Use of an 18s rRNA metagenomics approach as a method of detection of multiple infections in field blood samples collected on FTA cards from Cattle.

Dominic West

School of Environment and Life Sciences, University of Salford, UK

Submitted in Partial Fulfilment of the Requirements of the Degree of Masters of Science by Research, 14th June 2018

Table of Contents

Abstract

Parasitic disease causes major health problems in humans and animals globally. Diseases in tropical and remote areas, sometimes known as neglected tropical diseases, pose some of the greater problems due to the difficulty in collecting and analysing samples. Collecting samples from remote and tropical areas is difficult, especially for epidemiological studies because of the problems associated with transport and storage of samples. The use of technologies such as FTA cards, that fix and store nucleic acids direct from tissues, have transformed the collection of field samples for PCR. However, samples from FTA cards have yet to be used for multiple diseases identification. The majority of studies look to identify a single disease or, using several techniques, laboriously identify a small number of diseases from a single sample.

With the use of Next Generation Sequencing, this study looked to break new ground by identifying multiple pathogenic DNA from FTA card samples taken from cattle by amplifying the variable the V9 region of the 18s sequence using primers designed to target multiple eukaryotic sequences. A number of protocols were modified and optimised in order to elute a sufficient quality and quantity of DNA to use for NGS. Due to the age of the FTA cards, the samples were more difficult to elute from. The results of the NGS showed that multiple diseases were successfully identified to a genus level, sequencing expected pathogens such as *Theileria* and *Eimeria*. Also identifying *Leptomonas*, a relative of the Trypanosomatidae family. There were also a large number of bacteria and fungus identified, which were unexpected and were as a result of the age of the FTA cards used. Some of these bacteria are known to be detritivores and could play a role in breaking down substances on the cards, shortening their shelf life. The age of the cards was a major limitation as it made it difficult to elute DNA. The success of this study will allow for many samples to be collected from remote areas and tested via FTA cards. The next step will be to conduct a similar experiment on a larger scale with experimentally infected FTA cards and then move forward onto testing FTA card samples collected from humans.

1.0 Introduction

1.1 Aims

The collection, transport and use of blood samples collected from animals and humans in remote and tropical areas of the world can be difficult, particularly when such samples are used for epidemiological studies. FTA cards have revolutionised the collection and preservation of blood samples taken from the field and PCR based methods have been used to detect parasitic and other diseases. However, to date, these cards have only been used to detect single diseases or very arduously detect multiple diseases with limited success. With the use of Next Generation sequencing, the aim of this project is to create a series of protocols to allow the detection and amplification of DNA from multiple pathogens found in the blood samples collected on FTA cards. Alongside this, it is also an aim to identify whether control samples collected from around the blood sample on clean FTA cards contain the same pathogens contained within the blood sample.

1.2 African Cattle Diseases

A large number of parasites and pathogens impact farming and care of livestock in Sub-Saharan Africa, resulting in economic losses to the impacted communities. A study by Kristjanson *et al* (1999) estimated that, without veterinary assistance, up to 25% of livestock are killed by disease-causing damages of US$120-180 per annum per farmer. This is a significant amount as the majority of the farmers live in poor rural communities and live off as little as US$1 a day. Hemoprotozoan parasites, such as *Trypanosoma*, *Babesia,* and *Theilaria* are some of the major parasites, from the order Piroplasmida, that impact cattle farming in Eastern, Western and Southern Africa (Musinguzi *et al*, 2016). These diseases, along with others such as Brucellosis, are known as neglected tropical diseases. They have received little funding for research and prevention but still cause significant damage to economies as well as being a major public health risk (Ducrotoy *et al*, 2014). Many vector-borne diseases such as these are the cause of increased levels of mortality

and morbidity in cattle populations, with infections leading to decreased milk and meat production. Not only does this impact revenue, it also has further economic implications with the outlay on vector control measures (Makalaa, Manganic, Fujisakia, & Nagasawa, 2003).

Brucellosis is known to be one of the most widespread and common global zoonoses. Caused by the genus *Brucella*, it causes infertility and abortions in livestock and is highly infectious and is passed on by contact with aborted fetuses, placental fluid, vaginal fluid and milk (Ducrotoy *et al*, 2014). Data collected on *Brucella* species show that the financial losses are not only from the animal production but also impact on public health. Though no exact figures have been gathered for Africa, estimates of financial loss in Latin America due to Bovine Brucellosis is close to US$600 million. Bovine Brucellosis main symptoms are abortions, premature births of stillborn calves. There is also a 25% decrease in milk production by cows infected with *Brucella*. The Brucellae locate themselves in the mammary glands and so continually produce pathogens in the milk (Ducrotoy *et al*, 2014).

Tick-borne pathogens make up a large proportion of cattle diseases in Africa. Babesiosis and Anaplasmosis are two tick-borne diseases that are prevalent across a lot of sub-Saharan Africa. Babesiosis, transmitted by ticks of the genus *Boophilus,* is caused by caused by the parasites *Babesia bovis* and *B. bigemina.* Bovine Anaplasmosis is caused by *Anaptasma marginale,* which can be transmitted by multiple tick species (Jongejan el at, 1988). In areas such as Zambia and Zimbabwe, Babesiosis is not thought of as a major disease, however, if *B. bovis* is spread into farms outside of the distributional range of its vector *B. microplus* it can be fatal (Jongejan el at, 1988). A study carried out in Nigeria discovered that 82.6% of cattle were infected with tick-borne diseases. The results also showed a large variety of pathogens, often occurring alongside one another in complex co-infection situations. The most common pathogens were *Theilaria* species and *Anaplasma* species, occurring at a prevalence of at least 30%, with *Babesia bovis* and a number of other zoonotic pathogens being less common, occurring below 10% (Lorusso *et al*, 2016). *Theileria mutans* and *T. velifera* were discovered to have the highest prevalence in the cattle population and was discovered in all age ranges of cattle. Even with a high

number of tick-borne pathogens located in the cattle populations, there is still limited use of acaricides by the Fulani pastoralists. Farmers rely on the removal of ticks by hand from the more obvious areas of the cow, such as the udder. This combined with the lack of consideration of co-infection may have an impact on the control of tick-borne diseases in the area (Lorusso *et al*, 2016).

Echinococcus granulosus sensu lato, the parasite that causes Cystic Echinococcosis, has a worldwide distribution with an incidence of 200 cases per 100,000 in highly endemic areas (Eckert, Gemmell, Meslin, & Pawłowski, 2001). The disease is predominantly found in farming communities due to its intermediate hosts mainly infecting livestock such as cattle, sheep, and pigs. The definitive host is the domestic dog, which is commonly found in farming communities (Bouéa *et al*, 2017).

1.2.1 Trypanosome Disease

Trypanosomiasis is a disease caused by a protozoan parasite of the genus *Trypanosoma* and affects both humans and animals. There are several species of trypanosomes, *Trypanosoma congolense*, and *Trypanosoma vivax* that parasitise animals, and *Trypanosoma brucei* which has 3 sub-species. *T. brucei brucei that* infects animals and *T. brucei rhodesiense* and *T. brucei gambiense* that both infect humans, causing sleeping sickness (Laohasinnarong *et al*, 2011). Trypanosomiasis is a neglected tropical disease (NTD) transmitted by the Tsetse fly, which belongs to the genus *Glossina* (Fèvre *et al*, 2005). Of the species that infect humans, *T.b gambiense* is known to be a chronic disease that is transmitted to humans by vertical transmission through human-tsetse interaction. The other human African trypanosomiasis (HAT) species *T. b. rhodesiense* is an acute disease and has a more complicated epidemiology that involves transmission by tsetse flies from a range of wildlife, human and domestic animal reservoirs. If untreated, both species of HAT can be fatal, with the acute *T. b. rhodesiense* causing death in 6-8 months post initial detection of infection (Fyfe *et al*, 2016), however, the chronic *T. b. gambiense,* if untreated can be fatal, but there is evidence that infected individuals can harbor the parasite at low levels for more than 5 years (Checchi *et al*, 2008). Trypanosomiasis has a huge impact on rural and poor communities in developing countries as the disease only occurs remote rural areas. It is estimated that around 1.6 million

disability-adjusted life years (DALYs) are lost per year. In Soroti, Uganda, sleeping sickness caused by the *T. brucei rhodesiense* causes a load per case that is 78 times greater than that of malaria. Human and animal trypanosomes also increase their range along with livestock movements, with outbreaks occurring as a result of untested livestock movements via markets and trading (Welburn *et al*, 2006). It has been valued that animal Trypanosomiasis costs US$1.3 Billion annually to both the producers as well as the customers who consume the livestock product (Kristjanson *et al*, 1999). In areas where tsetse flies are located, the density of cattle populations is reduced by 37-70%. This impact on the meat and milk production is as much as 50% (Swallow, 2000). Current methods of screening include techniques such as dark ground microscopy and epidemiological testing of buffy coat samples. In recent times, methods have moved towards molecular-based testing making use of Polymerase Chain Reaction (PCR) techniques. Microscopy techniques are less commonly used as they lack sensitivity when working in the field as a result of the low parasitemia found in infected animals and require large amounts of labour. Setbacks such as the lack of sensitivity have been countered by using diagnostic techniques involving PCR. The increased sensitivity and specificity allow for detailed diagnosis of a number of different sub-species, however, this complex process, which can involve a number of different PCRs for each sample and so increases the expense of each diagnosis and is a difficult technique to use for diagnosis of bigger populations (Cox *et al*, 2005). Samples used in these methods are collected through direct blood samples and buffy coat samples. Some are collected and stored as they were removed, and others are collected by applying them to filter paper. This is seen as a useful way to collect samples on a large scale and doing DNA extraction *in situ*. However, this method of using filter paper does not allow for long-term storage and/or transport to laboratories away from the field. This is because standard filter paper does not prevent degradation of DNA and doesn't prevent the sample from spoiling (Ahmed *et al*, 2011). Experimental studies have shown that samples do not have to be collected directly from a mammalian host. A study by Boid, Jones, & Munro (1999) extracted trypanosome DNA from blood meals of *Stomoxyscalcitrans calcitrans* that had been fed on donkey blood infected with *T. evansi.* The flies were euthanized by snap freezing at -20 degrees centigrade and their gut contents were

smeared onto Whatman No.1 filter paper. DNA was eluted from the filter paper and the product was assayed to detect *T. evansi* by PCR.

Due to low parasitemia levels, diagnosis of African Trypanosomiasis is difficult. Looking for antibodies is a simple way of screening for the parasite however it cannot distinguish between old infections and currently active infection (de Almeida, Ndao, Van Meirvenne & Geerts, 1998). With new developments in PCR diagnosis, specific DNA hybridization probes have now been produced to diagnose specific Trypanosome species (Delespauxa, Ayralb, Geysena, & Geertsa, 2003). A study by Cox *et al* (2005) developed a nested PCR that includes both internal transcribed spacer (ITS) regions of the ribosomal RNA genes. The ITS PCR is capable of finding Trypanosomes in a sample that contains host DNA as well as PCR inhibitors found in blood. This PCR has been shown to be able to distinguish between Trypanosome species and some of its sub-species as well as being sensitive enough to detect a single parasite found in a sample.

1.3 FTA Cards used to collect DNA samples

FTA cards are a tool that has replaced the use of regular filter paper to collect DNA samples from the field. The application of samples such as blood or other tissues to common filter paper has been used as a method of sampling on a sizeable scale because it is more effective than other sampling methods, however, it is not long-lasting, and samples are not safeguarded from spoiling and degradation (Ahmed *et al*, 2011). FTA cards are devised to trap and collect DNA from tissue that is deposited on the treated area of the card. The set-up of the filter paper FTA card is permeated with a chaotropic mediator that denatures infectious agents. Chemicals already in the paper lyse any cells and hold and DNA and RNA within the filter paper matrix. This makes the samples non-infectious. This is a further benefit because transport it allows for simple transport through customs and avoids having to acquire government permits (Picard-Meyer, Barrat & Cliquet, 2006). The use of FTA cards allows for long-term storage of field samples and the ability to maintain the viability of nucleic acids for up to 10 years at room temperature (Picozzi *et al*, 2002).

In order to use the sample collected on FTA cards, they need to be prepared in accordance with specific protocols. Small punches are cut from the cards and washed in chemicals to aid the removal of and DNA present. The punches are then added to a PCR tube ready for PCR amplification of the target DNA (Cox *et al*, 2005).

FTA cards are used to store many different infectious agents. Studies have shown that FTA cards are suitable for immobilizing rabies virus in suspension. A study by Picard-Meyer, Barrat & Cliquet (2006) found that using FTA cards provides a non-complex and accurate device that complements the molecular studies involving rabies diagnosis. The same study discovered that isolating RNA from samples found to be infected with rabies was easily achieved permitting further molecular studies. A study on African trypanosomes has shown some minor limitations in the use of FTA. One of which is the sporadic covering of genetic material over the card (Ahmed *et al*, 2011). Using more punches from the sample increased the sensitivity from 5.9% when one to 5 discs were used to 35.6% when 10 discs were used, however, the use of lysed blood from the cards showed an increase to 73.3% when 10 discs were analysed. Though using more cuts increases the sensitivity, the use of 10 discs per PCR is thought to be excessive use of limited samples as well as being costly. It also would take longer to gather results (Ahmed *et al*, 2011). Such problems can be avoided however by the elution of DNA from samples on FTA cards. Chelex 100, a chelating resin that has a pH 11 aids the removal of material that can act as an inhibitor to PCR reactions and so enables the extraction of very pure template DNA. Chelex 100 also supports the disassociation of DNA into its single strands, in turn aiding the PCR reaction (Becker *et al*, 2004).

FTA cards have been used in areas where collecting and maintaining samples in a suitable condition from areas where the environment causes the main issue. In North and West Africa, researchers found that keeping collected samples from slaughterhouses in good condition when looking for *Echinococcus* species was difficult due to the high temperature and humidity, as well as long distances that the samples would need to travel. The use of FTA cards allowed for samples of the larval stages of *Echinococcus* species to be collected and transported to laboratories where PCR-based technologies allowed for the identification of the species (Bouéa *et al*, 2017). There is limited epidemiological data on Cystic Echinococcosis as a result of

the difficulties encountered when collecting samples. The use of FTA cards provided researchers to transport and store samples reliably and allow for analysis of these samples more accurately (Bouéa *et al*, 2017).

1.4 Molecular Detection

1.4.1 Protein-Based detection of Trypanosomes

Detection of trypanosomes has developed over time. The use of gel electrophoretic separation of isoenzymes was used to distinguish between morphologically similar parasites and identify different species and subspecies of both *Trypanosoma* and *Plasmodium* (Hide & Tait, 2009). Other techniques involved using DEAE-cellulose columns to separate trypanosomes from the blood by ion-exchange. Often the parasite was isolated from mice that had been experimentally infected (Lanham, & Godfrey, 1970). It was from these techniques that identification of parasites developed into more readily used protocols seen today.

1.4.2 Polymerase Chain Reaction

Traditional diagnostic methods for parasites and pathogens such as examination of buffy coat samples and microscopy-based techniques are becoming out dated due to their labour-intensive requirements and lack of sensitivity in more difficult environments like conditions found where tropical samples are collected. With levels of parasite presence often being low (Cox *et al*, 2005). Serology testing is also limited to being unable to accurately decipher between past and present infections of parasites such as Trypanosomes. They also can result in a large number of false-positive results leading to the prescription of treatment drugs with possible harmful side effects (Becker *et al*, 2004). Polymerase Chain Reaction (PCR) has become one of the more sensitive methods of detection for a diverse number of protozoan pathogens. However, the equipment needed for PCR does make it costly to use in the field (Laohasinnarong *et al*, 2011). PCR detection is used in the identification of pathogens by using primers designed to match a conserved region of DNA that is previously known (Bonnet *et al*, 2014). The use of PCR methods to amplify genes have also assisted in the detection of parasites because they offer greater sensitivity and can also target specific genes of individual species. Not only

this, molecular-based techniques targeting nucleic acids, such as nested PCR and real-time PCR also enable discovery of a parasite when there are low levels of parasitemia in the blood. PCR can also detect pathogens in patients that show little or no symptoms (Ndao, 2009). Developments in molecular diagnostic techniques have allowed for specific species identification, for example, nested PCR such as Internal Transcribed spacer (ITS) which target both the Internal Transcribed Spacers of ribosomal RNA of Trypanosomal sequences. This allows for the parasite to be amplified despite the presence of host DNA and the naturally occurring PCR inhibitors that are present in blood such as lactoferrin, heme and non-target DNA (Cox *et al*, 2005). PCR can also be used in a more general setting, to identify variance of species within a sample with a possible multitude of DNA within it (Jarman, Deagle & Gales, 2004).

1.4.3 PCR Blocking Primer

PCR reactions usually target specific sequences of DNA. Individual species DNA sequences are relatively simple to identify due to the unique identity of the sequence. However, when using universal primers to identify multiple DNA sequences, there is often a higher prevalence of a single sequence such as host DNA sequences in an individual sample. This can cause bias in results when testing samples using PCR or other molecular detection methods (Vestheim & Jarman, 2008). PCR based detection methods using universal primers are not sensitive enough to pick up low levels of DNA present in samples. Host DNA is more abundant and so inhibits the binding of less abundant DNA of non-host material to the primers (Vestheim & Jarman, 2008). When looking for eukaryotic microbial 18s sequences, it is difficult to identify and amplify such sequences as the eukaryotic host has its own 18s gene. As this gene is more abundant in the sample, inhibits the binding of microbial 18s sequences by binding to the primers more readily (Belda *et al*, 2017). To combat this, a blocking primer, also known as annealing blockers, can be used to prevent amplification of host DNA. A study on the diet of Australian Fur Seals used an adapted non-extendable primer that binds to the Fur Seal mtDNA only, allowing target DNA to be amplified. The blocking oligonucleotide must be specific to the host or group of hosts DNA that it is inhibiting. (Deagle, Kirkwood & Jarman, 2009). In this

study, the blocking primer is used to prevent the binding and amplification of mammalian 18s gene sequences. The primer (Mammal_block_I-short_1391f) was designed by the Earth Microbiome project (Earth Microbiome Project, 2016). These blocking primers have two separate regions that are connected by a polydeoxyinosine link. The 5' region is larger and binds to its complimentary region on the target DNA as the universal primer would. The 3' region is smaller and binds to the non-target DNA (Belda *et al*, 2017). In this case, the mammalian DNA is the non-target sequence. Mammalian sequences able to be blocked due to the highly conserved primer flanking sites of the V9 region of the 18s rRNA gene (Amaral-Zettler *et al,* 2009). The mammal blocking primer stops the extension of the primer and blocks the amplification of the host DNA. A 3-carbon spacer region on the 3' end of the sequence, which prevents the amplification, causes this. Using blocking primers allows for the discovery of low-level DNA, by blocking host sequences, DNA that is only present in small quantities is then able to be amplified (Shehzad *et al*, 2012). Annealing blockers are more efficient at preventing amplification of host DNA than other techniques such as using restriction enzymes to cut out and remove host DNA prior to PCR amplification because it prevents the amplification of host DNA instantly during the PCR process where as restriction digestion does not completely remove all host DNA from a sample (Leray *et al*, 2013).

1.4.4 Next Generation Sequencing

The use of next-generation sequencing (NGS) has begun to revolutionize detection and identification of recognized and novel pathogens. NGS uses the DNA that is to be sequenced to create a library of short fragments (Mardis, 2013). These fragments have adaptors attached which often have barcodes to identify individual samples. A PCR reaction then amplifies fragments of DNA using primers also containing the same adaptors. During sequencing, fragments are immobilized on a flow cell and clonally amplified. This immobilization is also to ensure that only one fragment of DNA occupies a space on the flow cell where sequences hybridize to part of the adaptors. This clonal amplification makes clusters of identical copies of the original DNA. Fluorescently tagged nucleotides are added and compete for space next to the separated DNA fragments and bind. A laser is used to excite the

fluorescent nucleotides and the emission is recorded by a computer for each cluster. Bioinformatics is used to interpret the raw data produced and create a visual representation of the sequences (Yohe & Thyagarajan, 2017). NGS has been used to identify variations in pathogens such as *Theileria* in Cape Buffalo and cattle. Studies conducted in Southern Africa used NGS to identify whether there are any new and unknown genotypes within populations of bovine species in Southern Africa. Following this, the experiment also identifies geographical locations of these pathogens (Mansa *et al*, 2016). The use of PCR in the identification of new pathogen species or species that are unexpected within a sample is restricted by the specificity of primers to an individual sequence of DNA. This can be overcome by the amplification of 18s sequences, which are highly conserved within among pathogens, especially parasitic pathogens (Bonnet *et al*, 2014). NGS has also been used in identifying new bovine pathogens. A study into the death of a cow with encephalitis was unable to identify the causative agent using diagnostic techniques more commonly used. The use of NGS identified a sequence that was similar to other astrovirus sequences. It was subsequently identified as a new bovine astrovirus and named BoAstV-BH89/14 (Schlottau *et al*, 2016). Without the use of NGS, such diseases are difficult to identify. Improvements in NGS have allowed for a greater ability to detect a single sequence of DNA sequences in a large volume of host DNA. Using NGS to target gene regions such as 18s rRNA region using universal primers, allows for detection of pathogens without the need for a species-specific primer, which in order to work, requires the target pathogen to be known in advance. This also allows for multiple pathogens to be detected from samples from complicated samples. This form of detection is also very efficient and becoming less expensive and labour-intensive (Cooper *et al*, 2016). Using NGS approaches, techniques such as shotgun metagenomics can identify all microbes present in a sample by sequencing all nucleic acids obtained from a sample. Shotgun sequencing is also less bias and provides a more accurate representation of what is present within a sample. (Muller *et al*, 2013). NGS is not only used for detection of pathogens but also as a tool to discover their genetic diversity and look at their evolution. This is vital in the constant battle against drug resistance of the likes of malaria (Rao *et al*, 2016). It has been used to characterize microsatellite loci of

Fasciola hepatica, a parasitic liver fluke of economic importance infecting cattle and sheep. *Fasciola hepatica* has a complicated reproduction cycle in the mammalian host which allows for large amounts of gene flow and as a result, creates the chance for large amounts of genetic variation within a population. Using NGS meant that the mitochondrial genome of two geographically different *F. hepatica* isolates has given a greater understanding into the intra-specific variation of the parasite (Cwiklinski *et al*, 2015). Metagenomic approaches have shown a number of examples where detection of pathogens has been missed by conventional detection methods. These normally occur when the pathogen found is not suspected to be the causative agent and so was not identified by the test used. Also, it can be the case that the pathogen is present in very low quantities and only detectable by NGS (Muller *et al*, 2013). Sequencing whole genomes is vital in the development of new drugs through the identification of new regions of genomic material, which can be progressed into targets for drug development. It also can be used to map and control the spread of drug resistance (Auburn *et al*, 2011). Recent research by Hino, Maruyama & Kikuchi, 2016 successfully identified various parasites from rat faeces. In doing so, the study also found that this method of identification was equally sensitive as methods previously used and was also a faster, more dependable method of detection. In turn, this will be used in the current study.

The use of 18s rRNA as a universal primer opposed to 16s is due to the higher copy number in 18s sequences. Copy numbers are greater in 18s rRNA found in the Eukaryote genome compared to the 16s Prokaryote genome. The copy number of an 18s sequence is related to the eukaryotic genome, whereas the copy number of 16s can vary depending on required resources (Wang *et al*, 2014). 18s sequences are highly conserved in all living cells, which makes it a useful target of universal primers (Hadziavdic *et al*, 2014).

NGS has been used on samples collected on FTA cards for genotyping of cattle in the USA. A study was conducted to see whether samples collected on FTA cards would produce DNA that could be used for high-throughput genotyping using the Illumina iSelect BeadChips. It was proven that FTA cards stored and protected DNA sufficiently for High-throughput genotyping to the same level as DNA directly sourced from a sample (McClure *et al*, 2009). A similar experiment was successfully

conducted on cell material taken from tumours looking at cancer treatment (Saieg *et al*, 2012).

1.5 Objectives

In order to complete this project, a number of objectives were established. The main objectives were as follows:

- To extract host DNA from Bovine blood stored onto FTA cards using Chelex 100 protocol of a quality to amplify by PCR the Tubulin housekeeping gene.
- To extract Trypanosomal DNA from Bovine blood stored onto FTA cards using Chelex 100 protocol of a quality to amplify by ITS PCR.
- To extract DNA from multiple pathogens present in Bovine blood stored onto FTA cards, using Chelex 100 protocol, of a quality and quantity to amplify via PCR.
- To extract DNA from multiple pathogens present in Bovine blood stored onto FTA cards, using Chelex 100 protocol, of a quality and quantity to amplify via Next Generation Sequencing.
- To analyse NGS sequence data to identify Operational Taxonomic Units (OTUs) and also identify similarities between locations on the FTA cards the samples were taken from.

2.0 Methods

2.1 Ethics

For this study to be carried out, an ethics checklist was carried out with the supervisor, Professor Geoff Hide, and as there was no collection or handling biohazardous samples, there were no ethical issues so further levels of approval were not needed. This was also completed alongside the risk assessment.

2.2 Collection of Cattle Samples from Nigeria.

A total of 80 bovine blood samples were collected from Ahoada, Nigeria, by Eze Justin Ideozu as part of his Ph.D. In this study only 70 samples were used. All of the samples were of *Bos indicus* subspecies and were identified by their humpback (Loftus *et al*, 1994). The samples were collected by butchers from an abattoir in Ahoada and stored on Whatman FTA cards and prepared using the protocol described in Chapter 2.2.

Table 1: Sample codes for 70 Bulls. 5 of the samples were duplicated and analysis was carried out 4 times each (Sample numbers 24, 25, 28, 30 and 66). 10 control samples collected from the area surrounding the blood sample.

Sample Code									
AH01	AH19	AH25	AH29	AH35	AH44	AH53	AH62	AH69	N30
AH10	AH02	AH25a	AH03	AH36	AH45	AH54	AH63	AH07	N41
AH11	AH20	AH25b	AH30	AH37	AH46	AH55	AH64	AH70	N05
AH12	AH21	AH25c	AH30a	AH38	AH47	AH56	AH65	AH08	N60
AH13	AH22	AH26	AH30b	AH39	AH48	AH57	AH66	AH09	N68
AH14	AH23	AH27	AH30c	AH04	AH49	AH58	AH66a	N01	
AH15	AH24	AH28	AH31	AH40	AH05	AH59	AH66b	N10	
AH16	AH24a	AH28a	AH32	AH41	AH50	AH06	AH66c	N21	
AH17	AH24b	AH28b	AH33	AH42	AH51	AH60	AH67	N24	
AH18	AH24c	AH28c	AH34	AH43	AH52	AH61	AH68	N28	

2.3 DNA Extraction from FTA cards containing Bovine blood samples.

2.2.1 Purification of DNA from FTA cards to be used in PCR.

Initially, all equipment was cleaned with 70% ethanol and put under UV light for 20 minutes. A punch was taken from each sample 10 times at a size of 2mm using a Harris punch and placed in 1.5ml Eppendorf tube. The Harris Punch cutter was cleaned between each sample by cutting out 5 punches from clean Whatman filter paper on a Harris cutting mat to avoid any cross-contamination. A control sample was cut from filter paper after every 10 samples. 1ml of FTA wash was added to each Eppendorf and then all the samples were placed on a shaker for 15 minutes at ambient temperature. This FTA wash was then aspirated off and then another 1ml of fresh FTA wash was added and the samples were placed back on a shaker for a further 15 minutes and aspirated off again. A volume of 1ml of 1x TE buffer was added to the samples and they were placed on a shaker for 15 minutes. The TE buffer was then aspirated off and the wash is repeated. Each of the 10 punches from all the samples was then transferred into PCR tubes using a 100µl filtered pipette tip and placed in an incubator at 37°C for 40 minutes or was left overnight at room temperature to dry.

2.2.2 DNA Elution in 5% Chelex from FTA cards

One ml of sterile water was added to 0.05g Chelex resin. In a sterile tube, to create a stock 5% Chelex solution. Using a sterile 200µl filtered pipette with the tip cut off, 100µl of 5% Chelex was added to the dried FTA samples and placed in a PCR machine at 90°C for 30 minutes. The eluted samples were then stored at -20°C before PCR analysis.

2.4 Amplification of DNA eluted from FTA cards.

2.3.1 Polymerase Chain Reaction.

To test the success of the extraction process, Mammalian Tubulin DNA was amplified using Tubulin primers, MTUBFOR (CGTGAGTGCATCTCCATCCAT) and MTUBREV (GCCCTCACCCACATACCAGTG) targeting the alpha-tubulin gene. The PCR process involved an initial denaturation phase, 1 cycle (94°C, 5 mins), 40 cycles of

Denaturation (94°C, 40 sec), annealing (60°C, 40 sec) and extension (72°C, 1 min 30 sec). The final phase of the PCR was 72°C for 10 mins.

In order to test the sensitivity of the extraction, a nested ITS PCR was carried out as described by Cox *et al* (2005) with additional modifications. The Primers for the first round were ITS 1 (5' GAT TAC GTC CCT GCC ATT TG 3') and ITS 2 (5' TTG TTC GCT ATC GGT CTT CC 3'). The second round of the PCR used ITS 3 (5' GGA AGC AAA AGT CGT AAC AAG G 3') and ITS4 (5' TGT TTT CTT TTC CTC CGC TG 3'). The reaction volume of 25µl was made up of 12.5µl of My Taq red mix (Bioline LTD), 0.5µl of each outer primer ITS 1 and ITS 2 and 9.5µl of PCR grade Water (Insert Manufacturer). 2µl of DNA eluted from FTA cards was added to the reaction mixture. The reaction process consisted of: 1 cycle (95°C, 1 min) 35 cycles (95°C-30 secs, 58°C-30 secs,72°C-30 secs) followed by 1 cycle (72°C-5 mins). 1µl of PCR product from the 1st round PCR was added to 24µl of new reaction mix (containing 10.5µl of water rather than 9.5µl) and containing inner primers ITS 3 and ITS 4. The reaction process was the same as previously detailed. Of the resulting PCR product, 15µl were added to a 1% agarose gel stained with GelRed and ran at a power of 110V. The gel was visualised under a transilluminator and shown using the GelSnap software.

After the ITS PCR, a further PCR was conducted to amplify the V9 region of the 18s sequence using primers designed to target multiple eukaryotic sequences in the FTA samples The PCR was carried out as described by Earth Microbiome Project. (2016).The forward primer used was Illumina_Euk_1391f (5'TCG TCG GCA GCG TCA GAT GTG TAT AAG AGA CAG GTA CAC ACC GCC CGT C 3') and the reverse primer was Illumina_EukBr (5' GTC TCG TGG GCT CGG AGA TGT GTA TAA GAG ACA GTG ATC CTT CTG CAG GTT CAC CTA C 3'). A mammalian blocking primer was also used, Mammal_block_I-short_1391f (5'GCC CGT CGC TAC TAC CGA TTG G /ideoxyI//ideoxyI//ideoxyI//ideoxyI//ideoxyI/ TTA GTG AGG CCC T/3SpC3/ 3') with a modified C3 spacer region. The PCR was carried out as described by Vestheim & Jarman, (2008) and Earth Microbiome Project. The reaction volume was 45µl, made up of 12.5µl of My Taq mix 2x (Bioline LTD), 5µl of the Illumina_Euk_1391f forward primer (1µM), 5µl of Illumina_EukBr reverse primer (1µM), 20µl of Mammalian blocking primer and 2.5µl of template DNA from the FTA cards. The reaction process consisted of: 1 cycle (95°C- 3 mins) 30 cycles (95°C- 30 secs, 65°C-30 secs, 58°C- 30

secs, 72°C- 30 secs) followed by 1 cycle (72°C- 5 mins). Of the resulting PCR product, 15µl were added to a 2% gel stained with GelRed and ran at a power of 110V. The gel was visualised under a transilluminator and shown using the GelSnap software.

2.3.2 Next Generation Sequencing from DNA amplified from PCR.

Following the amplicon PCR, the PCR product required clean up steps to prepare the samples for NGS. The clean-up used AMPure XP beads, with 20µl of the beads added to each sample. The beads were mixed into the sample by pipetting up and down 10 times. The samples were then incubated at ambient temperature for 5 minutes before being placed on a magnetic stand for a further 2 minutes until the supernatant was clear of beads. The supernatant was then removed and thrown away. After the removal of the supernatant, the beads were washed with 200µl of fresh 80% ethanol and left to stand for 30 seconds. Following this, the supernatant was removed and thrown away. A second identical ethanol wash was conducted, after which all the remaining supernatant was removed using a 10µl pipette. The samples containing the AMPure beads were then left open on the magnetic stand for 10 minutes to allow the beads to dry. Once dry, the plate was then removed from the magnetic stand and 52.5µl of PCR grade water was then added to the samples and pipetted up and down 10 times to mix the beads through the water, then left at ambient temperature for a further 2 minutes. The plate was then returned to the magnetic stand for 2 minutes until the supernatant was clear, 50µl of the supernatant was then transferred to a clean 96-well plate, ready for the index PCR.

To complete index PCR, 5µl of each sample was added to a new 96-well plate. i7 Index 1 primers were arranged from 1-12 (N716-N729) horizontally and i5 Index 2 primers were arranged A-H (S513-S522) vertically on an Illumina 96-well plate rack. Then in each well, a reaction mixture of 5µl DNA, 5µl of the corresponding Nextera XT Index Primer 1 and 2, 25µl Q5 High fidelity 2x Master mix and 10µl PCR grade water. The PCR program on a thermal cycler was 95°C for 3 minutes followed by 8 cycles of 95°C- 30 seconds, 55°C-30 seconds, 72°C for 30 seconds and a final extension of 72°C for 5 minutes. Once the PCR is complete, a second clean-up stage is required.

The same AMPure beads were used for the second clean up, with 56µl of the beads added to each sample. The beads were mixed into the sample by pipetting up and down 10 times to mix. The samples were then incubated at ambient temperature for 5 minutes before being placed on a magnetic stand for a further 2 minutes, until the supernatant was clear of beads. The supernatant was then removed and thrown away. After the removal of the supernatant, the beads were washed with 200µl of fresh 80% ethanol and left to stand for 30 seconds. Following this, the supernatant was removed and thrown away. A second identical ethanol wash was conducted, after which all the remaining supernatant was removed using a 10µl pipette. The samples containing the AMPure beads were then left open on the magnetic stand for 10 minutes to allow the beads to dry. Once dry, the plate was then removed from the magnetic stand and 52.5µl of PCR grade water was then added to the samples and pipetted up and down 10 times to mix the beads through the water, then left at ambient temperature for a further 2 minutes. The plate was then returned to the magnetic stand for 2 minutes until the supernatant was clear. Once this was achieved, 25µl of the supernatant was then transferred to a clean 96-well plate.

Each sample was quantified using a Qubit Assay, following manufacturer's instructions, providing the concentration of DNA. This allows for normalisation of the DNA and calculates volumes of each sample for pooling.

Prior to cluster generation and sequencing, the pooled DNA libraries are denatured. Initially, the concentration of the pooled sample is measured on a Qubit again; from which the result is converted into nM. A diluted volume of the 4nM-pooled library is combined with molecular H_2O to create a volume of 5µl. This is then combined with 5µl 0.2N NaOH in a microcentrifuge tube. The solution is vortexed, then centrifuged for 1 minute and left for 5 minutes at ambient temperature. Following this, 990µl of pre-chilled HT1 is added to the denatured library, creating 1ml of 20pM denatured library, and the entire solution is placed on ice. This sample was once again diluted further to produce 600µl of 12.5pM solution. This is done by adding 375µl of 20pM library to 225µl of pre-chilled HT1 in a microcentrifuge tube, then vortex and centrifuge.

2.5 Analysis of Next Generation Sequencing Data.

The data produced by NGS was analysed using R-studio and used the packages Phyloseq. Phyloseq is a software tool used for phylogenetic statistical analysis. This package contains tools that simplify reading of the data output of the most common operational taxonomic units (McMurdie, & Holmes 2013). The graphics used in Phyloseq are created by ggplot2 to produce the plots used to visualise the data (Wickham, 2009). All statistical analysis was done using excel. In order to analyse the difference in location from which the blood samples and control samples were taken, Non-metric multidimensional scaling (NMDS) was used based on the Bray-Curtis dissimilarity.

3.0 Results

3.1 The Development of a Suitable Positive Control DNA sample using a commercial source of Human genomic DNA for Tubulin PCR

3.1.1 Tubulin Positive Control

A tubulin positive control was used in a PCR to establish whether tubulin could be detected from the FTA cards. Tubulin is structurally important in eukaryotic cells and so should be present in all mammalian blood samples (Hammond, Cai, & Verhey, 2008), making it the ideal control sample to identify whether the Tubulin present in the collected samples has degraded by showing if the PCR is working.

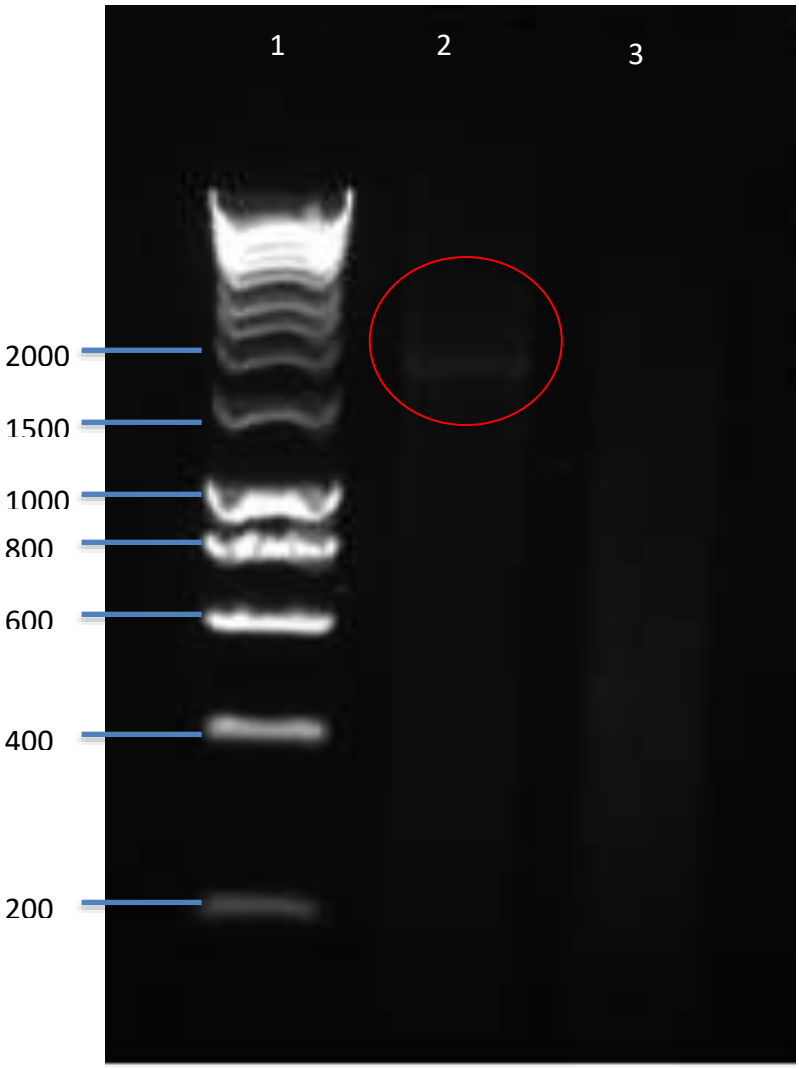


Figure 1: Electrophoresis 1% Gel of Tubulin Control PCR (2, 1600bp) and Negative Control (3), as well as 1Kb Hyperladder (1), Bioline Ltd.

Using a commercial source of complete human genomic DNA sample for PCR in Figure 1, it shows the success of the PCR. The result of this PCR meant that this control DNA was suitable to use for the control when testing for tubulin from the FTA cards. Moving forward, the task was to amplify tubulin from the FTA card blood samples. The hyperladder is not come out well and makes it difficult to read.

3.1.2 Tubulin DNA Elution and PCR from FTA cards

This experiment was conducted to attempt to elute tubulin DNA from the FTA cards and amplify by PCR. Tubulin is found in blood and so should be abundant in the FTA card samples. Finding tubulin would show whether the cards still had DNA on them or if it had degraded over time.

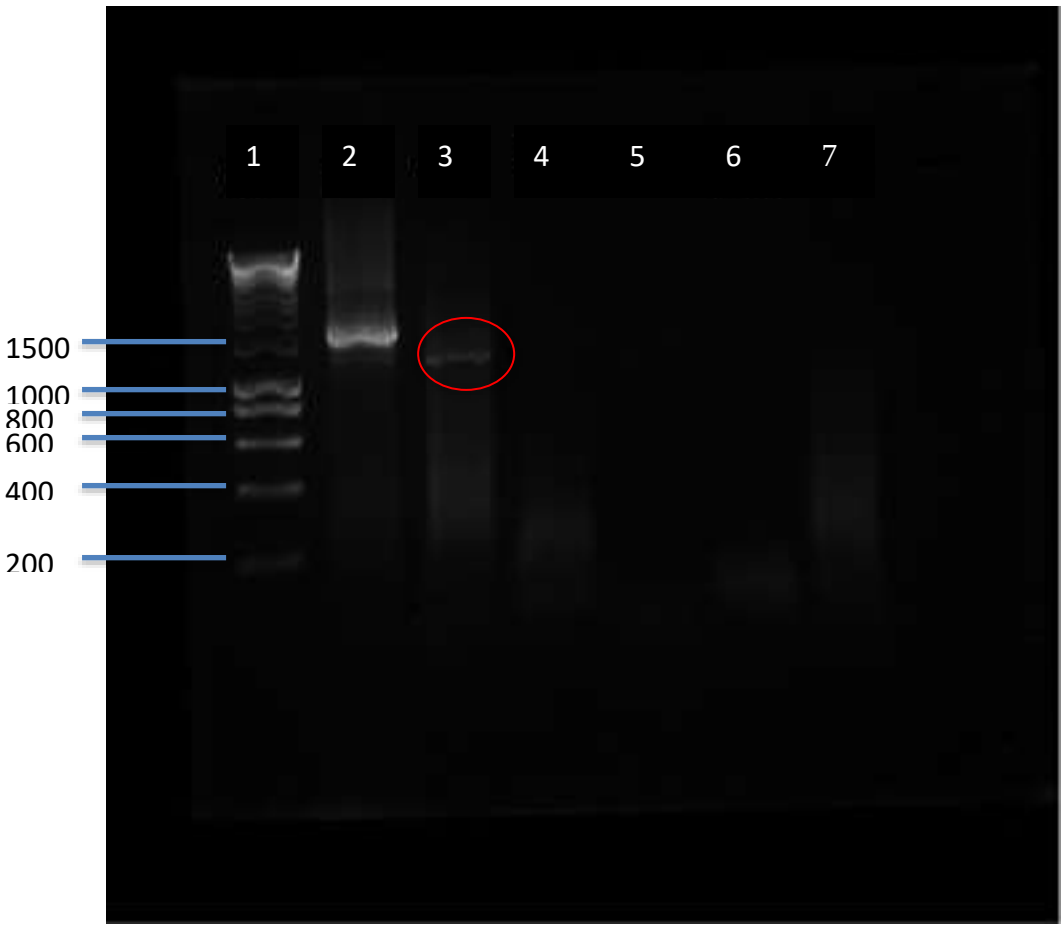


Figure 2: An electrophoresis 1% gel showing bands of a single positive sample (3) for tubulin and a positive control (2, 1600bp). from the amplificaiton of eluted DNA from FTA cards using tubulin primers. Lane 4,5,6 and 7 are also samples collected from FTA cards that were negative for the Tubulin gene.

The results in Figure 2 show that one card had DNA present on it. Sample 3 showed that tubulin was amplified from the sample and the band size was approximately 1250bp in length. This is a different size from the control as the tubulin gene varies from species to species, therefore the band size of bovine tubulin and human control tubulin will be different. Having eluted DNA off of a single card, it shows the DNA is still useable, however; optimisation was needed in order to successfully elute DNA from all samples. To improve the results, Chelex 100 was used. Studies by Ahmed *et al* (2011) showed that using Chelex 100 increased the effectiveness of DNA elution by six-fold. This was, therefore, the next process to utilize in order to obtain sufficient DNA elution.

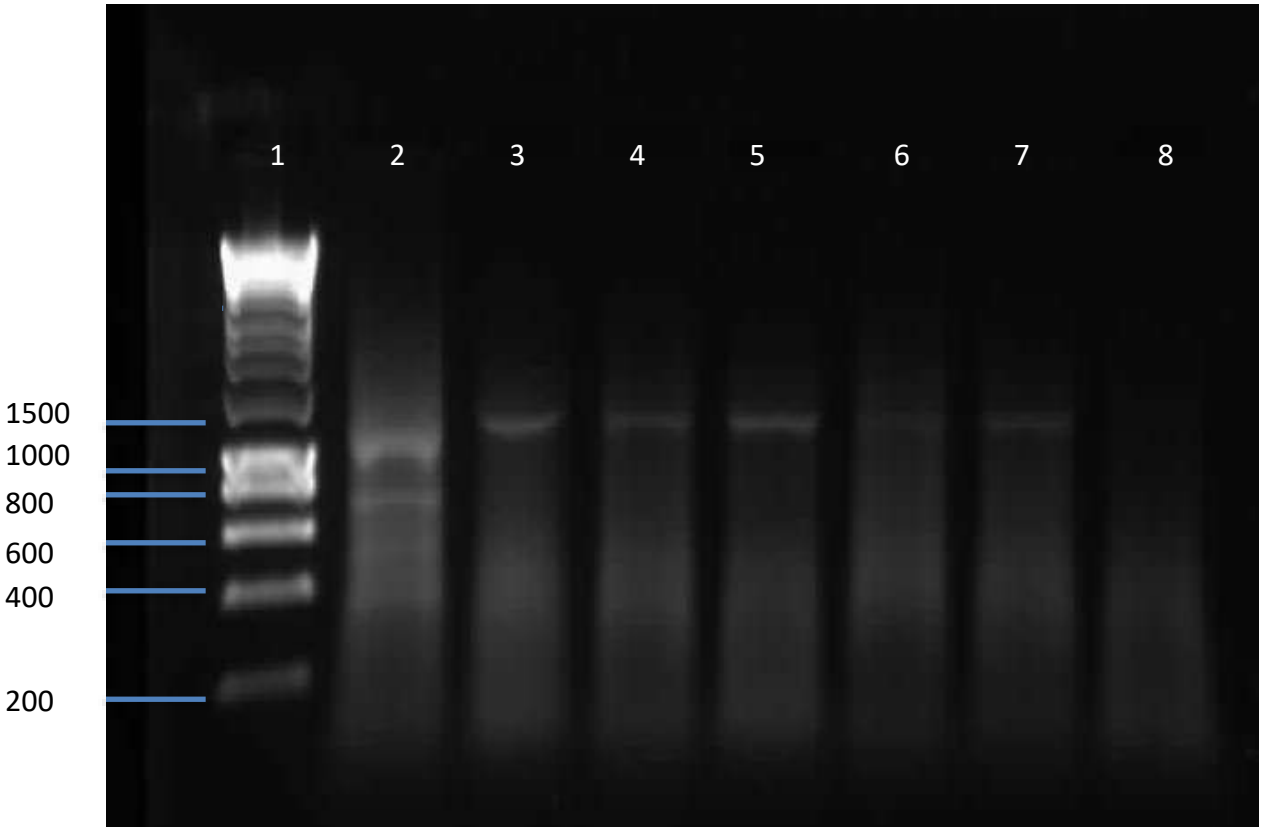


Figure 3: Electrophoresis 1% gel showing bands of DNA amplified by tubulin PCR after being eluted from FTA cards using a Chelex 100 protocol. Tubulin DNA was successfully amplified from sample 2,3,4,5 and 7. Sample 6 was negative. Lane 1 is 1Kb Hyperladder (Bioline Ltd) and lane 2 was a Tubulin positive control.

Optimisation of the elution of DNA from the FTA cards by using Chelex 100 was undertaken to amplify a better quality and quantity of DNA from the cards. Chelex 100 protects DNA at high temperatures, avoiding the degradation of DNA as well as stopping the carryover of inhibiting components of blood such as heme and lactoferrin while stabilizing the reaction (Ahmed *et al*, 2011). The outcome of using Chelex 100 was successful and DNA was eluted from 4 out of 5 cards tested. Following this success, it was possible to amplify tubulin DNA from the cards, therefore the next step was to attempt to amplify parasitic DNA from the cards.

3.2 Trypanosome PCR

3.2.1 Identifying a single parasite from FTA cards using Internal Transcribed Spacer PCR.

Internal transcribed spacer (ITS) PCR targets the most conserved region of the genome, the ribosomal DNA. It targets the ITS1 and ITS2 subunits that are located between the small subunit rRNA and 5.8S rRNA and between 5.8S rRNA and the large subunit rRNA (Kumar, & Shukla, 2005). Detecting trypanosomes from samples on FTA cards can be difficult. The parasite will remain where it is first placed on the card and does not spread uniformly. This means that a single punch may not contain the parasite and so result in a false negative when tested (Cox *et al*, 2010). For this reason, it was decided to increase the number of punches taken from the cards from 3 to 10. This increases the chance of individual parasites being detected and their DNA being amplified.

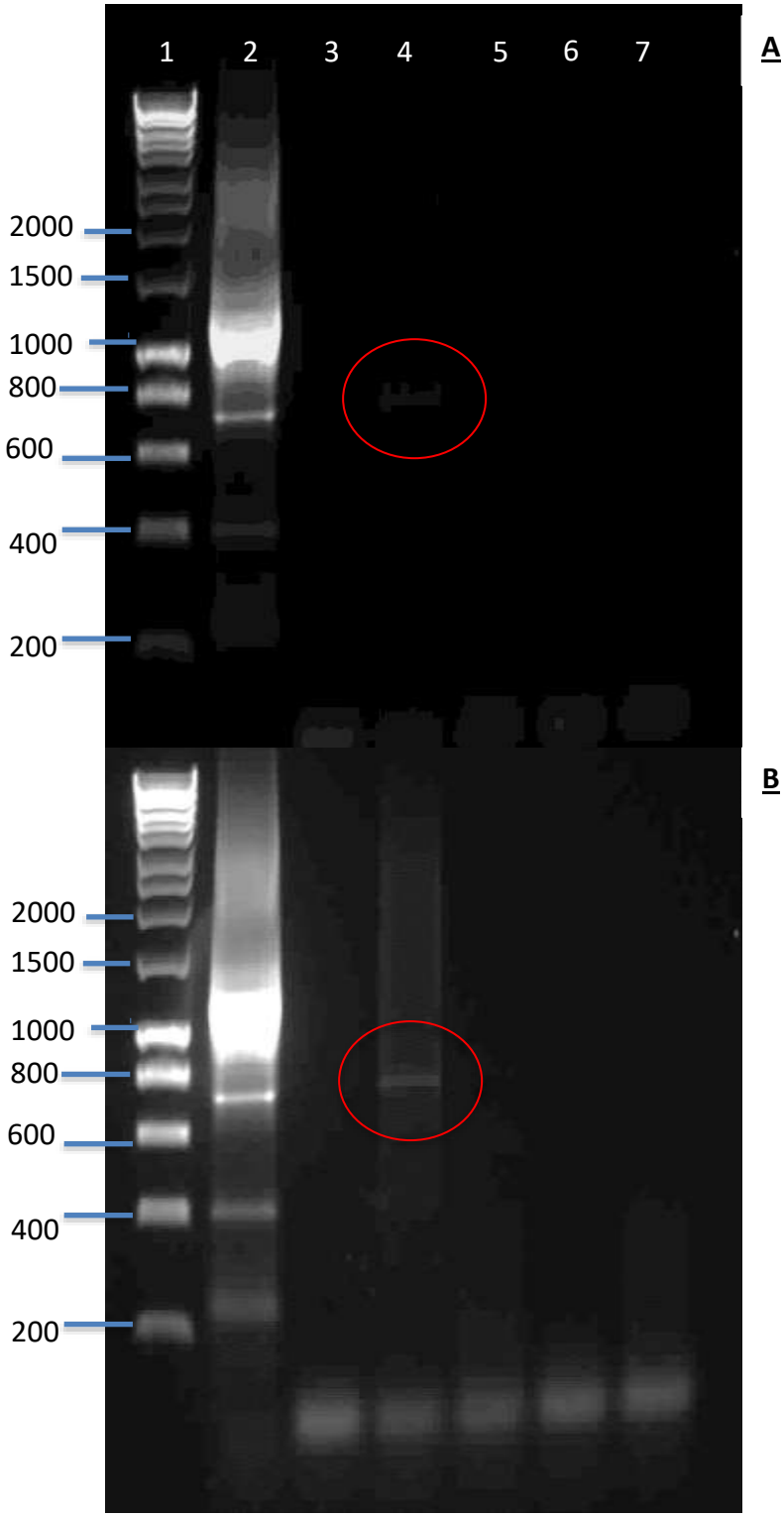


Figure 4: A. Electrophoresis 1% gel showing a band of DNA in lane 4 amplified by ITS PCR of samples taken from FTA cards. Lane 1 was 1Kb Hyperladder (Bioline Ltd) and lane 2 was *T. brucei* positive control. Lanes 3, 5, 6 and 7 are blood samples that were negative. B. The same gel over exposed to make the band visible.

The result of the early ITS PCRs yielded limited success. Only one sample was positive for trypanosome DNA and showed a very faint band. The samples selected had already tested positive for trypanosomes in previous studies. From this it was important to optimise this PCR to try to improve the results and amplify a greater quantity of DNA. The expected result of increasing the number of denaturing, annealing and extension cycles was that it would allow more trypanosome DNA to be amplified.

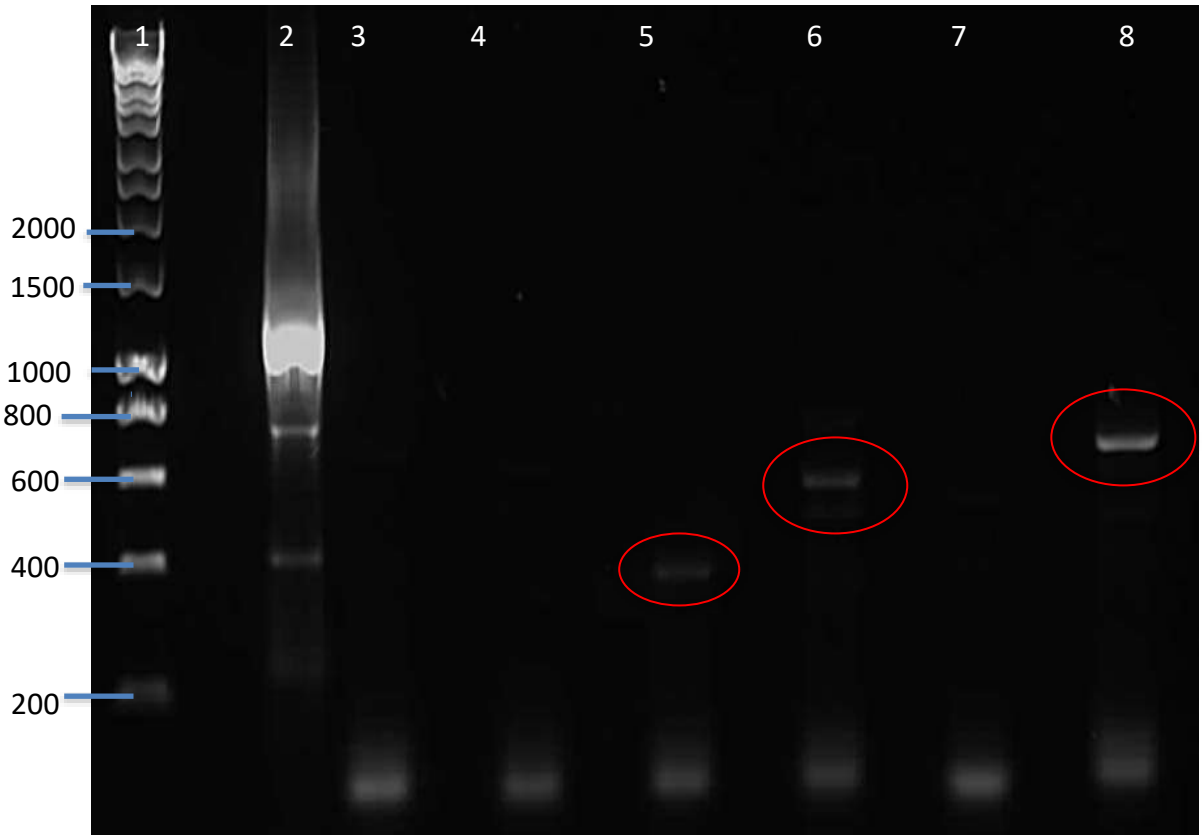


Figure 5: Electrophoresis 1% gel showing bands of amplified DNA from a single FTA card, increasing the number of extension cycles each time (35-40 cycles, lane 3- 35 cycles, lane 4-36 cycles, lane 5-37 cycles, 390bp, lane 6-38 cycles, 580bp, lane 7-39 cycles and lane 8-40 cycles, 620bp). Lane 1 was 1Kb Hyperladder (Bioline Ltd) and lane 2 was *T. brucei* positive control.

The number of extension cycles was increased one by one to allow more opportunity for the primers to bind to low levels of DNA in the sample. An individual sample was used (AH26) and in lane 3, the sample was left for 35 cycles. The number of cycles was increased by 1 each time. It was not until 37 cycles in lane 5 that a

band was first seen. After 38 cycles, in lane six, a band of a different size was seen more prominently. At 39 cycles, no band was seen but at 40 cycles, a bright band was seen, once again at a different size. Though increasing the cycles has provided more DNA. It has shown that from the same sample, multiple sizes of DNA are present, however, only one has been amplified sufficiently. This may highlight that increasing the samples may cause bias towards bands of a particular size. As there is no band visible at 39 cycles, this also suggests that there is a bias involved. The size of the band in lane 8 is of similar size to *T. vivax* (611bp) which would support the findings from previous studies done on these cards where *T. vivax* was found in this sample. The next step in improving this experiment is to increase the volume of the eluted DNA solution in the PCR reaction. When each band was measured to compare it to known band sizes of trypanosome species, the band found in lane 3 was not the same size as any known trypanosome species lengths according to data collected by Cox *et al* (2005), however, the size of the band in lane 4 was 603bp which suggests it could be *Trypanosoma vivax* (611bp, NCBI database), and the length of the band in lane 6 was 630bp which is 10bp different to the size of the band Cox *et al* (2005) found to be *T. vivax*.

Increasing the volume of eluted DNA solution from 1µl to 3µl results in more target DNA being added to the PCR reaction, therefore increased DNA sequences for primers to bind and so a greater chance of it being amplified.

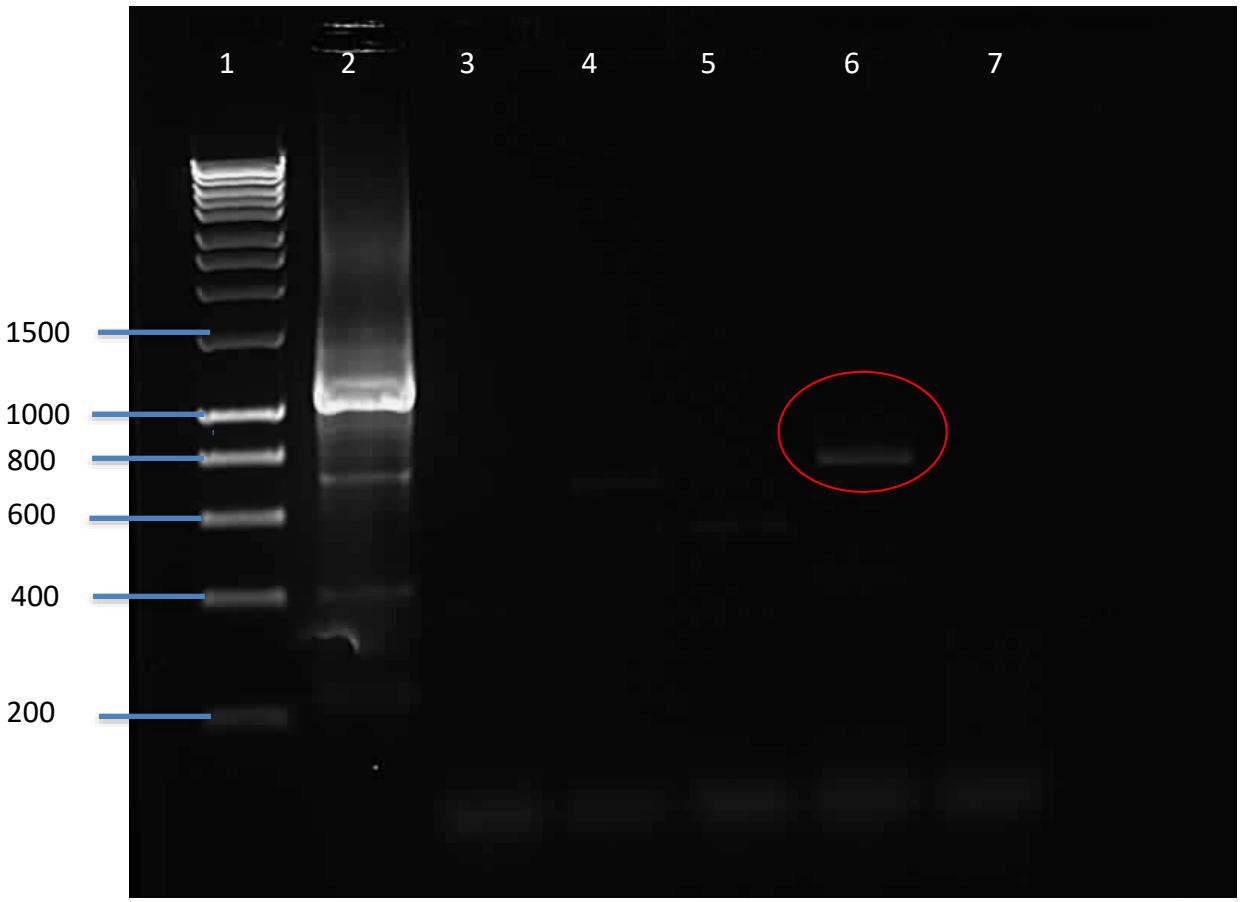


Figure 6: Electrophoresis 1% gel showing a band of DNA (6) amplified from FTA card using 3µl of eluted DNA solution. Lane 1 was 1Kb Hyperladder (Bioline Ltd) and lane 2 was *T. brucei* positive control. Lanes 3, 4 and 5 showed to be negative while lane 7 was a negative control.

This increase resulted in trypanosome DNA being amplified in one sample (AH26). From this outcome, it showed that a combination of both increasing the number of cycles and increasing the volume of eluted DNA solution could possibly bring about the required results to move forward.

By combining the optimisation techniques, it was hoped to improve the results of the ITS PCR. A greater volume of eluted DNA solution would increase the volume of target DNA in the reaction mix resulting in more binding of primers and greater amplification. Increasing the number of extension cycles increases the amplification phase of the PCR and so more DNA can be amplified.

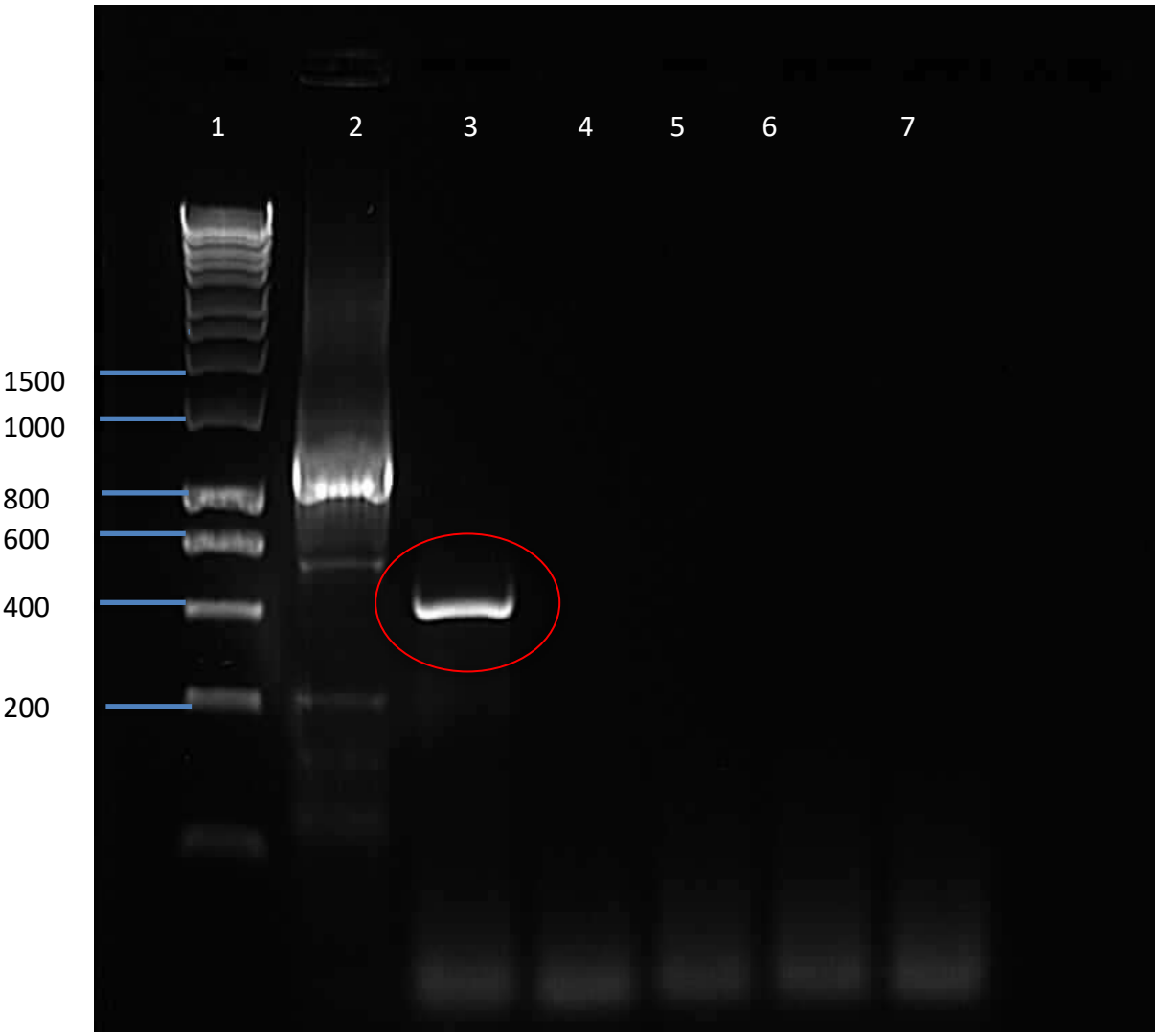


Figure 7: Electrophoresis 1% gel showing a band of DNA (3) amplified by ITS PCR from FTA card samples using 40 extension cycles and 3µl of eluted DNA solution. Lane 1 was 1Kb Hyperladder (Bioline Ltd) and lane 2 was *T. brucei* positive control. Lanes 4, 5 and 6 were negative while 7 was a negative control.

This combination proved to be a success; the gel image shows a band of DNA approximately 400bp in length. This proves that Trypanosome DNA could be amplified sufficiently from FTA cards in this experiment. However, as this band size does not correspond with any Trypanosome species identified by Cox *et al* (2005), this band would need to be sequenced in order to prove that this is in fact Trypanosome DNA. Progressing from this, more samples needed to be tested to repeat the experiment and identify samples that are known to be positive for trypanosome and also samples with potentially multiple species. It was also important to investigate whether any bias was caused by the increase in amplification cycles. To do this, samples tested by Eze Justin Ideozu (2015) that were positive for multiple species of trypanosome were tested to see if bias occurred within one sample.

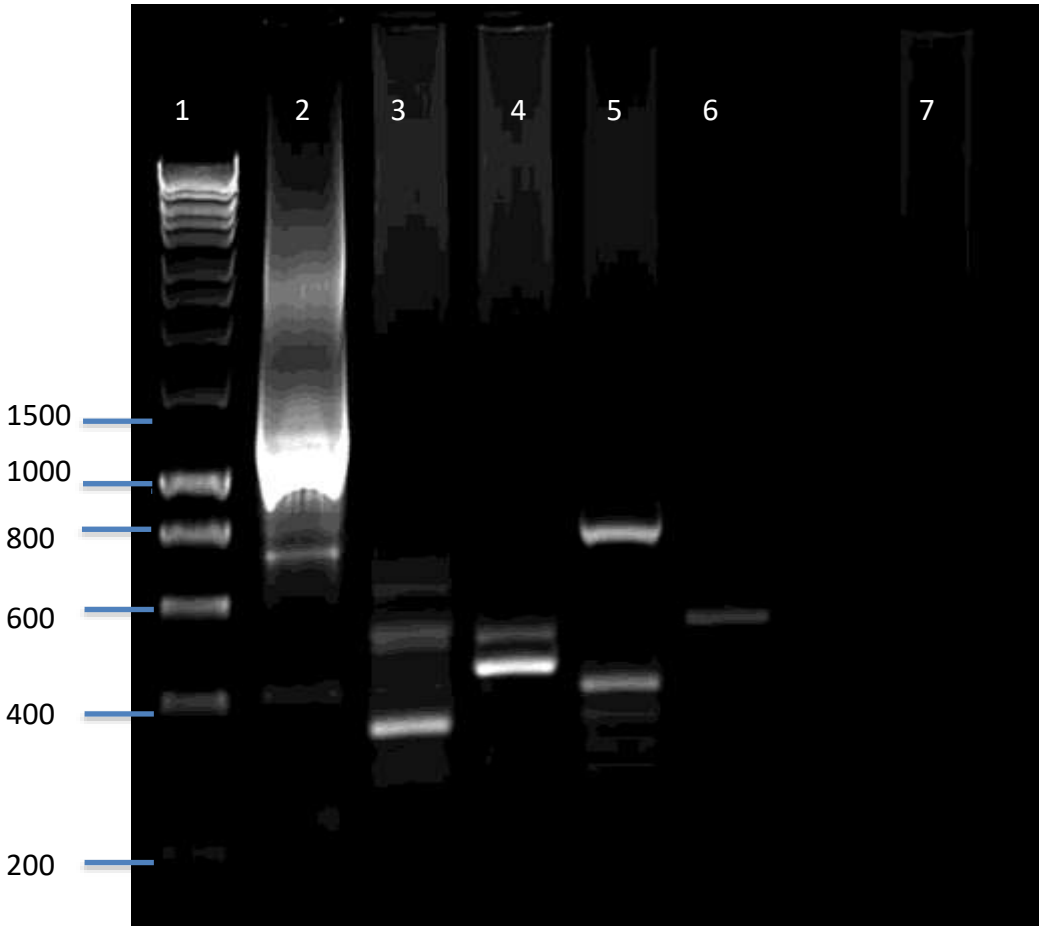


Figure 8: Electrophoresis 1% gel showing bands amplified by ITS PCR from FTA card samples using 40 extension cycles and 3µl of eluted DNA solution. Lane 3, 4 and 5

show multiple band lengths. Lane 6 shows a single band and lane 1 was 1Kb Hyperladder (Bioline Ltd) and lane 2 was *T. brucei* positive control.

Once DNA could be amplified to produce clear visible bands via ITS PCR, it was important to test samples that had previously been identified as positive, as well as samples that had multiple infections, this experiment resulted in all the samples tested showing positive for Trypanosomes of varying band lengths. Lane 3 had one clear band and one faint band, the clear band was 355bp in length and the faint band was 530bp in length, both of which did not match any known species found in the NCBI Bioinformatics database. Lane 4 showed multiple bands of lengths 446bp and 530bp in length which again didn't match any known species while lane 6 had a single band of 595bp, the closest known species to this length was *T. vivax*. Lane 5, however, had 2 clear bands showing multiple species trypanosomes present, the smallest band was 422bp in length which did not match any known species, and the larger band was 841bp which is closest to matching *T. simiae*. This outcome shows multiple trypanosome species in the sample set and the experiment has been optimised to amplify DNA of a single parasite from the FTA cards. From this, a PCR using primers for multiple Eukaryotic pathogens was used as well as a blocking primer to prevent the amplification of host 18s sequences and amplify all other 18s sequences that could be present.

3.3 PCR using primers that target Multiple Eukaryotic sequences.

3.3.1 Multiple Eukaryote PCR

When detecting multiple eukaryote pathogens by PCR, universal primers that detect 18s sequences of eukaryotes are used. The problem is that host DNA can also be amplified. In order to prevent this, a blocking primer is used. The blocking primer is designed to block a known oligonucleotide target, in this case, the mammalian 18s sequence. This is possible because mammalian 18s sequences are already sequenced and also distinct enough from other sequences to be specifically targeted.

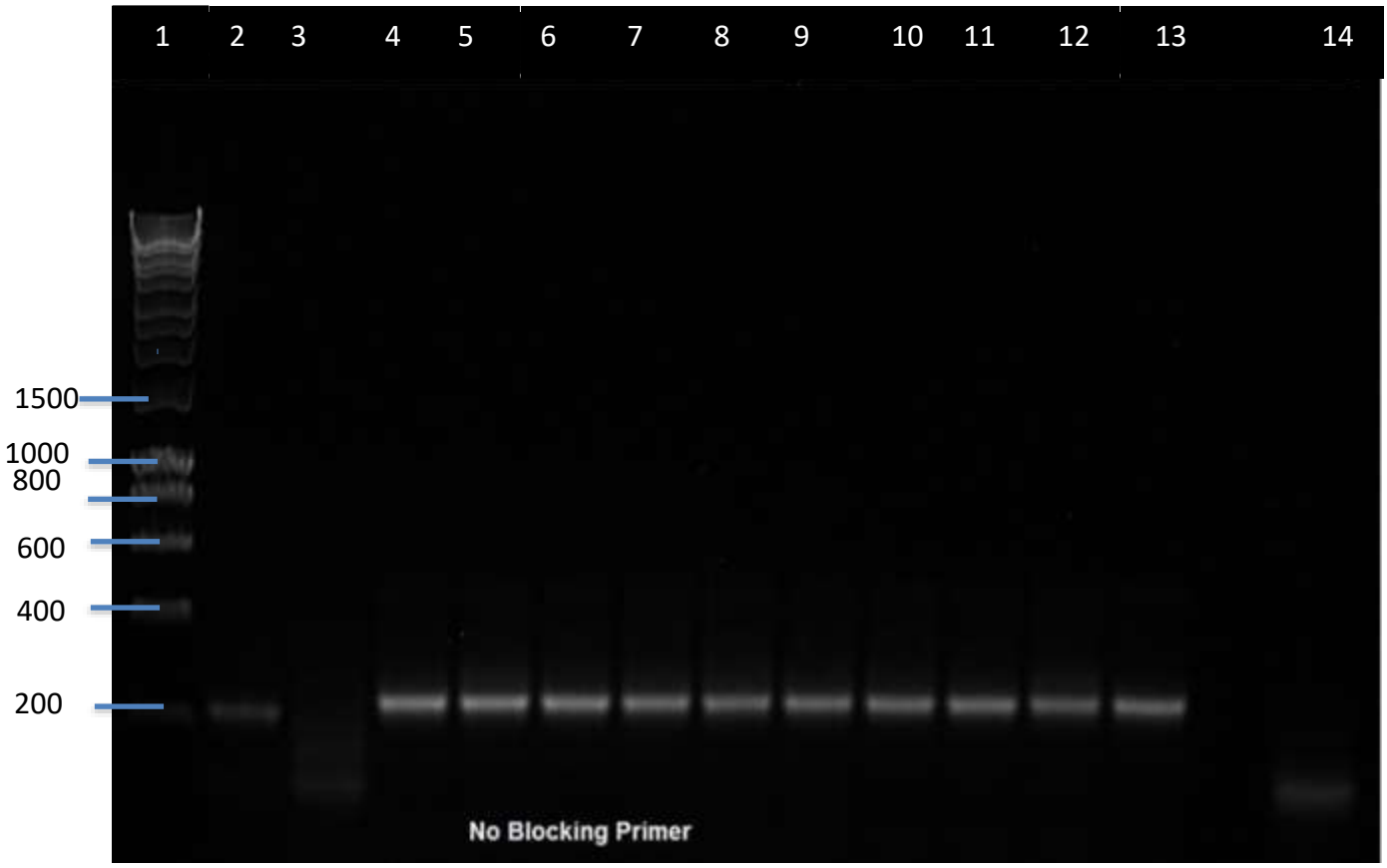


Figure 9: Electrophoresis 1% gel showing bands of DNA amplified using universal 18s eukaryotic primers without mammalian annealing blocking primer. Lane 1 was 1Kb Hyperladder (Bioline Ltd) and lane 2 was *T. brucei* positive control, lane 3 and 14 were negative controls and 2 to 12 were DNA samples eluted from FTA cards.

The universal eukaryotic primers amplify all 18s rRNA sequences within a sample, including host sequences. The results of the PCR show that without a blocking primer, there is the amplification of lots of DNA fragments at approximately 260bp, which is likely to be predominantly host DNA as there will be more host DNA present than pathogenic DNA. In the negative controls faint bands seen to be less than 100bpin length are primer dimer left over from the PCR process.

Adding an annealing blocking primer to a PCR that uses universal 18s eukaryotic primers allows the amplification of DNA that would otherwise be missed during PCR. This is due to mammalian host DNA binding to the primers. The 18s primers are designed to amplify all 18s sequences in a sample. As a result, host 18s sequences are the most likely to bind and be amplified by the primers because they are the most abundant DNA 18s sequence found. The annealing blocking primer has a

sequence that is complementary to the mammalian 18s sequence and so binds to any host 18s sequence DNA found in the sample. This, therefore, allows any other 18s sequences to bind to the primers and be amplified.

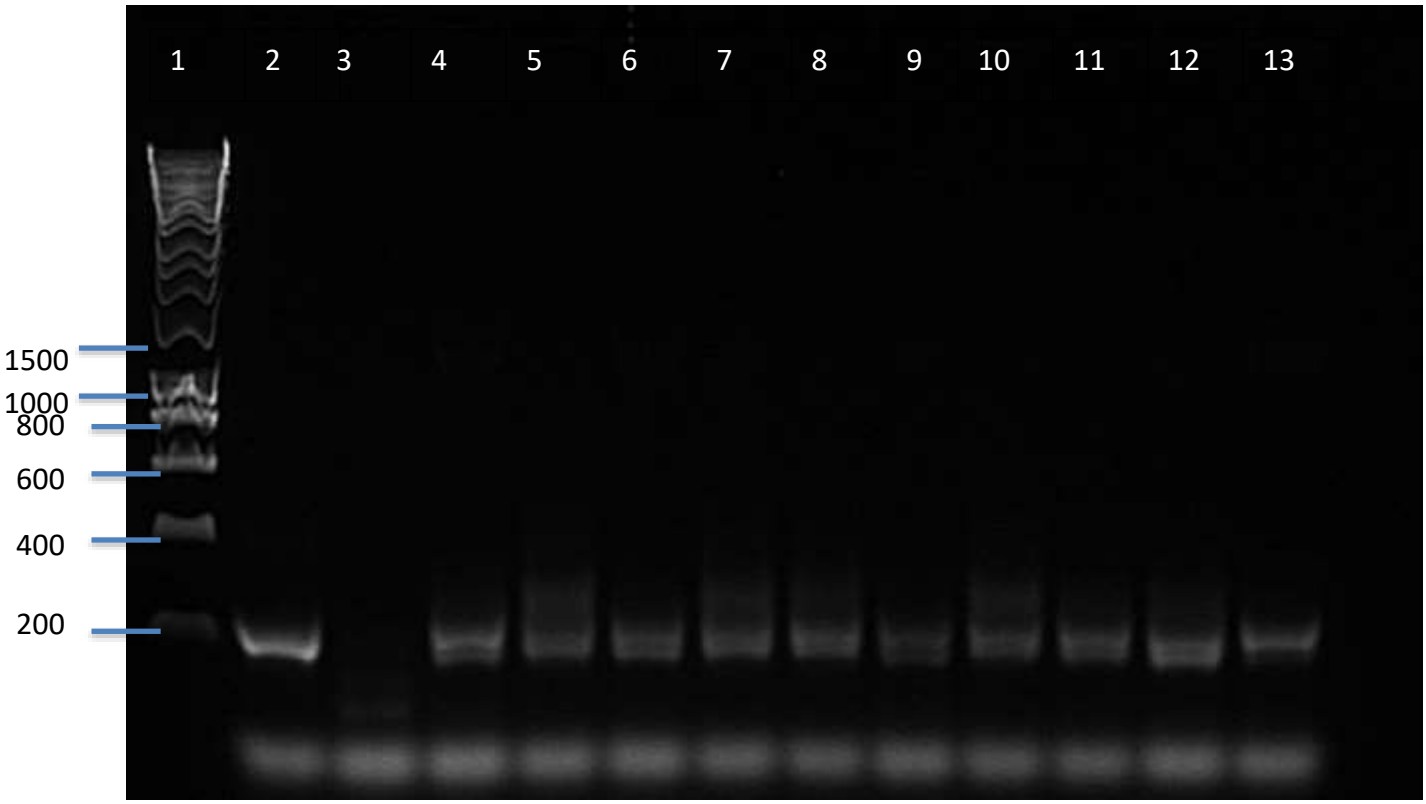


Figure 10: Electrophoresis 1% gel showing bands of DNA amplified using universal 18s eukaryotic primers along with a mammalian annealing blocking primer. Lane 1 was 1Kb Hyperladder (Bioline Ltd) and lane 2 was *T. brucei* positive control, lane 3 was a negative control and 2 to 12 were positive DNA samples eluted from FTA cards.

In figure 10, the annealing blocker has worked and prevented mammalian 18s sequences being amplified. This can be seen as there are more accurate bands shown in all the samples as well as lane 4 and 12 where there are possibly multiple bands visible. In order to measure the success of the annealing blocker, it needs to be tested to see how effective it is at blocking mammalian DNA.

To investigate how effective the annealing blocker was, it needed to be compared against known samples both with and without the blocker present using the universal primers. It also needed to be compared with how it would work with a spiked sample of both mammalian DNA and pathogen DNA.

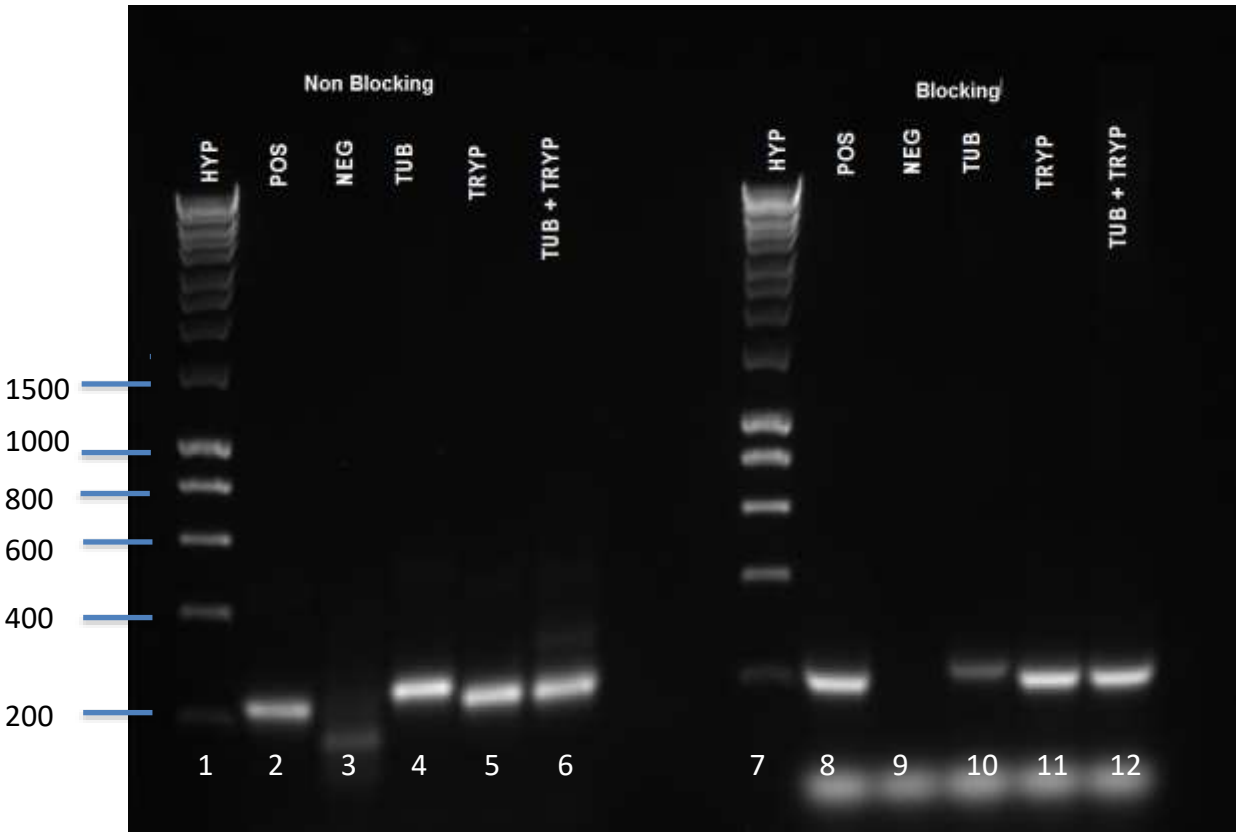


Figure 11: Electrophoresis 1% gel showing bands of tubulin DNA with and without annealing blocking primer. Lanes 2-6, annealing blocker was not present. Lane 8-12 annealing blocker was present. Lane 1 and 7 were 1Kb Hyperladder (Bioline Ltd), lane 2 and 8 were *T. brucei* positive control (260bp), lane 4 and 10 were tubulin (260bp), lane 5 and 11 were *T. brucei* again and lane 6 and 12 were a combination of tubulin and *T. brucei* DNA.

Figure 11 shows such a comparison, using human tubulin as the mammal DNA and the same *T. brucei* positive control used in the previous experiments. Without the annealing blocker, Human tubulin wasn't blocked at all. This can be seen in lane 4 where a bright tubulin band can be seen as well as in lane 6 where another bright band is visible with the combined tubulin and trypanosome DNA. With the annealing blocker, the Human tubulin in lane 10 is reduced significantly and is only slightly visible. The band of DNA in the combined sample is also less bright and is clearer showing that the tubulin in the sample has not been amplified. The primers used target the same 18S sequence found in all eukaryotic DNA. The

expected band size is 260bp which is why the controls are of different size to those used in previous experiments.

After establishing the success of the annealing blocker, it was then decided to look at the negative controls used. To this point, negative controls were made up of Whatman filter paper and went through the same process as the FTA card samples but were not taken directly from the FTA card where the sample was collected. It was decided that to ensure accurate controls, the negative samples must come from the same card that the sample has come from, collected from the clean part of the card where no blood was located. This ensures that the same material goes through the elution process, the samples containing the DNA and the outside of the sample on the card that should be negative for any DNA. Any contamination would show whether or not there is contamination occurring post sample collection.

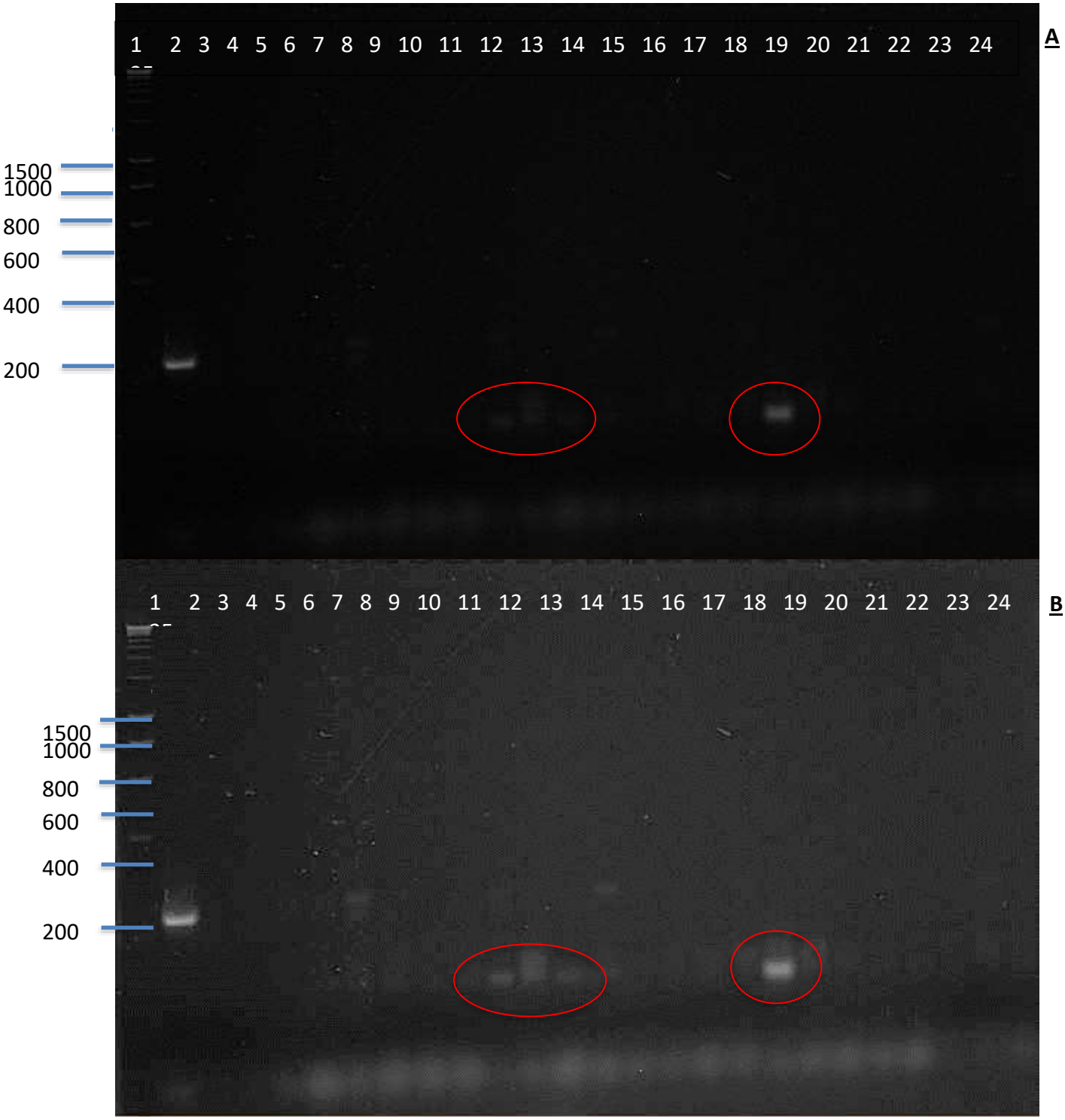


Figure 12: A. A representative section of an Electrophoresis 2% gel showing bands of DNA detected by universal microbial primers. Lane 2 was a *T. brucei* positive control, with lane 3, 9, 15 were negative controls taken from the same card as the sample was added to. Lanes 13, 14 and 20 showed positive results on the gel. B. The same gel over exposed to make the band visible.

In order to ensure negative controls were suitable and also further prove the success of the PCR, changes were made to the protocols. The negative controls were punched from the same cards as the samples, however, they were taken from the area surrounding the sample area that was free of blood. This would show whether there was any DNA present on the cards other than that in the blood sample. The PCR showed that these was nothing visible on the negative punches, however, NGS is more sensitive so will provide a more definitive result. The negative controls were in lane 3, 9 and 15. These samples were negative on the gel and further analysis using NGS is required. In order to view the PCR product better, the percentage of agarose in the gel was increased by 1% making it a 2% gel. This allowed for a clearer separation of the bands. In lanes 13, 14 and 20, positive samples are visible, however, NGS may show more due to higher sensitivity.

3.4 Next Generation Sequencing

Next-generation sequencing is the process in which a large number of small DNA fragments are sequenced in one go. Results from NGS are analysed by bioinformatics to combine the fragments and map them to known genomes (Behjati & Tarpey, 2013). In order to complete a NGS run, DNA was eluted from 70 cattle blood samples collected on FTA cards from Ahoada, Nigeria, with 5 samples having 3 extra duplicates taken from the, there was also 10 control samples taken from the blank area surrounding the blood sample. A total of 10 punches were taken for each sample and the DNA was eluted using Chelex 100 as described in the materials and methods. The samples were prepared for NGS by amplifying the DNA present in the sample by PCR using universal 18s primers as described in the materials and methods earlier. The PCR product is cleaned using AMPure beads with a series of washes and then put through an index PCR to barcode the samples. Once the samples have been barcoded they are then pooled together, and the end product is used for NGS. This sample is loaded onto the MiSeq and a run of 95bp was conducted. The results are given in excel format and which are loaded into R-studio where by the program Phyloseq is used to simplify the OTU sequence data and identify the taxa each read corresponds to and also compare samples and locations each sample was taken from.

3.4.1 The Abundance of Genera and the Variation of Genera between sample locations.

When attempting to use NGS to identify multiple species, not only can it show the number of pathogenic DNA found but also the abundance of DNA also. This can be used to identify active infections, level of parasitemia but also look at possible relationships of co-infection. Apicomplexa and Euglenozoa are 2 important phyla of parasitic organisms. Both contain many parasites that infect cattle and other mammals as well as zoonotic parasites. To investigate parasitic organisms located in the samples collected on FTA cards, the phyla Apicomplexa and Euglenozoa were analysed for the genera sequenced from the OTUs.

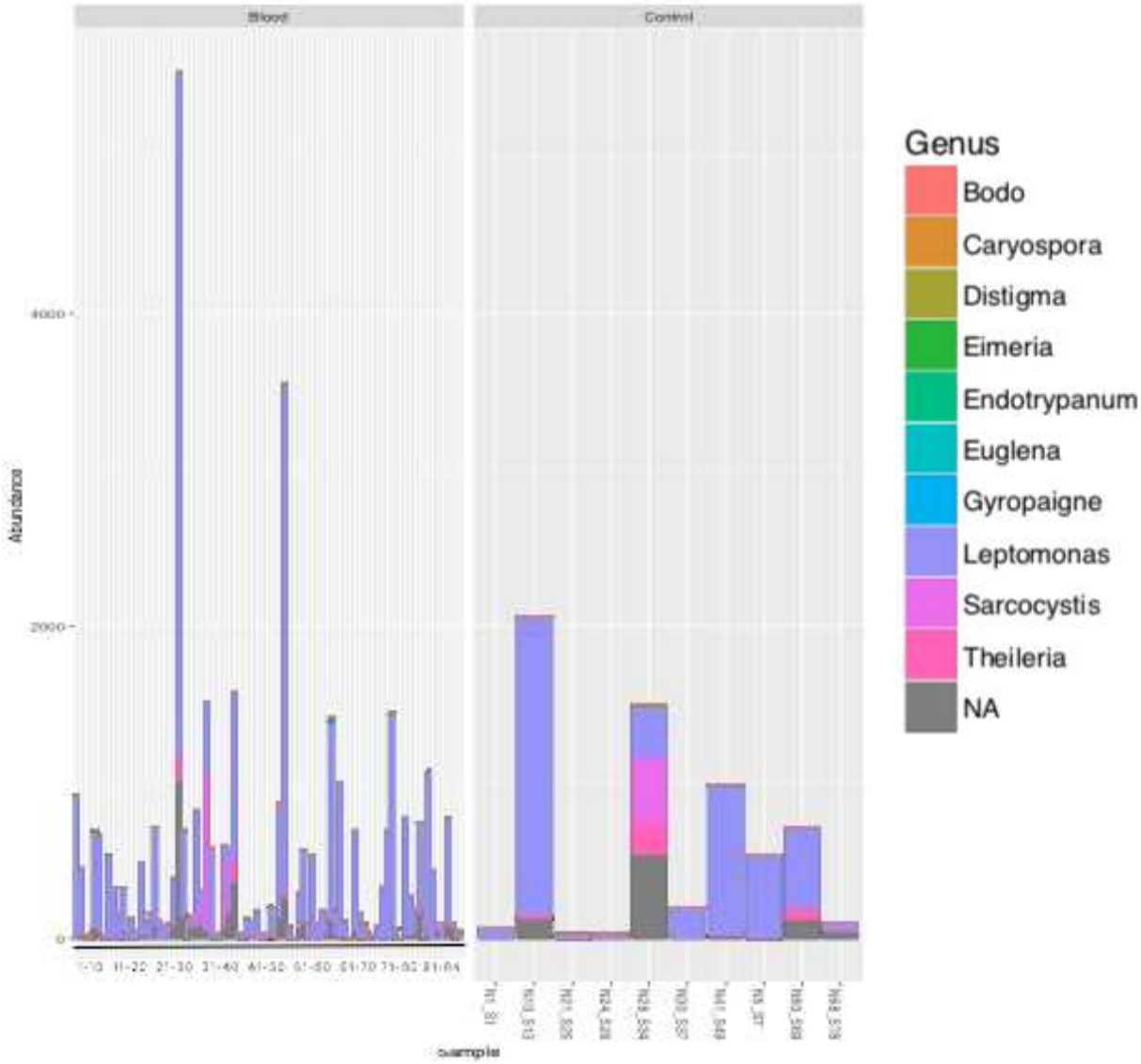


Figure 13: Abundance of Genera from the Phyla Apicomplexa and Euglenozoa in blood samples AH01-AH70 (left hand panel) and control samples N24-N66 (right hand panel). The control samples are taken from the non-blood coated area of the FTA card. The Y-axis represents the abundance of an individual sequence within each sample and the X-axis represents each sample. Both the samples and controls are on the same scale. Created using ggplot2/Phyloseq. For sample codes see table 2.

Table 2: Sample codes in an order corresponding to figure 13.

	Sample Code								
Location	1--10	11--20	21--30	31--40	41--50	51--60	61--70	71--80	81--84
	AH01	AH19	AH25	AH03	AH36	AH45	AH54	AH63	AH07
	AH10	AH02	AH25a	AH30	AH37	AH46	AH55	AH64	AH70
	AH11	AH20	AH25b	AH30a	AH38	AH47	AH56	AH65	AH08
	AH12	AH21	AH25c	AH30b	AH39	AH48	AH57	AH66	AH09
	AH13	AH22	AH27	AH30c	AH04	AH49	AH58	AH66a	
	AH14	AH23	AH28	AH31	AH40	AH05	AH59	AH66b	
	AH15	AH24	AH28a	AH32	AH41	AH50	AH6	AH66c	
	AH16	AH24a	AH28b	AH33	AH42	AH51	AH60	AH67	
	AH17	AH24b	AH28c	AH34	AH43	AH52	AH61	AH68	
	AH18	AH24c	AH29	AH35	AH44	AH53	AH62	AH69	

Figure 13 shows the 10 genera were sequenced from the phyla Apicomplexa and Euglenozoa. The Phyloseq program assigns the genera by matching sequence reads to a genus of species by attempting to match the DNA sequence it matches closest to. With the number of sequence reads this can only be matched accurately to a genus level. These are the most prominent genera found in the samples. *Theileria* and *Eimeria* are common parasites of cattle (Mans, Pienaar & Latif, 2015 & Matjila & Penzhorn, 2002). *Theileria* was found in a low abundance but high prevalence in the samples whereas *Eimeria* was only found in 2. *Leptomonas* was the species present in all samples as well as all the control samples. *Leptomonas* is a relative to the Trypanosomatidae family which has many protozoan parasites, of which there is a large number that infects cattle (Kraeva *et al*, 2015). *Sarcocystis* is an intracellular parasite commonly found in a range of mammals as well as humans. A number of species from the *Sarcocystis* genus are known to infect cattle (Fayer, 2004). It was found in a number of blood samples as well as some control samples. There was some overlap between the samples and the control samples. This may suggest that DNA material is not only found on the area stained with blood but also on the surrounding card area. *Leptomonas* was found most abundantly in the control samples, this could suggest that the most abundant DNA has been able to spread further across the card over time. To attempt to identify the true identity of the

samples labelled as *Leptomonas* found in samples, the sequences identified by NGS were analysed using the NCBI standard nucleotide search programme BLAST.

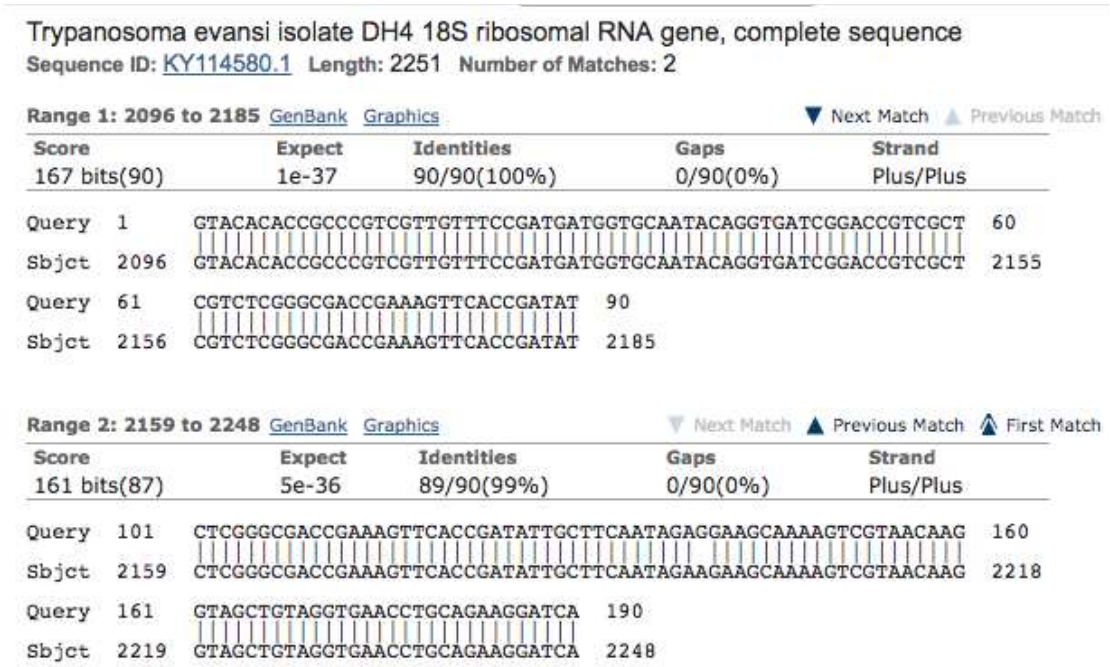


Figure 14: Blastn search result from the NCBI database. The sequence is *Trypanosoma evansi* and was the best match of the blast search.

The samples entered into the BLASTn database showed that *Leptomonas* shown in the NGS data corresponds to *Trypanosoma* species. *T. evansi* was the first hit on the BLAST searches. The query cover of the sequence was 94% for *T. evensi.*

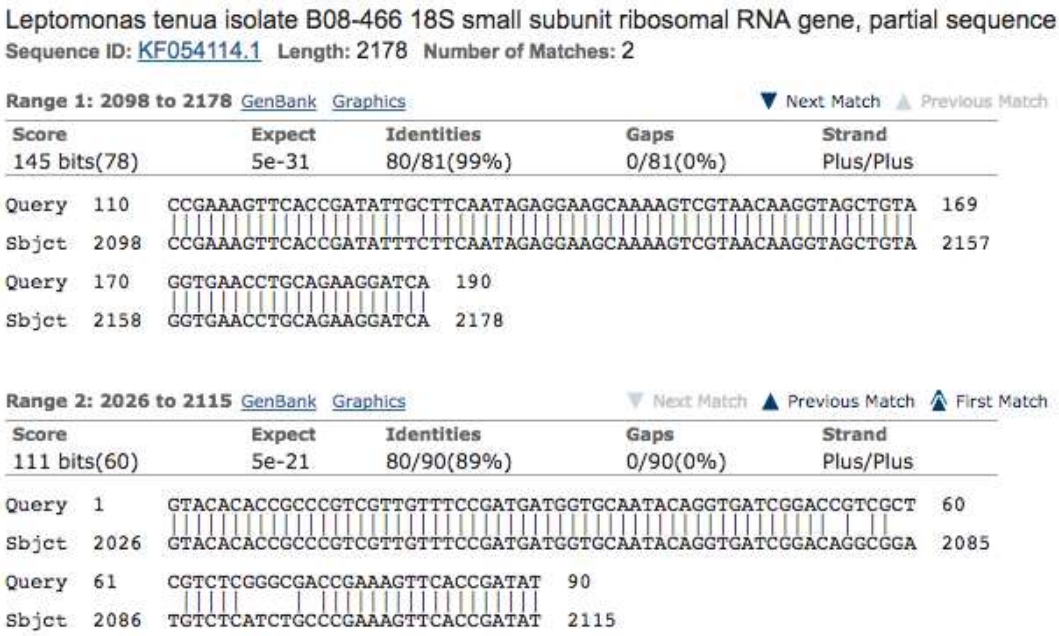


Figure 15: Blastn search result from NCBI database. The sequence is *Leptomonas tenua.* This match was the 73rd hit in the Blastn search.

The first match for a *Leptomonas* species was the 73 hits and there were only 3 matches in total. The matches were also only partial, and so it can be assumed with reasonable confidence that the NGS results showing *Leptomonas* are a species of *Trypanosoma.*

FTA cards have a long shelf life and have been known to store DNA successfully for up to 10 years (Picozzi *et al*, 2002). In an interest to look at what was on the card unrelated to the blood collected itself, it was decided to analyse the results of fungus found on the sample. This could be used later to look at what causes the FTA cards to no longer store DNA. Basidiomycota is a phylum of *Fungi* with approximately 30,000 species. It is believed that 8,400 of these species are pathogenic, with 40 yeast species known to infect humans and animals (Morrow & Fraser, 2009). This could suggest that there could be pathogenic species in the blood also.

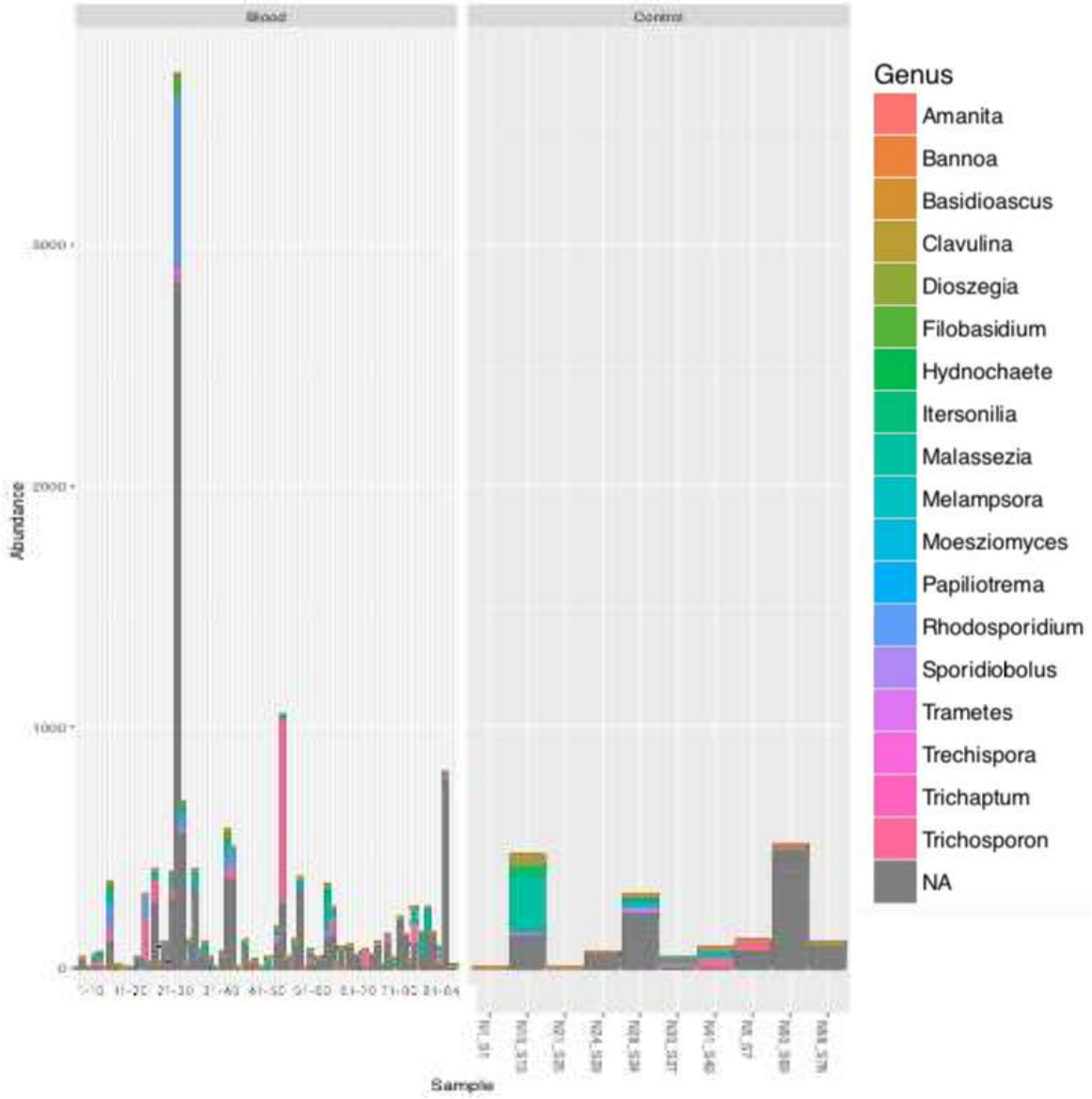


Figure 16: Abundance of Genera from the Phylum Basidiomycota in blood samples (left) and control samples (right). Created using ggplot2/Phyloseq. For sample codes see table 3.

Table 3: Sample codes in an order corresponding to figure 16.

	Sample Code								
Location	1--10	11--20	21--30	31--40	41--50	51--60	61--70	71--80	81--84
	AH01	AH19	AH25	AH03	AH36	AH45	AH54	AH63	AH07
	AH10	AH02	AH25a	AH30	AH37	AH46	AH55	AH64	AH70
	AH11	AH20	AH25b	AH30a	AH38	AH47	AH56	AH65	AH08
	AH12	AH21	AH25c	AH30b	AH39	AH48	AH57	AH66	AH09
	AH13	AH22	AH27	AH30c	AH04	AH49	AH58	AH66a	
	AH14	AH23	AH28	AH31	AH40	AH05	AH59	AH66b	
	AH15	AH24	AH28a	AH32	AH41	AH50	AH6	AH66c	
	AH16	AH24a	AH28b	AH33	AH42	AH51	AH60	AH67	
	AH17	AH24b	AH28c	AH34	AH43	AH52	AH61	AH68	
	AH18	AH24c	AH29	AH35	AH44	AH53	AH62	AH69	

The DNA of 18 different genera were sequenced from the phylum Basidiomycota with a number of overlaps in control samples and the blood samples. A number of Basidiomycota genus are known to be detritivores and could have grown post collection; however, pathogenic fungi could be present in the blood samples collected. This could suggest why there are similar genera on both the blood samples and control samples if they have grown post collection.

The number of sequence reads of each DNA sequence, which corresponds to the abundance of that taxon, identified in the blood sample and the controls that were taken from the surrounding area of the cards was compared to identify whether the control samples had the same pathogens as the blood sample itself. This was done to see whether, for Apicomplexan and Euglenozoan DNA sequences, to see if DNA could spread outside the discoloured part of the blood sample onto the blank card surrounding, and for the Basidiomycota this was compared to identify whether Basidiomycota species had grown post collection.

Table 4: The mean number of all DNA sequence reads per OTU identified per sample from the controls samples and the blood samples.

Control Sample	Mean Reads per OTU	Blood Sample	Mean Reads per OTU
N01	17.2	AH01	84.84615385
N05	57.30733945	AH05	47.67261905
N10	108.5805085	AH10	36.78571429
N21	28.3125	AH21	21.5
N24	21.27173913	AH24	55.31818182
N28	74.46666667	AH28	31.28712871
N30	18.74418605	AH30	39.31034483
N41	47.55384615	AH41	44.75862069
N60	60.09926471	AH60	27.31944444
N68	21.30693069	AH68	32.73568282

All the sequence reads were compared and, although there were differences between the blood samples and the control samples, it was not significant (P-value= 0.768). This was to be expected due to a large variety of different DNA sequences being found all over the card.

Table 5: The mean number of DNA sequence reads per OTU identified per sample from the Phylum Apicomplexa and Euglenozoa in control samples and the blood samples.

Control Samples	Mean Reads per OTU	Blood Samples	Mean Reads per OTU
N1_S1	34	AH01	304
N5_S7	269	AH05	142
N10_S13	258.125	AH10	111.75
N21_S25	13.5	AH21	27.5
N24_S29	11.66666667	AH24	83.5
N28_S34	213.7142857	AH28	50.33333333
N30_S37	33.66666667	AH30	20
N41_S49	65.53333333	AH41	41.5
N60_S69	101.8571429	AH60	62.45454545
N68_S78	34.66666667	AH68	33.66666667

A similar comparison of the number of sequence reads of DNA from the phylum Apicomplexa and Euglenozoa identified also revealed that there were also no significant differences in the number of sequence reads between the blood sample and the clean control sample (P-value= 0.712). From this it would suggest that there is potential for DNA to not be confined to the stained area of the card but that it may spread out over a larger area.

Table 6: The mean number of DNA sequence reads per OTU identified per sample from the Phylum Basidiomycota in control samples and the blood samples.

Control Samples	Mean Reads per OTU	Blood Samples	Mean Reads per OTU
N1_S1	0	1_S2	7
N5_S7	11.9	5_S6	20.11764706
N10_S13	25.88888889	10_S12	21.5
N21_S25	0	21_S24	9.2
N24_S29	22	24_S28	9.333333333
N28_S34	14.38095238	28_S33	10.36363636
N30_S37	11.75	30_S36	0
N41_S49	27.66666667	41_S48	10.75
N60_S69	34.13333333	60_S68	14
N68_S78	51.5	68_S77	37

A final comparison was made between the number of sequence reads from the phylum Basidiomycota found in the blood sample and the control sample. The results again show there is no significant difference in the mean number of sequence reads per OTU found in each location (P-value= 0.33). This suggests that the Basidiomycota species found in the blood sample are similar to the control sample and it suggest that these species grew after the samples were in storage and could potentially play a role in the spoiling of DNA on the cards.

When looking at the results of samples tested in this study, it was interesting to see the variety of genera identified. To gain an idea of what each sample was infected with and which was the most abundant genus per sample, the relative abundance of each sample was analysed.

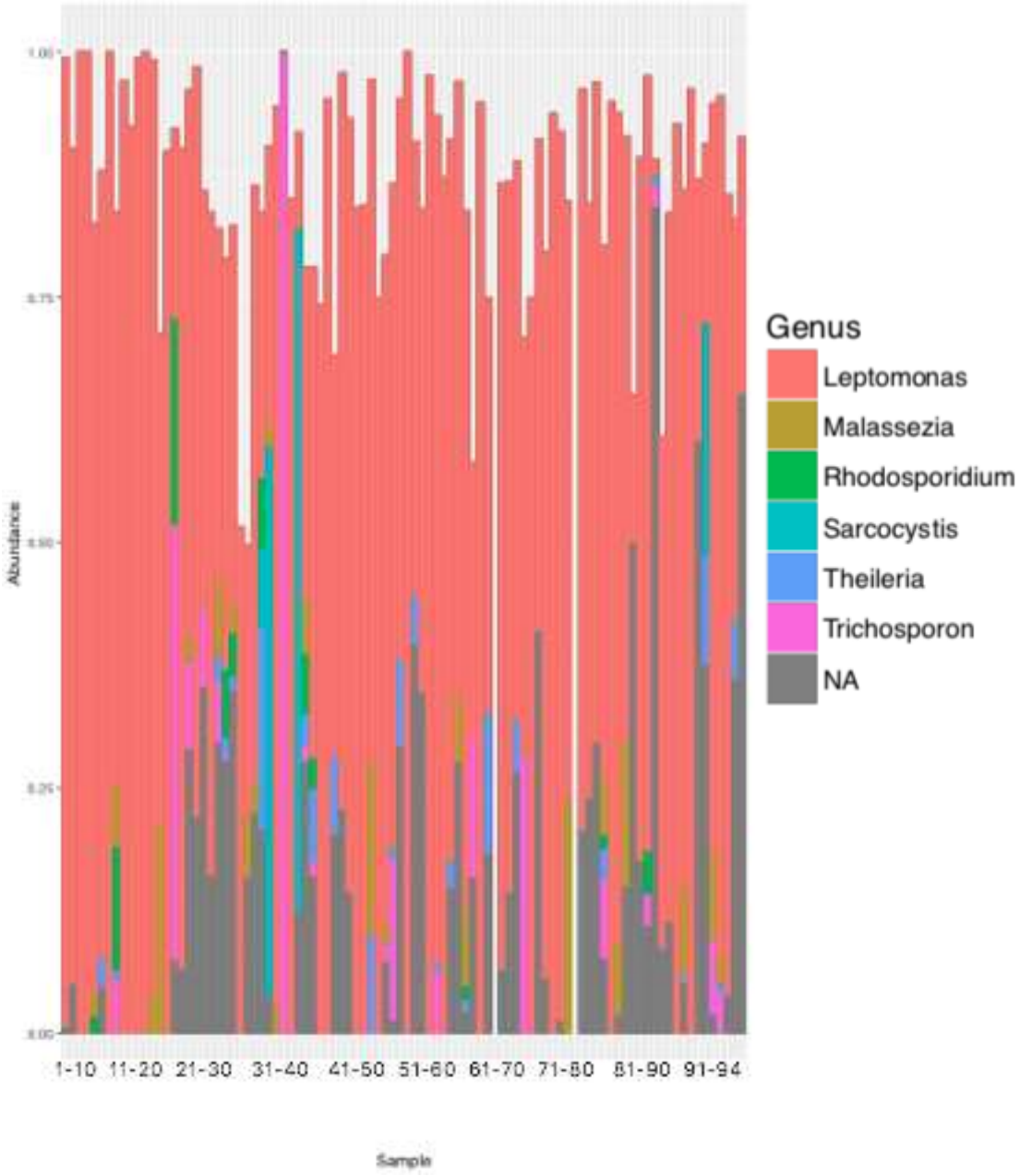


Figure 17: Relative abundance (%) of OTUs and their genera isolated from blood samples and control samples. Created using ggplot2/Phyloseq. For sample codes see table 7.

The relative abundance of each genera represented in figure 17 shows *Leptomonas* to be most abundant, with the majority of samples showing a minimum of 50% abundance. *Sarcocystis* and *Theileria* are present in varying abundance in a number of samples. This information could be used to identify potential co-infection routes and identify roles of individual species in the spread of diseases within a population.

Table 7: Sample codes in an order corresponding to figure 17.

	Sample Number									
Location	1--10	11--20	21--30	31--40	41--50	51--60	61--70	71--80	81--90	91--94
Sample Number	AH01	AH19	AH25	AH03	AH36	AH45	AH54	AH63	AH07	N41
	AH10	AH02	AH25a	AH30	AH37	AH46	AH55	AH64	AH70	N05
	AH11	AH20	AH25b	AH30a	AH38	AH47	AH56	AH65	AH08	N60
	AH12	AH21	AH25c	AH30b	AH39	AH48	AH57	AH66	AH09	N68
	AH13	AH22	AH27	AH30c	AH04	AH49	AH58	AH66a	N01	
	AH14	AH23	AH28	AH31	AH40	AH05	AH59	AH66b	N10	
	AH15	AH24	AH28a	AH32	AH41	AH50	AH6	AH66c	N21	
	AH16	AH24a	AH28b	AH33	AH42	AH51	AH60	AH67	N24	
	AH17	AH24b	AH28c	AH34	AH43	AH52	AH61	AH68	N28	
	AH18	AH24c	AH29	AH35	AH44	AH53	AH62	AH69	N30	

3.4.2 Analysis of 5 samples replicated 3 times to identify whether duplicate punches taken from FTA cards result in the same abundance of a parasite.

When samples collected on FTA cards are tested, 10 punches were taken from each card and sampled. This is due to the uneven coverage of parasite DNA on the cards. In order to investigate this using NGS, 10 punches were taken 4 times from 5 samples as replicates. The aim of doing this was to identify if there was a difference in the number of sequence reads between replicated samples.

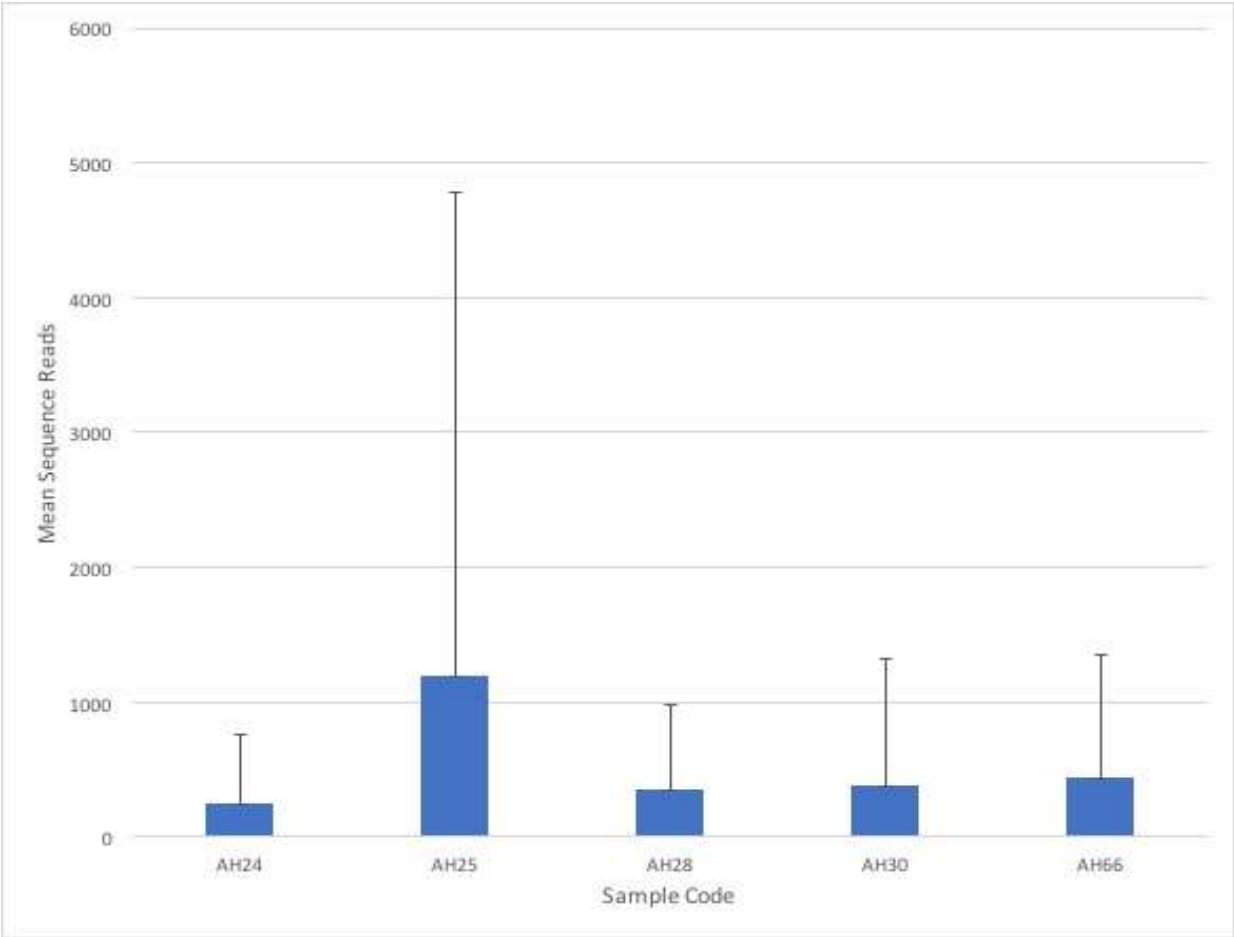


Figure 18: Histogram of the Mean sequence reads for *Leptomonas* from 4 replicates taken from 5 different FTA cards. 95% error bars were added, calculated by 2x standard deviation.

Figure 19 shows a histogram of the mean sequence reads from 4 replicates taken from 5 samples. The overlap of the error bars suggests that there is no significant difference in the mean number of sequences between samples, however, the size of the error bars suggests that there is a large variation in the number of sequence reads between the replicates in each sample. This means that, depending on where the punch is taken form on the cards, can influence the amount of DNA in the sample and so, sequenced by next-generation sequencing. Sample AH25 shows the greatest variation between replicates, as it has the largest error bar, it also has the highest standard deviation of 1798.64. In this case, one sample had a much greater number of sequence reads, again supporting the suggestion that the area the cards are taken from can influence the sequence results from the card.

3.4.3 A comparison of ITS tested samples and NGS tested samples identifying Trypanosome infection using Cohen's Kappa coefficient.

The samples used to conduct this study were previously tested using an ITS PCR as part of a Ph.D. thesis by Eze Justin Ideozu (2015). As part of this study, a selection of the samples was also tested using the same ITS PCR to identify is parasitic DNA was still useable on the FTA cards. In order to investigate how the ITS PCR compared with the NGS identification, Cohen's Kappa coefficient was used to identify the agreement between the results. Kappa statistics is a test that measures the agreement between two raters, in this case, whether ITS PCR identified samples to be positive or negative compared to the NGS results. Cohen's kappa coefficient also takes into account and corrects for chance agreement between raters (Perera, Heneghan & Badenoch, 2008).

In order to select a comparison for the NGS results against the ITS PCR, Basic models of infection were established. Initially, a conservative figure of 500 sequence reads for what can now be described as Trypanosome sequences were used and compared to the ITS PCR. By doing this, it describes any samples with less than 500 Trypanosome sequence reads as negative and was chosen as it appeared to be visually the top of the highest 'noise' peaks. Noise it the sequence reads too low to dictate a positive sample. A κ- Value of 0.23 suggests a fair agreement between the two processes. To make it less conservative, the average number of Trypanosome reads in the control samples was used as a level of noise. The average sequence read for Trypanosome was 70.5 reads per control sample and so anything above this in the blood samples was then seen as positive. A κ-value of -0.05 suggests a poor agreement between the ITS PCR results and the NGS results. This poor level of agreement suggests that this model is not suitable when comparing the two identification methods. The more conservative method showed a more realistic comparison, but also could suggest that the ITS PCR is less sensitive than the NGS.

3.4.4 Estimating the positivity of different parasite species within samples by applying the same conservative model that provided the best Cohen's Kappa Coefficient.

By using this conservative method and applying it to other known parasites of cattle, it can give an estimation of which samples are positive for other parasites

by the NGS data. The visual estimation of what represents a noise level in other parasites allows a representation of multiple pathogens present in the samples and the examples used are common cattle parasites.

Table 8: The infective status of each sample as established using conservative model tested by Cohen's Kappa coefficient comparing previously tested Trypanosome ITS PCR results with Trypanosome NGS results.

Sample	ITS-Tryp	NGS-Tryp	NGS-Theileria	NGS-Eimeria	NGS-Sarcocystis
AH01	Y	Y	N	N	N
AH02	Y	N	N	N	N
AH03	Y	N	N	N	N
AH04	Y	Y	N	N	N
AH05	Y	Y	N	Y	N
AH06	N	N	N	N	N
AH07	Y	Y	N	N	N
AH08	Y	N	N	N	N
AH09	Y	N	N	N	N
AH10	Y	N	N	N	N
AH11	Y	N	N	N	N
AH12	N	N	N	N	N
AH13	N	Y	N	N	N
AH14	N	Y	N	N	N
AH15	N	N	N	N	N
AH16	N	Y	N	N	N
AH17	N	N	N	N	N
AH18	N	N	N	N	N
AH19	N	N	N	N	N
AH20	N	N	N	N	N
AH21	Y	N	N	N	N
AH22	N	N	N	N	N
AH23	N	N	N	N	N
AH24	Y	Y	N	N	N
AH25	Y	Y	Y	N	N
AH26	Y	N	N	N	N
AH27	Y	N	N	N	N
AH28	Y	Y	Y	N	Y
AH29	Y	Y	N	N	N
AH30	Y	Y	Y	N	Y
AH31	N	N	N	N	N
AH32	N	N	N	N	N

AH33	N	N	N	N	N
AH34	N	N	N	N	N
AH35	N	N	N	N	N
AH36	N	N	N	N	N
AH37	N	N	N	N	N
AH38	Y	N	N	N	N
AH39	N	N	N	N	N
AH40	N	Y	N	N	N
AH41	Y	N	N	N	N
AH42	N	N	N	N	N
AH43	N	N	N	N	N
AH44	N	N	N	N	N
AH45	N	N	N	N	N
AH46	N	Y	N	N	N
AH47	N	N	N	N	N
AH48	N	N	N	N	N
AH49	N	N	N	N	N
AH50	N	N	N	N	N
AH51	N	Y	N	N	N
AH52	N	N	N	N	N
AH53	N	N	N	N	N
AH54	N	Y	N	N	N
AH55	N	N	N	N	N
AH56	N	N	N	N	N
AH57	N	N	N	N	N
AH58	N	N	N	N	N
AH59	N	N	N	N	N
AH60	N	Y	N	N	N
AH61	N	Y	N	N	N
AH62	N	N	N	N	N
AH63	N	N	N	N	N
AH64	N	Y	N	N	N
AH65	N	N	N	N	N
AH66	N	Y	N	N	N
AH67	N	N	N	N	N
AH68	N	N	N	N	N
AH69	N	N	N	N	N
AH70	N	N	N	N	N

The results show that 3 samples were infected with *Theileria,* 1 sample was infected with *Eimeria* and 2 samples were infected with *Sarcocystis.* This is lower

than 20 samples ITS positive Trypanosome and 20 NGS positive for Trypanosome, however, it shows that using this technique enables the amplification of low levels of pathogenic DNA which could be missed using conventional techniques.

3.4.5 Representation of species diversity in the blood samples as well as the control samples using Shannon Diversity Index.

The Shannon index is a tool to show the species diversity in a community. It shows the alpha diversity of a species. This is the diversity of a species within a local area. In the case of this research, this is the area on the FTA card where blood is present, and punches are taken from. The Shannon index also shows the beta diversity, this is the differentiation between habitats. This is the differentiation between the punches taken from the area of the blood sample and the punches taken as controls that are on the same cards but where no blood is present.

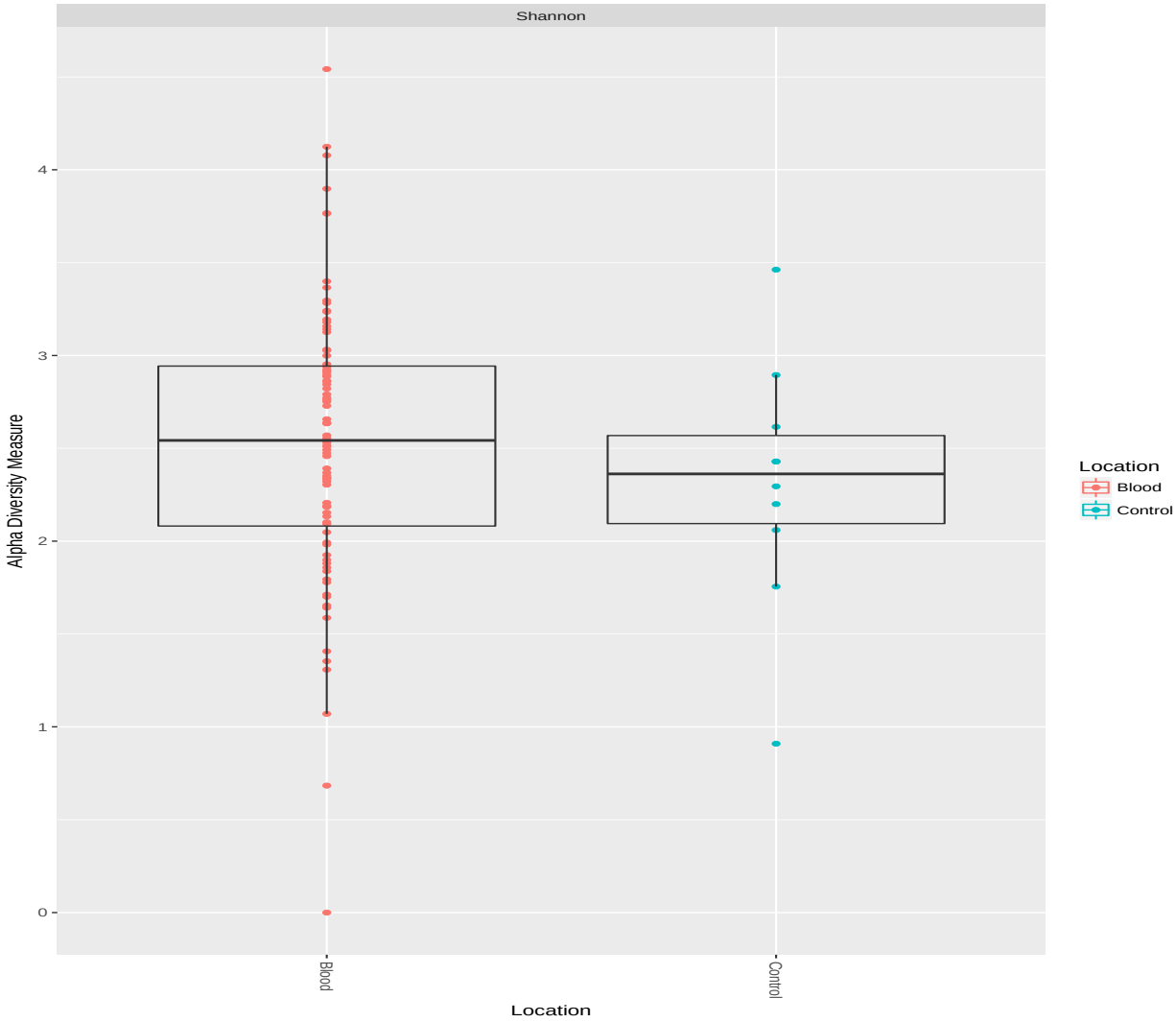


Figure 19: Box plot of Shannon Diversity Index comparing the location samples was collected from. X-axis shows the Alpha diversity and Y-axis shows the location. Red= Blood, Green= Control. Created using ggplot2/Phyloseq

The Shannon diversity index represents the variance in diversity between locations. Here it shows the difference in diversity between the blood sample and the negative control. The results show that the blood samples have a greater alpha diversity than the control samples however the controls also show a high diversity.

3.4.6 Identifying the variation of diversity between the location of which each sample is taken using Bray Non-metric multidimensional scaling

Bray Non-metric multidimensional scaling (NMDS) is a method that uses a distance matrix to show the dissimilarity between sites in which samples are taken.

This method breaks down data gathered from multiple dimensions, for example, different locations, communities or habitats, into 3 dimensions in order for it to be visualised and interpreted. NMDS uses rank orders, rather than the absolute abundance of a species within a community, which allows it to be more flexible and incorporate a greater variety of data.

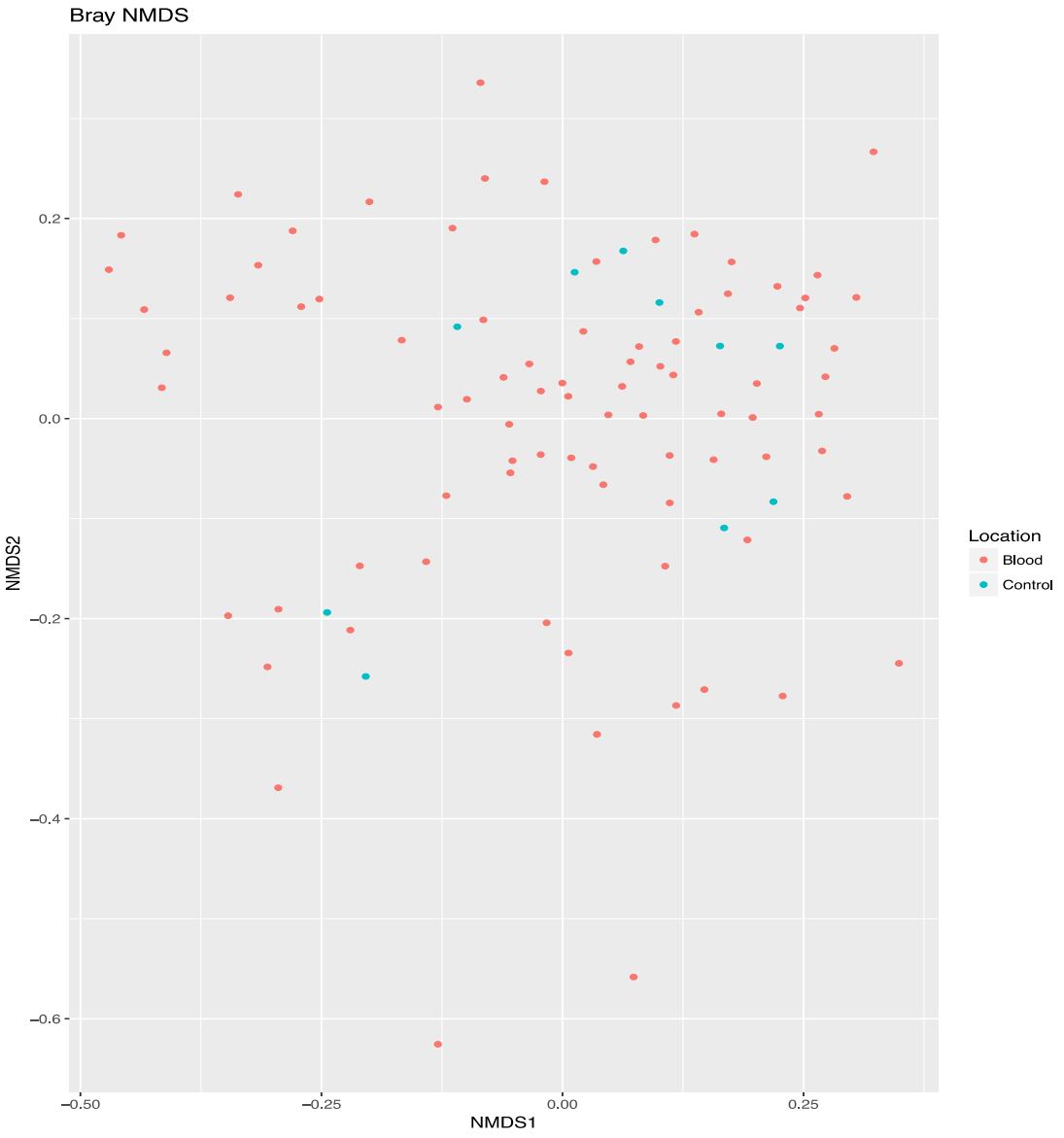


Figure 20: Bray NMDS showing the differences between the location of the blood sample and the control sample taken from the surrounding area. (Stress= 0.2) Created using ggplot2/Phyloseq.

The blood samples collected from FTA cards showed a similarity of species with the control samples. The majority of samples remain between -0.4 and 0.3 suggesting similar species found in both locations of samples. There is more variation within the blood samples, some of which reach over -0.6. The variation in the controls is less which is to be expected however the difference is not as great as first thought. Initial expectations were that the number of species found on the control samples taken from outside the blood on the cards would have a lot fewer species. This is due to the expected variation found in African cattle blood analysed in previous studies.

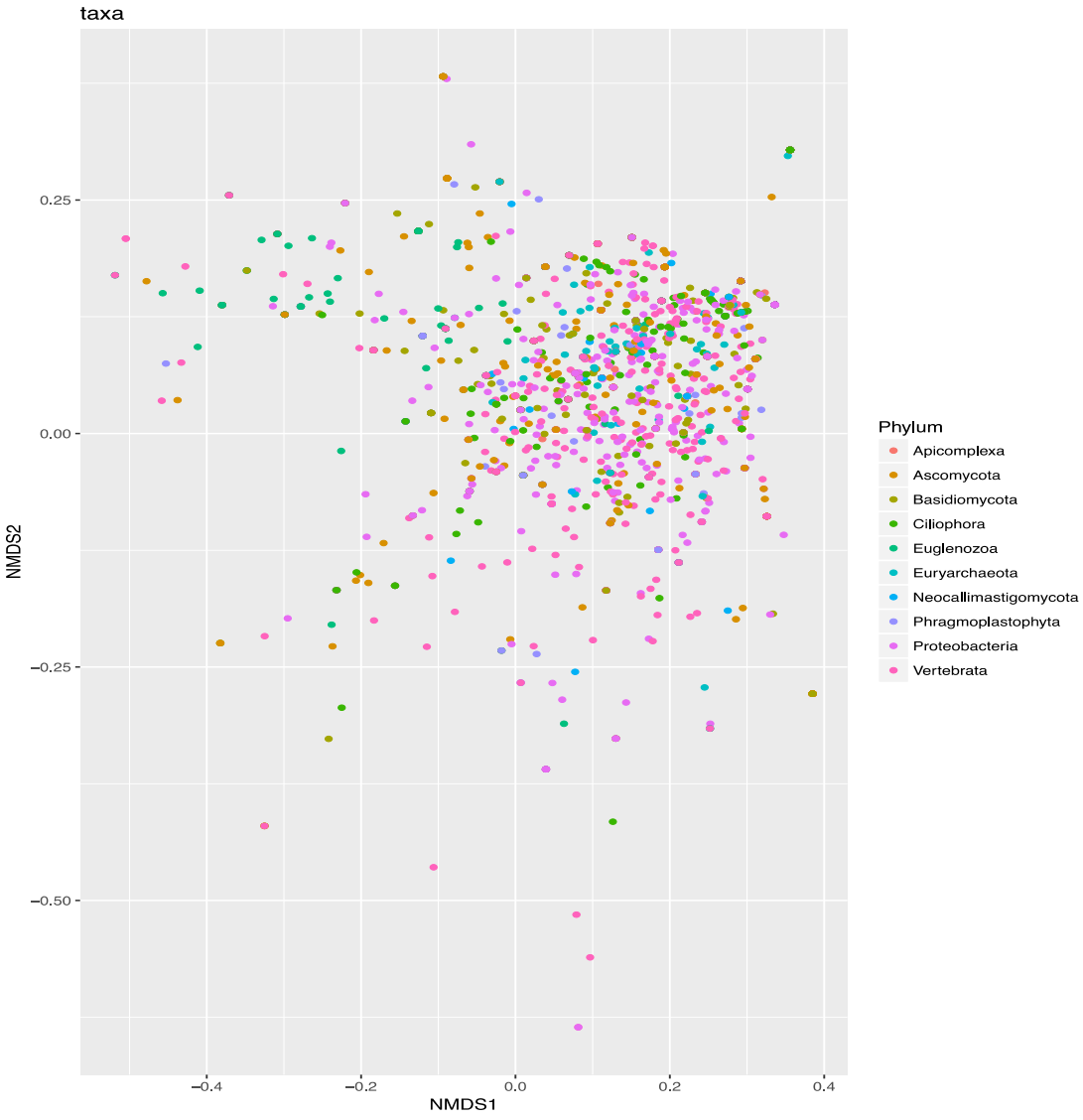


Figure 21: Bray NMDS showing the differences in Taxa found on each sample. (Stress= 0.2). Created using ggplot2/Phyloseq

The diversity of species found in each sample is represented in figure 14. It shows a wide variety of taxa found. It also supports the similarity of taxa within samples, as the majority of samples within -0.25 and 0.25. The presence of Vertebrata shows that the blocking primer was not 100% successful, though this was known already. The tight bunching of Apicomplexa shows the similarity of the phylum between the different samples where are there is slightly less bunching of the Euglenozoa. This suggests that the variation between samples of Euglenozoa is greater than Apicomplexa. This is perhaps due to a single more common genus of the Apicomplexa phylum being present in the samples. The most wide-spread phylum is the Basidiomycota, this suggests that the genera vary more than that of other phyla and that there is a greater variation between the samples.

4.0 Discussion

Neglected tropical diseases still have a major impact on health and the economy across the world. With a lot of such diseases occurring in rural communities in some of the poorest countries in the world, their impacts can be even greater (Welburn *et al*, 2006). The main aim of this study was to create a series of protocols to allow the detection and amplification of DNA from multiple pathogens found in blood samples collected on FTA cards. With this DNA, also conduct next-generation sequencing to identify the pathogens found. In order to succeed this aim, many different protocols were used and modified to optimise the results.

Through adaptation of FTA card elution protocols and optimisation of PCR protocols, a suitable quality and quantity of DNA were produced to use for next-generation sequencing. From this, multiple pathogen DNAs were successfully sequenced to the genus taxonomic level. A large variety of expected and unexpected genera were found, varying from common cattle parasites to a large range of bacterial and fungal DNA.

FTA cards were used to collect samples by binding DNA to them. They are saturated with chemicals designed to denature proteins and lyse cells that are placed on the card. The chemicals also ensure that the samples are non-hazardous

so can be safely handled and stored without the risk of spreading any pathogenic material on the cards. As a result, samples collected on FTA cards can be stored at room temperature for up to 10 years (Ahmed *et al*, 2011 & Picozzi *et al*, 2002). In other studies, samples have been used a lot earlier, for example, a study looking at geminivirus-diseased plant samples collected using FTA cards tested the cards 9 months after collection (Owor *et al*, 2007). Becker *et al* (2004) discovered that 9 months is the longest length of time that FTA cards can be stored and material can be successfully eluted and amplified from them. If the cards are used after this length of time, then genetic material isn't completely removed from the cards when washed and so accurate representation is then lost.

Samples from FTA cards have been tested to identify single diseases or identifying multiple diseases using time-consuming and labour-intensive techniques. Until this project, FTA cards had not been used along with next-generation sequencing to identify multiple pathogens from DNA eluted from the cards. In order to elute the DNA from the FTA cards using the standard protocol, it was important to use a Chelex 100 extraction method. This is a resin that when added to the eluted sample, helps remove and PCR inhibitors by binding to polar cellular components while leaving DNA and RNA in the solution (Al-Griw *et al*, 2017). Different protocols described boiling Chelex in the samples at 100°C in a water bath (Becker *et al*, 2004), however, in this study, this was not successful. A possibility for this is that the temperature is not accurate in all water baths. To solve this problem, the samples were put on a single cycle on a PCR machine set to 95°C for 30 minutes to ensure a constant temperature was maintained. The application of Chelex 100 to elution process was vital in ensuring the DNA from the FTA cards was of a suitable quality and free of inhibitors so it could be successfully amplified by PCR. Without suitable DNA, the entire process would be limited. Removing DNA from the cards also has its own limiting factors. A study by Cox *et al* (2010) exposed that when a sample is collected, parasitic material, in this case, Trypanosome DNA, is not distributed evenly on the FTA cards and therefore, a single punch is not representative of the entire sample. With this in mind, it was decided that 10 punches were to be taken from each card randomly to increase the representation. This reduced the chance of false negatives and when looking for multiple pathogens, by increasing the number of

punches, reduce the chance of missing low amounts of DNA. In order to pick up low levels of DNA in the samples tested via PCR, the protocols for each kind of PCR had to be optimised. Initially, with the ITS PCR looking for Trypanosomal DNA, the initial few PCRs showed limited success. To increase the opportunity for the DNA added to the PCR reaction mixture to be amplified, the number of PCR cycles was increased from 35-40 cycles. Each time increasing the number of cycles by one gave an idea of what point the optimum number of cycles would be. At 40 cycles a bright band was found. To help increase the yield of PCR product further, the volume of eluted DNA was increased from 1μl to 3μl in the first round of PCR. This will allow more binding between DNA and primer, which will result in more DNA being amplified. Both these modifications to the protocol were successful with a combination of them both was deemed to provide the best results. A similar approach was used when conducting the PCR targeting multiple eukaryote DNA. Both increasing the number of cycles as well as the volume of eluted DNA in the PCR reaction mixture proved the best results. The method used in this experiment used Chelex 100 to elute DNA from the FTA card and then adding it to a PCR mixture. There are other possible methods that use direct amplification from the FTA cards using PCR (Scherer *et al*, 2013). This kit allows for quicker amplification direct from the FTA cards which are added to a pre-made master mix. This also reduces the amount of steps needed to complete the extraction and amplification of DNA from the FTA cards, reducing the risk of contamination but also increasing the yield of DNA. This method is also inhibitor resistant which makes the PCR more efficient in amplifying DNA. Using a kit such as this Investigator® STR GO! Kit (QIAGEN Strasse 1, 40724 Hilden, Germany) on older cards may help ensure more DNA is taken from the cards and amplified due to its direct amplification process (Scherer *et al*, 2013).

The success of the NGS has shown a large range of Phyla from which DNA has been amplified. Of the phyla identified, the areas of interest are those that contain common zoonotic and enzoonotic parasites. Apicomplexa and Euglenozoa showed a variety of expected genera such as *Theileria* and *Eimeria*, common parasites of cattle (Mans, Pienaar & Latif, 2015 & Matjila & Penzhorn, 2002), as well as *Leptomonas* and *Bodo* which are relatives of the Trypanosomatidae family (Kraeva *et al*, 2015). This family contains many protozoan parasites, including a number of *Trypanosoma*

species which have been identified in the used samples during another study. Due to the NGS run only sequencing to 95 base pair reads, the accuracy of the results can only be given to genus level, however, with greater read length the accuracy could be increased to a higher taxonomic level, as it was identified that the sequences described as *Leptomonas* are actually *Trypanosoma*. *Sarcocystis*, another genus known to infect mammals, including humans, was also found. Literature supports that this genus contains species of parasite known to infect cattle (Fayer, 2004). These findings are supportive of the success of the study, as parasites like these have regularly been found to infect cattle. An interesting discovery is that of the presence of Basidiomycota genera on the FTA cards. Though it is possible for these bacteria to be in the blood samples as an infection, there is little evidence in literature to suggest cattle are regularly infected with Basidiomycota species. Due to the low abundance of all samples, it is possible that these genera have attached to the cards after the original samples were taken. This is a possible conclusion as Basidiomycota species are known to be present in the majority of land habitats and also contain species of decomposers (Edgcomb *et al*, 2011). Further evidence to support the theory that bacterial genera, such as those of the Basidiomycota, have grown on the FTA cards post sample collection is that there are the same genera found on the control samples surrounding the blood sample as found on the blood sample itself. Another explanation for this could also be the conditions in which the samples were collected. The samples were collected from an abattoir in Nigeria, where the hygiene and cleanliness of the environment may result in more than just cattle blood attaching to the cards. *Malassezia* is a genus of species that include a number of yeast species. These are opportunistic infections and have been seen in high prevalence in Zebu cattle (Duarte *et al*, 2003). Another yeast like organism found is the genus *Trichosporon*. Known to infect a range of environments such as soil, lakes, stream, as well as bird faeces and cattle. It is also known to be found on humans as a benign infection of the hair, known as White piedra (Colombo *et al*, 2011). This again suggests that there may be contamination from such organisms during collection if hygiene standards are low.

The location from where the samples were taken off the FTA card was chosen to identify if there were similar sequences of DNA in the blood stain on the FTA card

as well as the unstained card surrounding the blood stain. It had been expected that the controls taken from the area surrounding the blood stain would be free of pathogenic DNA. It was thought that there may be bacterial DNA present on the controls due to the regular handling of the cards and also due to the age of the cards, however, the presence of any pathogenic or parasitic material was not expected. The results showed that there was DNA from the phylum Apicomplexa and Euglenozoa such as *Trypanosoma, Eimeria,* and *Theileria.* Though not in as great abundance as in the blood samples, the presence suggests that when samples are applied to an FTA card, the genetic material is not isolated to the area where the blood stain is but may spread across the card. This shows the ability of the entire card to hold and fix DNA, but also shows the sensitivity of the cards by their ability to store small amounts of DNA but also for this amount to be amplified and identified. The Shannon diversity index box plot also shows there is diversity found in the control samples as well as the blood samples. With more samples compared it is expected that the alpha diversity would be greater in the blood samples, however, the analysis of data showed no significant difference in the mean number of OTUs identified from the blood samples and the control samples. The NMDS analysis shows an even distribution with the majority of samples remaining between 0.4 and 0.3, also suggesting that there is a similarity in species between the two sample locations.

By taking replicates from a number of samples, it was hoped that it would be possible to identify via NGS whether the location the punches were taken from the FTA card would influence the results. A study by Cox *et al*, (2010) identified that when trying to identify Trypanosome DNA from FTA cards, the positivity of a sample can be influenced by the area that the punch is taken from the FTA card. This is due to the uneven deposition of parasitic material on the cards when the sample is originally taken. Using NGS the results from this study support these findings, as it was seen that the number of sequence reads of a Trypanosome sequence had high degrees of variance between replicate samples, however, there was no significant difference in the mean sequence reads between the replicated samples. In order to further investigate this, a larger number of replicates would be needed with more samples. In this study, 10 punches were taken from each card. With the NGS data

showing such variance of Trypanosome DNA supporting the findings of Cox *et al*, 2010, it is also possible that other pathogens are also not accurately represented. This is not a reflection of the FTA cards ability to store DNA but shows a difficulty in the method of DNA extraction to produce a true representation of the genetic material stored on the FTA cards. Further investigation into the most efficient way of getting this would require taking punches from the whole of the card, however, the material on the cards is often high in value. In order to identify a possible solution to this issue, the FTA cards used could be deliberately infected with known pathogens and map the distribution so that the representation of DNA found by NGS can be compared with the known representation on the cards.

The samples used in this study have been previously tested for Trypanosome DNA using ITS PCR. When tested using ITS PCR, some previously positive samples came up negative. As a result, the protocol was successfully adapted by increasing the number of cycles and the volume of eluted FTA card solution to allow for DNA to be amplified. This decreases the comparability of the results due to the difference in protocols, however it did help identify previously positive samples were still able to be used. The need to change the protocol to find the positive samples is possibly as a result of the age of the cards. As previously discussed, FTA cards become less useable after 9 months as the DNA stored on them becomes more difficult to elute DNA completely (Becker *et al,* 2004). However, as part of the analysis of results, a comparison was made between the results of the ITS PCR and the NGS results using Cohen's Kappa statistics. Using a conservative model gave a fair agreement (κ-value= 0.23) between the two sets of results. From this it may suggest that ITS PCR is less sensitive than NGS, however, it must also be considered that evidence from previous studies by Cox *et al*, (2010) has shown that location on the card where the punches are taken can influence the result which is the more likely conclusion. In order to further strengthen the comparison, the same card locations must be tested with both NGS and ITS PCR in order to identify the most reliable and sensitive test. Another way to do this would be to also create a series of FTA cards with a known amount of Trypanosome DNA in the blood and have negative controls that contain uninfected blood to increase the accuracy of both experiments. It must also be calculated what the level of noise is and what is an actual positive sample. By

establishing an accurate level, it will then make the diagnosis of either positive or negative more accurate and reliable.

The ability to identify a large diversity and abundance of genera from FTA cards can help enable further study into a number of areas. Using NGS in this way can allow for the identification of pathogens from large populations of a sampled species, but also understand a relative abundance and diversity accurately. In order to increase the accuracy, an accurate representation of the FTA card should be taken to ensure DNA is not miss represented in the final pooled sample.

By identifying pathogenic DNA in samples from FTA cards using NGS, it identifies the diversity of DNA as well as the abundance. However, this does not explain the role of the pathogen infection. The presence of pathogenic or parasitic DNA may be due to a chronic or acute infection or perhaps its presence may only represent a past infection with a low level of DNA remaining. To identify the role of the infection further, a more specific analysis would be needed.

There has been limited work on using next-generation sequencing to identify 18s sequences from field samples collected on FTA cards targeting pathogenic DNA of multiple species. A study by Montmayeur *et al* (2017) experimentally infected FTA cards with Poliovirus to identify full genomes quickly using NGS. FTA cards have mainly been used in conjunction with FTA cards to identify genomes of a specific target species, another example of using FTA cards on Genotyping is a study by McClure *et al* (2009), attempting to genotype cattle in the USA. Prior to this study, there was nothing available in the literature identifying multiple pathogens using 18s sequences from field blood samples collected on FTA cards. This shows an area in which this work can be combined to improve the collection and identification of samples in other studies. Being able to identify pathogens from FTA cards using NGS could be applied to many areas of research. A study in South Africa looking at the diversity of the 18S SSU rRNA hyper-variable region of *Theileria* in cattle and Cape buffalo (*Syncerus caffer*) using NGS. In this study, EDTA-blood samples were collected (Mansa *et al*, 2016). Research investigating the quantity and quality of genomic DNA collected from blood samples stored at varying temperatures after long-term storage (up to 1 year), found that the amount and quality of genomic DNA collected from EDTA-blood samples decreased over time at +4°C. The lower the

temperature, the longer the DNA was preserved up to a year (Bulla *et al*, 2016). The lifespan of the DNA collected from buffalo and cattle could be extended with storage on FTA cards, and now that 18s sequences can be identified on these cards using NGS, the same identification can be done. The application of this technique could also help expand similar projects like the one conducted by Mansa *et al* (2011) by targeting all Apicomplexa DNA found in Cape Buffalo.

In a number of studies on FTA cards, a number of discs are taken from each card and tested individually for the presence of parasite DNA. It has been proven in the literature that the use of Chelex 100 resin can increase the efficiency of DNA elution from blood samples and also more cost-effective than using DNA extraction kits (Becker *et al*, 2004). It increases the volume of DNA removed from the cards which can then be used for PCR. Though this will help show if the sample is positive or not, it does not give a representation of prevalence of a parasite. In the case of Trypanosomiasis, parasitemia can vary due to the variation of antigens expressed by the trypanosome surface proteins (Ahmed *et al*, 2011). The use of NGS on samples such as these can show the abundance of a parasite or pathogen within the sample and can even detect low levels of DNA present in a sample. This could then be used to monitor the prevalence within a population as well as individual samples.

Often in parasitic infections, parasite levels can fluctuate. This makes it difficult to identify a population's prevalence as these fluctuations can result in the density of a parasite dropping to a level that is not detectable (O'Meara *et al*, 2007). Using NGS we have been able to detect low levels of parasite infection within a population of samples. This could give a more accurate representation of prevalence of parasitic infection, however, could it also represent a level of infection that is naturally present and shows no pathological symptoms? Levels of malaria in individuals vary depending on exposure to the parasite. A subject that has limited or no prior exposure to the malaria parasite will exhibit a greater parasite load with larger fluctuations than an individual who has had malaria in the past and may have developed slight immunity. These individuals will show a lower parasite load and their clinical symptoms will be reduced or non-existent (O'Meara *et al*, 2007). The data gathered in this study shows low levels of a number of parasites in the samples. What the data does not show is the role of these parasites on the individual the

sample is taken from, but also the role of the parasite within the population. Looking forward from this study, it would be interesting to investigate the low levels of pathogens found in the samples and identify whether they are an active infection, if they have an effect on the host or whether the low levels of a parasite represent an old infection.

In conclusion of this study, there are some possible enhancements to be made. To increase the accuracy of the overall results, a large number of replicates must be taken from each of the cards to identify a larger coverage of the DNA present within the sample, but also to be able to make a more accurate comparison and statistical analysis of the results. Along with this, during the next generation process, a greater number of base pair reads must be taken. By doing this it will allow the taxonomy of the sequenced DNA to be more accurate and give a greater understanding of the sequences present in a sample and their relationship to one another. In general, to increase this already successful experiment, the scale of the project should be increased. Another area of improvement would be to use a set of FTA cards with a known infection status. By applying blood samples that have been mixed with a known volume of pathogenic material, it will allow for predicted results to be compared to actual results and more accurate deductions can be made from the results. Alongside this, more control samples are needed but this need to be improved. Controls must still be taken from surround the infected sample but having controls taken from an uninfected blood sample on FTA card would help expose any inaccuracies in the process. The same uninfected blood sample must follow the same protocols as the infected samples as a quality control for the protocol.

5.0 References

- Ahmed, H.A., MacLeod, E.T., Hide, G., Welburn, S.C. and Picozzi, K. (2011). The best practice for preparation of samples from FTA®cards for diagnosis of blood borne infections using African trypanosomes as a model system. *Parasites and Vectors*. **4**, 68.
- Al-Griw, H. H., Zraba, Z. A., Al-Muntaser, S. K., Draid, M. M., Zaidi, A. M., Tabagh, R. M., & Al-Griw, M. A. (2017). Effects of storage temperature on the quantity and integrity of genomic DNA extracted from mice tissues: A comparison of recovery methods. *Open Veterinary Journal*, **7**, 239–243. http://doi.org/10.4314/ovj.v7i3.7
- Amaral-Zettler, L. A., McCliment, E. A., Ducklow, H. W., & Huse, S. M. (2009). A Method for Studying Protistan Diversity Using Massively Parallel Sequencing of V9 Hypervariable Regions of Small-Subunit Ribosomal RNA Genes. *PLoS ONE*, *4*(7), e6372. http://doi.org/10.1371/journal.pone.0006372
- Auburn, S., Campino, S., Clark, T., Djimde, A., Zongo, I., Pinches, R., Manske, M., Mangano, V., Alcock, D., Anastasi, E., Maslen, G., MacInnis, B., Rockett, K., Modiano, D., Newbold, C., Doumbo, O., Ouédraogo, J., & Kwiatkowski, D. (2011). An Effective Method to Purify Plasmodium falciparum DNA Directly from Clinical Blood Samples for Whole Genome High-Throughput Sequencing. *PLoS One*. **6**, e22213.
- Becker, S., Franco, J.R., Simarro, P.P., Stich, A., Abel, P.M., Steverding, D. (2004). Real-time PCR for detection of Trypanosoma brucei in human blood samples. *Diagnostic Microbiology and Infectious Disease*. **50**, 193-9.
- Behjati, S., & Tarpey, P. S. (2013). What is next generation sequencing? *Archives of Disease in Childhood. Education and Practice Edition*, **98**, 236–238. http://doi.org/10.1136/archdischild-2013-304340
- Belda, E., Boubacar Coulibaly, B., Fofana, A., Beavogui, A., S., Gohl, D., Vernick, K., & Riehle, M. (2017). Preferential suppression of Anopheles

gambiae host sequences allows detection of the mosquito eukaryotic microbiome. *Scientific Reports*. **5**.

- Boid, R., Jones, T.W., & Munro, A. (1999). A simple procedure for the extraction of trypanosome DNA and host protein from dried blood meal residues of haematophagous Diptera. *Veterinary Parasitology*. **85**, 313–317.
- Bonnet, S., Michelet, L., Moutailler, S., Cheval, J., Hébert, C., Vayssier-Taussat, M., & Eloit, M. (2014). Identification of Parasitic Communities within European Ticks Using Next-Generation Sequencing. *PLoS Neglected Tropical Diseases*, 8, e2753.
- Bouéa, F., Berbrib, I.E., Hormaza, V., Bouchera, J.-M., Mamyc, A.B.E., Traored, A., Fihrib, O.F., Petavye, A.-F., Dakkakb, A., & Umhanga, G. (2017). Use of FTA® card methodology for sampling and molecular characterization of *Echinococcus granulosus sensu lato* in Africa. *Experimental Parasitology*. **173**, 29–33.
- Boukary, A. R., Thys, E., Rigouts, L., Matthys, F., Berkvens, D., Mahamadou, I., Yenikoye, A. and Saegerman, C. (2012), Risk Factors Associated with Bovine Tuberculosis and Molecular Characterization of *Mycobacterium bovis* Strains in Urban Settings in Niger. *Transboundary and Emerging Diseases*, **59**: 490–502.
- Bulla, A., De Witt, B., Ammerlaan, W., Betsou, F., & Lescuyer, P. (2016). Blood DNA Yield but Not Integrity or Methylation Is Impacted After Long-Term Storage. *Biopreserv Biobank*. **14**, 29-38.
- Checchi, F., Filipe, J. A., Haydon, D. T., Chandramohan, D., & Chappuis, F. (2008). Estimates of the duration of the early and late stage of gambiense sleeping sickness. *BMC Infectious Diseases*, **8**, 16. http://doi.org/10.1186/1471-2334-8-16
- Colombo, A. L., Padovan, A. C. B., & Chaves, G. M. (2011). Current Knowledge of Trichosporon spp. and Trichosporonosis. *Clinical Microbiology Reviews*, **24**, 682–700. http://doi.org/10.1128/CMR.00003-11
- Cooper, M.K., Phalena, D. N., Donahoea, S.L., Roseb, K., Šlapetaa, J. (2016). The utility of diversity profiling using Illumina 18S rRNA gene amplicon deep

sequencing to detect and discriminate Toxoplasma gondii among the cyst-forming coccidia. *Veterinary Parasitology*. **216**, 38–45.

- Cox, A. P., Tosas, O., Tilley, A., Picozzi, K., Coleman, P., Hide, G., & Welburn, S. C. (2010). Constraints to estimating the prevalence of trypanosome infections in East African zebu cattle. *Parasites & Vectors*, **3**, 82.
- Cox, A. Tilley, A., McOdimba, F., Fyfe, J., Hide, G. and Welburn, S.C. (2005). A PCR based assay for detection and differentiation of African trypanosome species in blood. *Experimental Parasitology*. **111**, 24-29.
- Cwiklinski, K., Allen, K., LaCourse, J., Williams, D. J., Paterson, S., & Hodgkinson, J. E. (2015). Characterisation of a novel panel of polymorphic microsatellite loci for the liver fluke, Fasciola hepatica, using a next generation sequencing approach. *Infection, Genetics and Evolution*, **32**, 298–304. http://doi.org/10.1016/j.meegid.2015.03.014
- de Almeida, P.P., Ndao, M., Van Meirvenne, N., & Geerts, S. (1998). Diagnostic evaluation of PCR on dried blood samples from goats experimentally infected with Trypanosoma brucei brucei. *Acta Tropica*. **70**, 269–276.
- Deagle, B., Kirkwood, R., & Jarman, S. (2009) Analysis of Australian fur seal diet by pyrosequencing prey DNA in faeces. *Molecular Ecology*, **18**, 2022–2038.
- Delespauxa, V., Ayralb, F., Geysena, D., & Geertsa, S. (2003). PCR-RFLP using Ssu-rDNA amplification: applicability for the diagnosis of mixed infections with different trypanosome species in cattle. *Veterinary Parasitology*. **117**, 185–193.
- Duarte, E., Batista. R., Hahn R., & Hamdan, J. (2003). Factors associated with the prevalence of Malassezia species in the external ears of cattle from the state of Minas Gerais, Brazil. *Medical Mycology*. **41**, 137–142.
- Ducrotoy, M. J., Bertu, W. J., Ocholi, R. A., Gusi, A. M., Bryssinckx, W., Welburn, S., & Moriyón, I. (2014). Brucellosis as an Emerging Threat in Developing Economies: Lessons from Nigeria. *PLoS Neglected Tropical Diseases*, **8**, e3008.

- Earth Microbiome Project. (2016). 18S Illumina Amplicon Protocol. Available: http://www.earthmicrobiome.org/protocols-and-standards/18s/. [Accessed 3rd May 2018].
- Eckert, J., Gemmell, M.A., Meslin, F.-X., & Pawłowski, Z.S (2001). *WHO/OIE Manual on Echinococcosis in Humans and Animals: a Public Health Problem of Global Concern*. Paris: World Organisation for Animal Health. 101-120.
- Edgcomb, V. P., Beaudoin, D., Gast, R., Biddle, J. F. & Teske, A (2011). Marine subsurface eukaryotes: the fungal majority. *Environ. Microbiol.* **13**, 172–183
- Fayer, R. (2004). *Sarcocystis* spp. in Human Infections. *Clinical Microbiology Reviews*, 17, 894–902. http://doi.org/10.1128/CMR.17.4.894-902.2004
- Fèvre, E., Picozzi, K., Fyfe, J., Waiswa, C., Odiit, M., Coleman, P., & Welburn, S. (2005). A burgeoning epidemic of sleeping sickness in Uganda. *The Lancet*. **366**, 747.
- Fyfe, J., Picozzi, K., Charles Waiswa, C., Bardosh, K. L., Welburn, S. C. (2016). Impact of mass chemotherapy in domestic livestock for control of zoonotic T. b. rhodesiense human African trypanosomiasis in Eastern Uganda. *Acta Tropica*. 1-14. http://dx.doi.org/10.1016/j.actatropica.2016.08.022.
- Hadziavdic, K., Lekang, K., Lanzen, A., Jonassen, I., Thompson, E. M., & Troedsson, C. (2014). Characterization of the 18S rRNA Gene for Designing Universal Eukaryote Specific Primers. *PLoS ONE*, **9**, e87624.
- Hammond, J., Cai, D., & Verhey, K. (2008). Tubulin modifications and their cellular functions. *Current Opinion in Cell Biology*, **20**, 71–76. http://doi.org/10.1016/j.ceb.2007.11.010
- Hide, G., & Tait, A. (2009). Molecular epidemiology of African sleeping sickness. *Parasitology*. 136 (0), 1491–1500.
- Hino, A., Maruyama, H., & Kikuchi, T. (2016). A novel method to assess the biodiversity of parasites using 18S rDNA Illumina sequencing; parasitome analysis method. *Parasitology International*. **65**, 572-575.

- Jarman, S., Deagle, B., Gales, N. (2004). Group-specific polymerase chain reaction for DNA-based analysis of species diversity and identity in dietary samples. *Molecular Ecology*. **13**, 1313–1322.
- Jongejan, F., Perry, B.D., Moorhouse, P.D.S., Musisi, F. L.,Pegram, R.G., Snacken, M. (1988). Epidemiology of bovine babesiosis and anaplasmosis in Zambia. *Tropical Animal Health and Production*. **20**, 234–242.
- Kraeva, N., Butenko, A., Hlaváčová, J., Kostygov, A., Myškova, J., Grybchuk, D., Leštinová, T., Votýpka, J., Volf, P., Opperdoes, F., Flegontov, P., Lukeš, J., & Yurchenko, V. (2015). Leptomonas seymouri: Adaptations to the Dixenous Life Cycle Analyzed by Genome Sequencing, Transcriptome Profiling and Co-infection with Leishmania donovani. *Plos Pathogens*. **11**, e1005127.
- Kristjanson, P., Swallow, B., Rowlands, G., Kruska, R., & de Leeuw, P. (1999). Measuring the costs of African animal trypanosomosis, the potential benefits of control and returns to research. *Agricultural Systems*. **59**, 79-98.
- Kumar, M., & Shukla, P. K. (2005). Use of PCR Targeting of Internal Transcribed Spacer Regions and Single-Stranded Conformation Polymorphism Analysis of Sequence Variation in Different Regions of rRNA Genes in Fungi for Rapid Diagnosis of Mycotic Keratitis. *Journal of Clinical Microbiology*, **43**, 662–668. http://doi.org/10.1128/JCM.43.2.662-668.2005
- Lanham, S., & Godfrey, D., (1970) Isolation of salivarian trypanossomes from man and mammals using DEAE-cellu lose. Exp. Parasitol. 28, 521-534.
- Laohasinnarong, D., Thekisoe, O.M.M., Malele, I., Namangala, B., Ishii, A., Goto, Y., Kawazu, S., Sugimoto, C., Inoue, N. (2011). Prevalence of Trypanosoma sp. in cattle from Tanzania estimated by conventional PCR and loop-mediated isothermal amplification (LAMP). *Parasitology Research*. **109**, 1735–1739.
- Leray, M., Agudelo, N., Mills, S., & Meyer, C. (2013). Effectiveness of Annealing Blocking Primers versus Restriction Enzymes for Characterization of Generalist Diets: Unexpected Prey Revealed in the Gut Contents of Two Coral Reef Fish Species. *PLoS ONE*, **8**.
- Levitt, B., Obala, A., Langdon, S., Corcoran, D., O'Meara, W. P., & Taylor, S. M.

(2017). Overlap Extension Barcoding for the Next Generation Sequencing and Genotyping of Plasmodium falciparum in Individual Patients in Western Kenya. *Scientific Reports*, *7*,

- Loftus, R. T., MacHugh, D. E., Bradley, D. G., Sharp, P. M., & Cunningham, P. (1994). Evidence for two independent domestications of cattle. *Proceedings of the National Academy of Sciences of the United States of America*, **91**, 2757–2761.
- Lorusso, V., Wijnveld, M., Majekodunmi, A. O., Dongkum, C., Fajinmi, A., Dogo, A. G., Thrusfield, M., Mugenyi, A., Vaumourin, E., Igweh, A. C., Jongejan, F., Welburn, S. C., & Picozzi, K. (2016). Tick-borne pathogens of zoonotic and veterinary importance in Nigerian cattle. *Parasit Vectors*. **9**, 217.
- Makalaa, L.H., Manganic, P., Fujisakia, K., & Nagasawa, H. (2003). The current status of major tick-borne diseases in Zambia. *Veterinary Research*. **34**, 27 - 45.
- Mans, B., Pienaar, R., & Latif, A. (2015). A review of Theileria diagnostics and epidemiology. *International Journal for Parasitology: Parasites and Wildlife*. **4**, 104-118.
- Mansa, B., Pienaar, R., Ratabane, J., Pule, B., Latif, A. (2016). Investigating the diversity of the 18S SSU rRNA hyper-variable region of Theileria in cattle and Cape buffalo (Syncerus caffer) from southern Africa using a next generation sequencing approach. *Ticks and Tick-borne Diseases*. **7**, 869–879.
- Mardis, E. (2013). Next-Generation Sequencing Platforms. Annual Review of Analytical Chemistry. **6**, 287-303.
- Matjila, P., & Penzhorn, B. (2002). Occurrence and diversity of bovine coccidia at three localities in South Africa. *Veterinary Parasitology*. **104**, 93-102.
- McClure, M., McKay, S., Schnabel, R., & Taylor, J. (2009). Assessment of DNA extracted from FTA® cards for use on the Illumina iSelect BeadChip. *BMC Research Notes*, **2**, 107. http://doi.org/10.1186/1756-0500-2-107

- McMurdie, P., & Holmes S. (2013) phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. PLoS ONE. 8(4):e61217
- Miller, R., Montoya, V., Gardy, J.,Patrick, D., & Tang, P. (2013). Metagenomics for pathogen detection in public health. *Genome Medicine*. **5**, 1-14.
- Montmayeur, A., Ng, T., Schmidt, A., Zhao, K., Magaña, L., Iber, J., Castro, C., Chen, Q., Henderson, E., Ramos, E., Shaw, J., Tatusov, R., Dybdahl-Sissoko, N., Endegue-Zanga, M., Adeniji, J., Oberste, M., Burns, C. 2017. High-throughput next-generation sequencing of polioviruses. *J Clin Microbiol* **55**, 606–615.
- Morrow, C., & Fraser, J. (2009). Sexual reproduction and dimorphism in the pathogenic basidiomycetes. *FEMS Yeast Research*. **9**, 161–177.
- Musinguzi, S.P., Suganuma, K., Asada, M., Laohasinnarong, D., Sivakumar, T., Yokoyama, N., Namangala, B., Sugimoto, C., Suzuki, Y., Xuan, X., & Inoue, N. (2016). A PCR-based survey of animal African trypanosomosis and selected piroplasm parasites of cattle and goats in Zambia. *The Journal of Veterinary Medical Science*.
- Ndao, M. (2009). Diagnosis of Parasitic Diseases: Old and New Approaches. *Interdisciplinary Perspectives on Infectious Diseases*. **2009**, 1-15.
- O'Meara, W. P., Collins, W. E., & McKenzie, F. E. (2007). Parasite Prevalence: A Static Measure of Dynamic Infections. *The American Journal of Tropical Medicine and Hygiene*, **77**, 246–249.
- Owor, B., Shepherd, D., Taylor, N., Edema, R., Monjane, A., Thomson, J., Martin, D., & Varsani, A. (2007). Successful application of FTA® Classic Card technology and use of bacteriophage φ29 DNA polymerase for large-scale field sampling and cloning of complete maize streak virus genomes. *Journal of Virological Methods*. **140**, 100-105.
- Perera, R., Heneghan, C., & Badenoch, D. (2008). Questionnaire studies 2:inter-rater agreement. In: Banks, M., Brindley, L., & Pittman V. *Statistics Toolkit*. Massachusetts: Blackwell Publishing. 59-64.
- Picard-Meyer, E., Barrat, J., Cliquet, F. (2006). Use of filter paper (FTA) technology for sampling, recovery and molecular characterisation of rabies viruses. *Journal of Virological Methods*. **140**, 174-82.

- Picozzi, K., Tilley, A., Fèvre, E., Coleman, P., Magona, J., Odiit, M., Eisler, M., & Welburn, S. (2002). The diagnosis of trypanosome infections: applications of novel technology for reducing disease risk. *African Journal of Biotechnology*. **1**, 39–45.
- Rao, P., Uplekar, S., Kayal, S., Mallick, P., Bandyopadhyay, N., Kale, S., Singh, O., Akshaya Mohanty, A., Mohanty, S., Wassmer, S., & Carltonco, J. (2016). A Method for Amplicon Deep Sequencing of Drug Resistance Genes in *Plasmodium falciparum* Clinical Isolates from India. *Journal of Clinical Microbiology*. **54**, 1500–1511.
- Saieg, M. A., Geddie, W. R., Boerner, S. L., Liu, N., Tsao, M., Zhang, T., Kamel-Reid, S. and da Cunha Santos, G. (2012), The use of FTA cards for preserving unfixed cytological material for high-throughput molecular analysis. Cancer Cytopathology, 120: 206–214. doi:10.1002/cncy.20205
- Scherer, M., Alsdorf, B., Bochmann, L., Prochnow, A., & Engel, H. (2013). Development of the Investigator® STR GO! Kits for direct amplification of reference samples. *QIAamp® DNA Investigator Handbook*. **1**, 1.
- Schlottau, K., Schulze, C., Bilk, S., Hanke, D., Höper, D., Beer, M., & Hoffmann, B. (2016). Detection of a Novel Bovine Astrovirus in a Cow with Encephalitis.. Transbound Emerg Dis. **63**, 253-9.
- Shehzad, W., Riaz, T., Nawaz, M., Miquel, C., Poillot, C., Shah, S., Pompanon, F., Coissac, E., Taberlet, P. (2012). Carnivore diet analysis based on next-generation sequencing: application to the leopard cat (Prionailurus bengalensis) in Pakistan. *Molecular Ecology*. **21**, 1951-1965.
- Swallow, B. (2000). Impacts of trypanosomiasis on African agriculture. *PAAT Technical and Scientific.* **Series 2**. 52.
- Vestheim, H., & Jarman, S. (2008). Blocking primers to enhance PCR amplification of rare sequences in mixed samples – a case study on prey DNA in Antarctic krill stomachs. *Frontiers in Zoology*. **5**, 12.
- Wang, Y., Tian, R. M., Gao, Z. M., Bougouffa, S.,Qian, P-Y. (2014). Optimal Eukaryotic 18S and Universal 16S/18S Ribosomal RNA Primers and Their Application in a Study of Symbiosis. *PLoS ONE*. 9 (0), e90053.

- Welburn, S., Coleman, P., Maudlin, I., Fèvre, E.,Odiit, M., Eisler, M. (2006). Crisis, what crisis? Control of Rhodesian sleeping sickness. *Trends in Parasitology*. **22**, 123–128.
- Wickham H (2009). *ggplot2: elegant graphics for data analysis*. New York: Springer.
- Yohe, S., Thyagarajan, B. (2017). Review of Clinical Next-Generation Sequencing. *Arch Pathol Lab Med*. **141**, 1544-1557.

6.0 Appendix

6.1 Appendix 1: Mean sequence reads for *Leptomonas* from 4 replicates taken from 5 different FTA cards. Also includes Standard Deviation.

Sample	Mean of Replicates	SD	SD x2
AH24	245	256.1575036	512.3150073
AH25	1190	1798.643378	3597.286755
AH28	347	320.2426164	640.4852327
AH30	382.25	466.6628869	933.3257738
AH66	437.25	453.0683355	906.1366711

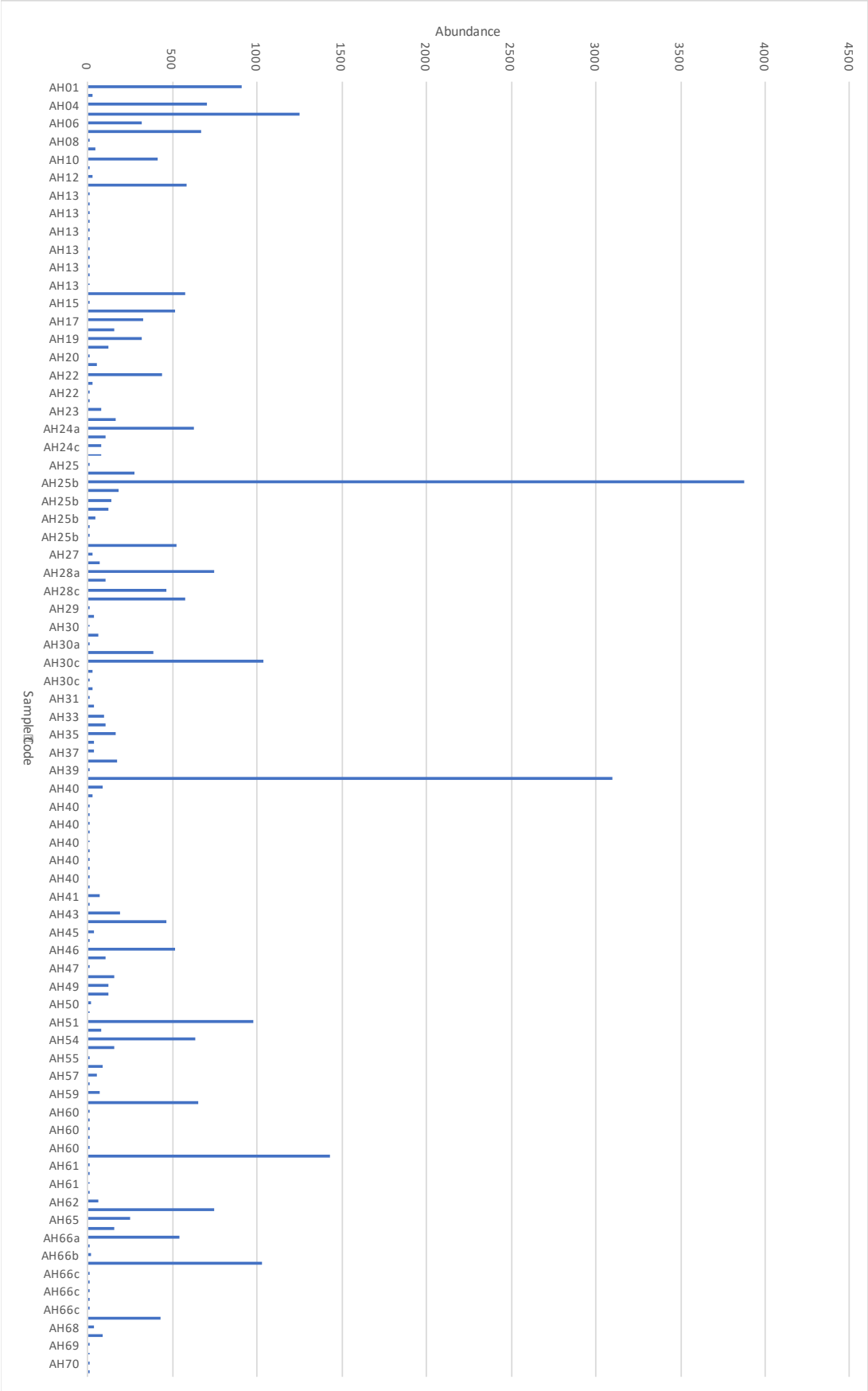
6.2 Appendix 2: A comparison of ITS tested samples and NGS tested samples identifying Trypanosome infection using Cohen's Kappa coefficient. Including the threshold number of reads chosen by a conservative estimate, mean sequence read of control samples and half the mean sequence read of control samples.

Sample	Location	Infected ITS	Tryp	NGS 500	NGS 70.5 (Average read)	NGS 35
AH01	Blood	Y	Unknown +	Y	Y	Y
AH02	Blood	Y	T. theileri	N	Y	Y
AH03	Blood	Y	T. simiae	N	N	N
AH04	Blood	Y	T. simiae	Y	Y	Y
AH05	Blood	Y	T. theileri/ T. vivax	Y	Y	Y
AH06	Blood	N		N	Y	Y

AH07	Blood	Y		Y	Y	Y
AH08	Blood	Y		N	N	N
AH09	Blood	Y		N	N	Y
AH10	Blood	Y	T. simiae	N	Y	Y
AH11	Blood	Y	T. theileri/ T. vivax	N	N	N
AH12	Blood	N		N	N	N
AH13	Blood	N		Y	Y	Y
AH14	Blood	N		Y	Y	Y
AH15	Blood	N		N	N	N
AH16	Blood	N		Y	Y	Y
AH17	Blood	N		N	Y	Y
AH18	Blood	N		N	Y	Y
AH19	Blood	N		N	Y	Y
AH20	Blood	N		N	Y	Y
AH21	Blood	Y	T. theileri	N	N	Y
AH22	Blood	N		N	Y	Y
AH23	Blood	N		N	Y	Y
AH24	Blood	Y	Unknown/T. theileri/ T. vivax	Y	Y	Y
AH25	Blood	Y	Unknown +	Y	Y	Y
AH26	Blood	Y	T. vivax	N	N	N
AH27	Blood	Y	T. vivax	N	N	N
AH28	Blood	Y	T. vivax	Y	Y	Y
AH29	Blood	Y	T. theileri	Y	Y	Y
AH30	Blood	Y	Unknown/T. theileri/ T. vivax/ T. simiae	Y	Y	Y
AH31	Blood	N		N	N	N
AH32	Blood	N		N	N	Y
AH33	Blood	N		N	Y	Y
AH34	Blood	N		N	Y	Y
AH35	Blood	N		N	Y	Y
AH36	Blood	N		N	N	Y
AH37	Blood	N		N	N	Y
AH38	Blood	Y	T. vivax	N	Y	Y
AH39	Blood	N		N	N	N
AH40	Blood	N		Y	Y	Y
AH41	Blood	Y	T. vivax	N	Y	Y
AH42	Blood	N		N	N	N
AH43	Blood	N		N	Y	Y
AH44	Blood	N		N	Y	Y
AH45	Blood	N		N	N	Y
AH46	Blood	N		Y	Y	Y

AH47	Blood	N		N	Y	Y
AH48	Blood	N		N	Y	Y
AH49	Blood	N		N	Y	Y
AH50	Blood	N		N	Y	Y
AH51	Blood	N		Y	Y	Y
AH52	Blood	N		N	Y	Y
AH53	Blood	N		N	N	N
AH54	Blood	N		Y	Y	Y
AH55	Blood	N		N	Y	Y
AH56	Blood	N		N	Y	Y
AH57	Blood	N		N	N	Y
AH58	Blood	N		N	N	N
AH59	Blood	N		N	Y	Y
AH60	Blood	N		Y	Y	Y
AH61	Blood	N		Y	Y	Y
AH62	Blood	N		N	N	Y
AH63	Blood	N		N	N	Y
AH64	Blood	N		Y	Y	Y
AH65	Blood	N		N	Y	Y
AH66	Blood	N		Y	Y	Y
AH67	Blood	N		N	Y	Y
AH68	Blood	N		N	N	Y
AH69	Blood	N		N	Y	Y
AH70	Blood	N		N	N	N

6.2.1 Graph showing the abundance of Trypanosome sequence reads, used to create conservative model estimates on infected and uninfected samples.



6.2.2 Cohen's Kappa coefficient for conservative estimate of 1000 sequence reads compared to ITS PCR results.

		ITS PCR			
		Positive	Negative		
NGS	Positive	9	11	20	28.57%
	Negative	11	39	50	71.43%
		20	50	70	
		28.57%	71.43%		
		k=(Pr(a)-Pr(e))/(1-Pr(e))			
		Pr(a)	0.685714286		
		Pr(e)	0.591832653		
		k	0.2300077		

6.2.3 Cohen's Kappa coefficient for mean sequence reads of control samples compared to ITS PCR results.

		ITS PCR			
		Positive	Negative		
NGS	Positive	12	34	46	65.71%
	Negative	8	16	24	34.29%
		20	50	70	
		28.57%	71.43%		
		k=(Pr(a)-Pr(e))/(1-Pr(e))			
		Pr(a)	0.4		
		Pr(e)	0.432653061		
		k	-0.057553957		

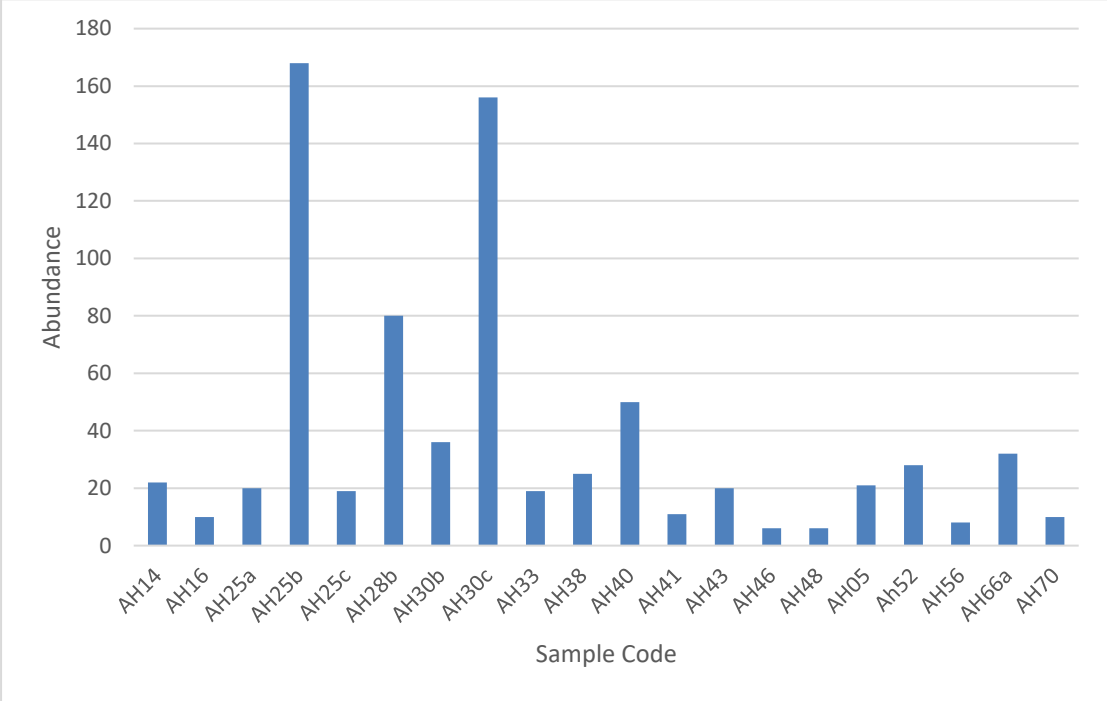
6.2.4 Cohen's Kappa coefficient for half the mean sequence reads of control samples compared to ITS PCR results.

		ITS PCR			
		Positive	Negative		
NGS	Positive	15	42	57	81.43%
	Negative	5	8	13	18.57%
		20	50	70	
		28.57%	71.43%		
		k=(Pr(a)-Pr(e))/(1-Pr(e))			
		Pr(a)	0.328571429		
		Pr(e)	0.365306122		

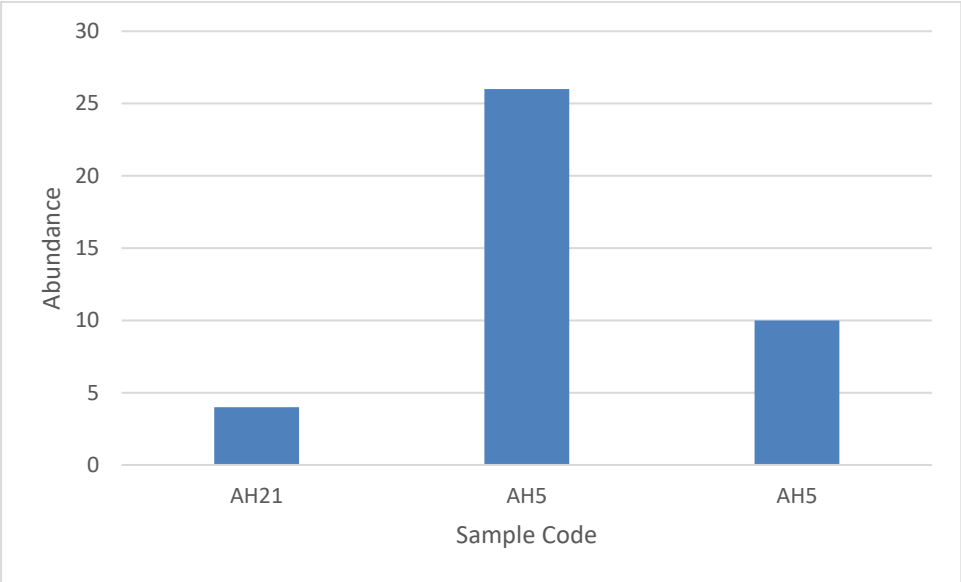
		k	-0.057877814		

6.3 Appendix 3: Graphs showing the abundance of *Theileria, Eimeria,* and *Sarcocystis* used to estimate infective status of each parasite using the conservative model described.

6.3.1 Graph showing the abundance of *Theileria* found in the tested samples. Above 50 sequence reads regarded positive.



6.3.2 Graph showing the abundance of *Eimeria* found in the tested samples. Above 10 sequence reads regarded positive.



6.3.2 Graph showing the abundance of Sarcocystis found in the tested samples. Above 100 sequence reads regarded positive.