BLAST results were amalgamated into a single taxon after phylogeny testing. This may be due to investigators being unable to determine the similarity between sequences generated in a study through BLAST searches alone, although this can be circumvented by employing a distance analysis such as nearest neighbour, which performed better than liberal but not strict phylogeny testing in tick identifications (Lv *et al.*, 2014). An additional advantage of phylogeny testing is the use of topotype assignment in separating geographic sites that contain the same species as has been used to separate

geographically distinct clades of the *Leishmania donovani* complex (Lukes *et al.*, 2007), FMD virus (Di Nardo *et al.*, 2011) and BTB (Vosloo *et al.*, 2001).

Phylogenetic taxonomic assignment does have its limitations including being more time consuming and labour intensive for a limited output. When generating data from clones, there is an added limitation of investigators having to create both sense and antisense datasets for sequences generated with the same primer due to the inability to control the direction of amplicon insertion into the cloning vector. One way of circumventing this is to sequence both sense and antisense strands for each clone. For a topotyping approach, sequencing both strands for higher accuracy in topotype assignment would have to be adopted considering the highly conservative nature of the 16S rRNA gene but increase time and cost for analysis (Case *et al.*, 2007). The cost and labour can be reduced by sequencing with the one primer first, identifying the clones that are sense or antisense and subsequently sequencing either one (sense or antisense) using the other primer to get all sequences in the same orientation. However, when comparing species identification of the *Mycobacterium* complex using the 16S rRNA gene, Turenne *et al.* (2001) state that numerous species in the complex will be seen as identical when only the 5' end of the gene (up to 510 bp) is used due to the base pair changes separating the species in this complex occurring near the 3' end. This was observed in this study with the *Clostridium haemolyticum/novyi* group monophyly in the reverse *Clostridium* tree (3'), due to a base pair change separating the species occurring near the 5' end of the gene (Sasaki *et al.*, 2001). Although BLAST searches are not immune to this if sequences are identical, the simplicity of taxonomic assignment adopted by this method (taking the highest scoring sequences regardless of what taxa follow) along with its limited accuracy often may overlook discrepancies if sequences are highly similar. One way to circumvent this is to use a universal primer set that is nested within the original amplicon used for cloning, and to purify and sequence that product instead. This will allow for single orientation sequencing as the primer

will target the inserted amplicon rather than the vector, provided that the primers are universal enough (see: Klindworth *et al.*, 2013).

With limitations of phylogeny testing and BLAST searches of GenBank realised, alternatives need to be considered. One alternative is to use BLAST searches on a number of databases in a continuum from more accurate, but smaller databases (curated) to less accurate, but larger databases (non-curated). When comparing BLAST results on the two largest public databases (GenBank and EMBL) and the curated RIDOM database, the latter database had a 100% match whilst the public databases had an accuracy of only 38% (Turenne *et al.*, 2001). RIDOM was found to be significantly better at species assignment compared to both curated and non-curated databases in many other studies, with the lowest assignment accuracy being just over 90% (Cloud *et al.*, 2002; Mellmann *et al.*, 2003; Becker *et al.*, 2004). By using databases like RIDOM first, the accuracy problem identified from BLAST searches in this study is reduced whilst the problem relating to databases like RIDOM only being used for identification of a small set taxa (medically relevant) is circumvented by using databases like GenBank to fill in the gaps at the end of the search continuum (Harmsen *et al.*, 2002).

Geographic traceability of cloned data

The results suggest that there is no significant difference between the microbial communities of Kalahari Oryx and Tswalu when analysed both statistically and through a cluster dendrogram. However, with only 40 microbial colonies being examined for each site, it must be stressed that whole community related comparisons can only be taken at face value until additional colonies have been sequenced. This is further exemplified by the nature of statistics and cluster dendrograms which may result in a simplified and averaged outlook of the data (Gauch Jr and Whittaker, 1981). A clearer divide is seen for specific taxa when examining the nMDS plot in-depth and tabulating taxon occurrence data (Figure 3.4b; Appendix 9), with

only five out of 17 taxa occurring in both sites. This method allows for an in-depth view of taxon-specific relationships and thus the ability to observe lower levels of microbial traceability. This includes identifying widespread taxa found in all sites which can be used for topotyping such as FMD in Cape Buffalo (Vosloo *et al.*, 2001) or taxa that are site-specific but abundant within the population for presence/absence keys, as in the case of Seabass (Pimentel *et al.*, 2017). The latter scenario was more prominent in this study whereby *C. sordellii* was found exclusively in Kalahari Oryx across all four individuals at a high abundance. This may be a good indicator species to determine whether samples came from Tswalu (negative screening result) or Kalahari Oryx (positive screening result), with taxon-specific primers, an approach which would reduce the cost of traceability screening dramatically (Pimentel *et al.*, 2017). In order to have higher confidence in this approach, the sample sizes from each site would need to be dramatically increased.

The viral topotyping approach as adopted by Vosloo *et al.* (2001) is more difficult to implement in this study due to no taxa being found in all sites across the entire population, and the lower levels of variation in bacteria compared to RNA viruses (Drake *et al.*, 1998). However, the genospecies *Clostridium novyi* sensu lato which contains both *C. haemolyticum* and *C. novyi* is a possible candidate considering that four of the eight individuals that underwent cloning were positive for this group, two from each site. Additionally, *C. novyi* has four type strains (A, B, C and D) whereby *C. haemolyticum* is type strain D (Willis, 1977; Smith and Williams, 1984) allowing for the possibility of strain typing within a geographic context. These species are also pathogenic which has the added benefit of faster evolutionary rates in genes coding for pathogenicity, *versus* that of the highly conserved 16S rRNA gene region (Hacker and Carniel, 2001; Rasko *et al.*, 2011). This is because genes coding for pathogenicity continuously evolve in an evolutionary arms race with the host but are also susceptible to horizontal gene transfer as many are often not part of the core gene pool located in

chromosomes (Hacker and Carniel, 2001). These genes may also undergo genomic structural variations, a feature which was important in determining the source of a diarrheagenic *E. coli* outbreak in Germany that caused 50 fatalities (Rasko *et al.*, 2011). Both *C. haemolyticum* and *C. novyi* are soil-borne, environmental species (Uzal *et al.*, 2016) which makes them a possible prominent source of infection to *S. temminckii* due to its close ecological relationship with soil through use of abandoned burrows and consumption of soil-centred prey (Jacobsen *et al.*, 1991; Heath and Coulson, 1997; Pietersen *et al.*, 2014b; 2016). The soil-borne nature of these microbes may provide an alternative approach to pangolin traceability sampling as soil samples in and around pangolin feeding sites and burrows can be collected which reduces the need to rely on material from deceased individuals or on invasive sampling procedures (Fierer and Jackson, 2006; Concheri *et al.*, 2011). Sampling directly from the soil dramatically increases the number of samples across the distributional range of the species at a finer resolution, whilst concurrently decreasing the sampling cost and effort required (Fierer and Jackson, 2006; Concheri *et al.*, 2011). The potential for environmental microbial analysis as proxies of host communities has already been shown in fish (Sousa and Silva-Souza, 2001) and humans (Hewitt *et al.*, 2012). It is important to note that Kalahari Oryx and Tswalu are only 100 km apart, which is within the assumed dispersal range of the species (van Aarde *et al.*, 1990). With this in mind, even though there was no significant or complete community separation between the two sites, this study shows the possible resolution that microbial traceability may have just by looking at low throughput cloning data. Further screening by using higher colony numbers or higher throughput approaches, increased pangolin sample sizes as well as additional sampling sites may help identify taxa that can be used as part of an absence/presence key or taxon-specific topotyping for site determination.

Additional variance factors of cloned data

It is important to consider other factors of variance in microbial composition other than geographic variation before drawing conclusions on traceability data. However, given the small sample size and the averaging effect of the statistical analyses, these results need to be interpreted with caution.

This study had a slight bias for earlier years of sample collection in Kalahari Oryx (2011–2012) and later years in Tswalu (2012–2015), and although there was no significance found, microbial community shifts have been documented with seahorse traceability across a period of only two months (Cohen *et al.*, 2018). However, sampling between two different seasons seemed to have no effect on the microbial profiles of fish in inland lake systems of Cameroon (Tatsadjieu *et al*., 2010). Apart from the small sample size in the present study, this incongruency between studies could be due to the fact that oceanic microbial communities are heavily influenced by ocean currents whereas inland lakes or soil of terrestrial environments do not undergo such dispersals (Shade *et al*., 2012; Brum *et al.*, 2015). Although time series studies in a variety of scenarios have shown that it is an important factor to consider, microbial communities do evolve to a stable state and if a perturbation occurs (such as a dietary shift, translocation, antibiotic treatment or abnormal weather) microbial communities can either revert back to the stable state that existed prior to the perturbation or shift to another stable state (Faust *et al.*, 2015). Many species of bacteria are also known to show variance in abundance during different seasons, although such variance is usually predictable (Faust *et al.*, 2015). From this I suggest that looking at the microbial community as a whole for geographic traceability may not be viable across a time series as a large amount of variation through complex interactions may influence its accuracy (Faust *et al.*, 2015), but that certain groups of taxa (those with known stability or high prevalence) or a single taxon should be considered.

One would expect that when comparing starvation (a prolonged death) with electrocution (a comparatively quick death), that microbial diversity would differ between these two events. Previous research has indicated that diet and quantity of food can affect the gut microbiome drastically (Flint *et al.*, 2007), however notwithstanding the small and biased sample size, it appears that organs such as lung and spleen are not as markedly influenced. Regardless of the outcome, causes of death that may result in an open carcass need to be analysed with the consideration of possible environmental contamination, particularly if the whole community level of traceability is undertaken (Howard *et al.*, 2010; Hyde *et al.*, 2013; 2015). This notion is further expressed in terms of time lapse of sampling after the animal is deceased because of the influence of decomposition on microbial communities as well as the increased risk of scavengers and decomposers contaminating the carcass (Howard *et al.*, 2010; Hyde *et al.*, 2013; 2015). The samples used in this study were collected within hours of death due to tracking the animals and regular fence-line patrols which nullifies the previously mentioned concerns, but this may not be possible for all studies. Additionally, these factors should only affect prevalence data and not presence/absence data as traces of microbial DNA and dead cells that were prevalent in live hosts will still be present in the carcass for a given time period.

In a study conducted on human cadavers which characterised the microbiomes of a number of internal organs, microbial community separation was based on the individual rather than organ type (Can *et al.*, 2014). This is probably related to the fact that internal organs are not as heavily influenced by external biological and environmental conditions (Can *et al.*, 2014). Although this was shown in the statistical output of this study, it was incongruent with what was suggested from the cluster dendrogram. When comparing organ data for STEM 53 and STEM 70, the cluster dendrogram showed that the microbial community of STEM 70 spleen was more dissimilar to the lung than it was to STEM 53. The test between STEM 70

and STEM 53 allowed for a more accurate comparison of the effect of organs on microbial diversity without the influence of the microbial variability between individuals and thus suggests that this may indeed be an important factor. However, STEM 70 had a bias in spleen (n = 24) to lung (n = 12) bacterial clones due to the lung sample having reduced colony formation and a larger proportion of small PCR amplicon sizes which were not processed further. This may have had an influence on the results and the need for additional colonies and samples is duly expressed for testing organ effects on community level structure of pangolins. Substantial microbial variation has been noted between body sites of humans, however these were in some sort of contact with the external environment such as skin, buccal or vaginal cavity (Huse *et al.*, 2012; Quaak *et al.*, 2018). The study by Can *et al.* (2014) also included blood, which aligned with the other organs screened during that study. This has important implications, as similar validation based on larger sample sizes in pangolins would permit authorities working on pangolin traceability to collect any of several sample types for microbial proxy tracing. This is provided that they are from internal organs from dead individuals or blood, which can be sampled from live individuals. If microbial community variability is larger between organs than between samples taken from different geographical localities, deciding on a specific organ for future traceability tests may need to be considered. Additionally, if only one or a few microbial taxa are chosen for traceability estimates, it is vital to choose taxa that are not organ-specific.

Another important factor of variation to consider is the gene as well as the database used for taxon identification (Owen, 2004; DeSantis *et al.*, 2006). The 16S rRNA gene is the most widely used gene for bacterial classification, and thus well-represented in databases. However, the gene has innate complications such as suggested recombination involving parts of the gene (Young, 2001), presence of intervening sequences within the gene (Hurtado *et al.*, 1997) and inter-operon variation within a strain of up to 5% (Clayton *et al.*, 1995; Ueda *et al.*,

1999; Case *et al.*, 2007). These innate complications are of great concern as they can unknowingly influence species and strain level identification but more dramatically, molecular typing (Young, 2001; Owen, 2004). If these methods are to be considered for geographic traceability, additional genes that may reduce innate issues found with the 16S rRNA gene (Owen, 2004), including genes that have a single copy across the genome such as rpoB (Dahllöf *et al.*, 2000; Case *et al.*, 2007), should be included. Issues relating to GenBank include the incorrect species name for certain sequences, use of unestablished species names, chimeras of sequences, a large number of poor quality sequences with degenerate or unscored sites, representative sequences that are of insufficient length to generate good scores and the inability to easily identify sequences of cultured specimens from published results (Turenne *et al.*, 2001; Patel *et al.*, 2004; DeSantis *et al.*, 2006). Additionally, ~43% of GenBank's full length 16S rRNA sequences are amalgamated into the pseudo divisions "environmental samples" and "unclassified" (DeSantis *et al.*, 2006). The use of multiple gene regions and multiple databases before assigning sequences to a specific species should be considered.

Cloning versus NGS

With such a drastic incongruency in microbial community composition between cloning and NGS approaches for the same samples, determining which approach to use for future analyses is integral. However, it is important to note that rarefaction curves of cloned samples did not reach the required plateau, suggesting additional colonies for each sample would need to be picked and sequenced for more reliable results. Considering that the BLAST rarefaction curves of STEM 70 (36 colonies) and STEM 53 (59 colonies), which had comparatively lower rarefaction gradients than the other cloning samples, had not yet reached plateaux, suggests a species' plateau is expected to only be reached above 60 colonies for cloning data. A study comparing various metagenomic approaches on the same microbial

community showed that even after 600 colonies were sequenced, the Ion Torrent technology still identified 105 more genera and 42 more families than did the cloning data (Samarajeewa *et al.*, 2015). This aligns with the present study and can be attributed to the significantly higher throughput of NGS and its ability to identify species at low abundances (Xu, 2006; Samarajeewa *et al.*, 2015; Zhou *et al.*, 2015). This may also explain the incongruency between cloning and NGS in this study, which was at 98.48% for STEM 70 and STEM 53 spleen samples. Other sources of incongruency may be attributed to taxonomic misidentifications as well as the use of different primers which can result in biases in species-specific binding and sequences of different lengths or gene regions (Torsvik and Øvreås, 2002; Kircher and Kelso, 2010; Ihrmark *et al.*, 2012; Samarajeewa *et al.*, 2015; Parada *et al.*, 2016). Additional evidence for the latter is provided by this study in the observed differences in taxonomic relationships across the 16S rRNA gene in the cloning data when comparing phylogenies of different orientated amplicons (opposite ends of the 16S rRNA gene). The same can be said across different primer sets targeting different regions of the gene on metagenomics platforms (Cai *et al.*, 2013; Fouhy *et al.*, 2016). Ion Torrent's bias for AT-rich genomes as well as its directionality and species-specific sequencing errors or truncations may have also played a role (Quail *et al.*, 2012; Salipante *et al.*, 2014). Although the level of incongruency was not measured in the commercial microbial consortia study, high levels of incongruency was noted in prevalence data when comparing Ion Torrent NGS with denaturing gradient gel electrophoresis (DGGE) and PCR-cloning using the TOPO® TA kit (Samarajeewa *et al.*, 2015). The most prevalent genus was congruent between cloning and DGGE, however the NGS technology identified another genus as most prevalent (Samarajeewa *et al.*, 2015). When comparing STEM 53 and STEM 70 cloned and NGS spleen data, a similar trend in species data was found but not in genus-specific data, suggesting that the smaller 140 bp targeted sequence under NGS for Samarajeewa *et al.* (2015) may have not allowed for the resolution obtained in

our study which consisted of six targeted 16S rRNA gene regions. With these incongruencies in mind, identifying which approach should be used in future studies is integral.

PCR-cloning has the drawback of not being able to accurately determine how many colonies to pick and sequence without prior knowledge of sample-specific diversity (Samarajeewa *et al.*, 2015). This may lead to an overestimation of colonies to process, which increases financial waste; or underestimation of colonies to process, which affects the accuracy of the data produced. Other drawbacks include the costs and labour required to get a comparably lower output, even after sequencing 600 colonies. In this study it cost just over ZAR 2 000 per sample without the inclusion of consumables (petri dishes, spreaders, tips and tubes) if 15 clones were picked and processed. In accordance with the rarefaction curves, in excess of 60 colonies would need to be sequenced per sample to identify all possible taxa in a sample, which would cost around ZAR 6 500 per sample. This price only includes unidirectional sequencing. The Ion Torrent run was done at ZAR 3 400 per sample off the 318 chip at ten samples per chip which yielded approximately 350 000 reads per sample. This was shown to be adequate for our samples as rarefaction curves showed a plateau at around 100 000 reads, suggesting that this price can be halved per sample if necessary. Cloning has the added drawback of human-related mistakes and biases in analyses which cannot be automated. However, these mistakes are small in comparison with the automated pipelines of some bioinformatics analyses and the added human influence allows one to not only identify taxon-specific mistakes but identify taxa at a more accurate and finer level due to longer, more accurate sequencing reads and phylogeny testing (Patel *et al.*, 2004; Hoff, 2009; Kircher and Kelso, 2010; Samarajeewa *et al.*, 2015; Muñoz-Colmenero *et al.*, 2017).

Accurate taxonomic assignment is one of the biggest downfalls relating to NGS metagenomics as sequence reads are often short and are assigned on OTU estimates which do not consider the biology or taxonomy of the species (Zhang *et al.*, 2013; Bálint *et al.*, 2016).

This results in lower confidence in the results compared to that of the large gene segments and phylogeny testing that cloning can provide (Zhang *et al.*, 2013; Bálint *et al.*, 2016). However, the ability to adjust assignment parameters in the bioinformatics pipeline for more comparable results (Muñoz-Colmenero *et al.*, 2017), use a variety of databases, shift to sequencing platforms that provide larger amplicons for sequencing (Rasko *et al.*, 2011; Quail *et al.*, 2012), produce custom primer panels designed for longer reads and other gene regions as well as recent advances in phylogeny-based assignments of NGS data (Zhang *et al.*, 2013) mean that this downfall is slowly becoming a thing of the past.

Study caveats

The biggest caveat in this study was the sample size with only ten colonies per individual across eight individuals being used to test the taxonomic assignment methods and geographic traceability (40 colonies from each site) whilst only two samples were used to compare NGS with cloning (STEM 53 and STEM 70). However, pangolins are difficult to locate and sample in nature considering their vulnerable conservation status (Pietersen *et al*., 2014a,c), CITES Appendix I listing (CITES, 2017) and their highly elusive nature (Swart, 2013; Pietersen *et al*., 2014a,c). The major source of fresh, intact samples were from accidental electrocution through electric fencing, which may be the best source for future studies (Pietersen *et al*., 2014a,b). The reason why ten microbial colonies per individual was taken as the standard was because three samples had significantly reduced colonies that arose from a storage malfunction of cloned samples. In order to keep samples comparable between the analyses involving geographic traceability and other forms of variance, colony numbers had to be comparable. However, increasing this to a satisfactory level would not have been feasible considering the costs involved with upstream lab work of over 60 colonies per sample (based on the estimated species rarefaction curve plateaux of STEM 53 and STEM 70). Another caveat

is the variability of taxonomic assignment between sense and antisense strands when under phylogeny testing, which was not expected. Other sources of variance may also have accounted for said difference and are caveats of the study. These include the difference in primers used which led to different amplicon sizes and regions of the 16S rRNA gene, use of different databases for taxonomic assignment and comparing the two metagenomics approaches when the species rarefaction curves of the cloned samples had not yet plateaued.

Conclusion and synthesis

This research has shown the potential for geographic traceability through microbial proxies of *S. temminckii* individuals sampled from two localities that are 100 km apart from each other. Although complete community separation was not realised, identification of a key taxon (*C. sordellii*) which could be used as a site indicator through its presence or absence in *S. temminckii* samples was detected. This provided insights into the potential for traceability through presence/absence keys. This study also identified taxa such as the *Clostridium novyi* sensu lato genospecies (*C. novyi* and *C. haemolyticum*) that could be screened for in *S. temminckii* for strain typing using virulence genes due to their high presence within the sample set. Screening of additional colonies of samples or use of NGS platforms, additional samples from Kalahari Oryx and Tswalu as well as additional sites may provide better insight into these hypotheses. The aforementioned conclusions would not have been realised without the use of metagenomics approaches which will be the first step in determining the best set of species for topotyping and presence/absence keys across *S. temminckii*'s range. The relative importance of different sources of variation in terms of their effect on microbial community variance, in decreasing order, are: variation between metagenomics approaches (NGS/cloning) > individuals > taxonomic assignment approach adopted (phylogeny testing/BLAST) > different organs. This suggests that the most important goal to address is that of which metagenomic

approach to use and keep as the gold standard for future analyses.

If cloning were to be used as the metagenomics approach in identifying key taxa for the traceability of pangolins, I suggest the use of strict phylogenetic analyses through bi-directional sequencing as the mode of taxon identification. This approach is more time-consuming compared to heuristic search tools, however it allows for identification of incongruencies between sites at lower taxonomic levels without the need for additional analyses, significantly lowers false positive rates (Ross *et al.*, 2008), has the ability to identify erroneous sequences and misidentifications as well as uncultured taxa that are similar or identical to one another (or to a described species). Provided the throughput is high enough and primers are unbiased during cloning, topotyping of key taxa for traceability purposes can already begin at this metagenomics stage, however the sheer cost and time needed to provide a sufficient throughput may be the downfall of this method.

With the countless drawcards related to NGS metagenomics that fit requirements of pangolin microbial proxy tracing and the limitations becoming a thing of the past, the scientific community has, and will continue to, shift to this era of sequencing. This research has highlighted the potential that NGS has for microbial diversity screening in forensic cases owing to its high throughput, facile and cost-effective approach in a fraction of the time. Under these characteristics, I suggest that that the metagenomics approach to be used to determine key taxa for the traceability of pangolins in general should be that of an NGS approach.

Regardless of the approach, the possible sources of variation described in this research and the knock-on effects that these may have on the feasibility of microbial proxy tracing as a forensic tool should be kept in mind.

Chapter 4:

Prevalence and diversity of *Clostridium novyi* sensu lato in the Temminck's Ground Pangolin (*Smutsia temminckii*): traceability potential and public health concern

Introduction

With the advent of the recent Ebola epidemic and Zika virus threat, it is becoming increasingly important to detect, prevent and eradicate potential zoonotic (animal to human) and enzootic (animal to animal) threats before they manifest into global crises (Karesh *et al*., 2012; Kilpatrick and Randolph, 2012). This necessity for pathogen screening in animals is further expressed in economics with direct and indirect economic losses caused by zoonoses amounting to US$ 20 billion and US$ 200 billion, respectively, over the ten-year period of 2000–2010 (World Bank, 2010).

One Health is a global approach that attempts to curb the drastic implications that disease epidemics can have on conservation as well as social and economic facets of society. The initiative adopts a holistic approach using a variety of methods, one of which entails pathogen discovery and surveillance across the globe by a number of stakeholders in order to better identify possible disease risks associated with the environment as well as human, wildlife and domestic animal populations (UNICEF, 2008; World Bank, 2010; Mackenzie *et al*., 2014).

Trade and bacterial pathogens

The legal trade in animals has had its fair share of disease outbreaks (Travis *et al*., 2011), most notably the rinderpest epidemic in Africa (Daszak *et al*., 2000). However, the illicit trade in wildlife bears a greater risk of disease transmission and spread due to the uncontrolled exploitation of new source populations, forced cross-species or cross-population exposure and exposure to naïve consumer populations (Bell *et al*., 2004; Fèvre *et al*., 2006). Some recent examples include the monkeypox epidemic in the United States of America (CDC, 2003; Reed *et al*., 2004) and Severe Acute Respiratory Syndrome (SARS) epidemic which started in China and rapidly attained a global presence (Bell *et al*., 2004).

The importance of screening for bacterial pathogens in particular is warranted given that an estimated 54.3% of all the emerging infectious disease events between 1940 and 2004 were due to bacterial and rickettsial pathogens (Jones *et al*., 2008). This, together with the many examples of devastating bacterial diseases that have a broad host range occurring in wildlife, domestic animal and human populations (e.g. brucellosis, leptospirosis, anthrax and more recently listeriosis), highlights the importance of bacterial screening for potential zoonotic and enzootic transmission events (Binder and Mermel, 1998; Buzgan *et al*., 2010; Gul and Erdem, 2014).

Clostridium novyi sensu lato

In areas where bursts of precipitation are few and far between, the survival of bacterial species takes on its own evolutionary pathway (Fierer *et al*., 2012; Makhalanyane e*t al*., 2015). The Kalahari savanna biome in southern Africa is characterised by semi-arid to arid ecosystems where the survival of soil-borne microbial communities has been shaped to a greater extent by the abiotic factors they encounter (including soils of high pH and salinity and low moisture), than by their competitive interactions (Skujinš, 1984; Fierer *et al*., 2012). These conditions lead to a higher relative abundance of soil bacterial taxa that possess osmoregulatory and dormancy processes (Fierer *et al*., 2012; Makhalanyane e*t al*., 2015). The formation of endospores is one such process, which the phylum Firmicutes is well known for (Onyenwoke *et al*., 2004; Galperin, 2013). Nested in this phylum is the class Clostridia which are bacillus-shaped, spore-forming, obligatory anaerobes that are generally gram-positive but are known to be gram variable as growth progresses (Cato *et al*., 1986; Beveridge, 1990; Hatheway, 1990). Organisms currently included in the genus *Clostridium* contain a genomic GC content ranging from 22–55%, and represent a highly diverse taxonomic group (Hatheway, 1990). As a result, this taxonomic group is notorious for misclassifications triggering calls for major taxonomic

revisions (Collins *et al*., 1994; Stackebrandt *et al*., 1999; Yutin and Glaperin, 2013; Lawson, 2016; Lawson and Rainey, 2016; Tindall, 2016). With such high levels of diversity and ubiquity, it is perhaps not surprising that many of the 200 known *Clostridium* species (Húngaro *et al*., 2016) have pathogenic properties pertaining to illness of humans and animals (Smith and Williams, 1984; Stackebrandt *et al*., 1999).

One such pathogenic group is *Clostridium novyi* sensu lato, the newly proposed name for the genospecies *C. botulinum* group III, *C. novyi*, and *C. haemolyticum*, owing to their highly analogous chromosomal core and plasmidome inter-lineage mixing (Eklund *et al*., 1974; Schallehn and Eklund, 1980; Skarin *et al*., 2011, 2014). The group is of major conservation, health and economic concern for human, domestic and wild animal populations alike as it contains the aetiological agents responsible for gas gangrene, infectious necrotic hepatitis (black disease), bacillary haemoglobinuria (red water disease) and animal botulism (Uzal *et al*., 2016).

Genospecies *C. botulinum* group III is a pathogenic group known to cause animal botulism, which is considered an emerging disease in Europe associated with substantial economic losses, particularly in the poultry industry (Lindberg *et al*., 2010; Skarin *et al*., 2013; Souillard *et al*., 2014). The disease is caused by botulinum neurotoxins (BoNt's) which result in severe flaccid paralysis through the inhibition of acetylcholine release at the neuromuscular junction after ingestion and subsequent transference of preformed toxins or microbes into the bloodstream (Uzal *et al*., 2016). Based on the neurotoxins, the group consists of two serotypes (C and D) as well as mosaic chimeras (C/D, D/C), making it relatively difficult to distinguish (Hatheway, 1990; Moriishi e*t al*., 1996). Botulism is present in water, plant material, animal carcasses as well as the invertebrates that feed on these carcasses, and has been implicated in epizootic outbreaks in populations of more than 264 bird species with losses of over 50 000 birds at one time (Duncan and Jensen, 1976; Brand e*t al*., 1988; Kellerman *et al*., 1988;

Hatheway, 1989; Uzal *et al*., 2016). Of 52 000 farmed foxes and minks infected with type C botulism by toxic feed in Finland, 44 130 animals (85%) succumbed to infection (Lindström *et al*., 2004) suggesting a characteristically high case fatality rate (Uzal *et al*., 2016). Similar mass die-outs have been reported in South Africa with over 429 sheep and goats succumbing to infection due to botulism type D and C within a period of nine days from one feedlot and three weeks from another (Van de Lugt *et al*., 1995). Although not on the same scale as some of the previous examples, the economic impact of losses such as these are severe in the developing world.

Genospecies *C. haemolyticum* (*C. novyi* type D), a soil-borne strict anaerobe, is the causative organism of bacillary hemoglobinuria, more commonly known as red water disease due to the characteristic blood-stained urine and faeces during infection (Vawter and Records, 1926; Lewis, 2007; Hussein *et al.*, 2013; Olson, 2013). The beta toxin, otherwise known as phospholipase C, is produced by the microbe which after ingestion, resides in the liver (Hatheway, 1990; Navarro *et al*., 2017). This results in the destruction of capillary endothelial cells as well as the erythrocytes, which causes haemorrhaging and release of haemoglobin into the blood plasma, respectively, with an anoxic death within 2–4 days (Hatheway, 1990; Navarro *et al*., 2017). The overall morbidity of the disease in cattle is around 2% whilst the case fatality rate is *ca.* 96% (Stogdale and Booth, 1984; Navarro *et al*., 2017). In cases where naïve herds are exposed to an infected area, the overall mortality rate can be as high as 25% (Stogdale and Booth, 1984). Randhawa *et al*. (1995) suggested *C. haemolyticum* as the causative agent for the death of 16 sheep and the infection of an additional 44 sheep in India. The first human case of the disease reported by Saeb *et al.* (2017) revealed that *C. haemolyticum* was the aetiological agent responsible for an infection of an 18-year-old female diagnosed with acute lymphoblastic leukemia. The isolates' genome was sequenced and showed a multitude of virulence factors, including multi-drug antibiotic resistant genes (Saeb *et al*., 2017).

Genospecies *C. novyi* is a soil-based strict anaerobe that has been subdivided into four different type strains (strains A–D; Willis, 1977; Smith and Williams, 1984). The type A strain is implicated in one third of all gas gangrene cases whilst type B is the sole causative agent of black disease (infectious necrotic hepatitis; Hatheway, 1990; Lewis, 2007; Uzal *et al*., 2016). Black disease is aetiologically similar to bacillary haemoglobinuria (Navarro *et al*., 2017; Nyaoke *et al*., 2018), due to type B producing a beta toxin that is serologically identical to the one produced by *C. haemolyticum*; however, it is produced in much smaller quantities (Smith and Williams, 1984; Songer, 2005). Since type B produces limited quantities of this toxin, the alpha toxin (TcnA) also produced in this type strain has been suggested to be the possible etiological agent of black disease (Oksche *et al*., 1992; Guttenberg e*t al*., 2011; Uzal *et al*., 2016). This is because the toxin affects cell morphological properties along with disruption of GTP-binding proteins (Oksche *et al*., 1992; Guttenberg e*t al*., 2011), resulting in extensive fluid infiltration into connective tissue, death within hours and the characteristic central darkening of the carcass (Uzal *et al*., 2016). Onset of gas gangrene (clostridial myonecrosis) leads to muscle necrosis, destruction of erythrocytes resulting in the release of haemoglobin into blood plasma, toxaemia and subsequent death dependent on the site of infection (MacLennan, 1962). The type C strain, which has been known to produce a non-lethal gamma toxin, is considered to be the only non-pathogenic strain in *C. novyi* sensu lato (Willis, 1977; Hatheway, 1990; Skarin and Segerman, 2014). In 2000, an outbreak of *C. novyi* type A in 23 injection drug users in Scotland had a case fatality of 87% (McGuigan *et al.*, 2002). This agent has since been implicated in many more cases with botulism (Palmateer *et al.*, 2013). *Clostridium novyi* type B has also been implicated in the infection of an 11-year-old boy after trauma to the liver, marking the first human case involving this pathogen (Abdel-Haq *et al*., 2007).

Considering the vast differences in clinical and aetiological characteristics of each species exemplified through *C. botulinum*'s BoNT's that affect the neural circuits resulting in

paralysis, *C. haemolyticum*'s beta toxin that affects the circulatory system resulting in internal bleeding and the characteristic blood staining of waste materials as well as *C. novyi*'s proposed alpha toxin that affects the connective tissue through fluid infiltration and the characteristic darkening of the carcass, one would expect these species to be easily discernible. However, the aforementioned plasmidome inter-lineage mixing as well as the almost identical chromosomal core of the genospecies group complicates species identification both at a molecular and clinical level (Saeb *et al*., 2017). In terms of the plasmidome inter-lineage mixing, the plasmidomes of *C. novyi* sensu lato consist of large plasmids and circular prophages that have been implicated in inter-species as well as intra-species transfer across a variety of *Clostridium* species, resulting in species lineages not necessarily correlating with species designations (Eklund *et al*., 1974; Schallehn and Eklund, 1980; Skarin *et al*., 2011, 2014). This not only nullifies molecular diagnostics involving Polymerase Chain Reaction (PCR) identification methods that target plasmidome genes, but clinical identification procedures as well. This is because a *C. novyi* sensu lato species that acquires virulence genes from the plasmidome of another species has been shown to emulate clinical signs of the latter species (Eklund *et al*., 1974; Schallehn and Eklund, 1980; Skarin *et al*., 2011). In terms of the almost identical chromosomal core of the genospecies group, chromosomal genes that are typically used for PCR-based identification of bacteria such as 16S rDNA and the 16S–23S spacer region have been unsuccessful in distinguishing between the *C. novyi* sensu lato species due to the high levels of sequence identity between these taxa (Sasaki *et al*., 2000, 2001). As a result, PCR-based identification methods that target chromosomal genes with relatively high levels of variation, or rapid whole genome sequencing techniques are needed to discern species in clinical cases (Saeb *et al*., 2017). The former approach was attempted by targeting the highly variable flagellin gene using group-specific primers capable of rapidly distinguishing between four *Clostridium* species including *C. haemolyticum* and *C. novyi* type B (Sasaki *et al.*, 2002). However, this primer set

as well as the multiplex assay reported in the same study had only been tested on bacterial cultures or applied to clinical case studies with high infection profiles (García *et al*., 2009; Takagi *et al*., 2009), respectively. The multiplex primers were also only designed to target a proportion of the genospecies in *C. novyi* sensu lato which increases the cost of analyses per species if more than one species were to be examined by PCR-sequencing. This highlights the need to establish a cost-effective, *C. novyi* sensu lato group-specific primer set (including *C. botulinum*) for screening environmental animal samples. Without the availability of reliable markers that can be used to accurately identify *C. novyi* sensu lato genospecies it will not be possible to determine which species are present in the world's most heavily trafficked mammal. The development of a species-informative PCR assay is integral to determining whether the *C. novyi* sensu lato presence previously confirmed through 16S rRNA (Chapter 3) represents an overlooked public health concern in addition to the assay's use as a possible tool for geographic traceability.

Methods

Study site and sample collection

Samples were collected from 34 *S. temminckii* specimens between 2009 and 2015 as part of a previous ecological study on the species (Appendix 1; Pietersen, 2013). The bulk of the individuals perished due to accidental electrocution by electric game fencing (79%), whilst vehicle-associated road kills, natural causes and confiscated poached individuals made up the remainder. Specimens were immediately dissected upon receipt or were individually sealed in sterile plastic bags and stored in a dedicated freezer at -20 °C until dissected. Only intact specimens were sampled to prevent potential exposure to environmental contaminants. Thirty-three specimens were obtained from the northern parts of the Northern Cape Province of South Africa, specifically Tswalu Kalahari Reserve (n = 8) and Kalahari Oryx Game Farm (n = 25),

while a single individual believed to originate from north-eastern South Africa was obtained from the illegal trade after it had perished (Figure 4.1).

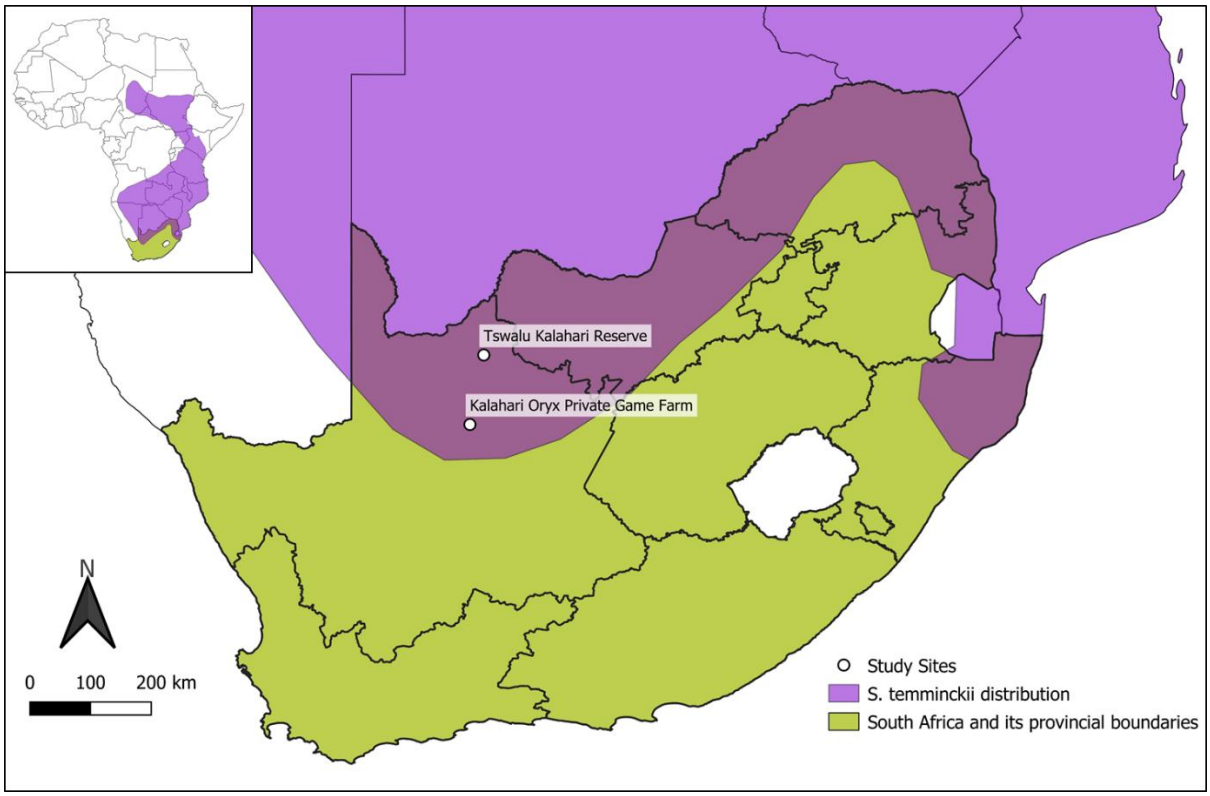


Figure 4.1: The two study sites in the Northern Cape Province of South Africa (green shaded area) and their relative positions within the known distribution of the Temminck's Ground Pangolin (*Smutsia temminckii*).

DNA Extraction

DNA extractions of spleen, liver and lung tissues were prepared using a ZR Genomic DNA™ Tissue MiniPrep DNA extraction kit (Zymo Research Corporation, California, U.S.A.), following the manufacturer's protocol for solid tissue. All extractions were performed within the Transboundary Animal Diseases Programme (TADP) Biosafety Level Three (BSL3) facility at the Onderstepoort Veterinary Institute, University of Pretoria, with section 20 approval (reference number: 12/11/1/1) from the South African Department of Agriculture, Fisheries and Forestry (DAFF).

PCR amplification and sequencing of the 16S rRNA and phospholipase C genes

Template DNA (3 µL) was added to a 37 µL Polymerase Chain Reaction (PCR) master mix containing DreamTaq™ DNA polymerase (Thermo Fisher Scientific Inc., Massachusetts, U.S.A) at a concentration of 1.5 units/40 µL, 10 µL DreamTaq™ buffer (Thermo Fisher Scientific), 0.2 µM dNTP's (Thermo Fisher Scientific) and 0.4 µM of each primer. The primers Clo-PhloCF 5'-ATG-GCA-CAG-GAA-CTC-ATG-CA-3' and Clo-PhloCR2 5'-ATG-GCA-CAG-GAA-CTC-ATG-CA-3' were designed specifically for this study and target a ~1 kbp region of the phospholipase C gene (beta toxin). The primers were designed on the basis of published phospholipase C nucleotide sequences from two species in the *C. novyi* sensu lato genospecies (12 *C. botulinum* sequences and one *C. haemolyticum* sequence) along with an outgroup species (*C. baratii*). These sequences were translated to amino acid residues and added to a database of all known clostridial species with the phospholipase C gene (114 amino acid sequences). The primer binding sites of the *C. novyi* sensu lato genospecies were determined and compared with the other clostridial species in the database for specificity purposes before developing them for use. Touchdown thermal cycling was performed with final annealing temperature of 51 °C (Table 4.1).

Table 4.1: Touchdown thermal cycling conditions used for the phospholipase C gene amplification with primers Clo-PhloCF and Clo-PhloCR2.

Stage	*Denaturation*	*Annealing*	*Elongation*	*Cycles*
[1]	96°C; 10s			1
[2]	96°C; 12s	55°C; 30s	72°C; 1:50s	2
[3]	96°C; 12s	53°C; 25s	72°C; 1:45s	3
[4]	96°C; 12s	51°C; 20s	72°C; 1:45s	35
[5]			72°C; 1min	1

PCR products were visualized on 1.5% agarose gel under ultraviolet (UV) light and positive samples of the expected size were purified directly from the tube using a Roche High Pure PCR Product purification kit (Roche Diagnostics GmbH, Mannheim, Germany) and were prepared for sequencing using the BigDye® Terminator v. 3.1 Cycle Sequencing Kit (Applied Biosystems, California, U.S.A). Unincorporated primers and dyes were removed by sodium-acetate precipitation, and the denatured DNA was analysed on an ABI PRISM™ 3100 Analyzer (Applied Biosystems, Foster City, California) located at the University of Pretoria's core sequencing facility. Reference sequences were obtained from GenBank through the use of the Basic Local Algorithm Search Tool (BLAST™) function in NCBI (Altschul *et al*., 1990).

Sequence analyses

Newly generated nucleotide sequences were visualized and edited using the Chromas package embedded in MEGA 7 (Kumar e*t al*., 2015), and aligned using ClustalW (Thompson *et al.*, 1994). As most of the reference sequences obtained from GenBank were amino acid sequences with limited corresponding nucleotide data, all nucleotide data were translated to

amino acid sequences in MEGA 7 prior to analysis. A Maximum Likelihood (ML) analysis was conducted using PhyML (Guindon e*t al*., 2010) online under the Institut Français de Bioinformatique (IFB) and France Génomique platform (http://www.atgc-montpellier.fr/phyml/) using an Akaike Information Criterion (AIC) best-fit model selection with nodal support assessed through non-parametric bootstrap resampling (10 000 replicates). GTR+I was selected as the best-fit model of sequence evolution for the nucleotide dataset, while CpREV+I+F was selected for the amino acid dataset for the ML analysis. Bayesian Inference (BI) analyses were conducted using MrBayes v.3.2.6 (Ronquist *et al*., 2012), under a GTR+InvG (nested model) over 10 million generations with a burn-in of 25% for both nucleotide and amino acid datasets. Trees were visualised and edited in FigTree v1.4.3 (http://tree.bio.ed.ac.uk).

Statistical analyses

A Stan-based (Carpenter *et al*., 2016) Bayesian Inference analysis using the package brms (Bürkner, 2016) was run to determine the effect that sex, organ type, the interaction between body mass and length (the latter used as a proxy of age), along with sampling date, type of death as well as sampling latitude and longitude (locality), had on the occurrence of *C. novyi* sensu lato. In a separate analysis temperature, precipitation and altitude data (1970–2000) were incorporated into the BI framework as a proxy for latitude and longitude. These climatic data were extracted from the WorldClim database (Fick and Hijmans, 2017) using the raster package in R (Hijmans, 2016) at a resolution of 10 arc-minutes. Analyses were run in the Rstudio interface of R v3.4.3 (Rstudio Inc., Massachusetts, U.S.A) using four chains at 2 000 iterations each, with a burn-in of 1 000 samples under default priors and a mixed effects model accounting for each individual across the three types of organs screened.

Results

The phospholipase C gene region yielded a total population incidence of 17.65% (n=34) for *C. novyi* sensu lato. The *C. novyi* sensu lato group sequences that were identified from *S. temminckii* clustered with *C. haemolyticum*, *C. novyi* type B and *C. botulinum* group III in the phylogenetic analyses with strong ML (84%) and BI (0.96–0.99) support (Figure 4.2). Additionally, strong support (ML: 84%, BI: 0.96) was found for an unnamed clade which is sister to *C. haemolyticum* and *C. novyi* type B, but within the *C. novyi* sensu lato group from both the nucleotide (Appendix 10) and amino acid phylogenies (Figure 4.2). Only one instance each was found for *C. novyi*, *C. botulinum* and *C. haemolyticum*. The latter two were found exclusively in Tswalu individuals whilst *C. novyi* and what appears to be a novel taxon were found in individuals from Kalahari Oryx.

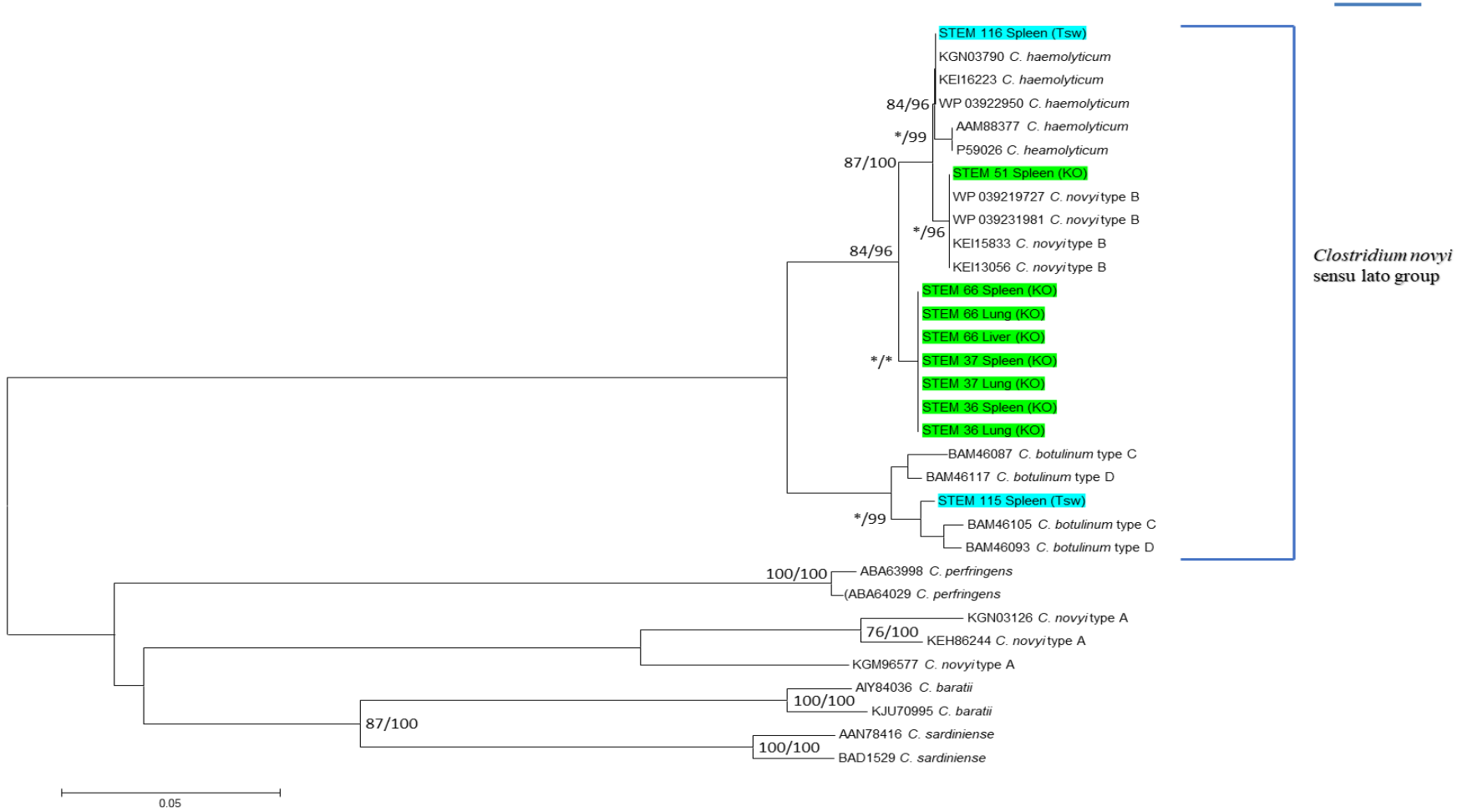


Figure 4.2: Uncorrected amino acid p-distance phylogeny of *Clostridium* inferred using the phospholipase C gene region (240 amino acids in length). All *Clostridium* species known to contain the phospholipase C gene, including representatives of the *Clostridium novyi* sensu lato group and sequences identified from Temminck's Ground Pangolins (*Smutsia temminckii*) from Tswalu Kalahari Reserve (blue) and Kalahari Oryx Private Game Farm (green) are included. Nodal support values from the Maximum Likelihood (ML) and Bayesian Inference (BI) analyses are presented ML/BI on the relevant nodes. Unsupported nodes (ML < 70% and BI < 0.95) are indicated by asterisks.

Clostridium novyi sensu lato was identified in the spleen samples of six individuals, whereas just three and one lung and liver sample, respectively, were PCR-sequencing positive (Table 4.2). This was mirrored by the statistical analyses with spleen having the strongest effect on *C. novyi* sensu lato occurrence (Estimate = 7.51 ± 2.73; $CI_{95\%}$ = 2.98−13.59). Additionally, the unnamed clade was consistently detected in more than one organ for each specimen (Table 4.2). Electrocution was the only source of mortality for positive samples (Table 4.2) which proved to be the only type of mortality interaction that was aligned with a positive screening result for *C. novyi* sensu lato (Estimate = 3.13 ± 4.52; $CI_{95\%}$ = -5.67−11.81). A similar trend was seen for sex, with males (Estimate = 2.51 ± 4.46, $CI_{95\%}$ = -6.15−10.99) being substantially more likely to test positive than females (Estimate = -2.67 ± 4.44, $CI_{95\%}$ = -11.59−6.18; Table 4.2). Mass had a relatively strong positive interaction with *C. novyi* sensu lato occurrence (Estimate = 2.33 ± 4.74; $CI_{95\%}$ = -6.98 – 11.67). Sampling date also had a relatively strong effect on *C. novyi sensu lato* occurrence, in 2011 (Estimate = 1.89 ± 4.08; $CI_{95\%}$ = -6.14 – 9.72) and 2015 (Estimate = 3.96 ± 4.74, $CI_{95\%}$ = -5.44 – 13.11), in particular. Body length as a proxy of age as well as latitude and longitude (locality) had no strong overall effects on *C. novyi* sensu lato occurrence, which from the latitude and longitude outcome meant that the influences of precipitation, temperature and altitude were relatively weak.

Table 4.2: Summary of the Temminck's Ground Pangolin (*Smutsia temminckii*) samples that tested positive for *Clostridium novyi* sensu lato using the phospholipase C gene region, detailing the organ that the bacteria was identified from. The pangolin's sex, age class, date of collection and cause of mortality are also indicated.

Specimen ID	***Organ***	***Sex***	***Locality***	***Sampling year***	***Cause of mortality***	***Clostridium species***
STEM 36	Spleen	Male	Kalahari Oryx	2011	Electrocution	Undescribed
STEM 36	Lung	Male	Kalahari Oryx	2011	Electrocution	Undescribed
STEM 37	Spleen	Male	Kalahari Oryx	2011	Electrocution	Undescribed
STEM 37	Lung	Male	Kalahari Oryx	2011	Electrocution	Undescribed
STEM 51	Spleen	Male	Kalahari Oryx	2011	Electrocution	*C. novyi* type B
STEM 66	Spleen	Male	Kalahari Oryx	2012	Electrocution	Undescribed
STEM 66	Lung	Male	Kalahari Oryx	2012	Electrocution	Undescribed
STEM 66	Liver	Male	Kalahari Oryx	2012	Electrocution	Undescribed
STEM 115	Spleen	Female	Tswalu	2015	Electrocution	*C. botulinum*
STEM 116	Spleen	Male	Tswalu	2015	Electrocution	*C. haemolyticum*

Discussion

A 17.65% prevalence of *C. novyi* sensu lato was recorded in a population of 34 *S. temminckii* individuals from two sites in the xeric, Kalahari savanna biome and one specimen confiscated from the illegal trade. This study is the first to detect the presence of *C. novyi* sensu lato in *S. temminckii* in particular as well as in pangolins in general. Previous studies have identified various ticks, a nematode (*Habronema hamospiculatum)* and a *Piroplasma* sp. in *S. temminckii* individuals (Baylis, 1931; Rewell, 1950; Mohapatra *et al.*, 2016).

There was strong support for an additional, undescribed, clade within the *C. novyi* sensu lato group. This clade has high nodal support in both the nucleotide and amino acid phylogenies and is genetically distinct from the remaining members of the *C. novyi* species group. The phospholipase C gene of *C. novyi* type B and *C. haemolyticum*, which plays an integral role in their respective pathogenesis as the beta toxin (Smith and Williams, 1984; Hatheway, 1990; Songer, 2005; Navarro *et al*., 2017), is highly similar to that of the purported novel taxon's phospholipase *C* gene. This together with the purported novel taxon's phylogenetic placement in the almost exclusively pathogenic *C. novyi* sensu lato group, suggests that this novel taxon may also be a pathogenic species, however further investigation through culturing and inoculation experiments are required to verify this. *Clostridium haemolyticum* was only found in the lung, possibly indicating that it is in a dormant spore phase and was picked up from the surrounding soil (Navarro *et al*., 2017). *Clostridium novyi* type B and *C. botulinum* were only found in spleen samples, suggesting active infections at the time of death (Mebius and Kraal, 2005).

To my knowledge, this is the first confirmed report of *C. haemolyticum* in Africa. This is highly probable due to this disease being easily misidentified as *C. novyi* type B both clinically (black disease; Lewis, 2007) or genetically through DNA–DNA hybridisation, 16S rRNA and 16S-23S linker screening (Nakamura *et al*., 1983; Sasaki *et al*., 2000, 2001). It may

also be misidentified as other *C. novyi* sensu lato species if genes in the plasmidome are screened due to plasmidome transfer between species (Eklund *et al*., 1974; Schallehn and Eklund, 1980; Skarin *et al*., 2011). Furthermore, *C. haemolyticum* is highly sporadic (Lewis, 2007) and has been found in several countries across all continents, with the exception of Africa and Antarctica (Olson, 2013; Navarro *et al*., 2017). The latter suggests that it has a high habitat tolerance and therefore it is unlikely that conditions in Africa are unsuitable for this species.

Using the newly designed phospholipase C primers, the identification of three pathogenic species within the *C. novyi* sensu lato group was possible, with phylogenetic analyses allowing for identification of three discrete lineages with high confidence. The potential utility of this PCR screening method for epidemiological purposes is further increased by the chromosomal location of the gene in the *C. novyi* sensu lato genome (Saeb *et al*., 2017). Phospholipase C is thus not prone to horizontal gene transfer between the species of the group, as other gene regions from the plasmidome are (Skarin *et al*., 2011, 2014; Saeb *et al*., 2017). The apparent absence of separation between the *C. botulinum* type strains when using the phospholipase C gene is as a result of these strains being defined on the basis of the botulinum neurotoxins that they produce due to the major role that they play in the aetiology of botulism (Oguma *et al*., 1995). Other flagellin-specific primers developed by Sasaki *et al*. (2002) have proven to work on *C. novyi* sensu lato clinical cases (García *et al*., 2009; Takagi *et al*., 2009). However, these primers do not include *C. botulinum* and because they are species-specific, multiple PCR screening assays need to be perfomed in order to identify the species in the *C. novyi* sensu lato group, thus increasing the cost per sample. This, along with the fact that DNA–DNA hybridisation, 16S rRNA and 16S–23S linker screening have proved unsuccessful in separating species within this group (Nakamura *et al*., 1983; Sasaki *et al*., 2000, 2001), suggests that the primer set developed in this study may be a valuable alternative assay for screening for *C. novyi* sensu lato presence in animal populations.

Results from Chapter 3 using universal 16S rRNA primers confirmed the presence of the *C. novyi* sensu lato group in STEM 36 and STEM 116. However, a positive screening result was not observed for STEM 115. The 16S rRNA amplification-cloning approach also provided evidence for *C. novyi* sensu lato in samples of STEM 41 and Pangolin 2 (Chapter 3), both of which were negative with the phospholipase C primers. This suggests that additional optimisation of the primers and/or reaction conditions may be warranted in order to improve the sensitivity of the test. Although comparing primers through cloned data with that of PCR screening data is different, additional research on primer specificity for the developed primers may still be required.

The results suggest that the main driving factors for *C. novyi* sensu lato occurrence and detection in the studied population of *S. temminckii* are the type of organs screened, sex and mass of the individuals. Although there was a slight bias in males in the sample set (62.09%), the proposed single-taxon approach may not be feasible for *C. novyi* sensu lato traceability purposes as the results suggest that spleen of large males offer the best chances of detection. These biases reduce the efficacy of using *C. novyi* sensu lato as a proxy for traceability purposes.

Strong, positive interactions between the spleen and lung organs with *C. novyi* sensu lato prevalence may be due to both organs being blood-rich (Carneiro and Donald, 1977). Lungs are also in direct contact with the soil-borne genospecies group through inhalation (Chakrabarty *et al.*, 2007), whilst spleens play an important role in clearing bacterial infections from the blood (Mebius and Kraal, 2005). The low occurrence of *C. novyi* sensu lato in the liver is surprising as the liver is known to harbour inactive *C. novyi* and *C. haemolyticum* spores before active infection occurs (Uzal *et al*., 2016) and processes a large volume of potentially contaminated blood from all parts of the gastrointestinal tract as well as the spleen itself (Grayson and Mendel, 1957). This suggests that the reduced level of positive screening results

in the liver may be related to the spleen's efficiency in removing blood-borne organisms before reaching the liver (Mebius and Kraal, 2005). Although the literature pertaining to an apparent male sex bias for *C. novyi* sensu lato is lacking, this apparent sex bias could be attributed to pangolin males having larger home ranges than females (Heath and Coulson, 1997) and longer roving periods (up to seven years, as opposed to approximately 1–2 years for females; Pietersen, 2013; Pietersen *et al*., 2014b), and thus greater probability of exposure to *C. novyi* sensu lato species. Pietersen *et al*. (2014b), however did not observe significant differences in home range size between males and females in a Kalahari study population (the same population used in this study) and thus the roving is a more likely candidate for the observed sex bias in this study. The apparent poor predictive potential of latitude and longitude may be as a result of the small variation in these variables between sampling localities (1°21' and 2°05', respectively), and explain the lack of predictive powers of the environmental variables that are linked to latitude and longitude (precipitation, temperature and altitude). In addition to incorporating samples from a wider geographic region, future studies should also consider soil conditions such as pH, temperature and moisture as the presence and prevalence of all *Clostridium* species appear to be correlated with soil conditions that fit that of the Kalahari biome (Skujinš, 1984; Hatheway, 1990; Fierer *et al*., 2012).

In order for spores to germinate, resulting in vegetative growth and toxin production of *C. haemolyticum* and *C. novyi* type B, there needs to be some sort of perturbation to the liver which creates a localized anaerobic environment (Bagadi and Sewell, 1973; Erwin, 1977; Hatheway, 1990). Perturbation is caused by liver trauma, necrosis and parasite invasion in the form of liver flukes or worms (Bagadi and Sewell, 1973; Erwin, 1977; Navarro *et al*., 2017; Nyaoke *et al*., 2018). Thus, the distribution and incidence of *C. haemolyticum* and *C. novyi* is often based on soil conditions as well as the presence and level of endo-parasites of the liver, with an emphasis on liver flukes (Uzal *et al*., 2016). However, liver flukes are suggested to be

uncommon in the arid areas of the Northern Cape Province based the distribution of its hosts - the four common *Lymnea* species found in South Africa (de Kock and Wolmarans, 2008). This mirrors a case study in Uruguay where only one out of 20 cases of *C. haemolyticum* were probably caused by liver fluke perturbation (Navarro *et al*., 2017), and also reflects the findings of a *C. novyi* type B case review (Nyaoke *et al*., 2018). Whilst other parasites have been implicated in cases of infectious necrotic hepatitis (see: Dmaresq, 1939) and bacillary haemoglobinuria (see: Janzen *et al*., 1981; Stogdale and Booth, 1984), no histological investigations were undertaken on specimens that were positive for *C. novyi* sensu lato in this study, thus preventing me from suggesting any possible parasites that may have resulted in proliferation of the possible *C. novyi* type B infection. Another potential cause of infectious growth could be from nutritional stress which has been suggested to occur in this population of *S. temminckii* (W. Panaino, *unpubl. data*) and is often the cause of non-alcohol induced liver damage in humans (Morgan *et al.*, 1976; Mezey, 1978). An additional source of liver perturbation may be from prolonged electrocution from electric fencing which can affect the normal functioning of vital organs and directly damage cells from heat or electrolysis (Anderson, 1957; Bikson, 2004). This may explain why electrocution was the only cause of death where a positive screening result was observed, however, this must be interpreted with caution as electrocution was the most prevalent cause of death in the studied population (77.89%).

Geographic traceability

The use of the *C. novyi* sensu lato genospecies group in its entirety as part of a site-specific presence/absence key is not possible due to there not being any substantial effect of locality (latitude and longitude) on a positive screening result. Geographic separation through the topotyping approach like that of the FMD virus (Bastos *et al.*, 2000; Vosloo *et al*., 2001;

Di Nardo *et al.*, 2011; Hall *et al.*, 2013) is visible for the *C. novyi* sensu lato genospecies. Although not technically topotyping considering these are genotypes found globally, the principle of molecular typing for geographic separation is shared between the study on FMD virus and this study. The findings are also a mixture of molecular strain typing and a presence/absence key. This is because through molecular typing *C. botulinum* and *C. haemolyticum* are found exclusively in Tswalu individuals whilst *C. novyi* type B and a possible novel taxon are found exclusively in Kalahari Oryx individuals. However, the two Tswalu individuals that tested positive were sampled at a different time (2015) to when all of the Kalahari Oryx individuals as well as the individuals from Tswalu that tested negative for *C. novyi* sensu lato were sampled (2011–2013). Without any evidence against a possible influence of sampling date, it cannot be concluded that these differences in genospecies were the result of geographic differences. This, together with the low prevalence of *C. novyi* sensu lato in this population, suggests that it is unlikely that this group will have any bearing on geographic traceability of *S. temminckii* and therefore will not be considered further. Instead this group does have a bearing on the potential risk of disease epidemics resulting from *S. temminckii* locally and internationally through the illicit trade.

Localised pangolin-related transmission

Botulism has a high direct epizootic potential due to its ability to infect animals in and around an infected carcass (Duncan and Jensen, 1976). The carcass of a deceased infected individual could essentially become a reservoir for the disease to spread, resulting in a potential outbreak emanating from the carcass. With known invertebrate and avian carriers of the disease, localised spread may not be bound by geographic (water bodies, mountain ranges, *et cetera*) or anthropogenic barriers (game fencing, urban infrastructure, *et cetera*). Necrophagous flies are one such carrier which can transfer botulism to other areas for up to 24 hours after

eating an infectious carcass (Uzal *et al.*, 2016). Spores that make their way into pastures, feed and water sources could result in major economic losses for both livestock and wild game industries (Hatheway, 1989; Lindström *et al*., 2004) as has previously occurred in South Africa (Van de Lugt *et al*., 1995).

Pietersen e*t al.* (2014b) indicated in a study conducted at Kalahari Oryx that *S. temminckii* density was estimated to be as high as 0.31 individuals/km^2 once transient individuals were taken into account. These transient roving individuals may transmit pathogens, including those of *C. novyi* sensu lato, from one home range to another along their roving routes. Dispersal may also be a factor of pathogen spread with distances of *S. temminckii* averaging around 49 km but known to reach up to 300 km, with both sexes known to disperse (van Aarde, 1990; Pietersen e*t al.*, 2014b). The aforementioned densities, dispersal distances and transient rovers may influence pathogen spread over large localised distances and to numerous individuals. However, the potential spread is significantly increased from localized areas to entire continents once the illegal trade in pangolins is considered.

Illegal trade in pangolins and potential for global disease spread

The possible spread of clostridial diseases may not seem to be of concern for the legal trade of livestock as disease policies, screening and vaccination programs are in place to prevent the spread of known diseases from a particular region (Scoones and Wolmer, 2006). However, these procedures are absent from the illegal wildlife trade and this has potentially far-reaching consequences for the spread of all *C. novyi* sensu lato species (Travis *et al*., 2011). With the level of illegal trade of pangolins in general known to be considerable (Challender *et al*., 2014; UNODC, 2016), along with the increase in illegal trade of African species, including *S. temminckii* (Challender and Hywood, 2012; Pietersen *et al*., 2014a,c), there is an added risk of pathogen spread via *S. temminckii* to other traded species, naïve end-destination populations

and consumers of its meat or other potentially infectious products (Bell *et al*., 2004; Fèvre *et al*., 2006; Smith *et al*., 2009b; Travis *et al*., 2011). The risk is heightened knowing that pangolins are often traded live or frozen for consumption purposes whilst *C. haemolyticum* and *C. novyi* spores remain dormant in animal liver tissue (Navarro *et al*., 2017; Nyaoke *et al*., 2018). The implications of this are substantial, particularly after recent reports of the first human cases of *C. novyi* type B (Abdel-Haq *et al*., 2007) and *C. haemolyticum* that was said to contain multi-drug antibiotic resistance genes (Saeb *et al*., 2017).

Study caveats

This study had a bias between sample sites in terms of number of samples with 23.53% of the sample set coming from Tswalu whilst Kalahari Oryx made up the remaining 76.47% (excluding the seized sample). The specificity and sensitivity of the primer set developed could not be determined, leaving the possibility for false negatives.

Conclusion and synthesis

The potential for geographic proxy tracing of *C. novyi* sensu lato as a molecular typing method was not realised in this study due to a low prevalence of the group and the potential effect of sampling date on genospecies variation between Tswalu and Kalahari Oryx. Furthermore, the variation of a positive screening result between different organs, a male bias and the effect of mass may reduce the efficacy of *C. novyi* sensu lato being used for geographic traceability of *S. temminckii*. This highlights the potential bias that small sample sets can have on determining key site-specific taxa or taxa for molecular typing procedures. More samples with higher throughput data during the metagenomics stage of microbial traceability as well as the use of one health related databases may better identify taxa for molecular typing procedures in general as well as for *S. temminckii* in particular.

Considering the potential epidemiological role that *C. novyi* sensu lato can play in human, domestic and wild animal populations, the fear of possible transmission and perturbation of these species with subsequent outbreaks and spread of the diseases they cause is real. This is heightened when a species such as *S. temminckii* is highly endangered and heavily traded. This along with climate change as well as increased interaction of human, domestic and wild animal populations exacerbates the situation. A call for increased understanding of these diseases and the immunological responses of *S. temminckii*, increased environmental screening of possible outbreak areas along with increased awareness of the plight of this animal may in turn reduce the risk of disease epidemics in the future. However, increased awareness of the possible pathogens that pangolins may harbour and the resultant epidemics that may arise from these pathogens may inadvertently become a tool to deter the trade and consumption of pangolins in the future.

Chapter 5: Summary and conclusions

With pangolins being labelled the most heavily trafficked wild mammal in the world (Challender *et al.*, 2014), and with data showing increased levels of trade (Heinrich *et al.*, 2017), the regulation of pangolin trade through regulatory bodies has had limited success in curbing this trade. Investigating geographic traceability as a direct and complementary approach is necessary. However, with a number of flaws relating to DNA markers (Arenas *et al.*, 2017), along with six independent research groups working on traceability through pangolin DNA, identifying and testing the feasibility of complementary traceability approaches is essential. The current study is the first to investigate the feasibility of using stable isotope and microbial techniques as traceability tools for pangolins. This included determining possible sources of variation specific to pangolin traceability for both of these alternative molecular approaches, whilst concurrently identifying the best approach for future traceability analyses. This groundwork is often not considered during traceability research which can lead to the use of expensive, non-universal and/or low-resolution techniques with limited potential for forensic investigations (Ogden, 2008). The data generated in this study are valuable for establishing baseline estimates of microbial diversity and prevalence in order to better identify the disease risks related to such a heavily trafficked mammal (Fèvre *et al.*, 2006; Smith *et al.*, 2009b).

General findings and further research

Stable isotope traceability

Although there are many examples of successful geographic traceability of wildlife through use of stable isotopes (van der Merwe *et al.*, 1990; Vogel *et al.*, 1990; Koch *et al.*,

1995; Ziegler *et al.*, 2016; Cerling *et al.*, 2018), a limited number of these have gone into testing possible factors that may reduce the accuracy and feasibility of this approach. This research tested whether variation within the scale, between different body regions within the same individual, between different individuals and scale harvesting methods were important sources of variability to consider for stable isotope traceability of Temminck's Ground Pangolins (*Smutsia temminckii*). The results indicated that the sampling site on individual scales influences the resultant $\delta^{13}C$ and $\delta^{15}N$ isotopic values, but that the anatomical position of the sampled scale and treatment thereof to mimic harvesting methods (boiling or direct flame), do not. The latter factor shows the robustness of isotopic analyses (Bocherens, 2009; Harbeck *et al.*, 2011), whilst the absence of an effect of the anatomical location of the sampled scale on the results is important for forensic cases considering that seizures often come in bags of already harvested, loose scales (Mwale *et al.*, 2016; Heinrich *et al.*, 2017). A cost- and time-effective sampling method to account for the observed inter-scale variability was tested, indicating that the use of homogenised duplicates of a scale sample may be sufficient to account for this observed variability.

Although the two Kalahari sites (xeric savanna) did not separate based on $\delta^{13}C$ and $\delta^{15}N$ values, an indication of isotopic separation between the Kalahari sites and a sample from the Lowveld was noted. However, due to inter-individual variability within the Kalahari Oryx population being higher than that of all three populations together, as well as the possible influence of date of sampling for the Klaserie sample, geographic partitioning was not realised through carbon and nitrogen alone. It is not surprising that no differences were observed between Tswalu and Kalahari Oryx samples (100 km apart), considering that *S. temminckii* juveniles may disperse in excess of 49 km from their natal home ranges in these areas (Pietersen *et al.*, 2014b) and that these dispersals may exceed 300 km (van Aarde *et al.*, 1990). Additionally, both sites have similar geology, climate and vegetation types (Pietersen, 2013;

van Rooyen and van Rooyen, 2017), whilst basal nitrogen isotope values are heavily influenced by soil-related factors (Evans, 2001; Aranibar *et al.*, 2008; Craine *et al.*, 2015) and basal carbon by vegetation type (Dawson *et al.*, 2002; Marshall *et al.*, 2007; Tipple and Pagani, 2007).

Future research using additional isotopes is therefore recommended for better geographic partitioning, once limitations of each have been considered. This has been shown to markedly increase the resolution of geographic traceability in elephant populations as these provide more levels on which to separate populations (Koch *et al.*, 1995; Ziegler *et al.*, 2016). Deuterium (δD) and oxygen ($\delta^{18}O$) basal variability through water is known for its gradients in altitude and latitude, with increased depletion in ^{2}H and ^{18}O noted the further away from the equator and the higher in elevation precipitation occurs (Hobson and Wassenaar, 1996; Hobson *et al.*, 2003; Ben-David and Flaherty, 2012). Water body type (rainfall, ground water, oceanic water and lakes) as well as the impact of the ocean on an area's hydrological processes also show δD and $\delta^{18}O$ variability (Gat, 1996). However, future analyses of δD and $\delta^{18}O$ values of pangolins must consider the absorptive behaviour of water in scales (Liu *et al.*, 2016b) and the influence that the hydrogen atom exchange in tissues may have on these values (Chesson *et al.*, 2018b). This is specifically related to the harvesting method of boiling as well as scale cleaning during sample preparation. Sulphur ($\delta^{34}S$) is another isotope with potential as it can be affected by wet or dry deposition of atmospheric sulphur in different plants, microbial influence on soil, presence of sea spray as well as the underlying geology itself (Richards *et al.*, 2003), which also influences lead ($\delta^{206,207,208}Pb$) isotopes (Vogel *et al.*, 1990). Additionally, $\delta^{34}S$ has been shown to have minimal fractionation values (−1‰) between diet and keratinous structures (Richards *et al.*, 2003). However future research using $\delta^{34}S$ must not only consider the fact that $\delta^{34}S$ has been shown to produce high localised variability in human and animal studies, probably due to high variability across a finite geographic scale, but that it can also result in similar values between different sites that are separated over large distances from one another

(Nehlich, 2015). Strontium ($\delta^{87}Sr$) can also be affected by atmospheric deposition as well as the age, weathering and type of rock substrate (Blum *et al.*, 2000), however like sulphur, future studies should take into account the high level of variability at localised scales due to high levels of heterogeneity within rock, water and plant sources (Bentley *et al*., 2004).

Further research should also delve into understanding pangolin scale growth rates through methods such as bomb-curve radiocarbon dating (Uno *et al.*, 2013) as this will aid in better interpretations of variability across the scale. Traceability estimates using alternative pangolin samples such as skin should be explored, however caution should be given to biochemically active samples (Viljoen *et al.*, 2016). The isotopic analysis of soil often attached to scales during seizures of live pangolins and bags of scales may be useful in traceability of pangolins, but requires in-depth testing first. This may allow investigators to use already developed soil isoscapes (Bataille and Bowen, 2012) or reference soil samples across pangolin distributions as a means to trace the origin of pangolins through the soil often attached to their scales. This method has the potential to reduce the ethical, time and sample number limitations relating to development of isotopic reference databases of pangolin samples for traceability purposes.

Microbial traceability

Generating the first baseline microbial diversity reference dataset for *S. temminckii* and pangolins in general is the first step to understanding the feasibility of geographic traceability through microbial proxies. Metagenomics is key to developing baseline estimates, however determining which method of taxonomic assignment should be considered (phylogeny testing or heuristic search tools such as BLAST) and which metagenomics approach (PCR-cloning or NGS) is more feasible for future traceability research is necessary. Both the choice of taxonomic assignment method and metagenomics approach significantly affected the resultant

microbial community composition with the use of count data rather than presence/absence data being more efficient in separating these factors. Phylogeny testing was suggested as the better taxonomic assignment method as it provides significantly lower false positives (Ross *et al.*, 2008), can identify taxa at lower levels of assignment (within species) and provide a mechanism for comparing the relatedness of sequences generated during the study. It also forces investigators to look more in-depth at the sequencing output, which allows for the discovery of bad quality or misidentified reference sequences, along with the identification of taxonomic name changes (Turenne *et al.*, 2001; Patel *et al.*, 2004; DeSantis *et al.*, 2006). It was suggested that NGS technology be used as the gold standard approach for metagenomic analyses as the technology is high-throughput which allows for identification of taxa at low prevalence in samples, it is more cost effective and has a faster turn-around time, particularly if phylogeny testing is to be included (Xu, 2006; Samarajeewa *et al.*, 2015; Zhou *et al.*, 2015). The technology is also decidedly adjustable whilst continuously being advanced in fronts that were initially its drawbacks (e.g. longer sequence reads, improved sequencing accuracy and phylogeny-based taxonomic assignments; Zhang *et al.*, 2013).

The feasibility of tracing the geographic origin of *S. temminckii* through microbial proxies across two study sites in South Africa (Kalahari Oryx and Tswalu) was tested. Although this goal was not realised at a whole microbial community level, a presence/absence approach for *C. sordellii* and possible topotyping for the genospecies group *C. novyi* sensu lato suggests that it is possible. Identifying these highly distinctive and ubiquitous taxa as opposed to traceability through whole community analyses alone reduces the impact that misidentifications, false positives and sequencing errors as well as other sources of innate variability such as sampling date, organ and cause of death may have on the accuracy of traceability estimates (Faust *et al.*, 2015). The sources of innate variability were tested with

organ type having a possible influence on microbial community composition which suggests that deciding on an organ type may be important for future traceability research.

Time is also an important factor to consider as microbial communities may change in an environment given a perturbation (Torsvik and Øvreås, 2002; Faust *et al.*, 2015), but may also change within an animal once it has died (Howard *et al.*, 2010; Hyde *et al.*, 2013; 2015). Time-lapse studies for both of these factors are therefore important when evaluating the feasibility of this approach, particularly if whole-microbial community traceability is to be performed (Cohen *et al.*, 2018).

Although no screening for viruses was conducted in this research, further research should delve into this as viruses evolve more rapidly than bacteria and are thus liklley to provide higher levels of resolution (Drake *et al.*, 1998). In addition, microbes that have a global presence, are environmental in origin and if possible have already been distinctly characterized geographically, are good candidates for microbial traceability studies looking at site-specific presence/absence or topotyping levels. These include species like the *Leishmania donovani* complex which causes Leishmaniasis (Lukes *et al.*, 2007; Alvar *et al.*, 2012), diarrheagenic *E. coli* and *Coxiella burnetii* which causes Q fever (Maurin and Raoult, 1999).

Clostridium novyi *sensu lato*

Based on the findings of the microbial traceability chapter (Chapter 3), an in-depth molecular typing approach of the genospecies *C. novyi* sensu lato was performed in order to determine whether this genospecies group would be a good candidate for geographic proxy tracing of *S. temminckii*. However, traceability was not realised as the group had a lower prevalence of 17.65% than what was initially anticipated from Chapter 3 (50%) as well as biases in organs screened, sex and mass of individuals.

Due to the genospecies group being notoriously difficult to characterise using previously published gene regions and methods (Nakamura *et al.*, 1983; Sasaki *et al.*, 2000; Sasaki *et al.*, 2001), the development of a single primer set targeting the phospholipase C gene (beta toxin) was integral in genospecies discrimination. This new primer set is the first to be successfully used on the *C. novyi* sensu lato group in a purportedly healthy animal population.

The identification of all three species within the genospecies group is also of epidemiological importance considering that *C. novyi* type B, *C. haemolyticum* and *C. botulinum* group III are respectively responsible for infectious necrotic hepatitis (black disease), bacillary haemoglobinuria (red water disease) and animal botulism (Uzal *et al.*, 2016). The population prevalence of 17.65% may be low for traceability testing purposes but is substantial when one considers the potential risk to public and animal health. This, along with the comparatively high alpha diversity in Chapter 3 when compared to human spleen diversity (Can *et al.*, 2014), suggests that there may be an underlying mechanism that explains these findings. The most likely, based on available data, is that genes coding for first-line defence immunity in skin epithelial and inner-mucosa protected tissues (intestine, lung and reproductive tissue) have been shown to be pseudonised and reduced in pangolins, compared to other mammals (Choo *et al.*, 2016). Despite the keratin-based scales possibly being the evolutionary trade-off for first-line defence immunity of skin cells as they act as a protective barrier for potential sources of infection (open wounds), the inner-mucosa protected tissues do not have this barrier (Meyer *et al.*, 2010; Choo *et al.*, 2016). These tissues may result in higher infection profiles particularly for soil-borne bacteria that rely on ingestion for infection, such as *C. novyi* sensu lato (Hatheway, 1990; Navarro *et al.*, 2017; Nyaoke *et al.*, 2018), as these myrmecophagous mammals (Jacobsen e*t al.*, 1991; Swart *et al.*, 1999; Pietersen e*t al.*, 2014b, 2016) are known to ingest soil during feeding.

If this is found to be the case, then the use of highly prevalent soil-borne microbial species in presence/absence and topotyping traceability analyses should be considered in future studies. One step further is to investigate traceability of *S. temminckii* through environmental samples (soil sampled from feeding and burrowing sites). With large microbial databases being created from both environmental samples in the Earth Microbiome Project (Thompson *et al.*, 2017) and disease-related samples in One Health centred databases (Lebov *et al.*, 2017), these may become vital georeferenced databases that complement future geographic traceability estimates of pangolins and other illegally traded animals. However, before these ideas can be realised, it is important to identify the level of microbial transmission from other sources to traded pangolins and the influence that this may have on microbial proxy traceability (Cohen *et al.*, 2018).

Conclusion

The potential for geographic traceability of *S. temminckii* through microbial and stable isotope signatures was assessed in this study. Chapters two, three and four not only dealt with identifying the potential that these approaches have for geographic traceability, but more importantly tested the assumptions relating to these techniques and technologies. This has led to me not only highlighting the strengths of each, but also drawing important considerations which in turn have provided best practice microbial and stable isotope protocols for future pangolin traceability research. These two approaches may complement each other, as exemplified by the forensic study that identified the source of the anthrax bioterrorism threat in 2001 (National Research Council, 2011). However, they also share some caveats such as the variation of geographic signatures over shorter time periods compared to DNA techniques as well as signature variability influenced by prevailing environmental conditions, tissue type and an animal's physiological status, age or sex.

In order to attain the ultimate aim of tracing seized pangolins and their derivatives to their geographic origins which can subsequently lead to the identification of poaching hotpots, syndicate networks or even spots for repatriation of live individuals (Fernandes and Caparroz, 2013; Ogden and Linacre, 2015), the current silver bullet approach of only using host DNA markers may need to be reconsidered as it may not provide the robustness and resolution needed (Clegg *et al.*, 2003). Although this method is well established and should not necessarily be replaced (Arenas *et al.*, 2017; Hampton-Marcell *et al.*, 2017), microbial and stable isotope traceability techniques have the potential to aid in identifying source populations of traded pangolins. However, substantial validation efforts are still required before these two approaches can be seen in the same light as host DNA forensics (Arenas *et al.*, 2017; Dunn *et al.*, 2017; Hampton-Marcell *et al.*, 2017).

It is integral that future research on pangolin traceability not only encapsulates the requirements of prosecutors in order for the outputs of such research to be used in a court of law (Ogden, 2008) but that all possible assumptions are considered and validated to permit future use of these tools. Regardless of the method used, any stage of data formation or interpretation (sample collection, transport, laboratory work, statistical analyses and database accuracy testing) will be heavily scrutinised in the court of law in order to find reasonable doubt of the evidence presented, especially if assumptions have not been validated (Ogden, 2008). This should be kept in mind when developing a method aimed to be used for prosecution purposes

The hope is that these methods will form part of a multifaceted approach in pangolin traceability research, conservation and prosecution in order to help curb the illegal trade of pangolins. Furthermore, including this as a direct conservation approach to the pangolin trade predicament may alleviate the burden that regulatory bodies and governments are facing with regards to this trade by complementing the flaws often aligned with regulatory conservation

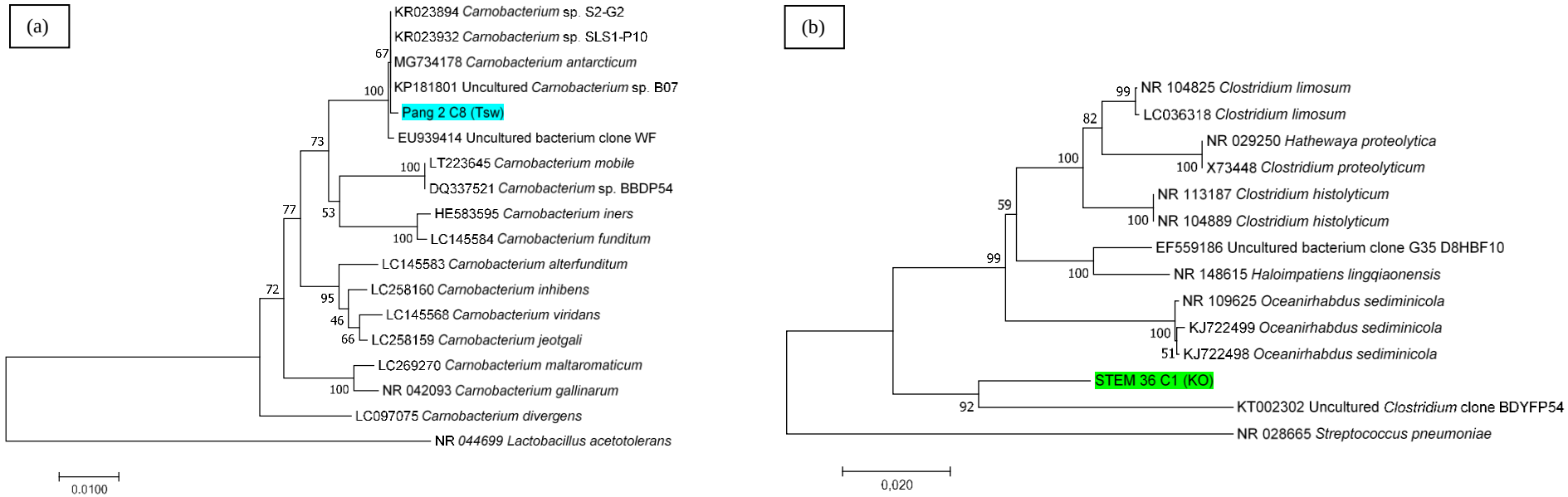
approaches (IUCN, 2000; Oldfield, 2003; Challender *et al.*, 2015). This can only help reduce the possible impact that the pangolin trade may have on ecological processes (Cooney *et al.*, 2015), the risk of disease epidemics (Fèvre *et al.*, 2006; Smith *et al.*, 2009b) and the resultant costs to the economy (Pimentel *et al.*, 2005; Westphal *et al.*, 2008; World Bank, 2010).

Appendices

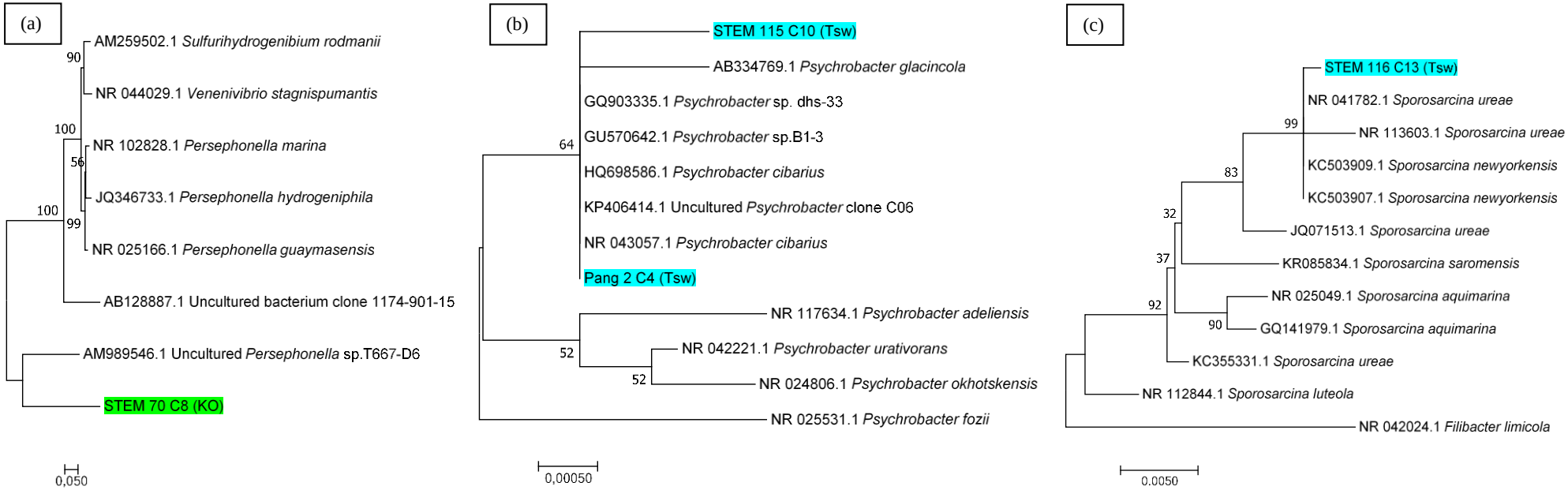
Appendix 1: Summary of the Temminck's Ground Pangolin (*Smutsia temminckii*) samples that were used in this study (sample ID), their origin, sampling date and method of traceability used.

Specimen	***Origin***	***Year of sampling***	***Stable isotopes***	***Microbial traceability***	**Clostridium novyi *sensu lato***
STEM 2	Kalahari Oryx	2009			Yes
STEM 8	Kalahari Oryx	2010	Yes		
STEM 9	Kalahari Oryx	2010			Yes
STEM 10	Kalahari Oryx	2010	Yes		Yes
STEM 12	Kalahari Oryx	2010	Yes		
STEM 13	Kalahari Oryx	2010	Yes		Yes
STEM 15	Kalahari Oryx	2010	Yes		
STEM 20	Kalahari Oryx	2010	Yes		
STEM 22	Kalahari Oryx	2010	Yes		
STEM 28	Kalahari Oryx	2010	Yes		
STEM 29	Kalahari Oryx	2010	Yes		
STEM 30	Kalahari Oryx	2010	Yes		Yes
STEM 32	Kalahari Oryx	2011	Yes		Yes
STEM 33	Kalahari Oryx	2011			Yes
STEM 34	Kalahari Oryx	2011	Yes		
STEM 36	Kalahari Oryx	2011	Yes	Yes	Yes
STEM 37	Kalahari Oryx	2011	Yes		Yes
STEM 41	Kalahari Oryx	2011		Yes	Yes
STEM 51	Kalahari Oryx	2011			Yes
STEM 52	Kalahari Oryx	2011	Yes		
STEM 53	Kalahari Oryx	2011	Yes	Yes	Yes
STEM 54	Kalahari Oryx	2011	Yes		Yes
STEM 55	Kalahari Oryx	2011	Yes		
STEM 56	Kalahari Oryx	2011	Yes		Yes
STEM 58	Kalahari Oryx	2011	Yes		
STEM 59	Kalahari Oryx	2011	Yes		

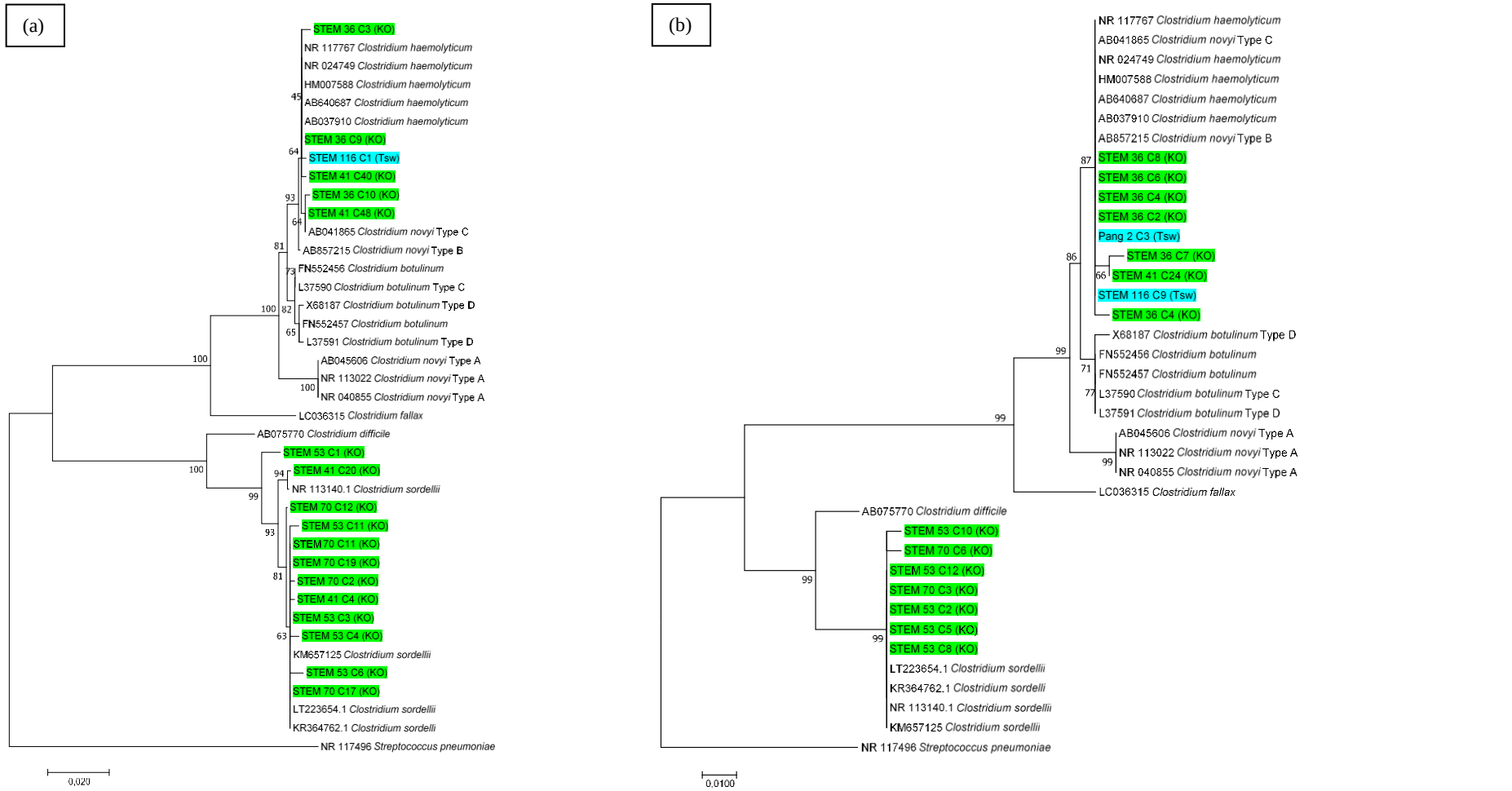
STEM 61	Kalahari Oryx	2011			Yes
STEM 62	Kalahari Oryx	2011	Yes		Yes
STEM 63	Kalahari Oryx	2011	Yes		
STEM 66	Kalahari Oryx	2012	Yes		Yes
STEM 67	Kalahari Oryx	2012	Yes		
STEM 70	Kalahari Oryx	2012	Yes	Yes	Yes
STEM 72	Kalahari Oryx	2012	Yes		Yes
STEM 81	Kalahari Oryx	2012	Yes		Yes
STEM 82	Kalahari Oryx	2012	Yes		
STEM 86	Kalahari Oryx	2012	Yes		
STEM 87	Tswalu	2012			Yes
STEM 88	Tswalu	2012			Yes
STEM 89	Tswalu	2012	Yes	Yes	Yes
STEM 90	Kalahari Oryx	2012			Yes
STEM 91	Kalahari Oryx	2012			Yes
STEM 102	Kalahari Oryx	2013			Yes
STEM 103	Kalahari Oryx	2013			Yes
STEM 115	Tswalu	2015		Yes	Yes
STEM 116	Tswalu	2015		Yes	Yes
STEM 117	Klaserie	1980	Yes		
Ex- Ray	Confiscated	2014			Yes
Pangolin 1	Unknown	2014			
Pangolin 2	Unknown	2014		Yes	Yes
STEM S3	Tswalu	2017	Yes		
STEM S4	Tswalu	2017	Yes		
STEM S5	Tswalu	2017	Yes		
STEM S6	Tswalu	2017	Yes		
Tswalu 5	Tswalu	Unknown			
Tswalu 6	Tswalu	Unknown			
Tswalu 7	Tswalu	Unknown			



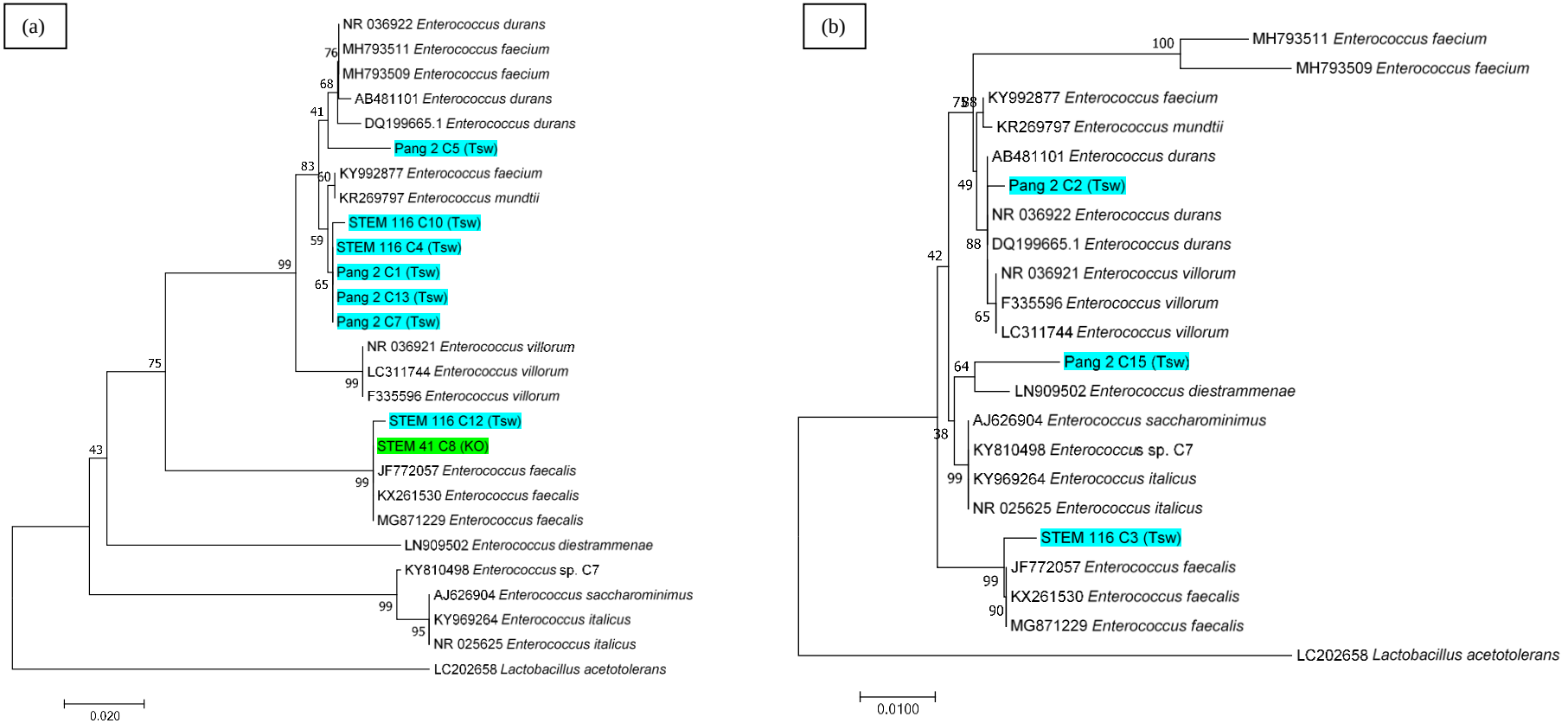
Appendix 2: Neighbour Joining p-distance trees of the 16S rRNA gene, showing the evolutionary relatedness of bacterial taxa obtained from cloned Temminck's Ground Pangolin (*Smutsia temminckii*) lung and spleen samples (highlighted) to reference taxa. Samples come from Tswalu Kalahari Reserve (blue) and Kalahari Oryx Private Game Farm (green). Nodal support was assessed with 5 000 non-parametric bootstrap replicates. These trees represent genera that had only one sequence direction (sense or antisense). The taxa identified using these trees were as follows: a) *Carnobacterium* (806 bp) with species ID of *C. antarticum*, b) *Oceanirhabdus* (1 088 bp) with species ID of Uncultured *Clostridium* clone BDYFP54.



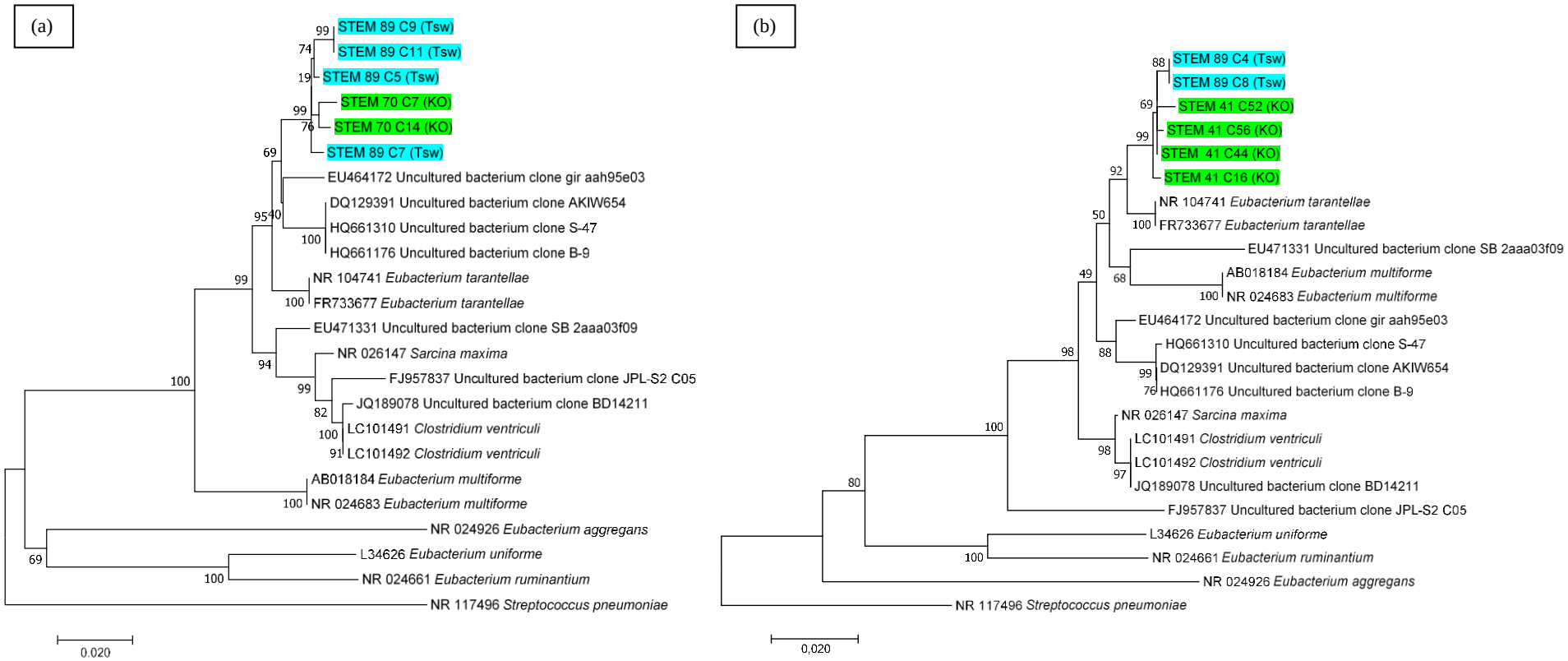
Appendix 3: Neighbour Joining p-distance trees of the 16S rRNA gene, showing the evolutionary relatedness of bacterial taxa obtained from cloned Temminck's Ground Pangolin (*Smutsia temminckii*) lung and spleen samples (highlighted) to reference taxa. Samples come from Tswalu Kalahari Reserve (blue) and Kalahari Oryx Private Game Farm (green). Nodal support was assessed with 5 000 non-parametric bootstrap replicates. These trees represent genera that had only one sequence direction (sense or antisense). The taxa identified using these trees were as follows: a) *Persephonella* (488 bp) with species ID of Uncultured bacterium aff. *Persephonella*, b) *Psychrobacter* (915 bp) with species ID of *P. cibarius*, c) *Sporosarcina* (898 bp) with species ID of *S. ureae / newyorkensis*.



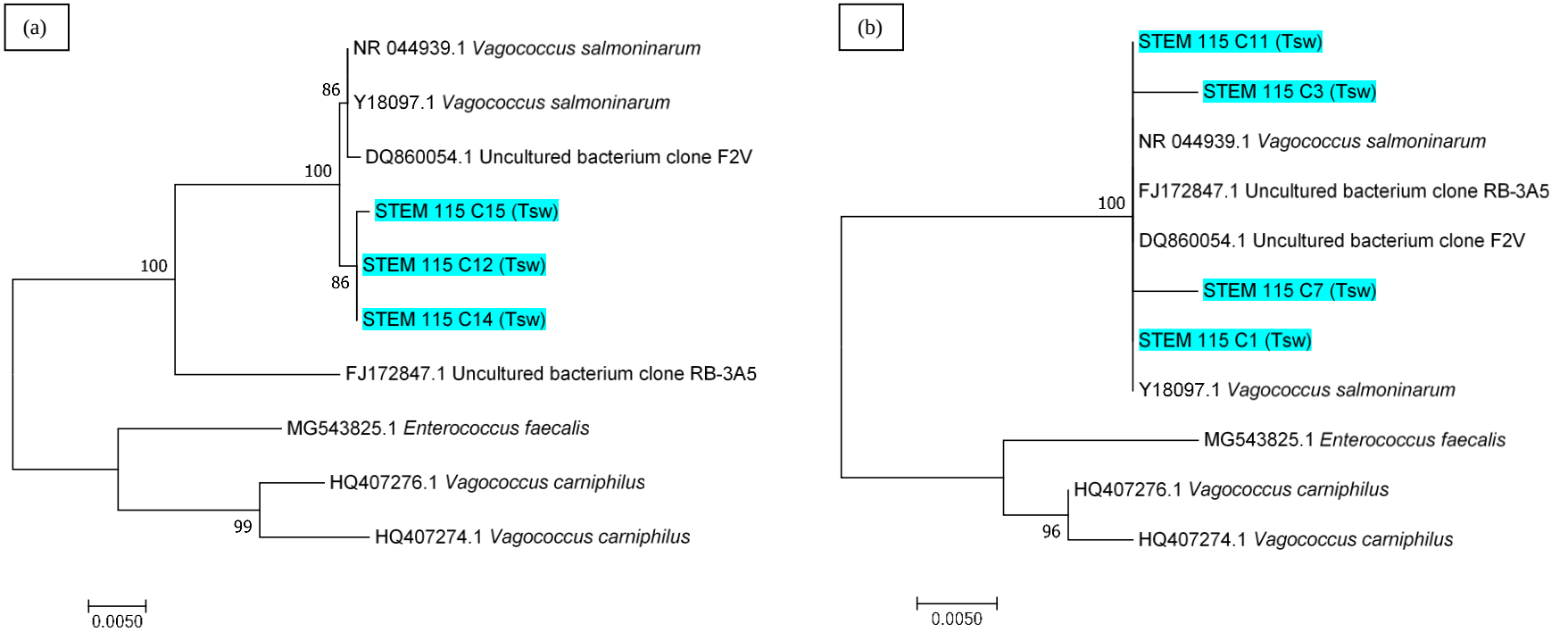
Appendix 4: Neighbour Joining p-distance trees of the 16S rRNA gene assessed using 5 000 non-parametric bootstrap replicates, showing the evolutionary relatedness of *Clostridium* taxa obtained from cloned Temminck's Ground Pangolin (*Smutsia temminckii*) lung and spleen samples (highlighted) to reference taxa. Samples were sourced from Tswalu Kalahari Reserve (blue) and Kalahari Oryx Private Game Farm (green). These trees represent genera that either had sequences starting at the a) 5' end (698 bp), or b) 3' end (488 bp) of the 16S rRNA gene. The taxa identified using these trees were as follows: a) *C. haemolyticum*, *C. novyi* type C and *C. sordellii*; b) *C. haemolyticum* / *novyi* and *C. sordellii*



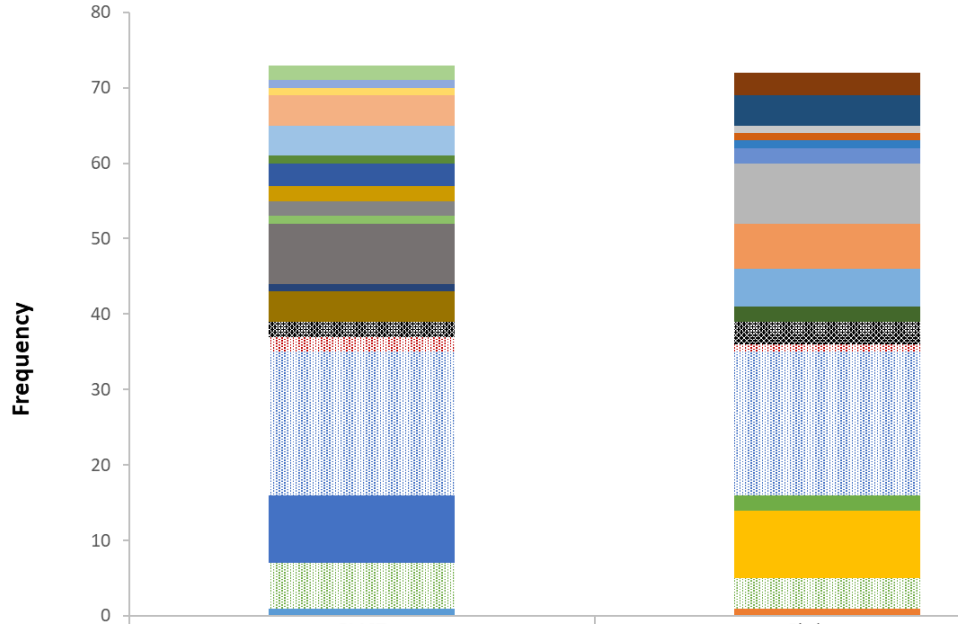
Appendix 5: Neighbour Joining p-distance trees of the 16S rRNA gene assessed using 5 000 non-parametric bootstrap replicates, showing the evolutionary relatedness of *Enterococcus* taxa obtained from cloned Temminck's Ground Pangolin (*Smutsia temminckii*) lung and spleen samples (highlighted) to reference taxa. Samples come from Tswalu Kalahari Reserve (blue) and Kalahari Oryx Private Game Farm (green). These trees represent genera that either had sequences starting at the a) 5' end (869 bp), or b) 3' end (881 bp) of the 16S rRNA gene. The taxa identified using these trees were as follows: a) *Enterococcus* sp. 1, *Enterococcus* sp. 2 and *E. faecalis*; b) *E. durans*, *E. faecalis* and *E. diestrammenae*.



Appendix 6: Neighbour Joining p-distance trees of the 16S rRNA gene assessed using 5 000 non-parametric bootstrap replicates, showing the evolutionary relatedness of *Eubacterium* taxa obtained from cloned Temminck's Ground Pangolin (*Smutsia temminckii*) lung and spleen samples (highlighted) to reference taxa. Samples originate from Tswalu Kalahari Reserve (blue) and Kalahari Oryx Private Game Farm (green). These trees represent genera that either had sequences starting at the a) 5' end (759 bp), or b) 3' end (736 bp) of the 16S rRNA gene. The taxa identified using these trees were as follows: a) *Eubacterium* sp. 1; b) *E.* sp. aff. *tarantellae*.



Appendix 7: Neighbour Joining p-distance trees of the 16S rRNA gene assess using 5 000 non-parametric bootstrap replicates, showing the evolutionary relatedness of *Vagococcus* taxa obtained from cloned Temminck's Ground Pangolin (*Smutsia temminckii*) lung and spleen samples (highlighted) to reference taxa. Samples come from Tswalu Kalahari Reserve (blue) and Kalahari Oryx Private Game Farm (green). These trees represent genera that either had sequences starting at the a) 5' end (921 bp), or b) 3' end (583 bp) of the 16S rRNA gene. The taxa identified using these trees were as follows: a) *V.* sp. aff. *salmoninarum*; b) *V. salmoninarum*.

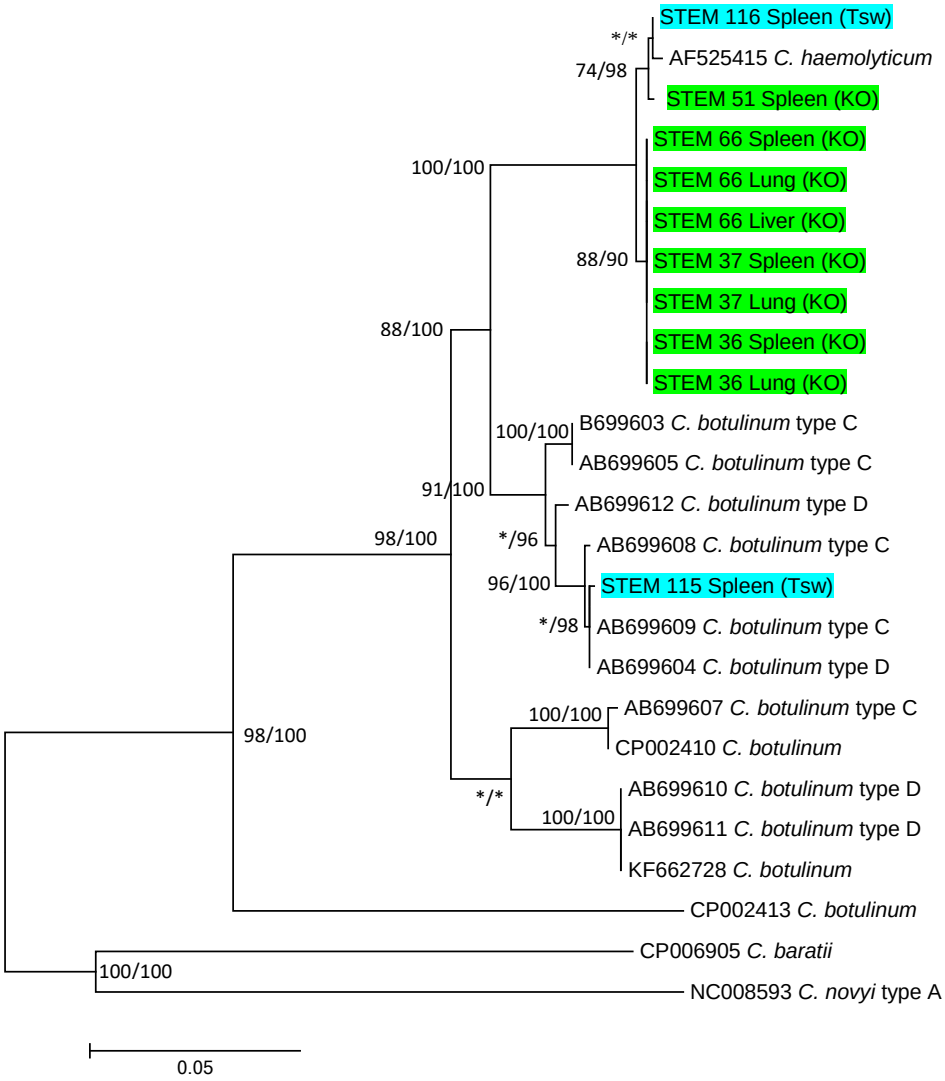


	BLAST	Phylogeny
Vagococcus sp. aff. salmoninarum		3
Vagococcus salmoninarum		4
Uncultured Psychrobacter clone C06	2	
Uncultured Persephonella sp.	1	
Uncultured Firmicutes bacterium clone M0027	1	
Uncultured Clostridium clone BDYFP54		1
Uncultured bacterium clone RB-3A5	4	
Uncultured bacterium clone gir_aah95e03	4	
Uncultured bacterium clone G35	1	
Uncultured bacterium clone F2V	3	
Uncultured bacterium clone AKIW654	2	
Uncultured bacterium clone 0182	2	
Uncultered bacterium aff. Persephonella		1
Sporosarcina ureae / newyorkensis		1
Sporosarcina ureae	1	
Psychobacter cibarius		2
Eubacterium tarantellae	8	
Eubacterium sp. aff. tarantellae		8
Eubacterium sp. 1		6
Enterococcus sp. 2		5
Enterococcus sp. 1		2
Enterococcus italicus	1	
Enterococcus faecium	4	
Enterococcus faecalis	2	3
Enterococcus durans	2	1
Clostridium sordellii	19	19
Clostridium novyi type C		2
Clostridium novyi	9	
Clostridium haemolyticum / novyi		9
Clostridium haemolyticum	6	4
Carnobacterium antarticum		1
Carnobacterium sp.	1	

Appendix 8: Stacked bar graph and associated frequency table comparing the difference in taxon identification and counts between the Basic Local Alignment Search Tool (BLAST) and phylogenetic taxonomic assignment methods of microbial diversity for eight cloned samples of Temminck's Ground Pangolins (*Smutsia temminckii*). Patterned segments indicate taxa that were identified in both methodologies.

Appendix 9: Microbial taxa together with their corresponding prevalence counts identified in clones of eight Temminck's Ground Pangolins (*Smutsia temminckii*). Taxon names were assigned using neighbour joining p-distance trees with 5 000 non-parametric bootstrap replicates. Taxa highlighted in red were found exclusively in Tswalu Kalahari Reserve, those highlighted in blue were found exclusively in Kalahari Oryx Private Game Farm, whilst orange refers to taxa that were found at both sites.

Taxon	Total count	Tswalu	Kalahari Oryx	No. Individuals
Carnobacterium antarticum	1	1		1
Clostridium haemolyticum	4	1	3	3
Clostridium haemolyticum/novyi	9	2	7	4
Clostridium novyi type C	2		2	2
Clostridium sordellii	19		19	4
Enterococcus durans	1	1		1
Enterococcus faecalis	3	2	1	2
Enterococcus sp. 1	2	2		1
Enterococcus sp. 2	5	5		2
Eubacterium sp. 1	6	4	2	2
Eubacterium sp. aff. *tarantellae*	8	4	4	2
Psychobacter cibarius	2	2		2
Sporosarcina ureae/newyorkensis	1	1		1
Uncultured bacterium aff. *Persephonella* sp.	1		1	1
Uncultured *Clostridium* clone BDYFP54	1		1	1
Vagococcus salmoninarum	4	4		1
Vagococcus sp. aff. *salmoninarum*	3	3		1



Appendix 10: Uncorrected nucleotide p-distance phylogeny of *Clostridium* inferred using the phospholipase C gene region (920 base pairs in length). All *Clostridium* species known to contain the phospholipase C gene, including representatives of the *Clostridium novyi* sensu lato group and sequences identified from Temminck's Ground Pangolins (*Smutsia temminckii*) from Tswalu Kalahari Reserve (blue) and Kalahari Oryx Private Game Farm (green) are included. Nodal support values from the Maximum Likelihood (ML) and Bayesian Inference (BI) analyses are presented ML/BI on the relevant nodes. Unsupported nodes (ML < 70% and BI < 0.95) are indicated by asterisks.

References

Abdel-Haq N.M., Chearskul P., Salimnia H. and Asmar B.I. (2007). Clostridial liver abscess following blunt abdominal trauma: case report and review of the literature. *Scandinavian Journal of Infectious Diseases*, *39*(8), pp.734-737.

Abensperg-Traun M. (2009). CITES, sustainable use of wild species and incentive-driven conservation in developing countries, with an emphasis on southern Africa. *Biological Conservation*, *142*(5), pp.948-963.

Abolnik C., Cornelius E., Bisschop S., Romito M. and Verwoerd D. (2006). Phylogenetic analyses of genes from South African LPAI viruses isolated in 2004 from wild aquatic birds suggests introduction by Eurasian migrants. *Developments in Biologicals*, *124*, pp.189-199.

Abubakar I., Irvine L., Aldus C., Wyatt G., Fordham R., Schelenz S., Shepstone L., Howe A., Peck M. and Hunter P. (2007). A systematic review of the clinical, public health and cost-effectiveness of rapid diagnostic tests for the detection and identification of bacterial intestinal pathogens in faeces and food. *Health Technology Assessment (Winchester, England)*, *11*(36), pp.1-216.

Acocks J.P.H. (1988). Veld types of South Africa. *Memoirs of the Botanical Survey of South Africa*, *57*, pp.1-146.

Ainur A., Sayang M., Jannoo Z. and Yap B. (2017). Sample size and non-normality effects on goodness of fit measures in structural equation models. *Pertanika Journal of Science & Technology*, *25*(2), pp.575-586.

Alexander J., Downs C., Butler M., Woodborne S. and Symes C. (2018). Stable isotope analyses as a forensic tool to monitor illegally traded African grey parrots. *Animal Conservation*, *21*, pp.1-10.

Allendorf, F.W., England, P.R., Luikart, G., Ritchie, P.A. and Ryman, N. (2008). Genetic effects of harvest on wild animal populations. *Trends in Ecology & Evolution*, *23*(6), pp.327-337.

Altschul S.F., Gish W., Miller W., Myers E.W. and Lipman D.J. (1990). Basic local alignment search tool. *Journal of molecular biology*, *215*(3), pp.403-410.

Alvar J., Velez I.D., Bern C., Herrero M., Desjeux P., Cano J., Jannin J., den Boer M. and Team W.L.C. (2012). Leishmaniasis worldwide and global estimates of its incidence. *PloS One*, *7*(5), p.e35671.

Ambrose S.H. and DeNiro M.J. (1986). The isotopic ecology of East African mammals. *Oecologia*, *69*(3), pp.395-406.

Aranibar J., Anderson I., Epstein H., Feral C., Swap R., Ramontsho J. and Macko S. (2008). Nitrogen isotope composition of soils, C3 and C4 plants along land use gradients in southern Africa. *Journal of Arid Environments*, *72*(4), pp.326-337.

Arenas M., Pereira F., Oliveira M., Pinto N., Lopes A.M., Gomes V., Carracedo A. and Amorim A. (2017). Forensic genetics and genomics: Much more than just a human affair. *PLoS Genetics*, *13*(9), p.e1006960.

Ayliffe L., Lister A. and Chivas A. (1992). The preservation of glacial-interglacial climatic signatures in the oxygen isotopes of elephant skeletal phosphate. *Palaeogeography, Palaeoclimatology, Palaeoecology*, *99*(3-4), pp.179-191.

Babrauskas V. (2006). Temperatures in flames and fires. *Fire Science and Technology Inc*. Available online at: www.doctorfire.com/flametmp.html. Accessed on 16 October 2018.

Baer C.F., Miyamoto M.M. and Denver D.R. (2007). Mutation rate variation in multicellular eukaryotes: causes and consequences. *Nature Reviews Genetics*, *8*(8), p.619.

Bálint M., Bahram M., Eren A.M., Faust K., Fuhrman J.A., Lindahl B., O'hara R.B., Öpik M., Sogin M.L. and Unterseher M. (2016). Millions of reads, thousands of taxa: microbial community structure and associations analyzed via marker genes. *FEMS Microbiology Reviews*, *40*(5), pp.686-700.

Barnes C., Jennings S., Polunin N.V. and Lancaster J.E. (2008). The importance of quantifying inherent variability when interpreting stable isotope field data. *Oecologia*, *155*(2), pp.227-235.

Barnes C., Sweeting C.J., Jennings S., Barry J.T. and Polunin N.V. (2007). Effect of temperature and ration size on carbon and nitrogen stable isotope trophic fractionation. *Functional Ecology*, *21*(2), pp.356-362.

Bartelink E.J., Berg G.E., Beasley M.M. and Chesson L.A. (2014). Application of stable isotope forensics for predicting region of origin of human remains from past wars and conflicts. *Annals of Anthropological Practice*, *38*(1), pp.124-136.

Bastos A., Boshoff C., Keet D., Bengis R. and Thomson G. (2000). Natural transmission of foot-and-mouth disease virus between African buffalo (Syncerus caffer) and impala (Aepyceros melampus) in the Kruger National Park, South Africa. *Epidemiology & Infection*, *124*(3), pp.591-598.

Bastos A.D., Penrith M.-L., Cruciere C., Edrich J., Hutchings G., Roger F., Couacy-Hymann E. and Thomson G.R. (2003). Genotyping field strains of African swine fever virus by partial p72 gene characterisation. *Archives of Virology*, *148*(4), pp.693-706.

Bataille C.P. and Bowen G.J. (2012). Mapping 87Sr/86Sr variations in bedrock and water for large scale provenance studies. *Chemical Geology*, *304*, pp.39-52.

Baylis H. (1931). On a nematode parasite of pangolins. *Journal of Natural History*, *8*(44), pp.191-194.

Becker K., Harmsen D., Mellmann A., Meier C., Schumann P., Peters G. and Von Eiff C. (2004). Development and evaluation of a quality-controlled ribosomal sequence database for 16S ribosomal DNA-based identification of Staphylococcus species. *Journal of Clinical Microbiology*, *42*(11), pp.4988-4995.

Bell D., Roberton, S. and Hunter P.R. (2004). Animal origins of SARS coronavirus: possible links with the international trade in small carnivores. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *359*(1447), pp.1107-1114.

Ben-David M. and Flaherty E.A. (2012). Stable isotopes in mammalian research: a beginner's guide. *Journal of Mammalogy*, *93*(2), pp.312-328.

Beres S.B., Carroll R.K., Shea P.R., Sitkiewicz I., Martinez-Gutierrez J.C., Low D.E., McGeer A., Willey B.M., Green K. and Tyrrell G.J. (2010). Molecular complexity of successive bacterial epidemics deconvoluted by comparative pathogenomics. *Proceedings of the National Academy of Sciences*, *107*(9), pp.4371-4376.

Beveridge T.J. (1990). Mechanism of gram variability in select bacteria. *Journal of Bacteriology*, *172*(3), pp.1609-1620.

Binder W. and Mermel L. (1998). Leptospirosis in an urban setting: case report and review of an emerging infectious disease. *The Journal of Emergency Medicine*, 16(6), pp.851-856.

BirdLife International. (2000). *Threatened birds of the world*. Lynx Edicions, Barcelona and Birdlife International, Cambridge, p.852.

Blum J.D., Taliaferro E.H., Weisse M.T. and Holmes R.T. (2000). Changes in Sr/Ca, Ba/Ca and 87Sr/86Sr ratios between trophic levels in two forest ecosystems in the northeastern USA. *Biogeochemistry*, *49*(1), pp.87-101.

Boakye M.K., Pietersen D.W., Kotzé A., Dalton D.L. and Jansen R. (2014). Ethnomedicinal use of African pangolins by traditional medical practitioners in Sierra Leone. *Journal of Ethnobiology and Ethnomedicine*, *10*(1), p.76.

Bocherens, H. (2009). Neanderthal dietary habits: review of the isotopic evidence. In: J. Hublin, M. Richards (eds.), *The evolution of Hominin diets*. Springer, Dordrecht, pp.241-250.

Boom R., Sol C., Salimans M., Jansen C., Wertheim-van Dillen P. and Van der Noordaa J. (1990). Rapid and simple method for purification of nucleic acids. *Journal of Clinical Microbiology*, *28*(3), pp.495-503.

Booth A.L., Wooller M.J., Howe T. and Haubenstock N. (2010). Tracing geographic and temporal trafficking patterns for marijuana in Alaska using stable isotopes (C, N, O and H). *Forensic Science International*, *202*(1-3), pp.45-53.

Bowen G.J., Ehleringer J.R., Chesson L.A., Stange E. and Cerling T.E. (2007). Stable isotope ratios of tap water in the contiguous United States. *Water Resources Research*, *43*(3), pp.1-12.

Bowen G.J., Wassenaar L.I. and Hobson K.A. (2005). Global application of stable hydrogen and oxygen isotopes to wildlife forensics. *Oecologia*, *143*(3), pp.337-348.

Brand C., Windingstad R., Siegfried L., Duncan R. and Cook R. (1988). Avian morbidity and mortality from botulism, aspergillosis, and salmonellosis at Jamaica Bay Wildlife Refuge, New York, USA. *Colonial Waterbirds*, pp.284-292.

Britton K., McManus-Fry E., Nehlich O., Richards M., Ledger P.M. and Knecht R. (2016). Stable carbon, nitrogen and sulphur isotope analysis of permafrost preserved human hair from rescue excavations (2009, 2010) at the precontact site of Nunalleq, Alaska. *Journal of Archaeological Science: Reports*, *17*, pp.950-963.

Brook S.M., van Coeverden de Groot P., Scott C., Boag P., Long B., Ley R.E., Reischer G.H., Williams A.C., Mahood S.P., Hien T.M., Polet G., Cox N. and Hai B.T. (2012). Integrated and novel survey methods for rhinoceros populations confirm the extinction of Rhinoceros sondaicus annamiticus from Vietnam. *Biological Conservation*, *155*(2), pp.59-67.

Brum J.R., Ignacio-Espinoza J.C., Roux S., Doulcier G., Acinas S.G., Alberti A., Chaffron S., Cruaud C., De Vargas C. and Gasol J.M. (2015). Patterns and ecological drivers of ocean viral communities. *Science*, *348*(6237).

Bürkner P.C. (2016). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1), pp.1-28.

Buzgan T., Karahocagil M., Irmak H., Baran A., Karsen H., Evirgen O. and Akdeniz H. (2010). Clinical manifestations and complications in 1028 cases of brucellosis: a retrospective evaluation and review of the literature. *International Journal of Infectious Diseases*, *14*, pp.e469-e478.

Cai L., Ye L., Tong A.H.Y., Lok S. and Zhang T. (2013). Biased diversity metrics revealed by bacterial 16S pyrotags derived from different primer sets. *PloS One*, *8*(1), p.e53649.

Can I., Javan G.T., Pozhitkov A.E. and Noble P.A. (2014). Distinctive thanatomicrobiome signatures found in the blood and internal organs of humans. *Journal of Microbiological Methods*, *106*, pp.1-7.

Cardinale B.J., Duffy J.E., Gonzalez A., Hooper D.U., Perrings C., Venail P., Narwani A., Mace G.M., Tilman D., Wardle D.A. and Kinzig A.P. (2012). Biodiversity loss and its impact on humanity. *Nature*, *486*(7401), pp.59-67.

Carneiro J.J. and Donald D.E. (1977). Blood reservoir function of dog spleen, liver, and intestine. *American Journal of Physiology-Heart and Circulatory Physiology*, *232*(1), pp.67-72.

Carpenter B., Gelman A., Hoffman M., Lee D., Goodrich B., Betancourt M., Brubaker M.A., Guo J., Li P. and Riddell A. (2016). Stan: A probabilistic programming language. *Journal of Statistical Software*, *20*(2), pp.1-37.

Carter D.O., Metcalf J.L., Bibat A. and Knight R. (2015). Seasonal variation of postmortem microbial communities. *Forensic Science, Medicine, and Pathology*, *11*(2), pp.202-207.

Case R.J., Boucher Y., Dahllöf I., Holmström C., Doolittle W.F. and Kjelleberg S. (2007). Use of 16S rRNA and rpoB genes as molecular markers for microbial ecology studies. *Applied and Environmental Microbiology*, *73*(1), pp.278-288.

Cato E., George W. and Finegold S. (1986). Genus Clostridium Prazmowski 1880. In: P. Sneath, N. Mair, M. Sharpe, and J. Holt (eds.), *Bergey's manual of systematic bacteriology*. Williams & Wilkins Co., Baltimore, pp.1141-1200.

CDC (2003). Update: multistate outbreak of monkeypox—Illinois, Indiana, Kansas, Missouri, Ohio, and Wisconsin, 2003. *The Morbidity and Mortality Weekly Report*, *52*, pp.642-646.

Ceballos G., García A. and Ehrlich P.R. (2010). The sixth extinction crisis: loss of animal populations and species. *Journal of Cosmology*, *8*(1), pp.1821-1831.

Cerling T.E., Andanje S.A., Gakuya F., Kariuki J.M., Kariuki L., Kingoo J.W., Khayale C., Lekolool I., Macharia A.N. and Anderson C.R. (2018). Stable isotope ecology of black rhinos (Diceros bicornis) in Kenya. *Oecologia*, *187*(4), pp.1095-1105.

Cerling T.E., Barnette J.E., Bowen G.J., Chesson L.A., Ehleringer J.R., Remien C.H., Shea P., Tipple B.J. and West J.B. (2016). Forensic stable isotope biogeochemistry. *Annual Review of Earth and Planetary Sciences*, *44*, pp.175-206.

Cerling T.E. and Harris J.M. (1999). Carbon isotope fractionation between diet and bioapatite in ungulate mammals and implications for ecological and paleoecological studies. *Oecologia*, *120*(3), pp.347-363.

Cerling T.E., Harris J.M. and Leakey M.G. (1999). Browsing and grazing in elephants: the isotope record of modern and fossil proboscideans. *Oecologia*, *120*(3), pp.364-374.

Cerling T.E., Omondi P. and Macharia A.N. (2007). Diets of Kenyan elephants from stable isotopes and the origin of confiscated ivory in Kenya. *African Journal of Ecology*, *45*(4), pp.614-623.

Cerling T.E., Wittemyer G., Ehleringer J.R., Remien C.H. and Douglas-Hamilton I. (2009). History of animals using isotope records (HAIR): a 6-year dietary history of one family of African elephants. *Proceedings of the National Academy of Sciences*, *106*(20), pp.8093-8100.

Chakrabarty K., Wu W., Booth J.L., Duggan E.S., Nagle N.N., Coggeshall K.M. and Metcalf J.P. (2007). Human lung innate immune response to bacillus anthracis spore infection. *Infection and Immunity*, *75*(8), pp.3729-3738.

Challender D and Hywood L. (2012). African pangolins under increased pressure from poaching and intercontinental trade. *TRAFFIC Bulletin*, *24*(2), pp.53-55.

Challender D and Waterman C. (2017). Implementation of CITES decisions 17.239 b) and 17.240 on pangolins (*Manis* spp.). Prepared by IUCN for the CITES Secretariat.

Challender D.W., Harrop S.R. and MacMillan D.C. (2015). Towards informed and multi-faceted wildlife trade interventions. *Global Ecology and Conservation*, *3*, pp.129-148.

Challender D.W., Waterman C. and Baillie J.E. (2014). Scaling up pangolin conservation. *IUCN SSC Pangolin Specialist Group Conservation Action Plan. Zoological Society of London, London, UK*.

Chesson L.A., Barnette J.E., Bowen G.J., Brooks J.R., Casale J.F., Cerling T.E., Cook C.S., Douthitt C.B., Howa J.D. and Hurley J.M. (2018a). Applying the principles of isotope analysis in plant and animal ecology to forensic science in the Americas. *Oecologia*, *187*(4), pp.1077-1094.

Chesson L.A., Barnette J.E., Bowen G.J., Brooks J.R., Casale J.F., Cerling T.E., Cook C.S., Douthitt C.B., Howa J.D. and Hurley J.M. (2018b). Applying the principles of isotope analysis in plant and animal ecology to forensic science in the Americas. *Oecologia*, pp.1-18.

Choo S.W., Rayko M., Tan T.K., Hari R., Komissarov A., Wee W.Y., Yurchenko A.A., Kliver S., Tamazian G. and Antunes A. (2016). Pangolin genomes and the evolution of mammalian scales and immunity. *Genome research*, p.36.

CITES (2017a). What is CITES? Available online at: https://cites.org/eng/disc/what.php. Accessed on 18 April 2017.

CITES (2017b). Amendments to the conservation of and trade in pangolins. Seventeenth meeting of the Conference of the Parties Johannesburg, South Africa, 24 September – 5 October 2016. CITES Secretariat.

Clayton R.A., Sutton G., Hinkle Jr P.S., Bult C. and Fields C. (1995). Intraspecific variation in small-subunit rRNA sequences in GenBank: why single sequences may not adequately represent prokaryotic taxa. *International Journal of Systematic and Evolutionary Microbiology*, *45*(3), pp.595-599.

Cleaveland S., Appel M., Chalmers W., Chillingworth C., Kaare M. and Dye C. (2000). Serological and demographic evidence for domestic dogs as a source of canine distemper virus infection for Serengeti wildlife. *Veterinary Microbiology*, *72*(3-4), pp.217-227.

Clegg S.M., Kelly J.F., Kimura M. and Smith T.B. (2003). Combining genetic markers and stable isotopes to reveal population connectivity and migration patterns in a Neotropical migrant, Wilson's Warbler (Wilsonia pusilla). *Molecular Ecology*, *12*(4), pp.819-830.

Cloud J., Neal H., Rosenberry R., Turenne C., Jama M., Hillyard D. and Carroll K. (2002). Identification of Mycobacterium spp. by using a commercial 16S ribosomal DNA sequencing kit and additional sequencing libraries. *Journal of Clinical Microbiology*, *40*(2), pp.400-406.

Codron J., Codron D., Sponheimer M., Kirkman K., Duffy K.J., Raubenheimer E.J., Mélice J.L., Grant R., Clauss M. and Lee-Thorp J.A. (2012). Stable isotope series from elephant ivory reveal lifetime histories of a true dietary generalist. *Proceedings of the Royal Society B: Biological Sciences*, *279*(1737), pp.2433-2441.

Cohen F.P., Pimentel T., Valenti W.C. and Calado R. (2018). First insights on the bacterial fingerprints of live seahorse skin mucus and its relevance for traceability. *Aquaculture*, *492*, pp.259-264.

Collins M.D., Lawson P.A., Willems A., Cordoba J.J., Fernandez-Garayzabal J., Garcia P., Cai J., Hippe H. and Farrow J.A.E. (1994). The phylogeny of the genus Clostridium: proposal of five new genera and eleven new species combinations. *International Journal of Systematic and Evolutionary Microbiology*, *44*(4), pp.812-826.

Concheri G., Bertoldi D., Polone E., Otto S., Larcher R. and Squartini A. (2011). Chemical elemental distribution and soil DNA fingerprints provide the critical evidence in murder case investigation. *PloS One*, *6*(6), p.e20222.

Coonan T.J., Bakker V., Hudgens B., Boser C.L., Garcelon D.K. and Morrison S.A. (2014). On the fast track to recovery: island foxes on the northern Channel Islands. *Monographs of the Western North American Naturalist*, *7*(1), pp.373-381.

Cooney R., Kasterine A., MacMillan D., Milledge S.A., Nossal K., Roe D. and John't Sas-Rolfes M. (2015). *The trade in wildlife: a framework to improve biodiversity and livelihood outcomes*. International Trade Centre, Geneva.

Costello M.J., May R.M. and Stork N.E. (2013). Can we name Earth's species before they go extinct? *Science*, *339*(6118), pp.413-416.

Coulton D.W., Clark R.G., Hobson K.A., Wassenaar L.I. and Hebert C.E. (2009). Temporal sources of deuterium (δD) variability in waterfowl feathers across a prairie-to-boreal gradient. *The Condor*, *111*(2), pp.255-265.

Coutu A.N., Lee-Thorp J., Collins M.J. and Lane P.J. (2016). Mapping the elephants of the 19th Century East African Ivory trade with a multi-isotope approach. *PloS One*, *11*(10), p.e0163606.

Cox G. and Chitty R. (1982). Some stochastic properties of fire plumes. *Fire and Materials*, *6*(3-4), pp.127-134.

Craine J.M., Brookshire E., Cramer M.D., Hasselquist N.J., Koba K., Marin-Spiotta E. and Wang L. (2015). Ecological interpretations of nitrogen isotope ratios of terrestrial plants and soils. *Plant and Soil*, *396*(1-2), pp.1-26.

Dahllöf I., Baillie H. and Kjelleberg S. (2000). rpoB-based microbial community analysis avoids limitations inherent in 16S rRNA gene intraspecies heterogeneity. *Applied and Environmental Microbiology*, *66*(8), pp.3376-3380.

Davies R.G., Hernández L.M., Eggleton P., Didham R.K., Fagan L.L. and Winchester N.N. (2003). Environmental and spatial influences upon species composition of a termite assemblage across neotropical forest islands. *Journal of Tropical Ecology*, *19*(5), pp.509-524.

Dawson T.E., Mambelli S., Plamboeck A.H., Templer P.H. and Tu K.P. (2002). Stable isotopes in plant ecology. *Annual Review of Ecology and Systematics*, *33*(1), pp.507-559.

Daszak P., Cunningham A. and Hyatt A. (2000). Emerging infectious diseases of wildlife-threats to biodiversity and human health. *Science*, *287*(5452), pp.443-449.

DeNiro M.J. and Epstein S. (1978). Influence of diet on the distribution of carbon isotopes in animals. *Geochimica et Cosmochimica Acta*, *42*(5), pp.495-506.

DeNiro M.J. and Epstein S. (1981). Influence of diet on the distribution of nitrogen isotopes in animals. *Geochimica et Cosmochimica Acta*, *45*(3), pp.341-351.

DeNiro M.J., Schoeninger M.J. and Hastorf C.A. (1985). Effect of heating on the stable carbon and nitrogen isotope ratios of bone collagen. *Journal of Archaeological Science*, *12*(1), pp.1-7.

DeSantis T.Z., Hugenholtz P., Larsen N., Rojas M., Brodie E.L., Keller K., Huber T., Dalevi D., Hu P. and Andersen G.L. (2006). Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Applied and Environmental Microbiology*, *72*(7), pp.5069-5072.

Diefendorf A.F., Mueller K.E., Wing S.L., Koch P.L. and Freeman K.H. (2010). Global patterns in leaf 13C discrimination and implications for studies of past and future climate. *Proceedings of the National Academy of Sciences*, *107*(13), pp.5738-5743.

Di Nardo A., Knowles N. and Paton D. (2011). Combining livestock trade patterns with phylogenetics to help understand the spread of foot and mouth disease in sub-Saharan Africa, the Middle East and Southeast Asia. *Revue Scientifique et Technique-OIE*, *30*(1), p.63.

Drake J.W., Charlesworth B., Charlesworth D. and Crow J.F. (1998). Rates of spontaneous mutation. *Genetics*, *148*(4), pp.1667-1686.

Duarte-Quiroga A. and Estrada A. (2003). Primates as pets in Mexico City: an assessment of the species involved, source of origin, and general aspects of treatment. *American Journal of Primatology*, *61*(2), pp.53-60.

Duckworth J.W., Batters G., Belant J.L., Bennett E.L., Brunner J., Burton J., Challender D.W.S., Cowling V., Duplaix N., Harris J.D. and Hedges S. (2012). Why South-East Asia should be the world's priority for averting imminent species extinctions, and a call to join a developing cross-institutional programme to tackle this urgent issue. *S.A.P.I.EN.S. Surveys and Perspectives Integrating Environment and Society*, *5*(2), pp.77-95.

Duffy R. (2016). *EU trade policy and the wildlife trade*. The European Parliament's Committee on International Trade, p.45.

Duncan R. and Jensen W. (1976). A relationship between avian carcasses and living invertebrates in the epizootiology of avian botulism. *Journal of Wildlife Diseases*, *12*(1), pp.116-126.

Dunn P.J., Salouros H., Carter J.F. and Doyle S.P. (2017). Forensic application of stable isotope delta values: Proposed minimum requirements for method validation. *Rapid Communications in Mass Spectrometry*, *31*(17), pp.1476-1480.

du Toit, Z. (2014). Population genetic structure of the ground pangolin based on mitochondrial genomes. PhD thesis, University of the Free State, Bloemfontein, South Africa, p.149.

Eklund M.W., Poysky F.T., Meyers J.A. and Pelroy G.A. (1974). Interspecies conversion of Clostridium botulinum type C to Clostridium novyi type A by bacteriophage. *Science*, *186*(4162), pp.456-458.

Edwards E.J. and Ogburn R.M. (2012). Angiosperm responses to a low-CO2 world: CAM and C4 photosynthesis as parallel evolutionary trajectories. *International Journal of Plant Sciences*, *173*(6), pp.724-733.

Ehleringer J.R., Bowen G.J., Chesson L.A., West A.G., Podlesak D.W. and Cerling T.E. (2008). Hydrogen and oxygen isotope ratios in human hair are related to geography. *Proceedings of the National Academy of Sciences*, *105*(8), pp.2788-2793.

Ellis R., Vogel J.C. and Fuls A. (1980). Photosynthetic pathways and the geographical distribution of grasses in South West Africa/Namibia. *South African Journal of Science 76*(7), 307-314.

Emslie R. (2011). Diceros bicornis ssp. longipes. The IUCN Red List of Threatened Species 2011: e.T39319A10198340. Available online at: http://dx.doi.org/10.2305/IUCN.UK.2011-2.RLTS.T39319A10198340.en. Downloaded on 11 May 2017.

Engler M. (2008). The value of international wildlife trade. *TRAFFIC Bulletin*, *22*(1), pp.4-5.

Evans R.D. (2001). Physiological mechanisms influencing plant nitrogen isotope composition. *Trends in Plant Science*, *6*(3), pp.121-126.

Farquhar G.D., Ehleringer J.R. and Hubick K.T. (1989). Carbon isotope discrimination and photosynthesis. *Annual Review of Plant Biology*, *40*(1), pp.503-537.

Faust K., Lahti L., Gonze D., de Vos W.M. and Raes J. (2015). Metagenomics meets time series analysis: unraveling microbial community dynamics. *Current Opinion in Microbiology*, *25*, pp.56-66.

Fernandes G.A. and Caparroz R. (2013). DNA sequence analysis to guide the release of blue-and-yellow macaws (Ara ararauna, Psittaciformes, Aves) from the illegal trade back into the wild. *Molecular Biology Reports*, *40*(3), pp.2757-2762.

Fèvre E.M., Bronsvoort B.M.d.C., Hamilton K.A. and Cleaveland S. (2006). Animal movements and the spread of infectious diseases. *Trends in Microbiology*, *14*(3), pp.125-131.

Fick S.E. and Hijmans R.J. (2017) Worldclim 2: New 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, *37*(12), pp.302-4315.

Fierer N. and Jackson R.B. (2006). The diversity and biogeography of soil bacterial communities. *Proceedings of the National Academy of Sciences*, *103*(3), pp.626-631.

Fierer N., Leff J.W., Adams B.J., Nielsen U.N., Bates S.T., Lauber C.L., Owens S., Gilbert J.A., Wall D.H. and Caporaso J.G. (2012). Cross-biome metagenomic analyses of soil microbial communities and their functional attributes. *Proceedings of the National Academy of Sciences*, *109*(52), pp.21390-21395.

Flint H.J., Duncan S.H., Scott K.P. and Louis P. (2007). Interactions and competition within the microbial community of the human colon: links between diet and health. *Environmental Microbiology*, *9*(5), pp.1101-1111.

Food and Agriculture Organization (2004). Somalia: towards a livestock sector strategy – final report (Report No. 04/001 IC–SOM). FAO/World Bank/European Union. Available online at: http://siteresources.worldbank.org/SOMALIAEXTN/Resources/so_LS_final_rpt.pdf. Downloaded on 15 May 2017.

Foran, D.R. (2006). Relative degradation of nuclear and mitochondrial DNA: an experimental approach. *Journal of Forensic Sciences*, *51*(4), pp.766-770.

Fouhy F., Clooney A.G., Stanton C., Claesson M.J. and Cotter P.D. (2016). 16S rRNA gene sequencing of mock microbial populations-impact of DNA extraction method, primer choice and sequencing platform. *BMC Microbiology*, *16*(1), p.123.

Fuller B.T., Fuller J.L., Harris D.A. and Hedges R.E. (2006). Detection of breastfeeding and weaning in modern human infants with carbon and nitrogen stable isotope ratios. *American Journal of Physical Anthropology: The Official Publication of the American Association of Physical Anthropologists*, *129*(2), pp.279-293.

Fuller B.T., Fuller J.L., Sage N.E., Harris D.A., O'Connell T.C. and Hedges R.E. (2004). Nitrogen balance and δ15N: why you're not what you eat during pregnancy. *Rapid Communications in Mass Spectrometry*, *18*(23), pp.2889-2896.

Fuller B.T., Fuller J.L., Sage N.E., Harris D.A., O'Connell T.C. and Hedges R.E. (2005). Nitrogen balance and δ15N: why you're not what you eat during nutritional stress. *Rapid Communications in Mass Spectrometry: An International Journal Devoted to the Rapid Dissemination of Up-to-the-Minute Research in Mass Spectrometry*, *19*(18), pp.2497-2506.

Galperin M.Y. (2013). Genome diversity of spore-forming Firmicutes. *Microbiology Spectrum*, *1*(2), pp.TBS-0015.

Gao X. (2012). *Color illustrations of Chinese traditional medicine*. Traditional Chinese Medicine Classics Press, Beijing, China.

García A., Ayuso D., Benítez J.M., García W.L., Martínez R. and Sánchez S. (2009). Clostridium novyi infection causing sow mortality in an Iberian pig herd raised in an outdoor rearing system in Spain. *Journal of Swine Health and Production*, *17*(5), pp.264-268.

Gardy J.L., Johnston J.C., Sui S.J.H., Cook V.J., Shah L., Brodkin E., Rempel S., Moore R., Zhao Y. and Holt R. (2011). Whole-genome sequencing and social-network analysis of a tuberculosis outbreak. *New England Journal of Medicine*, *364*(8), pp.730-739.

Gat J.R. (1996). Oxygen and hydrogen isotopes in the hydrologic cycle. *Annual Review of Earth and Planetary Sciences*, *24*(1), pp.225-262.

Gaubert P., Antunes A., Meng H., Miao L., Peigné S., Justy F., Njiokou F., Dufour S., Danquah E., Alahakoon J. and Verheyen E. (2017). The complete phylogeny of pangolins: scaling up resources for the molecular tracing of the most trafficked mammals on Earth. *Journal of Heredity*, *109*(4), pp.347-359.

Gaubert P., Njiokou F., Ngua G., Afiademanyo K., Dufour S., Malekani J., Bi S.G., Tougard C., Olayemi A., Danquah E. and Djagoun C.A. (2016). Phylogeography of the heavily poached African common pangolin (Pholidota, Manis tricuspis) reveals six cryptic lineages as traceable signatures of Pleistocene diversification. *Molecular Ecology*, *25*(23), pp.5975-5993.

Gauch Jr H.G. and Whittaker R.H. (1981). Hierarchical classification of community data. *The Journal of Ecology*, pp.537-557.

Gaudin T.J., Emry R.J. and Wible J.R. (2009). The phylogeny of living and extinct pangolins (Mammalia, Pholidota) and associated taxa: a morphology based analysis. *Journal of Mammalian Evolution*, *16*(4), p.235.

Gentile N., Siegwolf R.T., Esseiva P., Doyle S., Zollinger K. and Delemont O. (2015). Isotope ratio mass spectrometry as a tool for source inference in forensic science: A critical review. *Forensic Science International*, *251*, pp.139-158.

Gilbert J.A., Quinn R.A., Debelius J., Xu Z.Z., Morton J., Garg N., Jansson J.K., Dorrestein P.C. and Knight R. (2016). Microbiome-wide association studies link dynamic microbial consortia to disease. *Nature*, *535*(7610), pp.94-103.

Gilmore S., Peakall R. and Robertson J. (2007). Organelle DNA haplotypes reflect crop-use characteristics and geographic origins of Cannabis sativa. *Forensic Science International*, *172*(2), pp.179-190.

Godoy S.N. and Matushima E.R. (2010). A survey of diseases in passeriform birds obtained from illegal wildlife trade in São Paulo City, Brazil. *Journal of Avian Medicine and Surgery*, *43*(3), pp.199-209.

Gonzalvez A., Armenta S. and De La Guardia M. (2009). Trace-element composition and stable-isotope ratio for discrimination of foods with protected designation of origin. *TrAC Trends in Analytical Chemistry*, *28*(11), pp.1295-1311.

Goodrich J., Lynam A., Miquelle D., Wibisono H., Kawanishi K., Pattanavibool A., Htun S., Tempa T., Karki J., Jhala Y. and Karanth U. (2015). Panthera tigris. The IUCN Red List of Threatened Species 2015: e.T15955A50659951. Available online at: http://dx.doi.org/10.2305/IUCN.UK.2015- 2.RLTS.T15955A50659951.en. Downloaded on 28 March 2017.

Graham-Rowe D. (2011). Biodiversity: endangered and in demand. *Nature*, *480*(7378), pp.S101-S103.

Grayson J. and Mendel D. (1957). The role of the spleen and the hepatic artery in the regulation of liver blood flow. *The Journal of Physiology*, *136*(1), pp.60-79.

Guindon S., Dufayard J.F., Lefort V., Anisimova M., Hordijk W. and Gascuel O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic Biology*, *59*(3), pp.307-321.

Gul H. and Erdem H. (2014). Brucellosis (Brucella species). In: J. Bennett, R. Dolin, and M. Blaser (eds.), *Principles and practice of infectious diseases* (8th edition). Elsevier Health Sciences, pp. 2584-2589.

Guttenberg G., Papatheodorou P., Genisyuerek S., Lü W., Jank T., Einsle O. and Aktories K. (2011). Inositol hexakisphosphate-dependent processing of Clostridium sordellii lethal toxin and Clostridium novyi α-toxin. *Journal of Biological Chemistry*, *286*(17), pp.14779-14786.

Hacker J. and Carniel E. (2001). Ecological fitness, genomic islands and bacterial pathogenicity: a Darwinian view of the evolution of microbes. *EMBO Reports*, *2*(5), pp.376-381.

Hall-Martin A., Van Der Merwe N., Lee-Thorp J., Armstrong R., Mehl C., Struben S. and Tykot R. (1993). Determination of species and geographic origin of rhino horn by isotopic analysis and its possible application to trade control. *Rhinoceros Biology and Conservation. Zoological Society of San Diego, San Diego*, pp.122-135.

Hall M.D., Knowles N.J., Wadsworth J., Rambaut A. and Woolhouse M.E. (2013). Reconstructing geographical movements and host species transitions of foot-and-mouth disease virus serotype SAT 2. *MBio*, *4*(5), pp.e00591-00513.

Hampton-Marcell J.T., Lopez J.V. and Gilbert J.A. (2017). The human microbiome: an emerging tool in forensics. *Microbial Biotechnology*, *10*(2), pp.228-230.

Harbeck M., Schleuder R., Schneider J., Wiechmann I., Schmahl W.W. and Grupe G. (2011). Research potential and limitations of trace analyses of cremated remains. *Forensic Science International*, *204*(1-3), pp.191-200.

Harmsen D., Rothgänger J., Frosch M. and Albert J. (2002). RIDOM: ribosomal differentiation of medical micro-organisms database. *Nucleic Acids Research*, *30*(1), pp.416-417.

Harper C.K., Vermeulen G.J., Clarke A.B., Jacobus I. and Guthrie A.J. (2013). Extraction of nuclear DNA from rhinoceros horn and characterization of DNA profiling systems for white (Ceratotherium simum) and black (Diceros bicornis) rhinoceros. *Forensic Science International: Genetics*, *7*(4), pp.428-433.

Harris S.R., Feil E.J., Holden M.T., Quail M.A., Nickerson E.K., Chantratita N., Gardete S., Tavares A., Day N. and Lindsay J.A. (2010). Evolution of MRSA during hospital transmission and intercontinental spread. *Science*, *327*(5964), pp.469-474.

Hartman G. (2011). Are elevated δ15N values in herbivores in hot and arid environments caused by diet or animal physiology? *Functional Ecology*, *25*(1), pp.122-131.

Hatheway C. (1989). Bacterial sources of clostridial neurotoxins. In: L. Simpson (ed.), *Botulinum neurotoxin and tetanus toxin*, Academic Press, London, pp.3-24.

Hatheway C.L. (1990). Toxigenic clostridia. *Clinical Microbiology Reviews*, *3*(1), pp.66-98.

He Z., Gentry T.J., Schadt C.W., Wu L., Liebich J., Chong S.C., Huang Z., Wu W., Gu B. and Jardine P. (2007). GeoChip: a comprehensive microarray for investigating biogeochemical, ecological and environmental processes. *The ISME Journal*, *1*(1), p.67.

Heath M. and Coulson I. (1997). Home range size and distribution in a wild population of Cape pangolins, Manis temminckii, in north-west Zimbabwe. *African Journal of Ecology*, *35*(2), pp.94-109.

Heath M.E. (1992). Manis temminckii. *Mammalian Species*(415), pp.1-5.

Heaton T.H., Vogel J.C., von La Chevallerie G. and Collett G. (1986). Climatic influence on the isotopic composition of bone nitrogen. *Nature*, *322*(6082), p.822.

Hedges R.E., Stevens R.E. and Richards M.P. (2004). Bone as a stable isotope archive for local climatic information. *Quaternary Science Reviews*, *23*(7-8), pp.959-965.

Heighton S. (2016). Scaly bacteria: the quest to identify pangolin pathogens. BSc (Hons) thesis, University of Pretoria, Pretoria, South Africa, p.54

Heinrich S., Wittmann T.A., Prowse T.A., Ross J.V., Delean S., Shepherd C.R. and Cassey P. (2016). Where did all the pangolins go? International CITES trade in pangolin species. *Global Ecology and Conservation*, *8*, pp.241-253.

Heinrich S., Wittman T.A., Ross J.V., Shepherd C.R., Challender D.W. and Cassey P. (2017). *The global trafficking of pangolins: A comprehensive summary of seizures and trafficking routes from 2010–2015*. TRAFFIC, Southeast Asia, Regional Office, Selangor, Malaysia, p.50.

Hesslein R.H., Hallard K. and Ramlal P. (1993). Replacement of sulfur, carbon, and nitrogen in tissue of growing broad whitefish (Coregonus nasus) in response to a change in diet traced by δ34S, δ13C, and δ15N. *Canadian Journal of Fisheries and Aquatic Sciences*, *50*(10), pp.2071-2076.

Hewitt K.M., Gerba C.P., Maxwell S.L. and Kelley S.T. (2012). Office space bacterial abundance and diversity in three metropolitan areas. *PloS One*, *7*(5), p.e37849.

Higgins R.M., Danilowicz B.S., Balbuena J.A., Daníelsdóttir A.K., Geffen A.J., Meijer W.G., Modin J., Montero F.E., Pampoulie C. and Perdiguero-Alonso D. (2010). Multi-disciplinary fingerprints reveal the harvest location of cod Gadus morhua in the northeast Atlantic. *Marine Ecology Progress Series*, *404*, pp.197-206.

Hijmans R.J. (2016). Raster: Geographic Data Analysis and Modeling. R package version 2.5-8. https://CRAN.R-project.org/package=raste.

Hilton-Taylor C., Pollock C.M., Chanson J.S., Butchart S.H., Oldfield T.E. and Katariya V. (2009). State of the world's species. *Wildlife in a Changing World–An Analysis of the 2008 IUCN Red List of Threatened Species*, pp.15-41.

Hitch S. (1998). Losing the elephant wars: CITES and the ivory ban. *Georgia Journal of International and Comparative Law*, *27*, pp.167-197.

Hobson K.A. (1999). Tracing origins and migration of wildlife using stable isotopes: a review. *Oecologia*, *120*(3), pp.314-326.

Hobson K.A. (2007). Isotopic tracking of migrant wildlife. In: R. Michener and K. Lajtha (eds.), *Stable Isotopes in Ecology and Environmental Science*. Blackwell Publishing, Victoria, pp.155-175.

Hobson K.A., Alisauskas R.T. and Clark R.G. (1993). Stable-nitrogen isotope enrichment in avian tissues due to fasting and nutritional stress: implications for isotopic analyses of diet. *Condor*, *95*(2), pp.388-394.

Hobson K.A. and Clark R.G. (1992). Assessing avian diets using stable isotopes I: turnover of 13C in tissues. *Condor*, *94*(1), pp.181-188.

Hobson K.A. and Quirk T.W. (2014). Effect of age and ration on diet-tissue isotopic (Δ13C, Δ15N) discrimination in striped skunks (Mephitis mephitis). *Isotopes in Environmental and Health Studies*, *50*(3), pp.300-306.

Hobson K.A. and Wassenaar L.I. (1996). Linking breeding and wintering grounds of neotropical migrant songbirds using stable hydrogen isotopic analysis of feathers. *Oecologia*, *109*(1), pp.142-148.

Hobson K.A., Wassenaar L.I., Milá B., Lovette I., Dingle C. and Smith T.B. (2003). Stable isotopes as indicators of altitudinal distributions and movements in an Ecuadorean hummingbird community. *Oecologia*, *136*(2), pp.302-308.

Hoff K.J. (2009). The effect of sequencing errors on metagenomic gene prediction. *BMC Genomics*, *10*(1), p.520.

Hothorn T., Bretz F. and Westfall P. (2008). Simultaneous inference in general parametric models. *Biometrical Journal*, *50*(3), pp.346-363.

Howard G.T., Duos B. and Watson-Horzelski E.J. (2010). Characterization of the soil microbial community associated with the decomposition of a swine carcass. *International Biodeterioration & Biodegradation*, *64*(4), pp.300-304.

Howland M.R., Corr L.T., Young S.M., Jones V., Jim S., Van Der Merwe N.J., Mitchell A.D. and Evershed R.P. (2003). Expression of the dietary isotope signal in the compound-specific δ13C values of pig bone lipids and amino acids. *International Journal of Osteoarchaeology*, *13*(1-2), pp.54-65.

Hua L., Gong S., Wang F., Li W., Ge Y., Li X. and Hou F. (2015). Captive breeding of pangolins: current status, problems and future prospects. *ZooKeys*, *507*, p.99.

Huffman J.E. and Wallace J.R. (2012). *Wildlife forensics: methods and applications*. John Wiley & Sons, Sussex, UK.

Húngaro H., Caturla M., Horita C., Furtado M. and Sant'Ana A.S. (2016). Blown pack spoilage in vacuum-packaged meat: A review on clostridia as causative agents, sources, detection methods, contributing factors and mitigation strategies. *Trends in Food Science & Technology*, *52*, pp.123-138.

Huntley B. (1982). Southern African savannas. In: B. Huntley and B. Walker (eds.), *Ecology of Tropical Savannas*. Springer, New York, pp.101-119.

Hurtado A., Clewley J.P., Linton D., Owen R.J. and Stanley J. (1997). Sequence similarities between large subunit ribosomal RNA gene intervening sequences from different Helicobacter species. *Gene*, *194*(1), pp.69-75.

Huse S.M., Ye Y., Zhou Y. and Fodor A.A. (2012). A core human microbiome as viewed through 16S rRNA sequence clusters. *PloS One*, *7*(6), p.e34242.

Hussein H., El-Amir Y., Aamer A. and Elghaffar S. (2013). Bacillary hemoglobinuria in dairy cows: clinical, hematological, biochemical, and pathological alterations. *Comparative Clinical Pathology*, *22*(6), pp.1137-1143.

Hutton J. and Webb G. (2002). Legal trade snaps back: using the experience of crocodilians to draw lessons on regulation of the wildlife trade. *Crocodiles. Proceedings of the 16th Working Meeting of the CSG-IUCN-The World Conservation Union. Gland, Switzerland*, p.5.

Hyde E.R., Haarmann D.P., Lynne A.M., Bucheli S.R. and Petrosino J.F. (2013). The living dead: bacterial community structure of a cadaver at the onset and end of the bloat stage of decomposition. *PloS One*, *8*(10), p.e77733.

Hyde E.R., Haarmann D.P., Petrosino J.F., Lynne A.M. and Bucheli S.R. (2015). Initial insights into bacterial succession during human decomposition. *International Journal of Legal Medicine*, *129*(3), pp.661-671.

Ihrmark K., Bödeker I., Cruz-Martinez K., Friberg H., Kubartova A., Schenck J., Strid Y., Stenlid J., Brandström-Durling M. and Clemmensen K.E. (2012). New primers to amplify the fungal ITS2 region–evaluation by 454-sequencing of artificial and natural communities. *FEMS microbiology ecology*, *82*(3), pp.666-677.

Ishibashi H., Takeuchi T., Whyte I. and Koike H. (1999). δ^ 15N and δ^ 13C measurements from the African elephant, Loxodonta africana, used for ivory sourcing. *Bulletin of the Graduate School of Social and Cultural Studies, Kyushu University*, *5*, pp.1-8.

IUCN (2000). *Trade measures in multilateral environmental agreements*. A report by IUCN—the world conservation union on the effectiveness of trade measures contained in the convention on international trade in endangered species of wild fauna and flora (CITES). Prepared for the Economics, Trade and Environment Unit, United Nations Environment Programme. Available online at: https://cites.org/common/prog/economics/iucn-trademeasuresinCITES.pdf. Downloaded on 19 April 2017.

Jacobsen N., Newbery R., De Wet M., Viljoen P. and Pietersen E. (1991). A contribution of the ecology of the Steppe pangolin Manis temminckii in the Transvaal. *Zeitschrift für Säugetierkunde*, *56*(2), pp.94-100.

Jacquet J.L. and Pauly D. (2008). Trade secrets: renaming and mislabeling of seafood. *Marine Policy*, *32*(3), pp.309-318.

Jelić D., Toth T. and Verbanac D. (2003). Macromolecular databases–a background of bioinformatics. *Food Technology and Biotechnology*, *41*(3), pp.269-286.

Jones K., Patel N., Levy M., Storeygard A., Balk D., Gittleman J. and Daszak P. (2008). Global trends in emerging infectious diseases. *Nature*, 451, pp.990-993.

Juniper T. and Yamashita C. (1990). The conservation of Spix's macaw. *Oryx*, *24*(04), pp.224-228.

Jyothsna T.S., Tushar L., Sasikala C. and Ramana C.V. (2016). Paraclostridium benzoelyticum gen. nov., sp. nov., isolated from marine sediment and reclassification of Clostridium bifermentans as Paraclostridium bifermentans comb. nov. Proposal of a new genus Paeniclostridium gen. nov. to accommodate Clostridium sordellii and Clostridium ghonii. *International Journal of Systematic and Evolutionary Microbiology*, *66*(3), pp.1268-1274.

Karesh W., Dobson A., Lloyd-Smith J., Lubroth J., Dixon M., Bennett M., Aldrich S., Harrington T., Formenty P., Loh E., Machalaba C., Thomas M. and Heymann D. (2012). Ecology of zoonoses: natural and unnatural histories. *The Lancet*, 380(9857), pp.1936-1945.

Katuwal H.B., Neupane K.R., Adhikari D., Sharma M. and Thapa S. (2015). Pangolins in eastern Nepal: Trade and ethno-medicinal importance. *Journal of Threatened Taxa*, *7*(9), pp.7563-7567.

Keim P., Van Ert M.N., Pearson T., Vogler A.J., Huynh L.Y. and Wagner D.M. (2004). Anthrax molecular epidemiology and forensics: using the appropriate marker for different evolutionary scales. *Infection, Genetics and Evolution*, *4*(3), pp.205-213.

Kellerman T., Coetzer J. and Naudé T. (1988). *Plant poisonings and mycotoxicoses of livestock in Southern Africa*. Oxford University Press, Cape Town.

Kelly A., Thompson R. and Newton J. (2008). Stable hydrogen isotope analysis as a method to identify illegally trapped songbirds. *Science & Justice*, *48*(2), pp.67-70.

Kelly J.F. (2000). Stable isotopes of carbon and nitrogen in the study of avian and mammalian trophic ecology. *Canadian Journal of Zoology*, *78*(1), pp.1-27.

Kelly L.J. and Martínez del Rio C. (2010). The fate of carbon in growing fish: an experimental study of isotopic routing. *Physiological and Biochemical Zoology*, *83*(3), pp.473-480.

Kilpatrick A. and Randolph S. (2012). Drivers, dynamics, and control of emerging vector-borne zoonotic diseases. *The Lancet*, 380(9857), pp.1946-1955.

Kingdon J. (1971). *East African mammals–An atlas of evolution in Africa Vol. 1.(Primates, Hyraxes, Pangolins, Protoungulates, Sirenians)*. Academic Press, London, pp.365-369.

Kircher M. and Kelso J. (2010). High-throughput DNA sequencing–concepts and limitations. *Bioessays*, *32*(6), pp.524-536.

Klindworth A., Pruesse E., Schweer T., Peplies J., Quast C., Horn M. and Glöckner F.O. (2013). Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Research*, *41*(1), pp.e1-e1.

Koch P.L. (2007). Isotopic study of the biology of modern and fossil vertebrates. In: R. Michener and K. Lajtha (eds.), *Stable Isotopes in Ecology and Environmental Science*. Blackwell Publishing, Victoria, pp.99-154.

Koch P.L., Heisinger J., Moss C., Carlson R.W., Fogel M.L. and Behrensmeyer A.K. (1995). Isotopic tracking of change in diet and habitat use in African elephants. *Science*, *267*(5202), pp.1340-1343.

Kohn M.J. (2010). Carbon isotope compositions of terrestrial C3 plants as indicators of (paleo) ecology and (paleo) climate. *Proceedings of the National Academy of Sciences*, *107*(46), pp.19691-19695.

Körner C., Farquhar G. and Wong S. (1991). Carbon isotope discrimination by plants follows latitudinal and altitudinal trends. *Oecologia*, *88*(1), pp.30-40.

Kumar S., Stecher G. and Tamura K. (2016). MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular biology and evolution*, *33*(7), pp.1870-1874.

Lawson P. (2016). The taxonomy of the genus Clostridium: current status and future perspectives. *微生物学通报*, *43*(5), pp.1070-1074.

Lawson P.A. and Rainey F.A. (2016). Proposal to restrict the genus Clostridium Prazmowski to Clostridium butyricum and related species. *International Journal of Systematic and Evolutionary Microbiology*, *66*(2), pp.1009-1016.

Lebov J., Grieger K., Womack D., Zaccaro D., Whitehead N., Kowalcyk B. and MacDonald P. (2017). A framework for One Health research. *One Health*, *3*, pp.44-50.

Leitner M., Davies A.B., Parr C.L., Eggleton P. and Robertson M.P. (2018). Woody encroachment slows decomposition and termite activity in an African savanna. *Global Change Biology*, *24*(6), pp.2597-2606.

Lenzen M., Moran D., Kanemoto K., Foran B., Lobefaro L. and Geschke A. (2012). International trade drives biodiversity threats in developing nations. *Nature*, *486*(7401), pp.109-112.

Lewis C. (2007). Clostridial diseases. In: I. Aitken (ed.), *Diseases in sheep fourth edition*. Blackwell publishing, Oxford, pp.156-167.

Li S. (1982). *Ben Cao Gang Mu (Compendium of Materia Medica)*. People's Medical Publishing House, Beijing, China

Lindberg A., Skarin H., Knutsson R., Blomqvist G. and Båverud V. (2010). Real-time PCR for Clostridium botulinum type C neurotoxin (BoNTC) gene, also covering a chimeric C/D sequence—Application on outbreaks of botulism in poultry. *Veterinary microbiology*, *146*(1-2), pp.118-123.

Lindsey P. and Skinner J. (2001). Ant composition and activity patterns as determined by pitfall trapping and other methods in three habitats in the semi-arid Karoo. *Journal of Arid Environments*, *48*(4), pp.551-568.

Lindström M., Nevas M., Kurki J., Sauna-Aho R., Latvala-Kiesilä A., Pölönen I. and Korkeala H. (2004). Type C botulism due to toxic feed affecting 52,000 farmed foxes and minks in Finland. *Journal of Clinical Microbiology*, *42*(10), pp.4718-4725.

Liu Z., Jiao D., Weng Z. and Zhang Z. (2016a). Structure and mechanical behaviors of protective armored pangolin scales and effects of hydration and orientation. *Journal of the Mechanical Behavior of Biomedical Materials*, *56*, pp.165-174.

Liu Z., Jiao D., Weng Z. and Zhang Z. (2016b). Water-assisted self-healing and property recovery in a natural dermal armor of pangolin scales. *Journal of the Mechanical Behavior of Biomedical Materials*, *56*, pp.14-22.

Loy A., Schulz C., Lücker S., Schöpfer-Wendels A., Stoecker K., Baranyi C., Lehner A. and Wagner M. (2005). 16S rRNA gene-based oligonucleotide microarray for environmental monitoring of the betaproteobacterial order "Rhodocyclales". *Applied and environmental microbiology*, *71*(3), pp.1373-1386.

Lübcker N., Condit R., Beltran R.S., de Bruyn P.N. and Bester M.N. (2016). Vibrissal growth parameters of southern elephant seals Mirounga leonina: obtaining fine-scale, time-based stable isotope data. *Marine Ecology Progress Series*, *559*, pp.243-255.

Lukes J., Mauricio I.L., Schoenian G., Dujardin J.-C., Soteriadou K., Dedet J.-P., Kuhls K., Tintaya K., Jirků M. and Chocholová E. (2007). Evolutionary and geographical history of the Leishmania donovani complex with a revision of current taxonomy. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(22), pp.9375-9380.

Lv J., Wu S., Zhang Y., Chen Y., Feng C., Yuan X., Jia G., Deng J., Wang C. and Wang Q. (2014). Assessment of four DNA fragments (COI, 16S rDNA, ITS2, 12S rDNA) for species identification of the Ixodida (Acari: Ixodida). *Parasites & Vectors*, *7*(1), p.93.

Mackenzie J.S., McKinnon M. and Jeggo M. (2014). One health: From concept to practice. In: A. Yamada, L. Kahn, B. Kaplan, T. Monath, J. Woodall and L. Conti (eds.), *Confronting Emerging Zoonoses: The One Health Paradigm*. Springer, Tokyo, pp.163-189.

MacLennan J. (1962). The histotoxic clostridial infections of man. *Bacteriological Reviews*, *26*(2), p.177.

Makhalanyane T., Valverde A., Gunnigle E., Frossard A., Ramond J. and Cowan D. (2015). Microbial ecology of hot desert edaphic systems. *FEMS Microbiology Reviews*, *39*(2), pp.203-221.

Mallette J.R., Casale J.F., Jordan J., Morello D.R. and Beyer P.M. (2016). Geographically sourcing cocaine's origin–delineation of the nineteen major coca growing regions in South America. *Scientific Reports*, *6*, p.23520.

Mambeya M.M., Baker F., Momboua B.R., Koumba Pambo A.F., Hega M., Okouyi Okouyi V.J., Onanga M., Challender D.W., Ingram D.J., Wang H. and Abernethy K. (2018). The emergence of a commercial trade in pangolins from Gabon. *African Journal of Ecology*, *56*(3), pp.601-609.

Manwa L. and Ndamba G. (2011). The language of dress among the subcultural group of the Dzimbabwe people in Masvingo, Zimbabwe. *Journal of Emerging Trends in Educational Research and Policy Studies*, *2*(6), pp.436-442.

Marshall J.D., Brooks J.R. and Lajtha K. (2007). Sources of variation in the stable isotopic composition of plants. *Stable Isotopes in Ecology and Environmental Science*, *2*, pp.22-60.

Martínez del Rio C. and Carleton S.A. (2012). How fast and how faithful: the dynamics of isotopic incorporation into animal tissues. *Journal of Mammalogy*, *93*(2), pp.353-359.

Martínez del Rio C. and Wolf B.O. (2005). Mass-balance models for animal isotopic ecology. In: M. Stack and T. Wang (eds.). *Physiological and ecological adaptations to feeding in vertebrates*. Science Publishers, Enfield, New Hampshire, pp.141-174.

Maurin M. and Raoult D.f. (1999). Q fever. *Clinical Microbiology Reviews*, *12*(4), pp.518-553.

Maynard G. and Nowell D. (2009). Biosecurity and quarantine for preventing invasive species. In: M. Clout and P. Williams (eds.), *Invasive species management—A handbook of principles and techniques*. Oxford University Press, Oxford, pp.1-18.

Mazzaglia A., Studholme D.J., Taratufolo M.C., Cai R., Almeida N.F., Goodman T., Guttman D.S., Vinatzer B.A. and Balestra G.M. (2012). Pseudomonas syringae pv. actinidiae (PSA) isolates from recent bacterial canker of kiwifruit outbreaks belong to the same genetic lineage. *PloS One*, *7*(5), p.e36518.

McCarron J.K. and Knapp A.K. (2001). C3 woody plant expansion in a C4 grassland: are grasses and shrubs functionally distinct? *American Journal of Botany*, *88*(10), pp.1818-1823.

McCutchan J.H., Lewis W.M., Kendall C. and McGrath C.C. (2003). Variation in trophic shift for stable isotope ratios of carbon, nitrogen, and sulfur. *Oikos*, *102*(2), pp.378-390.

McGuigan C.C., Penrice G.M., Gruer L., Ahmed S., Goldberg D., Black M., Salmon J.E. and Hood J. (2002). Lethal outbreak of infection with clostridium novyi type a and other spore-forming organisms in Scottish injecting drug users. *Journal of Medical Microbiology*, *51*(11), pp.971-977.

McIlwee A. and Johnson C. (1998). The contribution of fungus to the diets of three mycophagous marsupials in Eucalyptus forests, revealed by stable isotope analysis. *Functional Ecology*, *12*(2), pp.223-231.

McKechnie A.E. (2004). Stable isotopes: powerful new tools for animal ecologists: news & views. *South African Journal of Science*, *100*(3-4), pp.131-134.

McNeill D. and Lichtenstein G. (2003). Local conflicts and international compromises: The sustainable use of vicuña in Argentina. *Journal of International Wildlife Law and Policy*, *6*(3), pp.233-253.

McRae L., Deinet S. and Freeman R. (2017). The diversity-weighted Living Planet Index: controlling for taxonomic bias in a global biodiversity indicator. *PloS One*, *12*(1), p.e0169156.

Mebius R.E. and Kraal G. (2005). Structure and function of the spleen. *Nature Reviews Immunology*, *5*(8), p.606.

Meier-Augenstein W. and Fraser I. (2008). Forensic isotope analysis leads to identification of a mutilated murder victim. *Science & Justice*, *48*(3), pp.153-159.

Meier R., Shiyang K., Vaidya G. and Ng P.K. (2006). DNA barcoding and taxonomy in Diptera: a tale of high intraspecific variability and low identification success. *Systematic Biology*, *55*(5), pp.715-728.

Mekota A.M., Grupe G., Ufer S. and Cuntz U. (2006). Serial analysis of stable nitrogen and carbon isotopes in hair: monitoring starvation and recovery phases of patients suffering from anorexia nervosa. *Rapid Communications in Mass Spectrometry: An International Journal Devoted to the Rapid Dissemination of Up-to-the-Minute Research in Mass Spectrometry*, *20*(10), pp.1604-1610.

Mellmann A., Cloud J.L., Andrees S., Blackwood K., Carroll K.C., Kabani A., Roth A. and Harmsen D. (2003). Evaluation of RIDOM, MicroSeq, and Genbank services in the molecular identification of Nocardia species. *International Journal of Medical Microbiology*, *293*(5), pp.359-370.

Meyer W., Liumsiricharoen M., Hornickel I., Suprasert A., Schnapper A. and Fleischer L.-G. (2010). Demonstration of substances of innate immunity in the integument of the Malayan pangolin (Manis javanica). *European Journal of Wildlife Research*, *56*(3), pp.287-296.

Millennium Ecosystem Assessment. (2005). *Ecosystems and Human Well-being: Biodiversity Synthesis*. World Resources Institute, Washington, DC, p.100.

Mohapatra R.K., Panda S., Nair M.V. and Acharjyo L.N. (2016). Check list of parasites and bacteria recorded from pangolins (manis sp.). *Journal of Parasitic Diseases*, *40*(4), pp.1109-1115.

Montet D., Le Nguyen D.D. and Kouakou A.C. (2012). Determination of fish origin by using 16S rDNA fingerprinting of microbial communities by PCR-DGGE: An application on fish from different tropical origins. In: Z. Muchlisin (ed.), *Aquaculture* InTech, Rijeka, pp.93-108.

Morecroft M. and Woodward F. (1990). Experimental investigations on the environmental determination of δ13C at different altitudes. *Journal of Experimental Botany*, *41*(10), pp.1303-1308.

Moriishi K., Koura M., Fujii N., Fujinaga Y., Inoue K., Syuto B. and Oguma K. (1996). Molecular cloning of the gene encoding the mosaic neurotoxin, composed of parts of botulinum neurotoxin types C1 and D, and PCR detection of this gene from Clostridium botulinum type C organisms. *Applied and Environmental Microbiology*, *62*(2), pp.662-667.

Mount D.W. (2007). Using the basic local alignment search tool (BLAST). *Cold Spring Harbor Protocols*, *2007*(7), p.pdb. top17.

Muellner P., Pleydell E., Pirie R., Baker M., Campbell D., Carter P. and French N. (2013). Molecular-based surveillance of campylobacteriosis in New Zealand–from source attribution to genomic epidemiology. *Eurosurveillance*, *18*(3), p.20365.

Muñoz-Colmenero M., Martínez J.L., Roca A. and Garcia-Vazquez E. (2017). NGS tools for traceability in candies as high processed food products: Ion Torrent PGM versus conventional PCR-cloning. *Food chemistry*, *214*, pp.631-636.

Murphy B.P. and Bowman D.M. (2009). The carbon and nitrogen isotope composition of Australian grasses in relation to climate. *Functional Ecology*, *23*(6), pp.1040-1049.

Mwale M., Dalton D.L., Jansen R., De Bruyn M., Pietersen D., Mokgokong P.S. and Kotzé A. (2016). Forensic application of DNA barcoding for identification of illegally traded African pangolin scales. *Genome*, *60*(3), pp.272-284.

Nakamura S., Kimura I., Yamakawa K. and Nishida S. (1983). Taxonomic relationships among Clostridium novyi types A and B, Clostridium haemolyticum and Clostridium botulinum type C. *Microbiology*, *129*(5), pp.1473-1479.

National Research Council (2011). *Review of the scientific approaches used during the FBI's investigation of the 2001 anthrax letters*. National Academies Press, Washington, D.C.

Navarro M., Dutra F., Briano C., Romero A., Persiani M., Freedman J., Morrell E., Beingesser J. and Uzal F. (2017). Pathology of Naturally Occurring Bacillary Hemoglobinuria in Cattle. *Veterinary Pathology*, *54*(3), pp.457-466.

Nehlich O. (2015). The application of sulphur isotope analyses in archaeological research: a review. *Earth-Science Reviews*, *142*, pp.1-17.

Nellemann C., Henriksen R., Kreilhuber A., Stewart D., Kotsovou M., Raxter P., Mrema E. and Barrat S. (2016). *The rise of environmental crime: a growing threat to natural resources, peace, development and security*. A UNEP- INTERPOL Rapid Response Assessment. United Nations Environment Programme and RHIPTO Rapid Response–Norwegian Center for Global Analyses.

Newsome S.D., Tinker M.T., Monson D.H., Oftedal O.T., Ralls K., Staedler M.M., Fogel M.L. and Estes J.A. (2009). Using stable isotopes to investigate individual diet specialization in California sea otters (Enhydra lutris nereis). *Ecology*, *90*(4), pp.961-974.

Newsome S.D., Yeakel J.D., Wheatley P.V. and Tinker M.T. (2012). Tools for quantifying isotopic niche space and dietary variation at the individual and population level. *Journal of Mammalogy*, *93*(2), pp.329-341.

Nyaoke A.C., Navarro M.A., Beingesser J. and Uzal F.A. (2018). Infectious necrotic hepatitis caused by Clostridium novyi type B in a horse: case report and review of the literature. *Journal of Veterinary Diagnostic Investigation*, *30*(2), pp.294-299.

Ogden R. (2008). Fisheries forensics: the use of DNA tools for improving compliance, traceability and enforcement in the fishing industry. *Fish and Fisheries*, *9*(4), pp.462-472.

Ogden R. and Linacre A. (2015). Wildlife forensic science: a review of genetic geographic origin assignment. *Forensic Science International: Genetics*, *18*, pp.152-159.

Oguma K., Fujinaga Y. and Inoue K. (1995). Structure and function of Clostridium botulinum toxins. *Microbiology and Immunology*, *39*(3), pp.161-168.

Oksche A., Nakov R. and Habermann E. (1992). Morphological and biochemical study of cytoskeletal changes in cultured cells after extracellular application of Clostridium novyi alpha-toxin. *Infection and Immunity*, *60*(7), pp.3002-3006.

Oldfield S. (2003). *The trade in wildlife: regulation for conservation*. Routledge. New York, p.232.

Olive P.J., Pinnegar J.K., Polunin N.V., Richards G. and Welch R. (2003). Isotope trophic-step fractionation: a dynamic equilibrium model. *Journal of Animal Ecology*, *72*(4), pp.608-617.

Olson D. (2013). Bacillary hemoglobinuria (redwater disease) in cattle. In: J. R. Adams (ed.), *Cattle producer's handbook third edition*. Western Beef Resource Committee, Idaho, pp.655-657.

Omrani A.S., Al-Tawfiq J.A. and Memish Z.A. (2015). Middle East respiratory syndrome coronavirus (MERS-CoV): animal to human interaction. *Pathogens and Global Health*, *109*(8), pp.354-362.

Onyenwoke R.U., Brill J.A., Farahi K. and Wiegel J. (2004). Sporulation genes in members of the low G+ C Gram-type-positive phylogenetic branch (Firmicutes). *Archives of Microbiology*, *182*(2-3), pp.182-192.

Ou C.-Y., Ciesielski C.A., Myers G., Bandea C.I., Luo C.-C., Korber B.T., Mullins J.I., Schochetman G., Berkelman R.L. and Economou A.N. (1992). Molecular epidemiology of HIV transmission in a dental practice. *Science*, *256*(5060), pp.1165-1171.

Owen R.J. (2004). *Bacterial taxonomics*. Routledge. New York, p. 232.

Parada A.E., Needham D.M. and Fuhrman J.A. (2016). Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. *Environmental Microbiology*, *18*(5), pp.1403-1414.

Patel J.B., Wallace R.J., Brown-Elliott B.A., Taylor T., Imperatrice C., Leonard D.G., Wilson R.W., Mann L., Jost K.C. and Nachamkin I. (2004). Sequence-based identification of aerobic actinomycetes. *Journal of Clinical Microbiology*, *42*(6), pp.2530-2540.

Patel R., Piper K.E., Rouse M.S., Steckelberg J.M., Uhl J.R., Kohner P., Hopkins M.K., Cockerill F.R. and Kline B.C. (1998). Determination of 16S rRNA sequences of enterococci and application to species identification of nonmotile Enterococcus gallinarum isolates. *Journal of Clinical Microbiology*, *36*(11), pp.3399-3407.

Palmateer N.E., Hope V.D., Roy K., Marongiu A., White J.M., Grant K.A., Ramsay C.N., Goldberg D.J. and Ncube F. (2013). Infections with spore-forming bacteria in persons who inject drugs, 2000–2009. *Emerging Infectious Diseases*, *19*(1), p.29

Pearson S.F., Levey D.J., Greenberg C.H. and Del Rio C.M. (2003). Effects of elemental composition on the incorporation of dietary nitrogen and carbon isotopic signatures in an omnivorous songbird. *Oecologia*, *135*(4), pp.516-523.

Perera P., Karawita K. and Pabasara M. (2017). Pangolins (Manis crassicaudata) in Sri Lanka: A Review of Current Knowledge, Threats and Research Priorities. *Journal of Tropical Forestry and Environment*, *7*(1), pp.1-14.

Pietersen D.W. (2013). Behavioural ecology and conservation biology of ground pangolins Smutsia temminckii in the Kalahari Desert. MSc thesis, University of Pretoria, Pretoria, South Africa, p.118.

Pietersen D., Hywood L., Heighton S., Connelly E., and Jansen R. (*In Prep*). Out of Africa: A review of the illicit trade and legislation pertaining to pangolins (Pholidota: Manidae) in Africa.

Pietersen D., McKechnie A. and Jansen R. (2014a). A review of the anthropogenic threats faced by Temminck's ground pangolin, Smutsia temminckii, in southern Africa. *South African Journal of Wildlife Research*, *44*(2), pp.167-178.

Pietersen D.W., McKechnie A.E. and Jansen R. (2014b). Home range, habitat selection and activity patterns of an arid-zone population of Temminck's ground pangolins, Smutsia temminckii. *African Zoology*, *49*(2), pp.265-276.

Pietersen D.W., Symes C.T., Woodborne S., McKechnie A.E. and Jansen R. (2016). Diet and prey selectivity of the specialist myrmecophage, Temminck's ground pangolin. *Journal of Zoology*, *298*(3), pp.198-208.

Pietersen D., Waterman C., Hywood L., Rankin P. and Soewu D. (2014c). Smutsia temminckii. The IUCN Red List of Threatened Species 2014. Available online at: http://www.iucnredlist.org/details/12765/0. Downloaded on 09 March 2016.

Pimm S.L., Russell G.J., Gittleman J.L. and Brooks T.M. (1995). The future of biodiversity. *Science*, *269*(5222), pp.347-350.

Pimentel D., Zuniga R. and Morrison D. (2005). Update on the environmental and economic costs associated with alien-invasive species in the United States. *Ecological Economics*, *52*(3), pp.273-288.

Pimentel T., Marcelino J., Ricardo F., Soares A.M. and Calado R. (2017). Bacterial communities 16S rDNA fingerprinting as a potential tracing tool for cultured seabass Dicentrarchus labrax. *Scientific reports*, *7*(1), p.11862.

Podlesak D.W. and McWilliams S.R. (2006). Metabolic routing of dietary nutrients in birds: effects of diet quality and macronutrient composition revealed using stable isotopes. *Physiological and Biochemical Zoology*, *79*(3), pp.534-549.

Post D.M. (2002). Using stable isotopes to estimate trophic position: models, methods, and assumptions. *Ecology*, *83*(3), pp.703-718.

Pough F.H., Andrews R.M., Cadle J.E., Crump M.L., Savitzky A.H. and Wells, K.D. (1998). Conservation and the future of amphibians and reptiles. In: L. Vitt and J. Cadwell (eds.), *Herpetology (3rd edition).* Prentice Hall, New Jersey, pp.595-633.

Qiu J. (1985). *Rare books: integration of ancient medical books*. Shanghai Science & Technology Press, Shanghai, China.

Quaak F.C., van Duijn T., Hoogenboom J., Kloosterman A.D. and Kuiper I. (2018). Human-associated microbial populations as evidence in forensic casework. *Forensic Science International: Genetics*, *36*, pp.176-185.

Quail M.A., Smith M., Coupland P., Otto T.D., Harris S.R., Connor T.R., Bertoni A., Swerdlow H.P. and Gu Y. (2012). A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics*, *13*(1), p.341.

Randhawa S.S., Sharma D.K., Randhawa C.S., Gill B.S., Brar R.S. and Singh J. (1995). An outbreak of bacillary haemoglobinuria in sheep in India. *Tropical Animal Health and Production*, *27*(1), pp.31-36.

Rasko D.A., Webster D.R., Sahl J.W., Bashir A., Boisen N., Scheutz F., Paxinos E.E., Sebra R., Chin C.-S. and Iliopoulos D. (2011). Origins of the E. coli strain causing an outbreak of hemolytic–uremic syndrome in Germany. *New England Journal of Medicine*, *365*(8), pp.709-717.

Reed K.D., Melski J.W., Graham M.B., Regnery R.L., Sotir M.J., Wegner M.V., Kazmierczak J.J., Stratman E.J., Li, Y., Fairley J.A. and Swain G.R. (2004). The detection of monkeypox in humans in the Western Hemisphere. *New England Journal of Medicine*, *350*(4), pp.342-350.

Retief K., West A.G. and Pfab M.F. (2014). Can stable isotopes and radiocarbon dating provide a forensic solution for curbing illegal harvesting of threatened cycads? *Journal of Forensic Sciences*, *59*(6), pp.1541-1551.

Rewell R. (1950). Report of the society's pathologist for the year 1949. *Proceedings of the Zoological Society of London*, *120*(3), pp.485-495.

Richards M.P., Fuller B.T., Sponheimer M., Robinson T. and Ayliffe L. (2003). Sulphur isotopes in palaeodietary studies: a review and results from a controlled feeding experiment. *International Journal of Osteoarchaeology*, *13*(1-2), pp.37-45.

Rivalan P., Delmas V., Angulo E., Bull L.S., Hall R.J., Courchamp F., Rosser A.M. and Leader-Williams N. (2007). Can bans stimulate wildlife trade? *Nature*, *447*(7144), pp.529-530.

Roe D. (2008). *Trading nature*: *A report, with case studies, on the contribution of wildlife trade management to sustainable livelihoods and the Millennium Development Goals*. TRAFFIC International and WWF International, Cambridge, p.84.

Roh S.W., Abell G.C., Kim K.-H., Nam Y.-D. and Bae J.-W. (2010). Comparing microarrays and next-generation sequencing technologies for microbial ecology research. *Trends in Biotechnology*, *28*(6), pp.291-299.

Ronquist F., Teslenko M., Van Der Mark P., Ayres D.L., Darling A., Höhna S., Larget B., Liu L., Suchard M.A. and Huelsenbeck J.P. (2012). MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology*, *61*(3), pp.539-542.

Ross H.A., Murugan S. and Sibon Li W.L. (2008). Testing the reliability of genetic methods of species identification via simulation. *Systematic Biology*, *57*(2), pp.216-230.

Rothberg J.M., Hinz W., Rearick T.M., Schultz J., Mileski W., Davey M., Leamon J.H., Johnson K., Milgrew M.J. and Edwards M. (2011). An integrated semiconductor device enabling non-optical genome sequencing. *Nature*, *475*(7356), p.348.

Rubenstein D.R. and Hobson K.A. (2004). From birds to butterflies: animal movement patterns and stable isotopes. *Trends in Ecology & Evolution*, *19*(5), pp.256-263.

Rutherford M.C. and Westfall R.H. (1994). Biomes of southern Africa: an objective characterization. *Memoirs of the Botanical Survey of South Africa*, *54*, pp.1-98

Saeb A.T., Abouelhoda M., Selvaraju M., Althawadi S.I., Mutabagani M., Adil M., Al Hokail A. and Tayeb H.T. (2017). The use of next-generation sequencing in the identification of a fastidious pathogen: A lesson from a clinical setup. *Evolutionary Bioinformatics*, *13*, pp.1-15.

Salipante S.J., Kawashima T., Rosenthal C., Hoogestraat D.R., Cummings L.A., Sengupta D.J., Harkins T.T., Cookson B.T. and Hoffman N.G. (2014). Performance comparison of Illumina and ion torrent next-generation sequencing platforms for 16S rRNA-based bacterial community profiling. *Applied and Environmental Microbiology*, pp.AEM. 02206-02214.

Samarajeewa A.D., Hammad A., Masson L., Khan I., Scroggins R. and Beaudette L. (2015). Comparative assessment of next-generation sequencing, denaturing gradient gel electrophoresis, clonal restriction fragment length polymorphism and cloning-sequencing as methods for characterizing commercial microbial consortia. *Journal of Microbiological Methods*, *108*, pp.103-111.

Santos A., Satchabut T. and Vigo Trauco G. (2011). Do wildlife trade bans enhance or undermine conservation efforts. *Applied Biodiversity Perspective Series*, *1*(3), pp.1-15.

Sargeant D., Kammin W.R. and Collyard S. (2011). *Review and critique of current microbial source tracking (mst) techniques*. Environmental Assessment Program, Washington State Department of Ecology.

Sasaki Y., Kojima A., Aoki H., Ogikubo Y., Takikawa N. and Tamura Y. (2002). Phylogenetic analysis and PCR detection of Clostridium chauvoei, Clostridium haemolyticum, Clostridium novyi types A and B, and Clostridium septicum based on the flagellin gene. *Veterinary Microbiology*, *86*(3), pp.257-267.

Sasaki Y., Takikawa N., Kojima A., Norimatsu M., Suzuki S. and Tamura Y. (2001). Phylogenetic positions of Clostridium novyi and Clostridium haemolyticum based on 16S rDNA sequences. *International Journal of Systematic and Evolutionary Microbiology*, *51*(3), pp.901-904.

Sasaki Y., Yamamoto K., Kojima A., Norimatsu M. and Tamura Y. (2000). Rapid identification and differentiation of pathogenic clostridia in gas gangrene by polymerase chain reaction based on the 16S-23S r DNA spacer region. *Research in Veterinary Science*, *69*(3), pp.289-294.

Schallehn G. and Eklund M.W. (1980). Conversion of Clostridium novyi type D (C. haemolyticum) to alpha toxin production by phages of C. novyi type A. *FEMS Microbiology Letters*, *7*(2), pp.83-86.

Schmedes S.E., Sajantila A. and Budowle B. (2016). Expansion of microbial forensics. *Journal of clinical microbiology*, pp.JCM. 00046-00016.

Scholes R.J. and Biggs R. (2004). *Ecosystem Services in Southern Africa: A Regional Assessment*. The Regional-Scale Component of the Southern African Millennium Ecosystem Assessment. CSIR, Pretoria, South Africa.

Schürch A.C. and Siezen R.J. (2010). Genomic tracing of epidemics and disease outbreaks. *Microbial Biotechnology*, *3*(6), pp.628-633.

Schurr M.R., Hayes R.G. and Cook D.C. (2008). Thermally induced changes in the stable carbon and nitrogen isotope ratios of charred bones. In: C. Schmidt and S. Symes (eds.), *The Analysis of Burned Human Remains (second edition)*. Elsevier, London, pp.105-118.

Schwarcz H.P., Dupras T.L. and Fairgrieve S.I. (1999). 15N enrichment in the Sahara: in search of a global relationship. *Journal of Archaeological Science*, *26*(6), pp.629-636.

Schwartz T.S. and Karl S.A. (2008). Population genetic assignment of confiscated gopher tortoises. *Journal of Wildlife Management*, *72*(1), pp.254-259.

Scoones I. and Wolmer W. (2006). *Livestock, disease, trade and markets: policy choices for the livestock sector in Africa*. Institute of Development Studies. University of Sussex, UK, p.53.

Scott T.M., Rose J.B., Jenkins T.M., Farrah S.R. and Lukasik J. (2002). Microbial source tracking: current methodology and future directions. *Applied and Environmental Microbiology*, *68*(12), pp.5796-5803.

Sealy J.C., Van Der Merwe N.J., Thorp J.A.L. and Lanham J.L. (1987). Nitrogen isotopic ecology in southern Africa: implications for environmental and dietary tracing. *Geochimica et Cosmochimica Acta*, *51*(10), pp.2707-2717.

Sender R., Fuchs S. and Milo R. (2016). Revised estimates for the number of human and bacteria cells in the body. *PLoS Biology*, *14*(8), p.e1002533.

Serrano R., Blanes M. and Orero L. (2007). Stable isotope determination in wild and farmed gilthead sea bream (Sparus aurata) tissues from the western Mediterranean. *Chemosphere*, *69*(7), pp.1075-1080.

Shade A., Read J.S., Youngblut N.D., Fierer N., Knight R., Kratz T.K., Lottig N.R., Roden E.E., Stanley E.H., Stombaugh J. and Whitaker R.J., (2012). Lake microbial communities are resilient after a whole-ecosystem disturbance. *The ISME Journal*, *6*(12), p.2153.

Skarin H., Håfström T., Westerberg J. and Segerman B. (2011). Clostridium botulinum group III: a group with dual identity shaped by plasmids, phages and mobile elements. *BMC Genomics*, *12*(1), p.185.

Skarin H. and Segerman B. (2014). Plasmidome interchange between Clostridium botulinum, Clostridium novyi and Clostridium haemolyticum converts strains of independent lineages into distinctly different pathogens. *PLoS One*, *9*(9), p.e107777.

Skarin H., Tevell Åberg A., Woudstra C., Hansen T., Löfström C., Koene M., Bano L., Hedeland M., Anniballi F., De Medici D. and Olsson Engvall E. (2013). The workshop on animal botulism in Europe. *Biosecurity and Bioterrorism: Biodefense Strategy, Practice, and Science*, *11*(S1), pp.S183-S190.

Skujinš J., (1984). Microbial ecology of desert soils. In: K. Marshall (ed.), *Advances in Microbial Ecology*. Springer, USA, pp.49-91.

Smith A.D. and Dufty Jr A.M. (2005). Variation in the stable-hydrogen isotope composition of Northern Goshawk feathers: relevance to the study of migratory origins. *The Condor*, *107*(3), pp.547-558.

Smith C.J., Danilowicz B.S. and Meijer W.G. (2009a). Bacteria associated with the mucus layer of Merlangius merlangus (whiting) as biological tags to determine harvest location. *Canadian Journal of Fisheries and Aquatic Sciences*, *66*(5), pp.713-716.

Smith K., Acevedo-Whitehouse K. and Pedersen A. (2009b). The role of infectious diseases in biological conservation. *Animal Conservation*, *12*(1), pp.1-12.

Smith R.K., Marais A., Chadwick P., Lloyd P.H. and Hill R.A. (2008). Monitoring and management of the endangered Cape mountain zebra Equus zebra zebra in the Western Cape, South Africa. *African Journal of Ecology*, *46*(2), pp.207-213.

Smith L. and Williams B.L. (1984). *The pathogenic anaerobic bacteria (third edition).* Charles C Thomas Printing Ltd, Springfield, p.331.

Smithers R.H. (1971). *Mammals of Botswana*. Trustees of the National Museums of Rhodesia, p.340.

Snitkin E.S., Zelazny A.M., Thomas P.J., Stock F., Henderson D.K., Palmore T.N., Segre J.A. and Program N.C.S. (2012). Tracking a hospital outbreak of carbapenem-resistant Klebsiella pneumoniae with whole-genome sequencing. *Science Translational Medicine*, *4*(148), pp.148ra116-148ra116.

Soewu D. and Adekanola T. (2011). Traditional-medical knowledge and perception of pangolins (*Manis sps*) among the Awori people, southwestern Nigeria. *Journal of Ethnobiology and Ethnomedicine*, 7(1), pp.1-11.

Songer, J.G. (2005). Clostridial diseases in domestic animals. In: P. Duerre (ed.) *Handbook on clostridia*. Taylor & Francis, New York.

Souillard R., Woudstra C., Le Maréchal C., Dia M., Bayon-Auboyer M.H., Chemaly M., Fach P. and Le Bouquin S. (2014). Investigation of Clostridium botulinum in commercial poultry farms in France between 2011 and 2013. *Avian Pathology*, *43*(5), pp.458-464.

Sousa J.A.d. and Silva-Souza Â.T. (2001). Bacterial community associated with fish and water from Congonhas River, Sertaneja, Paraná, Brazil. *Brazilian Archives of Biology and Technology*, *44*(4), pp.373-381.

Spearman R. (1967). On the nature of the horny scales of the pangolin. *Zoological Journal of the Linnean Society*, *46*(310), pp.267-273.

Sponheimer M., Lee-Thorp J., de Ruiter D., Codron D., Codron J., Baugh A.T. and Thackeray F. (2005). Hominins, sedges, and termites: new carbon isotope data from the Sterkfontein valley and Kruger National Park. *Journal of Human Evolution*, *48*(3), pp.301-312.

Sponheimer M., Robinson T., Ayliffe L., Passey B., Roeder B., Shipley L., Lopez E., Cerling T., Dearing D. and Ehleringer J. (2003a). An experimental study of carbon-isotope fractionation between diet, hair, and feces of mammalian herbivores. *Canadian Journal of Zoology*, *81*(5), pp.871-876.

Sponheimer M., Robinson T., Roeder B., Passey B., Ayliffe L., Cerling T.E., Dearing M.-D. and Ehleringer J.R. (2003b). An experimental study of nitrogen flux in llamas: is 14N preferentially excreted? *Journal of Archaeological Science*, *30*(12), pp.1649-1655.

Springer A.M., Estes J.A., van Vliet G.B., Williams T.M., Doak D.F., Danner E.M., Forney K.A. and Pfister B. (2003). Sequential megafaunal collapse in the North Pacific Ocean: An ongoing legacy of industrial whaling? *Proceedings of the National Academy of Sciences*, *100*(21), pp.12223-12228.

Stackebrandt E., Kramer I., Swiderski J. and Hippe H. (1999). Phylogenetic basis for a taxonomic dissection of the genus Clostridium. *FEMS Immunology & Medical Microbiology*, *24*(3), pp.253-258.

Sternberg L.O., Deniro M.J. and Johnson H.B. (1984). Isotope ratios of cellulose from plants having different photosynthetic pathways. *Plant Physiology*, *74*(3), pp.557-561.

Stogdale L. and Booth A. (1984). Bacillary Hemoglobinuria in cattle [Clostridium haemolyticum, veterinary education]. *The Compendium on Continuing Education for the Practicing Veterinarian*, *6*, pp.S284-S290.

Sullivan A., Ellis P. and Knight I. (2003). A review of radiant heat flux models used in bushfire applications. *International Journal of Wildland Fire*, *12*(1), pp.101-110.

Sulzman E.W. (2007). Stable isotope chemistry and measurement: a primer. In: *Stable Isotopes in Ecology and Environmental Science* (eds. by R. Michener andK. Lajtha). Blackwell Publishing, Victoria, Australia, pp.1-21.

Swaisgood R., Wang D. and Wei F. (2016). Ailuropoda melanoleuca. IUCN red list of threatened species. Available online at: http://wwwiucnredlistorg/details/712/0. Downloaded on 09 May 2017.

Swap R.J., Aranibar J.N., Dowty P.R., Gilhooly III W.P. and Macko S.A. (2004). Natural abundance of ^{13}C and ^{15}N in C_3 and C_4 vegetation of southern Africa: patterns and implications. *Global Change Biology*, *10*(3), pp.350-358.

Swart, J. (2013). Smutsia temminckii Ground pangolin (Temminck's ground pangolin, Cape pangolin). In: J. Kingdon and M. Hoffmann (eds.), *Mammals of Africa Volume 5: Carnivores, Pangolins, Equids and Rhinoceroses*. Bloomsbury Natural History, London, pp. 400-405.

Symes C., Skhosana F., Butler M., Gardner B. and Woodborne S. (2017). Isotope (δ 13C, δ 15N, δ 2H) diet–tissue discrimination in African grey parrot Psittacus erithacus: implications for forensic studies. *Isotopes in Environmental and Health Studies*, *53*(6), pp.580-596.

Takagi M., Yamato O., Sasaki Y., Mukai S., Fushimi Y., Yoshida T., Mizukami K., Shoubudani T., Amimoto K., Chuma T. and Shahada F. (2009). Successful treatment of bacillary hemoglobinuria in Japanese Black cows. *Journal of Veterinary Medical Science*, *71*(8), pp.1105-1108.

Tatsadjieu N.L., Maïworé J., Hadjia M., Loiseau G., Montet D. and Mbofung C. (2010). Study of the microbial diversity of Oreochromis niloticus of three lakes of Cameroon by PCR-DGGE: Application to the determination of the geographical origin. *Food Control*, *21*(5), pp.673-678.

Thompson L.R., Sanders J.G., McDonald D., Amir A., Ladau J., Locey K.J., Prill R.J., Tripathi A., Gibbons S.M. and Ackermann G. (2017). A communal catalogue reveals Earth's multiscale microbial diversity. *Nature, 551*(7681).

Thomsen J.B. and Munn C.A. (1988). Cyanopsitta spixii: a non-recovery report. *Parrotletter, 1*(1), pp.6-7.

Tieszen L.L., Senyimba M.M., Imbamba S.K. and Troughton J.H. (1979). The distribution of C 3 and C 4 grasses and carbon isotope discrimination along an altitudinal and moisture gradient in Kenya. *Oecologia*, *37*(3), pp.337-350.

Tillberg C.V., McCarthy D.P., Dolezal A.G. and Suarez A.V. (2006). Measuring the trophic ecology of ants using stable isotopes. *Insectes Sociaux*, *53*(1), pp.65-69.

Tipple B.J. and Pagani M. (2007). The early origins of terrestrial C4 photosynthesis. *Annual Review of Earth and Planetary Sciences*, *35*, pp.435-461.

Tindall B. J. (2016). Priority of the genus name Clostridium Prazmowski 1880 (Approved Lists 1980) vs Sarcina Goodsir 1842 (Approved Lists 1980) and the creation of the illegitimate combinations Clostridium maximum (Lindner 1888) Lawson and Rainey 2016 and Clostridium ventriculi (Goodsir 1842) Lawson and Rainey 2016 that may not be used. *International Journal of Systematic and Evolutionary Microbiology, 66* (11), pp.4890-4894.

Thompson J.D., Higgins D.G. and Gibson T.J. (1994). CLUSTALW: improving the sensitivity of progressive weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, *22*, pp.4673-4680.

Tong J., Lü T.-b., Ma Y.-h., Wang H.-k., Ren L.-q. and Arnell R. (2007). Two-body abrasive wear of the surfaces of pangolin scales. *Journal of Bionic Engineering*, *4*(2), pp.77-84.

Tong J., Ma Y.-H., Ren L.-Q. and Li J.-Q. (2000). Tribological characteristics of pangolin scales in dry sliding. *Journal of Materials Science Letters*, *19*(7), pp.569-572.

Tong J., Ren L.-Q. and Chen B.-C. (1995). Chemical constitution and abrasive wear behaviour of pangolin scales. *Journal of Materials Science Letters*, *14*(20), pp.1468-1470.

Torsvik V. and Øvreås L. (2002). Microbial diversity and function in soil: from genes to ecosystems. *Current Opinion in Microbiology*, *5*(3), pp.240-245.

TRAFFIC (2016). Out of Sequence: is wildlife forensics delivering as an illegal trade enforcement tool? Wildlife DNA Forensics, *TRAFFIC Bulletin*, *28*(2), pp.56-58.

Travis D.A., Watson R.P. and Tauer A. (2011). The spread of pathogens through trade in wildlife. *Revue Scientifique et Technique-OIE*, *30*(1), p.219.

Tsutaya T. and Yoneda M. (2015). Reconstruction of breastfeeding and weaning practices using stable isotope and trace element analyses: A review. *American Journal of Physical Anthropology*, *156*, pp.2-21.

Turenne C.Y., Tschetter L., Wolfe J. and Kabani A. (2001). Necessity of quality-controlled 16S rRNA gene sequence databases: identifying nontuberculous Mycobacteriumspecies. *Journal of Clinical Microbiology*, *39*(10), pp.3637-3648.

Udey L.R., Young E. and Sallman B. (1977). Isolation and characterization of an anaerobic bacterium, Eubacterium tarantellus sp. nov., associated with striped mullet (Mugil cephalus) mortality in Biscayne Bay, Florida. *Journal of the Fisheries Board of Canada, 34*(3), pp.402-409.

Ueda K., Seki T., Kudo T., Yoshida T. and Kataoka M. (1999). Two distinct mechanisms cause heterogeneity of 16S rRNA. *Journal of Bacteriology*, *181*(1), pp.78-82.

UNICEF (2008). *Contributing to one world, one health*. A strategic framework for reducing risks of infectious diseases at the animal-human-ecosystems interface, p.68.

Uno K.T., Quade J., Fisher D.C., Wittemyer G., Douglas-Hamilton I., Andanje S., Omondi P., Litoroh M. and Cerling T.E. (2013). Bomb-curve radiocarbon measurement of recent biologic tissues

and applications to wildlife forensics and stable isotope (paleo) ecology. *Proceedings of the National Academy of Sciences*, *110*(29), pp.11736-11741.

UNODC (2016). *World wildlife crime report: Trafficking in protected species*. United Nations Office on Drugs and Crime, Vienna, p.101.

Uscamaita M.R. and Bodmer R. (2010). Recovery of the endangered giant otter Pteronura brasiliensis on the Yavarí-Mirín and Yavarí Rivers: a success story for CITES. *Oryx*, *44*(01), pp.83-88.

Uzal F.A., Songer J.G., Prescott J.F. and Popoff M.R. (2016). *Clostridial diseases of animals*. John Wiley & Sons, Iowa, 336 pp.

van Aarde R., Richardson P. and Pietersen E. (1990). Report on the behavioural ecology of the Cape pangolin (Manis temminckii). Mammal Research Institute, University of Pretoria. Internal Report.

Van der Lugt J., De Wet S., Bastianello S., Kellerman T. and Van Jaarsveld L. (1995). Two outbreaks of type C and type D botulism in sheep and goats in South Africa. *South African Veterinary Association*, *66*, pp.77-77.

van der Merwe N.J., Lee-Thorp J.A., Thackeray J.F., Hall-Martin A., Kruger F.J., Coetzee H., Bell R.H.V. and Lindeque, M. (1990). Source-area determination of elephant ivory by isotopic analysis. *Nature*, *346*(6286), p.744.

van Rijsoort J. (2000). Non-timber forest products (NTFPs): Their role in sustainable forest management in the tropics. *Theme Studies Series-Forests, Forestry and Biological Diversity Support Group.* National Reference Centre for Nature Management, Wageningen, Netherlands.

van Rooyen N. and van Rooyen G. (2017). Ecological evaluation of Tswalu Kalahari Reserve. Ekotrust, p.143.

Vander Zanden H.B., Bjorndal K.A., Reich K.J. and Bolten A.B. (2010). Individual specialists in a generalist population: results from a long-term stable isotope series. *Biology Letters*, *6*(5), pp.711-714.

Vander Zanden H.B., Nelson D.M., Wunder M.B., Conkling T.J. and Katzner T. (2018a). Application of isoscapes to determine geographic origin of terrestrial wildlife for conservation and management. *Biological Conservation*, *228*, pp.268-280.

Vander Zanden H.B., Reid A., Katzner T. and Nelson D.M. (2018b). Effect of heat and singeing on stable hydrogen isotope ratios of bird feathers and implications for their use in determining geographic origins. *Rapid Communications in Mass Spectrometry*, *32*(21), pp.1859-1866.

Vander Zanden M.J., Clayton M.K., Moody E.K., Solomon C.T. and Weidel B.C. (2015). Stable isotope turnover and half-life in animal tissues: a literature synthesis. *PloS One*, *10*(1), p.e0116182.

Vawter L. and Records E. (1926). Recent studies of icterohemoglobinuria of cattle. *American Veterinary Medical Association*, *68*, pp.494-512.

Viljoen G., Luckins A. and Naletoski I. (eds.) (2016). Animal migration tracking methods. In: *Stable Isotopes to Trace Migratory Birds and to Identify Harmful Diseases*. Springer, Switzerland, pp.11-33.

Vitousek P.M., Mooney H.A., Lubchenco J. and Melillo J.M. (1997). Human domination of Earth's ecosystems. *Science*, *277*(5325), pp.494-499.

Vogel J.C., Eglington B. and Auret J. (1990). Isotope fingerprints in elephant bone and ivory. *Nature*, *346*(6286), p.747.

Vogel J.C., Fuls A., and Ellis R.P. (1978). The geographical distribution of Kranz grasses in South Africa. *South African Journal of Science 74*, 209-215.

von Holstein I.C., Von Tersch M., Coutu A.N., Penkman K.E., Makarewicz C.A. and Collins M.J. (2018). Collagen proteins exchange oxygen with demineralisation and gelatinisation reagents

and also with atmospheric moisture. *Rapid Communications in Mass Spectrometry*, *32*(6), pp.523-534.

Vos P., Garrity G., Jones D., Krieg N.R., Ludwig W., Rainey F.A., Schleifer K.-H. and Whitman W.B. (2011). *Bergey's manual of systematic bacteriology: Volume 3: The Firmicutes*. Springer Science & Business Media.

Vosloo W., Bastos A., Michel A. and Thomson G. (2001). Tracing movement of African buffalo in southern Africa. *Revue Scientifique et Technique-Office International des Epizooties*, *20*(2), pp.630-635.

Wagner H. (2004). *New evidence suggests that monkey thought extinct still exists*. Ohio State University, Research Communications. Available online at: https://researchnews.osu.edu/archive/redcolo.htm. Accessed on 10 May 2017.

Wang B., Yang W., Sherman V.R. and Meyers M.A. (2016). Pangolin armor: overlapping, structure, and mechanical properties of the keratinous scales. *Acta Biomaterialia*, *41*, pp.60-74.

Warchol G.L. (2004). The transnational illegal wildlife trade. *Criminal Justice Studies*, *17*(1), pp.57-73.

Wasser S.K., Shedlock A.M., Comstock K., Ostrander E.A., Mutayoba B. and Stephens M. (2004). Assigning African elephant DNA to geographic region of origin: applications to the ivory trade. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(41), pp.14847-14852.

Wassenaar L.I. and Hobson K.A. (2006). Stable-hydrogen isotope heterogeneity in keratinous materials: mass spectrometry and migratory wildlife tissue subsampling strategies. *Rapid Communications in Mass Spectrometry: An International Journal Devoted to the Rapid Dissemination of Up-to-the-Minute Research in Mass Spectrometry*, *20*(16), pp.2505-2510.

Webb E.C., Stewart A., Miller B., Tarlton J. and Evershed R.P. (2016). Age effects and the influence of varying proportions of terrestrial and marine dietary protein on the stable nitrogen-isotope compositions of pig bone collagen and soft tissues from a controlled feeding experiment. *STAR: Science & Technology of Archaeological Research*, *2*(1), pp.54-66.

West J.B., Bowen G.J., Cerling T.E. and Ehleringer J.R. (2006). Stable isotopes as one of nature's ecological recorders. *Trends in Ecology & Evolution*, *21*(7), pp.408-414.

Westphal M.I., Browne M., MacKinnon K. and Noble I. (2008). The link between international trade and the global distribution of invasive alien species. *Biological Invasions*, *10*(4), pp.391-398.

Willis A.T. (1977). *Anaerobic bacteriology: Clinical and laboratory practice*. Butterworth, London.

Wong W.W., Cochran W.J., Klish W.J., Smith E.O., Lee L.S. and Klein P.D. (1988). In vivo isotope-fractionation factors and the measurement of deuterium-and oxygen-18-dilution spaces from plasma, urine, saliva, respiratory water vapor, and carbon dioxide. *The American Journal of Clinical Nutrition*, *47*(1), pp.1-6.

Woo P., Lau S., Teng J., Tse H. and Yuen K.-Y. (2008). Then and now: use of 16S rDNA gene sequencing for bacterial identification and discovery of novel bacteria in clinical microbiology laboratories. *Clinical Microbiology and Infection*, *14*(10), pp.908-934.

Woodford M.H. (2000). *Quarantine and health screening protocols for wildlife prior to translocation and release into the wild*. Wildlife Disease and Zoonotics, p.88.

World Bank (2010). People, pathogens, and our planet. Volume one: towards a one health approach for controlling zoonotic diseases. World Bank, Washington DC.

Wu L., Thompson D.K., Liu X., Fields M.W., Bagwell C.E., Tiedje J.M. and Zhou J. (2004). Development and evaluation of microarray-based whole-genome hybridization for detection of microorganisms within the context of environmental applications. *Environmental Science & Technology*, *38*(24), pp.6775-6782.

Wu Z., Yang L., Ren X., Zhang J., Yang F., Zhang S. and Jin Q. (2015). ORF8-related genetic evidence for Chinese horseshoe bats as the source of human severe acute respiratory syndrome coronavirus. *The Journal of Infectious Diseases*, *213*(4), pp.579-583.

Xu J. (2006). Invited review: microbial ecology in the age of genomics and metagenomics: concepts, tools, and recent advances. *Molecular Ecology*, *15*(7), pp.1713-1731.

Xu L., Guan J., Lau W. and Xiao Y. (2016). An overview of pangolin trade in China. *TRAFFIC Briefing Paper*, p.10.

Young J. (2001). Implications of alternative classifications and horizontal gene transfer for bacterial taxonomy. *International Journal of Systematic and Evolutionary Microbiology*, *51*(3), pp.945-953.

Yutin, N. and Galperin, M. (2013). A genomic update on clostridial phylogeny: Gram-negative spore formers and other misplaced clostridia. *Environmental Microbiology*, *15*(10), pp.2631-2641.

Zhang H., Miller M.P., Yang F., Chan H.K., Gaubert P., Ades G. and Fischer G.A. (2015). Molecular tracing of confiscated pangolin scales for conservation and illegal trade monitoring in Southeast Asia. *Global Ecology and Conservation*, *4*, pp.414-422.

Zhang J., Kapli P., Pavlidis P. and Stamatakis A. (2013). A general species delimitation method with applications to phylogenetic placements. *Bioinformatics*, *29*(22), pp.2869-2876.

Zhou J., He Z., Yang Y., Deng Y., Tringe S.G. and Alvarez-Cohen L. (2015). High-throughput metagenomic technologies for complex microbial community analysis: open and closed formats. *MBio*, *6*(1), pp.e02288-02214.

Ziegler S., Merker S., Streit B., Boner M. and Jacob D.E. (2016). Towards understanding isotope variability in elephant ivory to establish isotopic profiling and source-area determination. *Biological Conservation*, *197*, pp.154-163.