The use of molecular biology and epidemiological techniques in managing and studying European Foulbrood in honey bees

Hollie Louise Pufal

Thesis submitted to the School of Natural and Environmental Sciences in fulfilment for the requirements of the degree of Doctor of Philosophy
Newcastle University

18 June 2024

Abstract

European foulbrood (EFB) is a major bacterial disease infecting honey bees internationally, caused by the gram positive bacteria *Melissococcus plutonius*. This study aimed to gain more understanding of the spread of this disease in the UK using molecular and epidemiological methods. There is a multi-locus sequence typing (MLST) scheme that uses four genes to identify sequence types. Assigning a sequence type allows spread to be assessed. In the UK 72% of cases are made up of three sequence types spread across the country. More clarity within these sequence groups would allow outbreak clusters to be identified and managed. A cost-effective whole genome sequencing method was developed, and a single nucleotide polymorphism (SNP) analysis was performed on a selection of samples from naturally infected larval samples from England and Wales. This analysis showed clear outbreak clusters within the sequence types, that were geographically linked. From the same sequencing data, virulence and antimicrobial resistance genes were found. Some bacterial species act as secondary invaders following infection of *M. plutonius*. Whole genome sequencing data found genes from some of these species. One particular bacteria of interest found in some samples was *Paenibacillus alvei* a common secondary invader of *M. plutonius*. A survey was carried out in two areas with high cases of European Foulbrood, Somerset and Cambridgeshire, to assess what the beekeepers' opinions were on why the disease was spreading. Structural equation modelling was used to investigate causes of disease spread. Overall, this study produced whole genome sequencing data that gave clarity to outbreaks, gave more genomic insight and used beekeeper's opinions to assess the risks causing European foulbrood to spread.

Acknowledgements

First I would like to thank my supervisors Giles and Steve for their continued support and guidance throughout this process. Your expertise, patience, and dedication have been instrumental in shaping this thesis and helping me towards becoming an independent researcher. Thank you to Isabel for the temporary supervision amnd support during supervisor absence. Also, to Mark for my putting with my constant bothering on coding issues.

A special thank you to my family, Mum, Dad, Helena, Tom, Nana and Grandad for your love and support throughout this process and constantly believing in me. I am grateful for fellow members of the Bee Group, Ben and Thea, for providing support and friendship, and many giggles throughout my PhD. A special mention goes to my boyfriend, Alex, for his unwavering love, patience, and understanding throughout this journey. Your support has been my rock. I could not have done this without my amazing friendship circle in Newcastle, York and Huddersfield, Heather, Becky, Kate, Chess and Ines to name a few.

It has been a pleasure to be part of the MEP research group, they provided such a positive and supportive environment.

Thank you to Somerset Beekeepers, Cambridgeshire Beekeepers, the Veterinary Medicines Directorate and the Bee Disease Insurance for funding this project and making it possible.

And last but not least to Malteaser bunnies for fuelling my writing process in the last few months, those 3pm coffee breaks got me through.

Declaration

LFD buffer bottles samples were received from Fera Science Ltd via DEFRA, along with the metadata, with a material transfer agreement in place.

All other work was done by Hollie Louise Pufal

Contents

List of Figures

List of Tables

List of Acronyms

Places

UK - United Kingdom

Organisations

NBU - National Bee Unit

DEFRA - Department for Environment, Food & Rural Affairs

BBKA - British Beekeepers Association

Diseases

AFB - American Foulbrood

EFB - European Foulbrood

MRSA - Methicillin-resistant *Staphylococcus aureus*

NCBI - National Center for Biotechnology Information

Laboatory

DNA - Deoxyribonucleic Acid

EDTA - Ethylenediamine tetraacetic acid

FAM - Fluorescein amidite

HTS - High throughput sequencing

LFD - Lateral Flow Device

NGS - Next Generation Sequencing

OTC - Oxytetracycline

PBS - Phosphate Buffered Saline

PCR - polymerase chain reaction

PMA - Propidium Monoazide

qPCR - Quantitative polymerase chain reaction

RNA - Ribonucleic Acid

TAMRA - Tetramethylrhodamine

WGS - Whole Genome Sequencing

Bioinformatics and Genetics

AMR- Antimicrobial Resistance

ARG - Antibiotic Resistance Genes

CC - Clonal Complex

MLST - Multi-loci sequence typing

SNP - Single Nucleotide Polymorphism

ST - Sequence Type

Modelling and Statistics

CFI - Comparative Fit Index

GLM - Generalised Linear Model

RMSEA - Root Mean Square Error of Approximation

SEM - Structural Equation Modelling

Chapter 1

Introduction

1.1 General background

In 2019 UK agriculture contributed around £10,408 million to the economy (Government, 2023) and with over 20% of agricultural crops in the UK reliant on insect pollination, the huge economic benefit of insect pollination is apparent (Breeze et al., 2011). The honey bee (*Apis mellifera*) has been described as the primary insect pollinator, making it an economic asset worth around £430 million (Smith et al., 2011). Along with its pollination contribution, honey bees also contribute to the economy by producing valuable products such as honey and beeswax, in 2020 the value of exported honey in the UK was £20 million (ITC, 2023). Bee population decline has appeared frequently as headline news, and the plight of pollinators and the impact of reduced pollination on food security has become common knowledge (Potts et al., 2010). An annual survey monitoring honey bee colony loss across thirty-five countries predicted a 16.8% winter colony loss in England during 2019/2020, compared to 9% the year before. Some colony losses were reportedly caused by natural disasters and queen problems (1.2% and 6.5% respectively), but causes were not specified for the remaining 9.1% (Gray et al., 2023). Possible causes of honey bee colony losses are changes of land use, monocultures, climate change, loss of habitat, pesticide use and disease (Christen et al., 2016; Conte & Navajas, 2008; Naug, 2009; Potts et al., 2010; Smart et al., 2016; Clair et al., 2020). It is becoming increasingly acknowledged that disease is a major player in causing global colony losses in honey bees (Ellis & Munn, 2005) and the most common bacterial disease in UK honey bees is European Foulbrood (EFB).

EFB is found on every continent where honey bees are kept (Forsgren, 2010). The causative agent of EFB was originally identified by White in 1908 as Bacillus Y, since then it has had many names, *Bacillus pluton*, *Streptococcus pluton*, *Melissococcus Pluton* and now, after increased knowledge of the chemical properties it is known as *Melissococcus plutonius* (Bailey, 1956; Bailey & Collins, 1982; White, 1912). *M. plutonius* is a Gram-positive, non-sporulating bacteria that grows in lanceolate shapes (Bailey, 1956). The bacteria usually grow exponentially in the larval mid-gut and remain there without infecting the larval tissue (Bailey, 1983). Clinical signs of infection are usually most obvious during late spring; signs include an irregular brood pattern, which contains twisted larvae with opaque or creamy white guts (Figure 1.1) (White, 1912). Larvae tend to die after three to five days of infection, and are cleaned out by nurse bees, preventing further infection. Some infected larvae survive to pupation, excrete their guts into the comb and the cells are cleaned out by adult nurse bees, potentially causing bacteria to be transferred onto their mouthparts, causing spread of the disease throughout the colony and to other larvae (Forsgren, 2010; Bailey, 1983). It is unclear why the infected larvae die; one idea is that the bacteria and larvae compete for food causing the larvae to starve. Linked to the competition, infection is most obvious in the spring, this could be due to the ratio of nurse bees to larvae being lower at this time causing increased starvation, whereas in the other seasons larvae receive more food and are more likely to survive and silently spread the pathogen (Forsgren, 2010; NBU, 2017). Possible causes of infection could be general beekeeping practices, spreading infection through beekeepers' PPE or tools, collecting swarms that could potentially be carrying EFB, robbing bees if they are from or visit an infected colony and migratory beekeeping (Jacques et al., 2017; McKee et al., 2004; NBU, 2017). A reduced risk of infection can be achieved if good beekeeping practice is maintained, symptoms of foulbrood are recognised early, and effective quarantine measures, in the cases of infection or swarm collection, are used (NBU, 2017). There are many unknowns about the epidemiology of the disease making it harder to control.

Figure 1.1: A frame infected with EFB, presenting with pepperpot brood and twisted larve. Courtesy The Animal and Plant Health Agency (APHA), Crown Copyright

One method to control and manage EFB is by promoting bee health to beekeepers, raising awareness of the disease, including the risks and prevention. On behalf of the government, the National Bee Unit (NBU) aims to promote bee health across England and Wales. They do this by providing information in multiple formats as well as delivering training programmes (Wilkins et al., 2007). A website is run by the NBU, BeeBase, this has a dual purpose. The first is to provide easy access to a portal of information for beekeepers, including factsheets, leaflets, booklets, and other resources. These resources cover everything from specific disease leaflets, like *Foulbrood Disease of Honey Bees* to generic beekeeping guides, *Starting right with bees* (NBU, 2016, 2017). The other function is that beekeepers register to BeeBase, so the NBU have a record of each beekeeper, this allows disease to be monitored. Registering to BeeBase also allows the beekeepers to receive alerts if EFB is in close proximity to their apiary (Thompson & Brown, 1999). BeeBase also contributes towards a £23 million, 10 year flagship implementation plan, the Healthy Bee Plan 2030. The healthy bee plan 2030 aims to improve the health of honey bees and was created by the Department of Environment and Rural Affairs (DEFRA) and the Welsh government. One of the initiatives is to encourage beekeepers to sign up to BeeBase, another is to review the literature provided by the NBU (Defra & Government,

2021). As well as contributing to the Healthy Bee Plan 2030, the NBU run the apiary inspection programme. The programme trains and deploys regional bee inspectors to test for EFB and American Foulbrood (AFB), another foulbrood disease (Thompson & Brown, 1999).

In the UK EFB is subject to statutory controls, and statutory inspections are carried out throughout England and Wales by the bee inspectors, trained by the NBU, to inspect for foulbrood. Beekeepers in the UK are legally required to report any suspected cases of EFB to their local bee inspector (NBU, 2009). Lateral flow devices (LFD) are used routinely to diagnose EFB and AFB (Tomkies et al., 2009). The LFD is an immunoassay, like the ones currently being used in the COVID-19 home testing kits (Iacobucci, 2020). A suspect larva is added to a sodium azide buffer and a drop of the sample is placed onto the LFD device. Through capillarity action the sample flows through the membrane, and antigen specific antibodies for EFB detect the presence of EFB and turn blue in a positive case. This technology demonstrates the advantages of using in field and quick diagnostics, as a result is presented in less than thirty minutes and action can be taken immediately (Health & Care, 2021; Tomkies et al., 2009).

The action taken with highly diseased colonies is destruction but the treatment of low level of disease has changed over time. Between 1967 and 1984 the advised treatment was oxytetracycline (OTC), a bacteria static, for all low level infected colonies and surrounding ones, known as colony contact treatment (Jones & Morrison, 1962; Katznelson et al., 1952; Waite et al., 2007). OTC is now only used in exceptional cases with permission required from the bee inspector (NBU, 2009). Widespread treatment with OTC has been withdrawn for a variety of reasons, one is that treatment with OTC can hide symptoms of another foulbrood disease, American Foulbrood (AFB), and overuse of the antibiotic could lead to resistance as demonstrated with some AFB strains (Oldroyd et al., 1989), and OTC treatment efficacy was poor (Budge et al., 2010). Before 1967 the advice to beekeepers with infected colonies was to destroy all colonies where EFB was present at any level. The method now most commonly used to treat low level infected colonies is the shook swarm method, this involves taking out the infected combs, scorching the boxes and shaking the bees off onto fresh foundation (Waite et al., 2007). Studies have looked at a combination of OTC with shook swarm and have shown that the re-occurrence is much lower than just OTC alone, molecular techniques were used to monitor the efficiency of such treatments (Budge et al., 2010). Many molecular biology techniques are being used

to monitor, diagnose, and study honeybee diseases.

Molecular biology is defined as "the branch of biology that deals with the nature of biological phenomena at the molecular level through the study of DNA and RNA, proteins, and other macromolecules involved in genetic information and cell function" (Dictionary.com, 2020). A relatively new discipline, with the term believed to be original coined by Walter Warren in 1938 (Weaver, 1970). Since its origin, molecular biology has seen major breakthroughs, in the 1940s the work of Avery led to an understanding of the role of DNA, the 1950s saw Franklin, Watson and Crick uncover the structure of DNA, in the 1960s the first nucleic acid was sequenced (Avery et al., 1944; Holley et al., 1965; Watson & Crick, 1953). A shift into genomic studies was seen in the 1970s with the development of Sanger sequencing, as well as the birth of modern biotechnology when restriction enzymes were discovered (Kelly & Smith, 1970; Sanger et al., 1977). The 1980s saw the origin of the Human Genome project which in 2001 published a complete human genome sequence (Venter et al., 2001). The invention of the PCR (polymerase chain reaction) is believed to be one of the most important scientific breakthroughs in recent human history (Eeles et al., 1992). PCR was invented in 1986 by Kary Mullis and he became the recipient of the Nobel Prize in Chemistry in 1993. PCR is a technique that allows the amplification of small DNA segments, using temperature cycling, DNA polymerase and specific oligonucleotides. During PCR short, specific DNA fragments are amplified to reach a level of detection. The PCR product (amplicon) is run through gel electrophoresis and the presence or absence of a sequence can be assessed (Mullis et al., 1986). This method is now known as conventional PCR and is being utilised for the study of EFB. Conventional PCR was used to detect EFB in hive debris. Hive debris comes off the bottom of the hive, making this a less invasive way to detect and monitor the pathogen, as there is no need to enter the hive (Biová et al., 2021). Conventional PCR is also utilised in the Multi-locus sequence typing (MLST) of EFB when an outbreak occurs (Haynes et al., 2013) along with metabarcoding.

MLST is a method that traditionally uses around 6-10 housekeeping genes to identify different strains of a pathogen. There is allelic variation between loci for each species, so when internal fragments of these chosen genes are sequenced, an allelic profile can be determined and the samples can be categorised into different sequence types. Housekeeping genes are used for MLST as they are conserved across all strains, as the genes are essential, but enough differences exist for strains to be distinguished (Maiden

et al., 1998). Developed by Haynes et al an MLST-like scheme for EFB was designed using a mixture of housekeeping genes and non-housekeeping genes, making this MLST unusual. The EFB MLST utilises four loci, two housekeeping genes galK and purR, and two non-housekeeping genes argE and gbpB, the internal fragments of these genes are amplified using conventional PCR. The amplicons are cleaned up and sequenced using Sanger sequencing, the sequences are then analysed, the allelic profile determined, and a sequence type (ST) is assigned (Haynes et al., 2013). Initially the EFB MLST scheme identified 11 STs internationally and 12 in the UK, studies on a wider dataset of UK samples identified 15 ST types across England and Wales (Budge et al., 2014; Haynes et al., 2013). At the time of writing there are currently 46 STs identified worldwide (Jolley et al., 2018). In the UK the most prominent STs are ST3, ST5 and ST23, distributed nationally and making up 72% of all positive cases (Budge et al., 2014). The STs are grouped into 3 clonal complexes (CCs), CC12, CC13 and CC3 (Budge et al., 2014). The STs are grouped into CCs based on sequence similarity. CCs have been studied to assess worldwide distribution of strains, and showed CC3 to be the most common and that CC12 covers the atypical strains like the one found in Japan (Arai et al., 2012; Budge et al., 2014; Takamatsu et al., 2013). MLST typing is an example of how molecular techniques can be used to study the epidemiology of EFB at a genetic level using sequencing techniques.

Sequencing is a relatively new concept, in 1964 a tRNA was sequenced by Holley et al, this was the first nucleic acid to be sequenced (Holley et al., 1965). In 1977 Frederick Sangar developed a sequencing method called 'chain termination' method. Sanger's method utilised ddNTPs which are dNTPs but terminate the chain when they are incorporated to the template, by preventing the formation of the phosphodiester bond. In four separate reactions, one for each ddNTP, fragments of different lengths are created as all dNTPs are incorporated, so extension will continue in each fragment until a ddNTP is encountered. All four reactions are run side by side on a gel and a sequence can be obtained through visualisation (Sanger et al., 1977). More modern advances introduce fluorescent ddNTPs, each label is a different colour depending on the nucleotide, allowing the sequence to be determined in one reaction rather than four (Smith et al., 1986). This sequencing process was fully automated by ABI in 1987 and in 1996 capillary electrophoresis was introduced, this method is used now, and as a commercial service in some companies such as Eurofins (Eurofins, Germany) (Hunkapiller et al., 1991; Ju et al., 1996). Sanger sequencing is

used for the EFB MLST typing and was one of the methods used in the human genome project (Haynes et al., 2013; Venter et al., 2001). It wasn't until 2005 when the playing field changed to a new generation of sequencing known as next generation sequencing (NGS) or becoming widely known as high throughput sequencing (HTS) (Voelkerding et al., 2009).

The first NGS platform was created by 454 Life Sciences, but now many platforms are commercially available such as Nanopore sequencing (Oxford Nanopore Technologies, ONT), Illumina Sequencing (Illumina, Irvine, CA, USA), PacBio sequencing (Pacific Biosciences, Menlo Park, CA, USA) and Ion torrent® (Thermo Fisher Scientific, Waltham, MA, USA). Each platform has unique strategies, and various sequencing types can be performed, but the overarching method is the same, the nucleotide sequences are fragmented, millions of copies are amplified in parallel and a computer is used to read the sequences and analyse them (Behjati & Tarpey, 2013). The development of NGS allowed for a more high throughput sequencing method than its predecessors, and as time is progressing sequencing is getting cheaper. The National Human Genome Research Institute (NHGRI) calculated that by using NGS the cost of sequencing the whole human genome in 2014 cost approximately $5,731 and in 2020 dropped to $689, showing how quickly this technology is developing and becoming more accessible for general use (KA, 2016). Sequencing has been used to study *M. plutonius* both using whole genome sequencing (WGS) and metabarcoding.

Metabarcoding involves PCR followed by the amplicons being sequenced rather than the whole genome. The 16s region of bacteria can be informative as it is a highly conserved region, essential for its survival, so primers can be designed to pick up a variety of bacterium types in one go or can be designed specifically to a bacteria type. For *M. plutonius* 16s metabarcoding has been used both to sequence purely *M. plutonius* 16s rRNA for surveying and identification, and to study associated bacteria, gut microbiome bacteria and *Melissococcus*-like bacteria types by using more generic primers (Ansari et al., 2017; Arai et al., 2012; Erban et al., 2017). Secondary invaders have been found in EFB positive larvae these were *Paenibacillus dendritiformis*, *Paenibacillus alvei*, *Enterococcus faecalis*, *Brevibacillus laterosporus*, *Bacillus pumilus* and *Achromobacter eurydice* (Forsgren, 2010). Using 16s metabarcoding not only were secondary invaders identified, but also changes in the gut microbiome when *M. plutonius* was present were observed (Anderson et al., 2023). The microbiome of an organism is important for the host

as it contributes towards its survival, including protection from disease. Understanding microbiome changes in larvae with *M. plutonius* present, can help to understand the infection process, and this could be a useful tool in prevention and control. As well as metabarcoding, whole genome sequencing has opened up a new avenue to study *M. plutonius* and a clearer understanding of EFB.

Whole genome sequencing involves fragments to be sequenced and then reassembled computationally. The first *M. plutonius* strain had its genome sequenced in 2011 (Okumura et al., 2011) using 454 sequencing (Roche). This is still the most complete genome and is referred to as the reference genome on the NCBI website, which includes a database of genomes (NCBI, 2021). The reference genome is 2.1 Mb long and is made up of the main chromosomal DNA and two plasmids (pMP1 and pMP19) (Okumura et al., 2011). More recently a strain with atypical growth was sequenced as well as another two strains using PacBio sequencing (Okumura et al., 2018, 2019), currently 27 strains of EFB have whole genome sequences reported on the NCBI database (NCBI, 2021). This whole genome information has allowed researchers to explore the genetics of *M. plutonius* further than ever before. Virulence factors, expressed by virulence genes, are a key component for pathogenic bacteria. Virulence factors allow the bacteria to enter the host, survive and cause disease (Johnson, 2018). Using whole genome sequencing data of *M. plutonius* potential virulence genes have been identified, such as the Melissotoxin A (MtxA) gene found on the pMP19 plasmid (Djukic et al., 2018; Grossar et al., 2020). Finding virulence genes can contribute to further understanding of the disease aetiology of EFB and could help with the development of treatments. Antimicrobial resistance is a huge concern internationally, as one treatment for EFB is OTC there is a potential for *M. plutonius* to possess antibiotic resistance genes (ARG) (Ruckert et al., 2024; Waite et al., 2007). As honey bee products are consumed by humans, identifying these genes is important to reduce the risk of these genes being transferred to bacteria that is pathogenic to humans (Prestinaci et al., 2015). Whole genome sequence data has been used to identify ARGs, packages such as the NCBI AMR finder, use a database of known ARGs to search within the whole genome sequence of the input species and identify potential ARGs (oniciuc et al) (Feldgarden et al., 2019). As well as identifying genes whole genome data can be used to compare individuals within the same species.

For *Paenibacillus larvae*, the causative agent of AFB, a core genome MLST (cgMLST) scheme has been developed, using whole genome sequences and 2419 core genes were

selected to use for sequence typing, this allowed for higher resolution than traditional MLST typing (Bertolotti et al., 2021). Another more intensive method uses whole genome MLST (wgMLST) for *P. larvae* sequence typing that used 5745 loci in the MLST scheme, this was compared to cgMLST and traditional MLST. The comparison showed that traditional MLST was unable to distinguish clusters in an outbreak unlike wgMLST and cgMLST, also wgMLST was able to identify genetic diversity within outbreak clusters (Papić et al., 2021). Single nucleotide polymorphism (SNP) profiles, are another method used in comparative genomics. A SNP is a single change of a nucleotide to another, usually compared to a reference genome. SNP profiles generated for a specific species can be used to assess phylogeny, as mutations occur through generations of bacteria (Faison et al., 2014). By whole genome sequencing EFB infected material in outbreak situations, and applying cgMLST, wgMLST or SNP profiles, could increase genetic understanding, not only of different outbreaks but genetic differences within outbreaks can be identified, highlighting the potential these molecular techniques could have on understanding and therefore management of EFB in outbreak situations. Having whole genome sequencing data of *M. plutonius* has led to the design of quantitative PCR assays.

Quantitative PCR (qPCR), originally called kinetic PCR, was developed by Higuchi et al and uses a fluorescent dye, originally ethidium bromide. The dye binds to the newly synthesised double stranded DNA and fluoresces under UV light. The fluorescence is measured and used to quantify the DNA after each cycle (Russel et al., 1993). More recent methods use strand specific probes rather than generic dye to increase the specificity, TaqMan® PCR is an example this. The probe is designed to bind to the centre of the target sequence and has a long-wave emission fluorophore attached at one end and a short-wave emission fluorophore at the other. In proximity the long wave fluorophore quenches the short-wave emission, fluorescence is prevented, this phenomenon is called fluorescence energy transfer (FRET). The DNA strand is elongated from the primer binding region by the polymerase, during PCR, and the probe is degraded once the enzyme reaches the probe strand releasing the fluorophore. The fluorophores are no longer in close proximity, interrupting the FRET, so the short-wave fluorophore fluoresces and can be detected as a measure of DNA concentration (Kalinina et al., 1997; Didenko, 2001). qPCR is now a widely used technique in the field of diagnostics, forensics and many more, the most recent example being the widespread application of qPCR to test symptomatic patients for Covid-19 (Wang et al., 2020). This technique has commonly been used to

monitor and study EFB.

Previously EFB was detected using microscopy and more recently LFDs, however these techniques are excellent for confirming presence but are not quantitative so cannot be used to monitor levels of infection, whereas qPCR can (Hornitzky & Smith, 1998; Tomkies et al., 2009). A qPCR assay to detect the presence of *M. plutonius* in larvae was developed in 1998. The assay has a target within the 16s rRNA gene, and the primers are specific for *M. plutonius*, allowing discrimination from other bacteria that will likely be present in the larvae (Govan et al., 1998). Another assay designed to the 16s rRNA region was used to monitor the *M. plutonius* levels to assess treatment effectiveness by comparing shook swarm to OTC treatment (Budge et al., 2010). A hemi nested PCR targeting the 16s region was designed to test for *M. plutonius* in honeybees and their products. Hemi nested PCR involves 2 consecutive PCR reactions, the template of the second being the amplicon from the first to improve specificity (Mckee et al., 2003). Taking a different approach an assay was designed to target a section within the SodA gene, to monitor the levels of *M. plutonius* after sanitisation. SodA is a highly conserved region much like the 16s region and in some species of bacteria SodA has shown to be able to discriminate closely related species more effectively than 16s (Poyart et al., 2001; Roetschi et al., 2008). More recently a triplex PCR assay to detect EFB, AFB and the internal control bee DNA in one reaction was designed, making a one pot detection method for foulbrood. The EFB assay within this has a target within the NapA pseudo gene region (Dainat et al., 2018). The use of qPCR has added another molecular tool to the research of EFB, allowing detection in different mediums such as honey and the quantification element allows treatments to be tested for efficiency.

Testing for pathogens outside the host using molecular methods is an important tool when trying to study transmission of a disease. But this information also needs to be paired with other information such as human behaviour or movement to gain a more complete epidemiological picture. Understanding human movement paired with genetic information played an important role in highlighting staff members as one of the transmission routes of *Clostridioidies difficile* in a Swiss hospital. By sequencing swabs from infected patients and the shoes of staff members, it was found that the sequence type of *C. difficile* the patient was infected with matched the sequence type found on the staff member working with them in 74% of cases (Büchler et al., 2022). Beliefs can also influence disease epidemiology, during the COVID-19 pandemic many people believed in

conspiracy theories. Those who believed in those theories were less likely to be compliant with government recommendations and were less likely to take safety precautions such as being vaccinated (Earnshaw et al., 2020). These actions and beliefs can heavily influence how a disease is spread. Perception of EFB by beekeepers and how this is influenced is important to know in order to control and manage the disease. A popular epidemiological tool for studying these interactions is statistical modelling.

Statistical modelling is a common tool used in epidemiological studies. One modelling technique is structural equation modelling (SEM). SEM involves quantifying relationships between variables stated in a hypothesised model called a conceptual model. The variables can be made up of observed variables, that are directly measured or latent variables. Latent variables are not directly measured, but through multiple observed variables are interpreted. For example, if there was a latent variable called good biosecuirty practices, this could be explained by whether the beekeeper is disinfecting hive tools, sharing equipment and wearing disposable gloves, all three of these are observed variables (Thakkar, 2020). In Malaysia, villagers that had experienced Dengue fever outbreaks carried out surveys and interviews about their opinions and actions with regard to health promotional campaigns. From these data a SEM was constructed, and it was found that to control future outbreaks areas that strengthened self-efficacy, people's belief in their ability to achieve specific goals, needed to be concentrated on (Isa et al., 2013). This methodology using SEM and information from the public through surveys, showing it to be an important tool to highlight key gaps in knowledge and perceptions. This method could be used to gain a deeper understanding of the perception of EFB by beekeepers.

Molecular biology is an ever-advancing discipline, this has been utilised in EFB research, by using molecular techniques as important tools in monitoring and managing EFB. Sequencing could open up possibilities within management of EFB research, as shown already with the MLST scheme, this has allowed sequence types to be identified but whole genome sequencing of infected material could give more genetic insight into outbreaks and how they started. However, molecular biology alone is not enough to study the epidemiology of EFB. Complex interactions are involved in the epidemiology of a disease. Human beliefs, movement, management practices and behaviours play a major role in the epidemiology of EFB. Studying both the molecular biology of *M. plutonius* and investigating human behaviour, beliefs and disease, will provide a more complete understanding of the epidemiology of EFB.

1.2 Objectives

With the advancement of sequencing techniques and the use of epidemiological techniques in studying disease, the primary aim of my thesis was to use whole genome sequencing and statistical modelling to deepen our understanding of the epidemiology of EFB and how to manage it.

The individual aims of this thesis were:

1. To develop an affordable whole genome sequencing pipeline to generate whole genome sequencing data from larval samples infected with EFB from England and Wales, as opposed to cultured bacteria.

2. To develop a bioinformatic pipeline to study the phylogeny and geography from the sequencing data of the larval samples infected with EFB from England and Wales.

3. To delve deeper into the sequencing data of the larval samples infected with EFB from England and Wales, and investigate what virulence genes, antimicrobial resistance genes and associated bacteria were present.

4. To carry out a survey investigating beekeepers behaviours and their opinions of EFB, and use the data to perform structural equation modelling to better understand the contribution of beekeeping behaviours and beliefs have on the epidemiology of EFB.

1.3 Thesis outline

In chapter 2 an affordable whole genome sequencing pipeline was developed to produce whole genome sequencing data from EFB infected larval samples. The development process involved developing a host depletion method to reduce the amount of honey bee DNA in the sample to increase the sequencing capacity of the target of interest *M. plutonius*. A homemade column extraction method was developed to reduce costs, and the sequencing technology used was Oxford Nanopore Technologies. Chapter 3 then took this developed pipeline and used it to sequence infected larval samples from England and Wales. With this sequencing information a bioinformatic pipeline was developed and the output used to study the phylogeny of the samples in relation to geography. Chapter 4 used the same sequencing data generated in chapter 3 to investigate what virulence and

antimicrobial resistance genes were present in the genomes of *M. plutonius*. Assessing the potential of this sequencing to identify associated bacteria was also performed. Chapter 5 went beyond the genetics and investigated behaviours and beliefs of beekeepers in highly infected areas, and the impact this has on epidemiology. This was achieved by creating a survey for the beekeepers about EFB and their behaviours and beliefs, and using structural equation modelling to highlight areas that influence perception of EFB.

Chapter 2

Developing a sequencing pipeline to investigate European foulbrood outbreak sites across the UK

2.1 Introduction

Insect pollination plays a major role in UK agriculture, with approximately 20% of crop production relying upon it (Breeze et al., 2011). The European honey bee (*Apis mellifera*) is classed as a primary insect pollinator and for UK agriculture its worth is approximately £430 million (Smith et al., 2011). Apart from pollination, honey bees also produce economically important goods such as beeswax and honey (Carreck & Williams, 1998). However, pollinator populations are declining (Potts et al., 2010). For example, the winter colony mortality rate of honey bees in 2018/2019 was predicted to be 9% in England and 10.3% in Wales. Such declines will have detrimental effects not only for national economies, but also food security, particularly in a growing human population (Carvalho, 2006; National Statistics, 2022; Potts et al., 2010).

One of the major causes of population decline in honey bees across the globe is European Foulbrood (EFB) (Gray et al., 2020; Smith et al., 2013). EFB is caused by *Melissococcus plutonius* (*M. plutonius*), a gram-positive bacteria that grows in the mid gut of infected honey bee larvae, and competes for food (Bailey, 1956; White, 1912). About three to five days after infection, most larvae die, and the disease can spread when the nurse bees clear out the dead larvae (Bailey, 1983). EFB is a notifiable disease in the UK, meaning that all

suspected cases should be reported to the local bee inspector (Government, 1982). Action taken for highly infectious colonies is destruction, so the disease can have a catastrophic impact on honey bee populations (Thompson & Brown, 2001). To combat this disease, it is first of all necessary to be able to properly diagnose it. To this end, serological and molecular tools are key to providing in field diagnosis and an improved understanding of landscape level epidemiology of this disease.

One of the key serological tools, lateral flow devices (LFDs) are currently used to confirm the presence of EFB in colonies on site (Tomkies et al., 2009). A single suspect larva is added to a buffer bottle containing sodium azide and ball bearings and disrupted by shaking. Two drops of the sample are added to the device, and through capillarity action, the sample runs over EFB specific antibodies to produce, in infected cases, a blue positive line, in a similar manor to pregnancy test kits. The buffer bottles of samples that test positive for EFB are sent to the National Bee Unit (NBU) for multi locus sequence typing (MLST). MLST is a molecular method that involves sequencing sections of specific genes to identify a strain type for each sample (Maiden et al., 1998). The EFB MLST involves four genes, that currently identifies 46 different strain types (Haynes et al., 2013; Jolley et al., 2018). Assigning a strain type helps to track the epidemiology of the disease and also monitors the development of newly evolved sequence types (Budge et al., 2014). Although MLST is useful, reports from the UK suggest that three sequence types (ST3, ST5 and ST23) dominate, accounting for 72% of all isolates characterised (Budge et al., 2014), limiting our understanding of local disease spread. The emergence of more affordable whole genome sequencing methods could provide substantially more genetic information than the current four gene MLST scheme, thus improving our resolution of EFB spread in the UK.

Whole genome sequencing has been used in epidemiology of bacterial pathogens particularly in healthcare setting. It can be very useful in tracing outbreak sources and transmission routes. For example, an outbreak of methicillin-resistant *Staphylococcus aureus* (MRSA) in a neonatal intensive care unit was studied using whole genome sequencing and the movements of staff and patients. The sequencing results showed that there were two MRSA outbreaks present in the unit, and they both originated from an infected patient, and spread through a staff member who had contact with all parties involved (Madera et al., 2023). Also, a possible transmission route of *Clostridioides difficile* in a Swiss hospital was tested by swabbing soles of shoes. Using whole genome

sequencing of the swabs and sequencing data from the patients infected, 74% of positive shoe swabs matched the strain with the patient that staff member was working with (Büchler et al., 2022). Both of these examples show how useful whole genome sequencing can be to trace outbreaks sources and transmission routes and could be useful for EFB outbreaks. If this method were to be employed on all EFB infected samples each year, like the MLST testing, it would require a cost-effective method; with flexibility in scale to account for seasonal outbreaks; and the ability to work within existing field sampling system. A successful whole genome sequencing pipeline requires multiple linked steps across different laboratory and data platforms, beginning with sample preparation.

Current methods used to prepare samples for whole genome sequencing on *M. plutonius* have used bacterial cultures, which is time consuming, costly and difficult to do at scale (Djukic et al., 2018; Okumura et al., 2019). Sequencing from a single larva would prevent the culturing step but comes with challenges. Sequencing DNA extracts directly from infected larval samples provide sequence data from all organisms present, including the honey bee host, which will dilute the bacterial reads. Host depletion methods need to be in place before extraction to decrease the levels of honey bee DNA. Commercial kits are available for this and have been compared, they found the HostZERO™ and QiAamp were successful in depleting host DNA. (Heravi et al., 2020). Some cheaper home made methods, originally used to detect viability of bacteria, have been tested using an initial osmotic lysis and a treatment including Propidium Monoazide (PMA), DNase and Benzonase® (Akerley et al., 2002; Cangelosi & Meschke, 2014; Marotz et al., 2018; Nocker et al., 2006; Bruggeling et al., 2021; Amar et al., 2021). The osmotic lysis step relies on the host DNA having a weaker cell structure than the bacteria, so in the lysis solution the host cells will burst and release its DNA, whilst the bacterial cell wall stay intact. For both the DNase and the Benzonase® treatment the free DNA is digested by the enzyme, for the PMA the dye binds to the free DNA and prevents any downstream processes (Figure 2.1) (Marotz et al., 2018; Nocker et al., 2006).

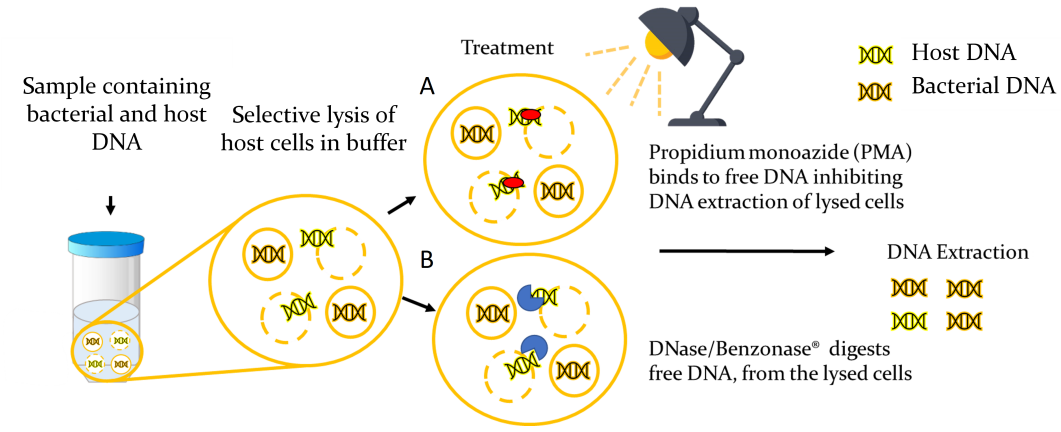


Figure 2.1: The methodology of host depletion methods previously used. The initial sample contains both bacterial and host DNA. The selective lysis of the host cells causes the cell wall of the host cells to become weaker and release the DNA, an example of a selective lysis treatment is water. Through osmotic lysis the host cells will burst and release the DNA, but the bacterial cells will remain intact A) For the PMA treatment when the sample is exposed to light the dye binds to the free DNA released during the lysis step, preventing the bound DNA from downstream processes. B) The DNase and Benzonase® treatment digest the free DNA and does not penetrate the bacterial cell wall.

The next step of the sequencing pipeline to consider is the DNA extraction method. The basic steps of DNA extraction involve cell disruption, protein denaturation, enzyme inactivation, contaminant removal and finally DNA precipitation and elution (Tan & Yiap, 2009). There are many different methods to perform DNA extraction such as using silica membranes and magnetic beads, with commercial kits available to do this (Chacon-Cortes & Griffiths, 2014). The silica membrane technique was initially used by Boom et al (Boom et al., 1990), using the idea that the positive charge of the silica attracts the negatively charged DNA backbone. Once bound to the silica the DNA is washed with buffers and then using a low salt buffer or water the DNA is released from the silica (Boom et al., 1990). The magnetic bead technique was first used by Hawkins et al using carboxyl coated magnetic particles. The method uses chemical groups on the magnetic beads to attract the DNA, when applied to a magnet the beads are pelleted and the unwanted material

is disposed. The beads are then washed and much like the silica method the DNA is removed from the beads with a low salt buffer (Hawkins et al., 1994). As well as magnetic beads coated with carboxyl groups other coatings are used such as silica (Oberacker et al., 2019).

The final consideration is which sequencing method and platform to use. The sequencing methods most recently used for *M. plutonius* was PacBio RS II (Okumura et al., 2018). The PacBio method requires a large number of samples to be tested to make it cost-effective, Cuber et al predicted that to make it £6 per sample 356 samples would be required to be sequenced for PacBio RS II. The cheapest platform was Oxford Nanopore Technologies, the figure to make £6 a sample was much lower with the MinION being 183 and the Flongle only 61 (Cuber et al., 2023). Oxford Nanopore Technologies use a chip that contains nanopores. When the single stranded DNA is loaded it passes through the nanopores, as it passes through the change in the conductivity of electricity is recorded. The conductivity of each base pair is different, allowing this information to be translated into a sequence (Laver et al., 2015). Instruments made by Oxford Nanopore Technologies include the MinION and Flongle adaptor, and both provide long read sequencing. Weighing approximately 90 g the MinION is a handheld device that can fit into any space, comes with a low instrument cost and can output up to 50 Gb of sequence (Lu et al., 2016). The Flongle is an adaptor for the MinION and provides cheaper flow cells with lower output, which is ideal for lower sample numbers, and uses the same sample preparation as MinION (Grädel et al., 2019). Having the same sample preparation method means the technology can be used interchangeably for different sample numbers. The cost of both the MinIon and Flongle instruments is significantly lower than other platforms (Cuber et al., 2023).

The overall aim of this chapter was to develop a cost-effective method, that can be done at a large scale, to produce whole genome sequencing data from the in-field samples present in the buffer bottles after LFD testing. The first step of achieving this was to test the host depletion methods currently used in other host materials to reduce the honey bee DNA. The second step was to develop a cheaper home-made DNA extraction method to those commercially available. To assess both the host depletion methods and extraction methods qPCR was used as this was cheaper and sequencing every test was unfeasible. The final step was to test the chosen methods on samples from the field on the cheaper of the Oxford Nanopore Technologies, the Flongle to ensure the methods were compatible

through the pipeline.

2.2 Methods

2.2.1 Introduction

The development of a cost-effective method required development at different stages of a sequencing pipeline (Figure 2.2). The sequencing process starts with the sample input. The input in this process was infected larval samples stored in the buffer bottles from the LFD kits. The next stage was host depletion, methods were tested in 2.2.2. The next step was to find an appropriate DNA extraction method in 2.2.3. The two developed methods were tested together and put through the sequencing library preparation, sequencing and sequence analysis stages in 2.2.4.

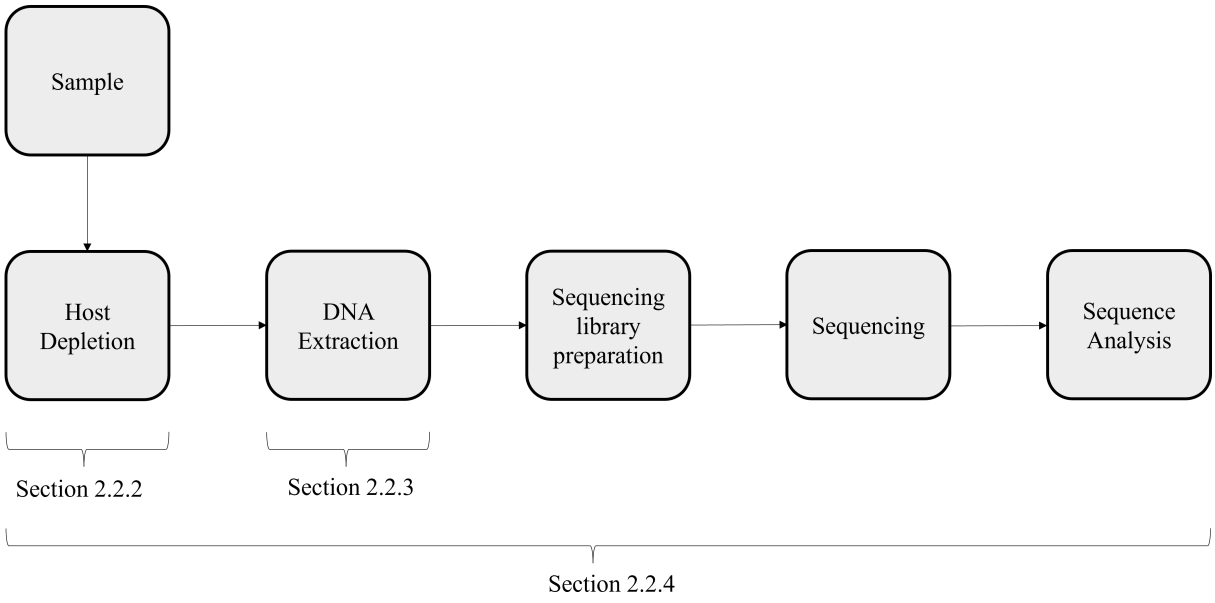


Figure 2.2: The stages involved in whole genome sequencing. The method section for the development of each of these stages is highlighted.

2.2.2 Host depletion

The first stage of developing a pipeline for sequencing *M. plutonius* from infected larva was to develop a successful host depletion method (Figure 2.3). This method needed to be both cost-effective and efficient at depleting the *A. mellifera* DNA.

Experiment	Lysis method	Host depletion method
Experiment 1: Section 2.2.2.3	No lysis	PMA DNase Nuclease for cell lysis No treatment
Experiment 2: Section 2.2.2.4	Water Blood buffer ATL buffer No lysis	No treatment PMA DNase
Experiment 3: Section 2.2.2.5	Water Blood buffer No lysis	PMA DNase
Experiment 4: Section 2.2.2.6	No lysis	PMA NEB® enrichment kit
Experiment 5: Section 2.2.2.7	No lysis	1% PMA 10% PMA 100 % PMA No treatment

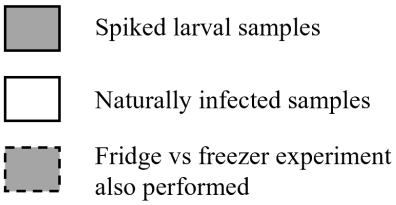


Figure 2.3: The experiments outlined for the host depletion development section of the pipeline. All combinations of lysis methods and host depletions method were performed.

2.2.2.1 Cost comparison

A cost comparison was performed on the different methods used for host depletion both home-made methods and commercially available kits. For the commercial kits, HostZERO™ microbial kit, QIAmp DNA microbiome kit and the NEBNext® microbiome DNA enrichment kit, the cost of the kits was sourced from the company's website. For the home-made methods, PMA, DNase and Thermo Scientific™ Pierce™ Universal Nuclease for Cell Lysis, the cost of the reagents was sourced from the internet. The cost per sample was calculated by dividing the cost of the kit or reagents by the estimated number of samples the kit was for, this cost did not include any additional consumables that could be required.

2.2.2.2 Bacteria growth and sample preparation

M110 agar plates and media were prepared following Frosgren et al's method (Forsgren et al., 2013). I obtined a plate of *M. plutonius*, ID 7521, on M110 agar from Fera Science Ltd. A colony was streaked onto a new M110 agar plate and stored in an anaerobic container with an anaerobic indicator (Oxoid BR055B) and Oxoid AnaeroGen anaerobic generator sachet (AN0020D). The bacteria were left at 35°C in anaerobic conditions for a week. A colony was re-suspended in 3 ml M110 liquid media in a 5 ml falcon tube sealed with parafilm and incubated at 35°C for a week. A plate count was performed using Forsgren et al's method (Forsgren et al., 2013). The 3 ml culture had a count of 80,000 CFU/ml. The culture was centrifuged at 5,000 g for 10 mins and resuspended in 300 μl of buffer E to generate 800,000 CFU/ml. Buffer E is a buffer containing sodium azide that comes in 5 ml bottles containing ball bearings from field test kits for European Foulbrood (Tomkies et al., 2009). Individual honey bee larvae, obtained from honey bee colonies at Newcastle University, were placed into six individual buffer E bottles from the field kits. The bottles were shaken for 2 minutes. Three of the bottles were spiked with 100 μl of the 800,000 CFU/ml *M. plutonius*, to mimic an infield sample from an infected apiary; bees were shown to have symptoms over 50,000 CFU per bee (Roetschi et al., 2008). All the buffer bottles were left at room temperature for a week to mimic the postage the samples would experience in the field. To investigate if storage affects the sample quality, one spiked bottle and one un-spiked buffer bottle were placed in the freezer for three days, another spiked and un-spiked bottles were stored in the fridge for three days and the final two bottles left at room temperature for three days. One bottle of buffer without larvae was spiked with 800,000 CFU/ml *M. plutonius* as a bacteria control.

2.2.2.3 Initial treatment comparison with spiked samples

The three chosen methods of host depletion were, Propidium Monoazide (PMA) treatment, DNase treatment and Thermo Scientific™ Pierce™ Universal Nuclease for Cell Lysis, a Benzonase® equivalent. For each treatment two aliquots were taken from the prepared buffer bottles at each storage condition and both spiked and un-spiked, as well as the bacteria control. All samples were centrifuged at 7,500 g for 10 minutes and the supernatant removed. For the PMA treatment, the pellets were re-suspended in 200 μl distilled water and 2 μl 1:10 Propidium Monoazide (PMA) (41105331, Biotium).

The samples were incubated for 5 minutes in a dark cupboard and then exposed to 60 lumens of LED light for 15 minutes. The PMA treated samples were centrifuged at 13,000 g for 10 minutes, supernatant removed, and pellet re-suspended in 200 μl Phosphate Buffered Saline (PBS). For DNase treatment pellets were re-suspended in 20 μl DNase buffer I, 1 μl DNase I (10649890, Fisher Scientific) and 180 μl distilled water, and incubated at 37°C for 60 minutes. To inactivate the DNase 20 μl 50 mM EDTA was added, and the samples incubated at 65°C for 10 minutes. For the Thermo Scientific™ Pierce™ Universal Nuclease for Cell Lysis treatment the pellets were re-suspended in 20 μl 10x buffer, 1 μl Thermo Scientific™ Pierce™ Universal Nuclease for Cell Lysis enzyme (1:50) and 180 μl distilled water and incubated at 37°C for 120 minutes. The treated samples along with the untreated samples were centrifuged at 10,000 g for 10 minutes, supernatant removed, and pellet re-suspended in 180 μl of TE buffer and 20 μl 50 mM EDTA. After treatment all the samples were centrifuged at 10,000 g for 5 minutes and the pellets were re-suspended in 180 μl Gram-positive lysis buffer(10 mM Tris-HCl, 1 mM EDTA, 20 mg/ml lysozyme and 1.2 % Triton-X 100). All the samples were extracted following the Qiagen DNeasy blood and tissue kit following the manufacturer's protocol for Gram-positive bacteria, eluting twice in 50 μl elution buffer (Qiagen, 2006). This method used a chaotropic lysis buffer to disrupt proteins and a silica membrane to capture nucleic acid. The level of *M. plutonius* in the samples was assessed using EFBFor (TGTTGTTAGAGAAGAATAGGGGAA), EFBRev2 (CGTGGCTTTCTGGTTAGA) and a dual labelled probe EFBProbe (FAM-AGAGTAACTGTTTTCCTCGTGACGGT-TAMRA) designed to a 16S rRNA fragment in *M. plutonius*. The level of *A. mellifera* in the samples was assessed using AJ307465-955F (TGTTTTCCCTGGCCGAAAG), 1016R (CCCCAATCCCTAGCACGAA) and a dual labelled probe 975T (FAM-CCCGGGTAACCCGCTGAACCTC-TAMRA) designed to the 18s rRNA region of *A. mellifera* (Budge et al., 2020). For each qPCR reaction both *M. plutonius* and *A. mellifera* assay, 1 μl of extracted DNA was added to 1x Universal Mastermix (Applied Biosystems™ :10733457), 375 nM forward primer, 375 nM reverse primer, 125 nM dual labelled probe, made up with molecular grade water. The reactions were run on an Applied Biosystems QuantStudio™ 5 PCR machine at 60°C for 2 minutes, 95°C for 10 minutes followed by 40 cycles of: 60°C for 1 minute and 95°C for 15 seconds. Relative concentrations were calculated using a dilution series and the ratio of *M. plutonius* to *A. mellifera* DNA. All analysis was performed in R (Version: R 4.2.1). To assess the host depletion using the ratios calculated for each condition,

a generalised linear model (GLM) analysis was performed using the laavan package (Version:0.6-17) (Rosseel, 2012) and plots were generated using ggplot2 (Version: 3.4.4) (Wickham, 2016) to visualise the data. The explanatory variables for the GLM were the host depletion and the storage method, as interaction terms, and the observed variable was the ratios of *M. plutonius* to *A. mellifera*. A GLM and a plot was also generated with the observed variable as *M. plutonius* DNA concentration at each condition to ensure there was sufficient bacteria being extracted.

2.2.2.4 Lysis treatment comparison with spiked samples

Different lysis treatments with the PMA and DNase treatment were combined on the same spiked samples based on the results from 2.2.2.3. The lysis treatments tested were ATL (939011, Qiagen) with proteinase K (19131, Qiagen) usually used to lyse samples in the Qiagen DNeasy kits, blood buffer (Promega) used in the Wizard® HMW DNA Extraction Kit to lyse blood cells and a final lysis condition of distilled water. From the buffer bottles prepared in 2.2.2.2, sixteen 200 μl aliquots were taken of the spiked larvae stored in the freezer. All samples were centrifuged at 20,000 g for 5 minutes, and the supernatant discarded, the pellets were taken forward to lysis. Four pellets were re-suspended in 180 µl ATL buffer and 20 μl proteinase K and incubated for 1 hour at 56°C. Four pellets were re-suspended in 200 μl water and left to incubate at room temperature for 1 hr. Four pellets were re-suspended in 200 μl blood buffer and left to incubate at room temperature for 10 minutes. The four remaining pellets were left unlysed. The samples were centrifuged at 10,000 g for 5 minutes and the supernatant discarded. For each host depletion method one pellet was taken from each lysis treatment including the unlysed pellets. The PMA treatment and DNase treatment were performed as described in 2.2.2.3. The samples were centrifuged at 10,000 g for 5 minutes and the supernatant removed. The remaining samples, two pellets from each lysis including the unlysed, were left untreated. All samples were re-suspended in 180 μl Gram-positive lysis buffer ready for extraction. The DNA extraction, qPCR and analysis were performed as described in 2.2.2.3. For the GLM the observed variable was the treatment and lysis combined as one variable and the explanatory variables were the same as in 2.2.2.3.

2.2.2.5 Testing host depletion methods on naturally infected samples

Naturally infected larvae from field samples were tested using the most promising lysis methods from 2.2.2.4, water and blood buffer, combined with the most successful treatment methods from 2.2.2.3, PMA and DNase. Samples of larvae in the LFD buffer bottles from infected sites in 2020 were received from Fera Science Ltd via DEFRA. Five of the samples were selected randomly to test the lysis and treatment methods. For each sample nine aliquots of 200 μl was transferred into a 2 ml tube and centrifuged at 20,000 g for 5 minutes, the supernatant was discarded. For each sample three pellets were lysed in water and three pellets were lysed in the blood lysis, as performed in 2.2.2.4. From each lysis and each sample, one pellet for PMA treatment was re-suspended in 200 μl water, one pellet for DNase treatment was re-suspended in 20 μl DNase buffer I, 1 μl DNase and 180 μl distilled water and the final two for the untreated controls were re-suspended in 180 μl Gram-positive lysis buffer ready for extraction. The PMA and DNase treatment were performed as described in 2.2.2.3. The DNA extraction, qPCR and analysis were performed as described in 2.2.2.3. For the GLM the observed variable was the treatment and lysis combined as one variable and the explanatory variables were the same as in 2.2.2.3.

2.2.2.6 Comparing PMA method to a commerically available kit

Based on the results from 2.2.2.5 PMA with no lysis was chosen as a potential method to compare to the NEBNext® microbial enrichment kit. For each method, triplicate 200 μl aliquots of each of the three 2020 buffer bottle samples were sampled. Each sample was centrifuged at 20,000 g for 5 minutes, and the supernatant discarded. A pellet from each sample was re-suspended in water and PMA treatment was performed as described in 2.2.2.3. Post treatment, the samples were spun at 10,000 g for 5 minutes. The pellets from each sample and the PMA treated pellets were re-suspended in 180 μl Gram-positive lysis buffer. A DNA extraction was performed as described in 2.2.2.3. From the extracted samples a 30 μl aliquot of each untreated sample, three from each sample, was put through the NEBNext® Microbiome DNA Enrichment kit protocol, following the manufacturer's protocol, but scaling down the volumes by 8 x to adjust for a smaller than recommended volume. AMPure XP beads (A63881, Beckman Coulter) were used to purify the samples. Beads were added at 1:1 volume ratio and mixed by pipetting. The samples were left

to incubate for 1 minute, and placed on a magnet, after 2 minutes the supernatant was removed. Whilst keeping the tube on the magnet, 80 % ethanol was added and then immediately removed, this was repeated once. The tube was removed from the magnet and the beads were re-suspended with 60 μl of DNase free water. After 2 minutes the tube was put on the magnet and the supernatant containing the sample transferred to a clean tube. A qPCR was performed on the enriched extracts and the PMA experiment extracts and analysed as described in 2.2.2.3. For the GLM the observed variable was the treatment combined as one variable and the explanatory variables were the same as in 2.2.2.3.

2.2.2.7 PMA scaled down

Based on the results from 2.2.2.5 and the costs of PMA treatment (Table 2.1), whether the amount of PMA could be reduced with no reduction in efficacy was explored. A single pool was created my mixing 1 ml of buffer from each of ten different 2020 buffer bottles. Ten aliquots of 200 μl was transferred into 2 ml tubes. The samples were centrifuged at 20,000 g for 5 minutes, and the supernatant discarded. For the untreated control one pellet was re-suspended in 180 μl Gram-positive lysis buffer ready for extraction. The remaining nine samples were re-suspended in 200 μl water. PMA treatment was performed as described in 2.2.2.3 but with varying amounts of PMA added in triplicate (Table 2.1), representing 1%, 10% and 100% of the recommended 1:10 PMA volume. The DNA extraction, qPCR and analysis were performed as described in 2.2.2.3. For the GLM the observed variable was the treatment and lysis combined as one variable and the explanatory variables were the same as in 2.2.2.3.

Table 2.1: Cost and volume of Propidium Monoazide (PMA) added to the samples at different percentages of the recommended 1:10 volume.

Percentage (%)	PMA Volume (µl)	Cost per sample (£)
100	1.00	0.1900
10	0.10	0.0190
1	0.01	0.0019

2.2.3 DNA extraction development

The second stage of developing a pipeline for sequencing *M. plutonius* from infected larva was to test various DNA extraction methods (Figure 2.4). The extraction method needed to be affordable and effective at extracting *M. plutonius*.

Experiment	Extraction method	Elution buffer
Experiment 6: Section 2.2.3.2	Qiagen DNeasy	AE Qiagen buffer
	Caterpillar method	Caterpillar buffer
	Home-made column method	Caterpillar buffer
Experiment 7: Section 2.2.3.3	Home-made bead method	TE buffer
	Qiagen DNeasy	AE Qiagen buffer
	Home-made column method	Caterpillar buffer
Experiment 8: Section 2.2.3.4	Home-made column method	AE Qiagen buffer
	Qiagen	TE buffer
		Caterpillar buffer
		Water

Figure 2.4: The experiments outlined for the DNA extraction method development section of the pipeline. For experiment 6 and 7 only the elution buffer in line with the extraction method was used, for experiment 8 all elution buffers were tested with the home-made column extraction, and only AE buffer was used to elute DNA in the Qiagen DNeasy method.

2.2.3.1 Cost comparison

A cost comparison was performed on the different methods used for DNA extraction both home-made methods and commercially available kits. For the commercial kits, Qiagen DNeasy blood and tissue kit, Promega® food kit and Promega® HMW kit, the cost of the kits was found from the company's website. The cost per sample was calculated dividing the cost of the kit by the estimated number of samples the kit was for, this cost did not include any additional consumables that could be required. For the home-made methods, home-made bead extraction, home-made column extraction and the caterpillar method the cost of the individual reagents was found through an internet search, and the cost of additional consumables was included in the per sample estimate.

2.2.3.2 Column comparison

Three DNA extraction methods were compared. The first was using the Qiagen DNeasy blood and tissue kit. The second was an extraction method taken from section 12 of a

published method, Tissue extraction from whole caterpillars V.1 method (Kitson, 2017), used with modified volumes and called the caterpillar method throughout this study. The third was a modified version of the published caterpillar method called the home-made column method. From three different 2020 buffer bottles 1.2 ml of each sample was transferred into individual 2 ml tubes. Each tube was centrifuged at 20,000 g for 5 minutes. The supernatant was discarded, and the pellets re-suspended in 1080 μl Gram-positive lysis buffer and incubated at for 30 minutes at 37°C. Each sample was aliquoted into six tubes of 180 μl. From each sample, two 180 μl aliquots were taken forward for Qiagen DNeasy blood and tissue extraction. The extraction was performed following the method described in 2.2.2.3. Another two 180 μl aliquots from each sample were taken forward for extraction using the caterpillar method. For the caterpillar method, 97 μl lysis solution 2 (120 mM sodium chloride, 50 mM Tris HCl, 20 mM EDTA, 3% SDS, distilled water to make up to volume) and 3 μl proteinase K (10 mg/μl) were added and the aliquots were incubated at 37°C for 1 hour. To each aliquot 280 μl Protein Denaturation buffer (5 M Guanidine HCl and made up to volume with distilled water) and 280 μl ethanol was added and vortexed. The aliquot was transferred to a column (nbs biologicals, SD5008 - EZ-10), and centrifuged at 6,000 g for 1 minute. The flow through was discarded and 500 μl wash buffer 1 (7 M Guanidine HCl, 56% Ethanol and made up to volume with water) was added. The column was centrifuged at 6,000 g for 1 minute. The flow through was discarded and 500 μl wash buffer 2 (10 mM Tris HCl, 70% ethanol and made up to volume with water) was added. The column was centrifuged at 20,000 g for 3 minutes. The flow through was discarded, the column transferred to a new 1.5 ml microtube and 50 μl elution buffer (10 mM Tris HCl, made up to volume with water). The column was left to stand for 1 minute and centrifuged at 10,000 g for 1 minute, the addition of elution buffer and subsequent steps was repeated once. Another two 180 μl aliquots from each sample were taken forward for extraction using the home-made method. For the home-made method, 200 µl PDB and 25 µl proteinase K were added to each aliquot and incubated for 30 minutes at 56°C. To each sample 200 μl 100% ethanol was added, and the solution was mixed using a pipette. The sample was transferred to a column (nbs biologicals, SD5008 - EZ-10) and centrifuged at 6,000 g for 1 minute. The flow through was discarded and 500 μl wash buffer 2 was added. The column was centrifuged at 20,000 g for 3 minutes. The flow through was discarded, the column transferred to a new 1.5 ml microtube and 50 μl elution buffer. The column was left to stand for 1 minute and centrifuged at 10,000 g for 1 minute, the addition of elution buffer and subsequent steps

was repeated once. This method used the solutions from the published Tissue extraction from whole caterpillars V.1 method (Kitson, 2017) but following a similar protocol to the DNeasy Qiagen kit. A qPCR was performed on the extracts and analysis was performed as described in 2.2.2.3, host depletion was still assessed for all DNA extraction development experiments, to assess if a method was biased towards the host DNA or naturally depleting it. For the GLM the observed variable was the extraction method and the explanatory variables were the same as in 2.2.2.3.

2.2.3.3 Home-made bead test

The silica membrane-based methods, Qiagen DNeasy method and the home-made column method from 2.2.3.2 were compared to a home-made bead based method. This method used carboxyl coated magnetic particles to bind to the DNA. A single pool was created by mixing 500 μl of buffer from each of ten different 2020 buffer bottles. For each extraction method 200 μl sample was extracted in triplicate. The two column-based extractions were performed on the prepared samples as described in 2.2.3.2. For the bead extraction, a modified BOMB Genomic DNA extraction protocol was used (Oberacker et al., 2019). The L6 lysis buffer (120 g guanidinium thiocyanate, 100 ml 0.1 M Tris hydrochloride at pH 6.4, 22 ml 0.2 M EDTA, using NaOH to adjust to pH 8.0 and 2.6 g Triton X-100) from Boom et al was used instead of the lysis buffer in the BOMB protocol (Boom et al., 1990). The elution buffer was replaced with TE buffer. Another modification to the protocol was step 1 was replaced with using 200 μl infected larvae instead of overnight culture as stated in the protocol. A qPCR was performed on the extracts and analysis was performed as described in 2.2.2.3. For the GLM the observed variable was the extraction method and the explanatory variables were the same as in 2.2.2.3.

2.2.3.4 Optimising DNA elution

Based on results from the 2.2.3.3, I wanted to optimise the home-made column method by testing different elution buffers, water, TE buffer, the buffer used in the caterpillar protocol (caterpillar buffer) and Qiagen AE buffer. A single pool was created my mixing 500 μl of buffer from each of ten different 2020 buffer bottles. The home-made column extraction method was performed on three 200 μl aliquots of the prepared pool as described in 2.2.3.2. This was repeated for another nine samples, but with the elution buffer replaced

in triplicate by water, TE and Qiagen AE buffer. To a final three 200 μl aliquots a Qiagen column extraction was performed as described in 2.2.2.3. A qPCR was performed on the extracts and analysis as described in 2.2.2.3. For the GLM the observed variable was the elution method and the explanatory variables were the same as in 2.2.2.3.

2.2.4 Combining host depletion methods with the selected DNA extraction method

The final stage of developing a pipeline for sequencing *M. plutonius* from infected larva was to combine the host depletion methods with the developed extraction method (Figure 2.5). This was to ensure both methods work together for maximum efficiency for both host depletion and effectiveness.

Experiment	Sample input volume	Treatments
Experiment 9: Section 2.2.4.1	200 µl 1 ml 2 ml	Water lysis only Water lysis with DNase 2 µl PMA 10 µl PMA 20 µl PMA No treatment
Experiment 10: Section 2.2.4.2	2 ml	Water lysis only Water lysis with DNase No treatment
Experiment 11: Section 2.2.4.3	2 ml	Water lysis with DNase No treatment

2 samples taken forward for Nanopore sequencing

Figure 2.5: The experiments outlined for the host depletion methods combined with the developed DNA extraction method (Home-made column extraction with a TE elution and a 2 ml sample input volume). For experiment 9 all treatments were performed for each sample input volume.

2.2.4.1 Combining the selected DNA extraction method with selected host depletion methods at different volumes

Based on the results from 2.2.3.4 the home-made column extraction with a TE elution buffer was taken forward as the standard extraction method. Using the standard extraction method, the best lysis, host depletion combinations were tested, chosen from

2.2.2.5, PMA with no lysis, DNase with a water lysis, as well as just water lysis with no treatment, at three different sample input volumes, 200 μl, 1 ml and 2 ml (Table 2.2). The volumes were tested to see if a larger input produced more DNA. The PMA treatment was tested at different volumes to account for the change in input volume. A single pool was created my mixing 1 ml of buffer from each of twenty different 2020 buffer bottles. Six aliquots of each input volume were taken. Each treatment was performed on an aliquot from each input volume. The PMA treatment was performed as described in 2.2.2.3. The two adapted PMA treatments were performed as described in 2.2.2.3 but with 10 μl 1:10 PMA and 20 μl 1:10 PMA. The DNase treatment with a water lysis treatment was performed as described in 2.2.2.4. The water lysis only was performed as in 2.2.2.4. An untreated sample from each volume was centrifuged at 20,000 g for 5 minutes and the supernatant discarded. All the pellets were re-suspended in 180 μl Gram-positive lysis buffer ready for extraction. The samples were extracted using the home-made column extraction with the elution buffer as TE buffer, as described in 2.2.3.4. A qPCR was performed on the extracts and an analysis, both as described in 2.2.2.3. For the GLM the observed variables were the host depletion methods and volume and the explanatory variables were the same as in 2.2.2.3.

Table 2.2: The lysis treatments, host depletion treatments and volume conditions tested. Each lysis treatment was tested with each host depletion method, and each combination was tested at all three sample input volumes.

Lysis	Treatment	Volume		
None	Untreated	200 µl	1 ml	2 ml
Water lysis	Untreated			
Water lysis	DNase			
None	PMA 1 µl			
None	PMA 10 µl			
None	PMA 20 µl			

2.2.4.2 Test of DNase Treatment with and without water lysis

All the method development was done using affordable qPCR methods for analysis, but the endpoint for the final samples is high throughput sequencing using Oxford Nanopore Technologies sequencing platforms. Based on the results from 2.2.4.1 the DNase treatment with a water lysis with a 2 ml sample input volume was chosen to test on the Oxford Nanopore Flongle. The DNase was also tested without the lysis step to see the effect of the water lysis on the ratio of *M. plutonius* reads to *A. mellifera* reads. Triplicate 2 ml samples were taken from the 2020 buffer bottles (E2683, E2666 and E2672) and each centrifuged at 20,000 g for 5 minutes. The supernatant was discarded, and the pellets retained for treatment by water lysis with DNase as described in 2.2.2.4 and no lysis with DNase as described in 2.2.2.3. The final replicate was the unlysed, untreated control. DNA extraction was performed using the home-made column extraction with TE elution method as described in 2.2.3.4. The nine extracts were prepared for sequencing on the flongle using the SQK-LSK109 (Oxford Nanopore) kit with the EXP-PBC096 (Oxford Nanopore) barcoding kit following the Ligation sequencing gDNA-PCR barcoding (Version:PBGE96_9068_v109_revT_14Aug2019). For the PCR step conditions the number of cycles were 18 and the extension time was 3 minutes, with a final extension of 3 minutes. The short fragment buffer was used for the final clean up. The protocol was not available for flongle so for the priming and loading section the protocol used was Ligation sequencing gDNA (Version:GDE_9063_v109_revAP_25May2022). The Flongle was run for 24 hours with live basecalling and live barcoding. All packages used for bioinformatics were through the Bioconda package manager (Grüning et al., 2018). Using Minimap2 (Version:2.24) the raw reads for each sample were mapped to the *M. plutonius* reference genome (Genbank:AP018492.1) and the *A. mellifera* genome (Genbank:GCA_003254395.2) obtained from NCBI (Li, 2018; Okumura et al., 2011; Wallberg et al., 2019). Using Samtools (Version:1.16.1) the alignment file generated from Minimap2 was converted and sorted into a BAM file (Danecek et al., 2021). The number of mapped reads to both the *M. plutonius* and *A. mellifera* were calculated using Samtools. The percentage of the total reads for both *M. plutonius* and *A. mellifera* were calculated for each sample.

2.2.4.3 Testing final method using qPCR and flongle sequencing

Based on the results in 2.2.4.2, the final method was tested, DNase treatment with a water lysis, extracted using the home-made column extraction with DNA eluted in TE buffer and an input volume of 2 ml, on more samples. Two 2 ml samples were taken from the 2020 buffer bottles (E2319, E2326, E2327, E2540 and E2559) and each centrifuged at 20,000 g for 5 minutes. The supernatant was discarded, and the pellets retained for treatment by water lysis with DNase with a DNA extraction using the home-made column extraction and a TE elution as described in 2.2.4.2 and an untreated replicate. The untreated replicate was extracted using the home-made column extraction and a TE elution as described in 2.2.3.2. A qPCR was performed on the extracts and analysis was performed as described in 2.2.2.3. The two samples with the highest concentrations, E2319 and E2327 were taken forward for Oxford Nanopore Flongle sequencing, performed and analysed as described in 2.2.4.2.

2.3 Results

2.3.1 Host depletion method development

The costs for commercial kits were far higher than those from the home-made methods, (Table 2.3). The Thermo Scientific™ Pierce™ Universal Nuclease for Cell Lysis was ~50p cheaper than Benzonase® nuclease. The cheapest method for host depletion was the DNase, at only 9p per sample.

Table 2.3: Cost per sample of both commercial bacterial enrichment kits and home-made methods. The cost of each method was calculated, the cost does not include additional consumables.

Kit	Type	Cost per sample (£)
HostZERO™ microbial DNA kit	Commercial	13.46
QIAamp DNA microbiome kit	Commercial	13.36
NEBNext® microbiome DNA enrichment kit	Commercial	34.04
DNase Treatment	Homemade	0.09
PMA Treatment	Homemade	0.38
Thermo Scientific™ Pierce™ Universal Nuclease	Homemade	3.76
Benzonase® Nuclease	Homemade	4.36

2.3.1.1 Initial host depletion treatment comparison with spiked samples

Host depletion was visualised as a change in the ratio of *M. plutonius* to *A. mellifera* DNA for each host depletion and storage method (Figure 2.6). A generalised linear model (GLM) was performed to assess the relationship between the ratio of untreated samples and the treated samples (Table 2.4). GLM showed that there was no significant difference between the ratio when comparing untreated samples to the three host depletion treatments, and no significant difference in ratio as a result of storage conditions, performed on the same GLM. For the GLM it was shown using a normality test that the residuals did not deviate from a normal distribution (Figure A.1).

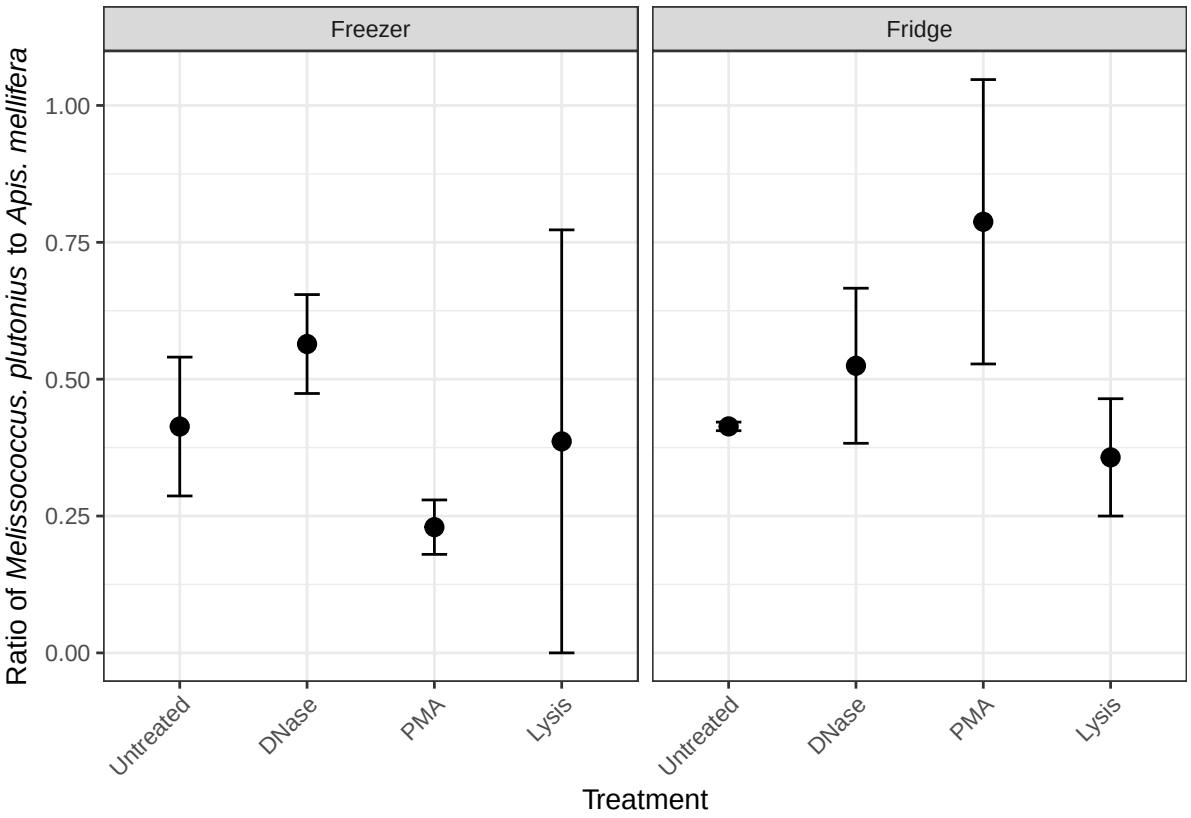


Figure 2.6: Effect of host depletion treatments (DNase, PMA and Thermo Scientific™ Pierce™ Universal Nuclease for Cell Lysis (Lysis)), and storage method (fridge or freezer) on the ratio of the concentration of pathogen target to host DNA in larvae containing a spiked culture of *Melissococcus plutonius*

Table 2.4: Results from a generalised linear model assessing the relationship of the ratio of the concentration of pathogen target to host DNA in larvae containing a spiked culture of bacteria, with samples treated with different host depletion treatments (DNase, PMA and Thermo Scientific™ Pierce™ Universal Nuclease for Cell Lysis (Lysis)) and storage method (fridge or freezer). The family used for the model was Gaussian. (*) represents a significant difference.

Variable	t-value	p-value	sig
(Intercept)	2.230	0.056	
DNase treatment	0.575	0.581	
PMA treatment	-0.701	0.503	
Lysis treatment	-0.103	0.920	
Fridge storage	0.001	0.999	
DNase treatment with fridge storage	-0.108	0.917	
PMA treatment with fridge storage	1.503	0.171	
Cell lysis treatment with fridge storage	-0.079	0.939	

The concentration of *M. plutonius* decreased in all the treated samples compared with the untreated samples (Figure 2.7). The samples treated with the lysis treatment had a *M. plutonius* concentration of nearly 0 ng/μl. A generalised linear model (GLM) of the *M. plutonius* concentration showed the concentration was significantly lower in all the DNase, PMA and lysis treatments when compared to the untreated samples, with no significant difference between storage conditions (Table 2.5). For the GLM it was shown using a normality test that the residuals did not deviate from a normal distribution (Figure A.2).

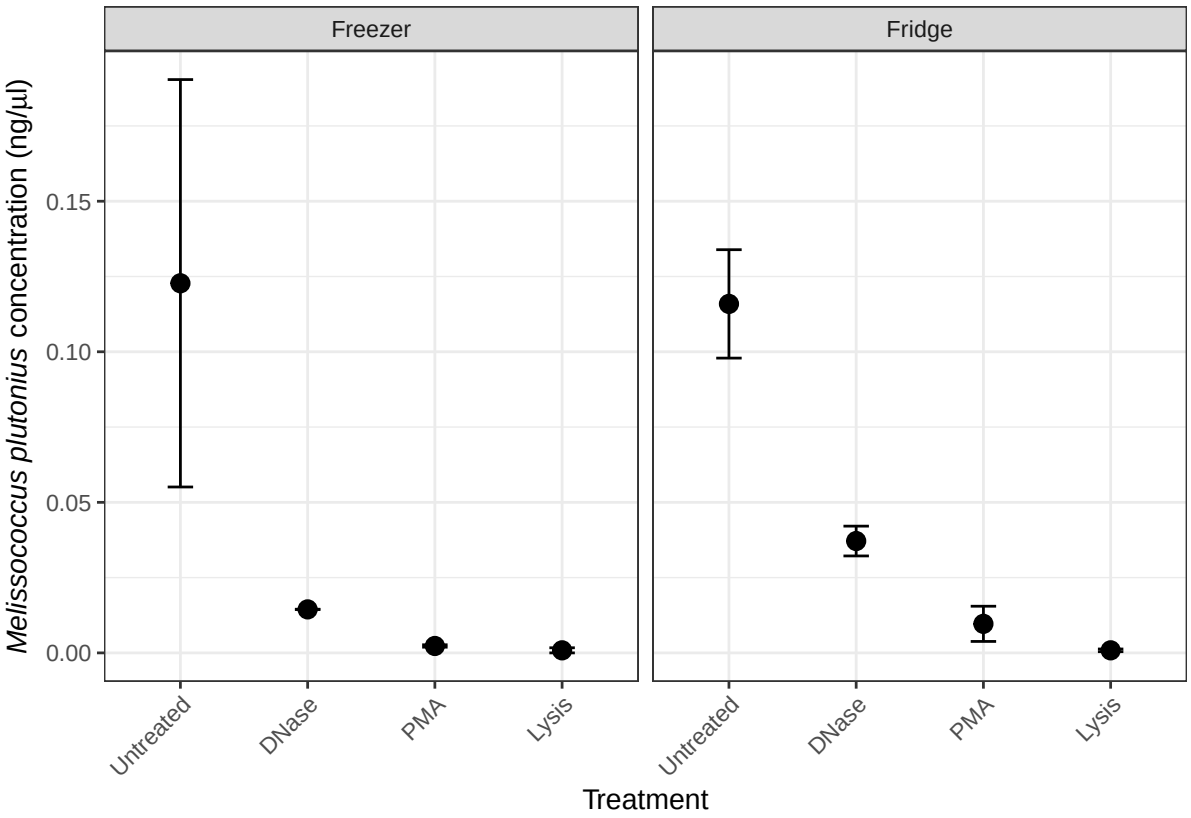


Figure 2.7: Effect of host depletion treatments (DNase, PMA and Thermo Scientific™ Pierce™ Universal Nuclease for Cell Lysis (Lysis)), and storage method (fridge or freezer) on the concentration of *Melissococcus plutonius* in larvae containing a spiked culture of *Melissococcus plutonius*

Table 2.5: Results from a generalised linear model assessing the relationship of *Melissococcus plutonius* concentration in larvae containing a spiked culture of *Melissococcus plutonius* with samples treated with different host depletion treatments (DNase, PMA and Thermo Scientific™ Pierce™ Universal Nuclease for Cell Lysis (lysis)) and storage method (fridge or freezer). The family used for the model was Gaussian. (*) represents a significant difference.

Variable	t-value	p-value	sig
(Intercept)	4.930	0.001	*
DNase Treatment	-3.075	0.015	*
PMA Treatment	-3.420	0.009	*
Lysis Treatment	-3.462	0.009	*
Fridge Storage	-0.195	0.851	
DNase Treatment with Fridge Storage	0.593	0.569	
PMA Treatment with Fridge Storage	0.285	0.783	
Cell Lysis Treatment with Fridge Storage	0.138	0.894	

2.3.1.2 Comparison of lysis treatments combined with host depletion methods with spiked samples

Host depletion was visualised as a change in the ratio of *M. plutonius* to *A. mellifera* DNA for each lysis treatment combined with host depletion treatment (Figure 2.8). A generalised linear model (GLM) was performed to assess the relationship between the ratio of unlysed, untreated samples and the lysed and treated samples (Table 2.6). GLM showed that there was a significant increase in the ratio of the concentration of pathogen to host when the samples were lysed with blood lysis and host depleted with PMA treatment than with no lysis and no host depletion treatment. The GLM also showed that there was a significant increase in the ratio of the concentration of pathogen to host when the samples were lysed with water and host depleted with PMA treatment than with no lysis and no host depletion treatment. There was no significant difference between all the other lysis, host depletion treatment ratios compared to the unlysed, untreated sample. For the

GLM it was shown using a normality test that the residuals did not deviate from a normal distribution (Figure A.3).

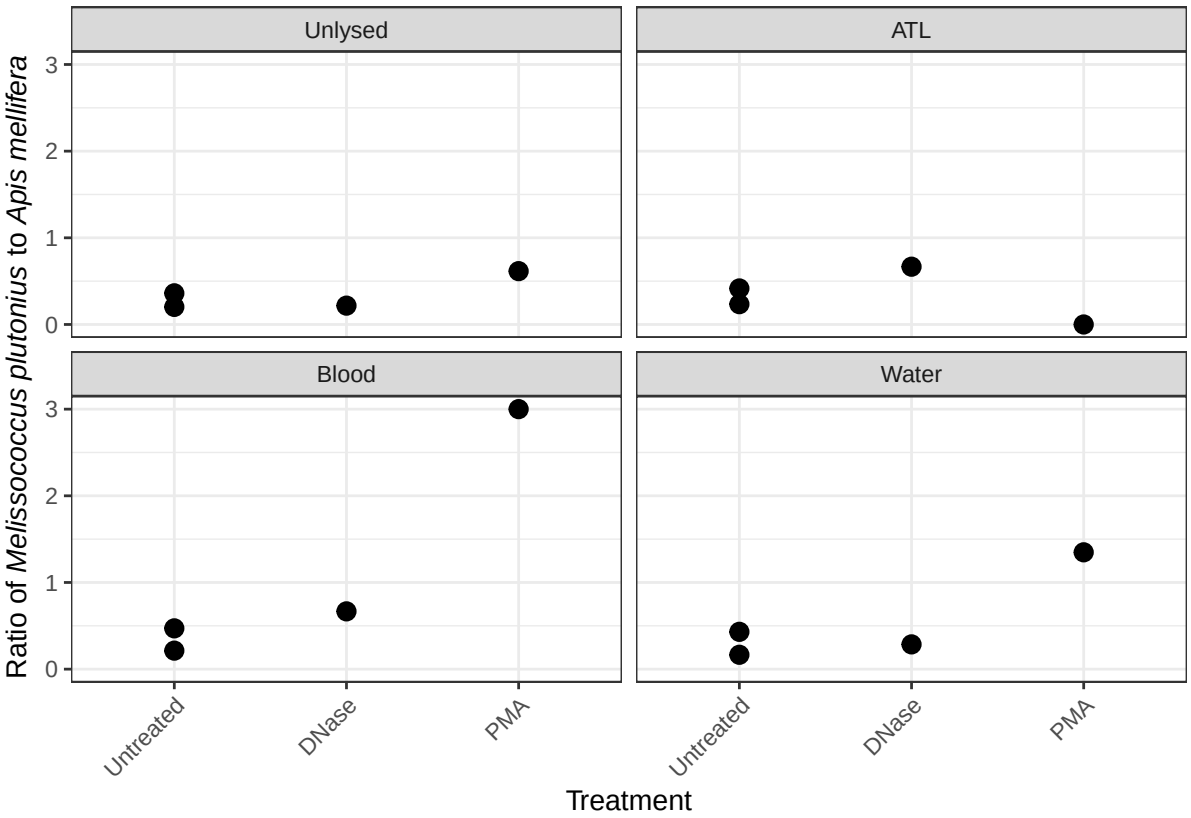


Figure 2.8: Effect of lysis treatments (blood buffer, ATL buffer and water) combined with host depletion treatments (DNase and PMA) on the ratio of the concentration of pathogen target to host DNA in larvae containing a spiked culture of *Melissococcus plutonius*

Table 2.6: Results from a generalised linear model assessing the relationship of the ratio of the concentration of pathogen target to host DNA in larvae containing a spiked culture of bacteria, with samples treated with a combination of different lysis treatments (blood buffer, ATL buffer and water) and different host depletion treatments (DNase and PMA). The family used for the model was Gamma. (*) represents a significant difference.

Variable	t-value	p-value	sig
(Intercept)	11.953	0.000	*
DNase treatment with ATL lysis	-1.872	0.135	
DNase treatment with blood lysis	-1.872	0.135	
DNase treatment with no lysis	0.349	0.744	
DNase treatment with water lysis	-0.024	0.982	
PMA treatment with ATL lysis	1.624	0.180	
PMA treatment with blood lysis	-7.400	0.002	*
PMA treatment with no lysis	-1.646	0.175	
PMA treatment with water lysis	-4.299	0.013	*
Untreated with ATL lysis	-0.284	0.790	
Untreated with blood lysis	-0.392	0.715	
Untreated with water lysis	-0.111	0.917	

The concentration of *M. plutonius* appeared to decrease in all the treated samples compared with the untreated samples (Figure 2.9). A generalised linear model (GLM) of the *M. plutonius* concentration showed the concentration was significantly lower in all lysis and host depletion treatment combinations when compared to the untreated samples, except with the untreated, water lysis combination (Table 2.7). For the GLM it was shown using a normality test that the residuals did not deviate from a normal distribution (Figure A.4).

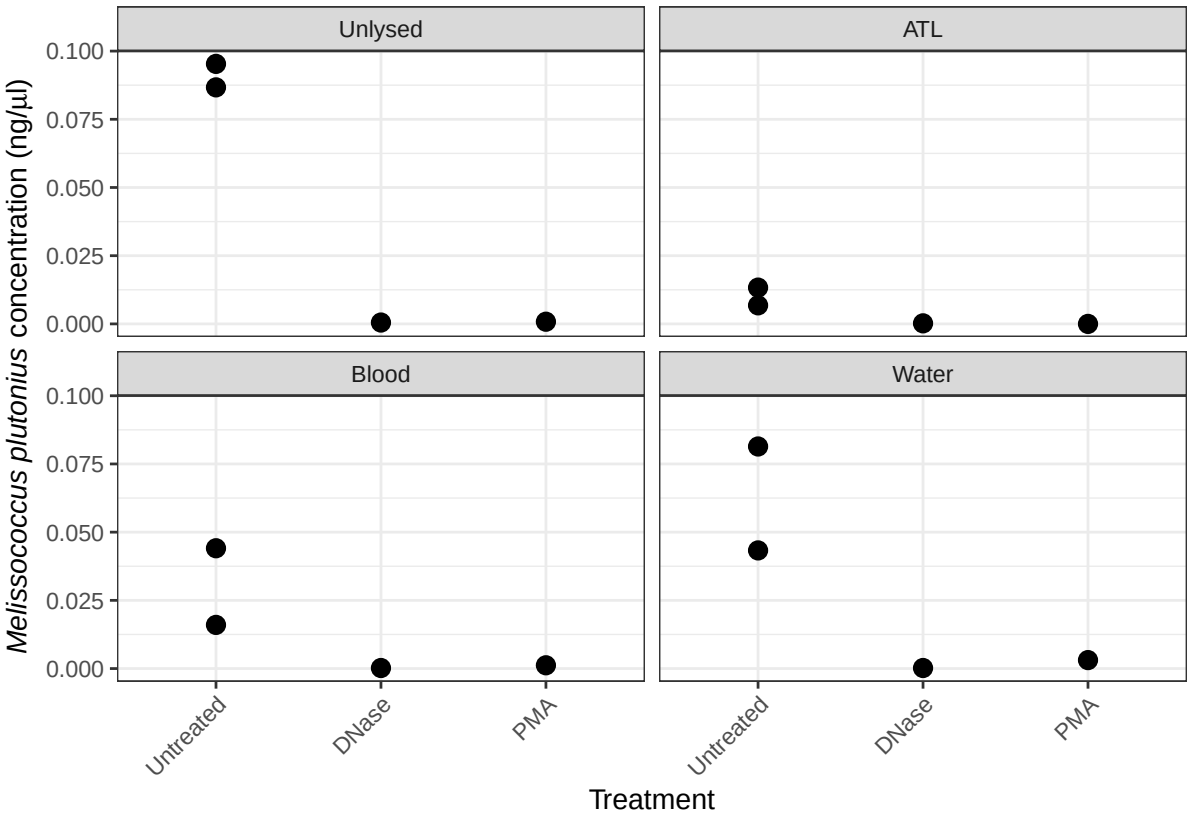


Figure 2.9: Effect of lysis treatments (blood buffer, ATL buffer and water) combined with host depletion treatments (DNase and PMA) on the concentration of the *Melissococcus plutonius* DNA in larvae spiked with cultured bacteria

Table 2.7: Results from a generalised linear model assessing the relationship of the ratio on the concentration of *Melissococcus plutonius,* in larvae containing a spiked culture of bacteria, with samples treated with a combination of different lysis treatments (blood buffer, ATL buffer and water) and different host depletion treatments (DNase and PMA). The family used for the model was Gamma. (*) represents a significant difference.

Variable	t-value	p-value	sig
(Intercept)	86.589	0.000	*
DNase treatment with ATL lysis	4.276	0.013	*
DNase treatment with blood lysis	4.276	0.013	*
DNase treatment with no lysis	4.261	0.013	*
DNase treatment with water lysis	4.276	0.013	*
PMA treatment with ATL lysis	4.286	0.013	*
PMA treatment with blood lysis	4.228	0.013	*
PMA treatment with no lysis	4.247	0.013	*
PMA treatment with water lysis	4.136	0.014	*
Untreated with ATL lysis	4.715	0.009	*
Untreated with blood lysis	3.517	0.025	*
Untreated with water lysis	1.629	0.179	

2.3.1.3 Testing lysis treatments combined with host depletion methods on naturally infected samples

Host depletion was visualised as a change in the ratio of *M. plutonius* to *A. mellifera* DNA for each lysis treatment combined with host depletion treatment on naturally infected samples (Figure 2.10). A generalised linear model (GLM) was performed to assess the relationship between the ratio of unlysed, untreated samples and the lysed and treated samples (Table 2.8). GLM showed that there was a significant increase in the ratio of the concentration of pathogen to host with the DNase treatment with blood lysis and

DNase treatment with water lysis, compared with the untreated, unlysed samples, but no significant difference with DNase treatment with no lysis. For PMA treatments the ratio of the concentration of pathogen to host was significantly higher with all lysis treatments including no lysis, compared with the untreated, unlysed samples. For no host depletion treatment with water lysis, and no host depletion with blood lysis, the GLM showed that the ratio of the concentration of pathogen to host was significantly higher compared with the untreated, unlysed samples. For the GLM it was shown using a normality test that the residuals did not deviate from a normal distribution (Figure A.5).

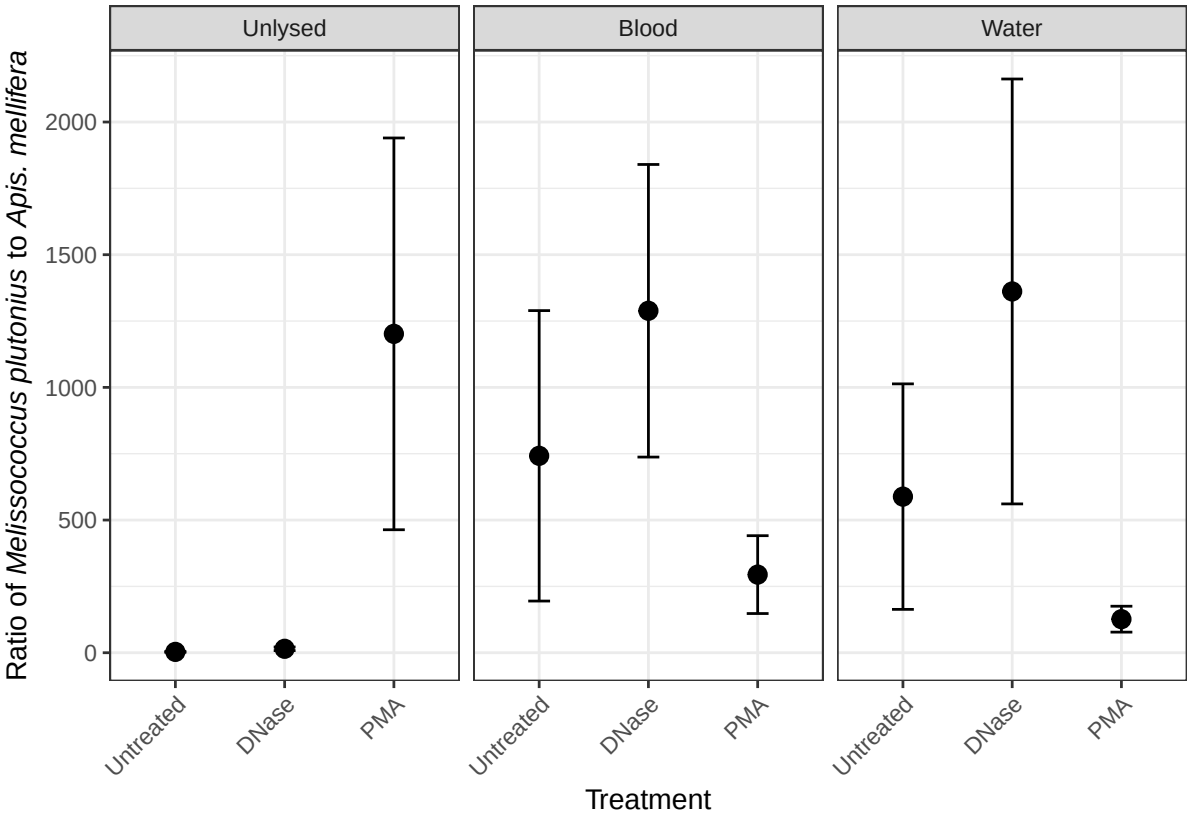


Figure 2.10: Effect of lysis treatments (blood buffer and water) combined with host depletion treatments (DNase and PMA) on the ratio of the concentration of pathogen target to host DNA in larvae naturally infected with *Melissococcus plutonius*

Table 2.8: Results from a generalised linear model assessing the relationship of the ratio of the concentration of pathogen target to host DNA in larvae naturally infected with *Melissococcus plutonius,* with samples treated with a combination of different lysis treatments (blood buffer and water) and different host depletion treatments (DNase and PMA). The family used for the model was Gaussian with a log link. (*) represents a significant difference.

Variable	t-value	p-value	sig
(Intercept)	1.351	0.186	
DNase treatment with blood lysis	3.994	0.000	*
DNase treatment with no lysis	0.813	0.422	
DNase treatment with water lysis	4.440	0.000	*
PMA treatment with blood lysis	2.759	0.009	*
PMA treatment with no lysis	3.971	0.000	*
PMA treatment with water lysis	2.750	0.009	*
Untreated with blood lysis	3.250	0.003	*
Untreated with water lysis	3.116	0.004	*

The concentration of *M. plutonius* appeared to increase in some treated samples and decrease in others (Figure 2.11). A generalised linear model (GLM) of the *M. plutonius* concentration showed there was no significant difference in the concentration between all lysis and treatment combinations when compared to the unlysed, untreated samples (Table 2.9). For the GLM it was shown using a normality test that the residuals did not deviate from a normal distribution (Figure A.6).

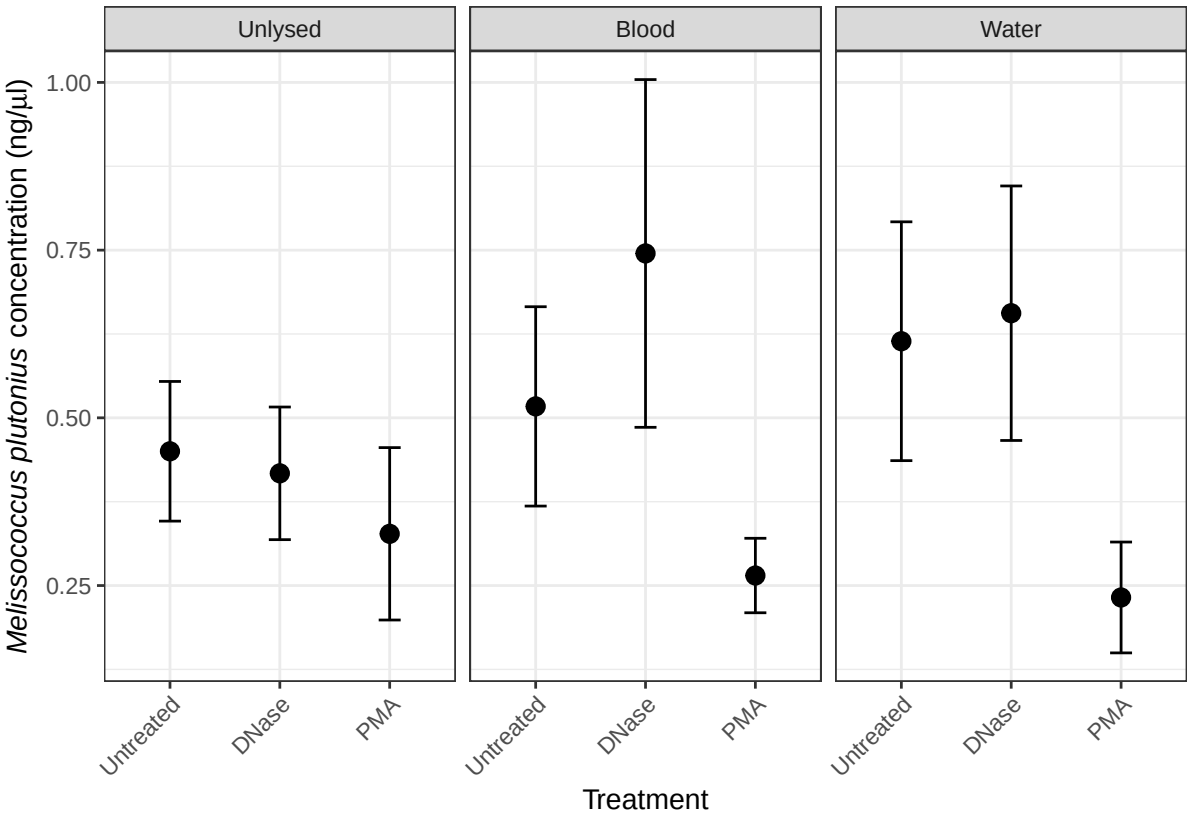


Figure 2.11: Effect of lysis treatments (blood buffer and water) combined with host depletion treatments (DNase and PMA) on the concentration of the *Melissococcus plutonius* DNA in larvae naturally infected with *Melissococcus plutonius*

Table 2.9: Results from a generalised linear model assessing the relationship of the ratio on the concentration of *Melissococcus plutonius* in larvae naturally infected with *Melissococcus plutonius,* with samples treated with a combination of different lysis treatments (blood buffer and water) and different host depletion treatments (DNase and PMA). The family used for the model was Gaussian. (*) represents a significant difference.

Variable	t-value	p-value	sig
(Intercept)	2.997	0.005	*
DNase treatment with blood lysis	1.389	0.174	
DNase treatment with no lysis	-0.155	0.878	
DNase treatment with water lysis	0.969	0.339	
PMA treatment with blood lysis	-0.872	0.389	
PMA treatment with no lysis	-0.546	0.588	
PMA treatment with water lysis	-1.026	0.312	
Untreated with blood lysis	0.315	0.755	
Untreated with water lysis	0.772	0.445	

2.3.1.4 Comparison of the PMA method and a commerically avaliable kit

Host depletion was visualised as a change in the ratio of *M. plutonius* to *A. mellifera* DNA for each host depletion treatment on naturally infected samples (Figure 2.12). A generalised linear model (GLM) was performed to assess the relationship between the ratio of untreated samples and the treated samples (Table 2.10). The GLM showed that the the ratio of the concentration of pathogen to host with the PMA treatment was significantly higher than the untreated samples. There was no significant difference in the ratio of the concentration of pathogen to host between the untreated samples and the commercially available NEBNext® enrichment kit. For the GLM it was shown using a normality test that the residuals did not deviate from a normal distribution (Figure A.7).

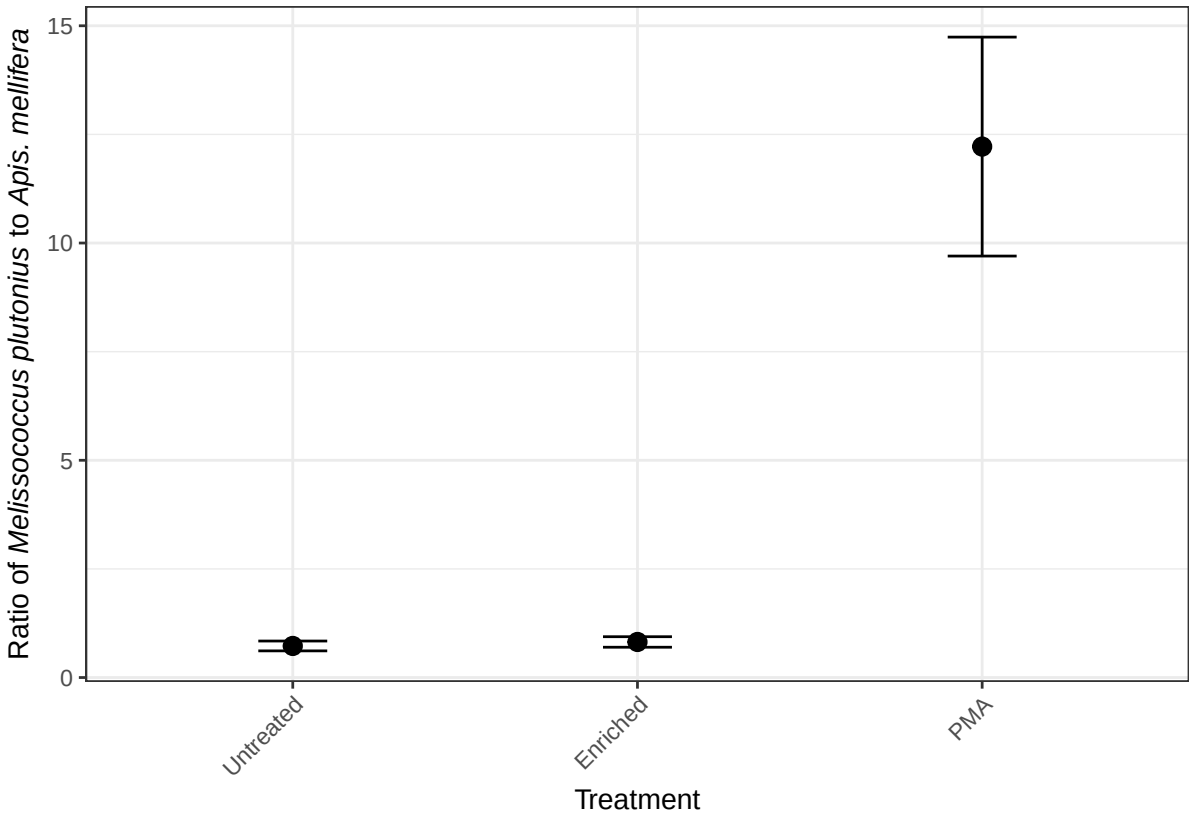


Figure 2.12: Effect of host depletion treatments (PMA and NEBNext® enrichment) on the ratio of the concentration of pathogen target to host DNA in larvae naturally infected with *Melissococcus plutonius*

Table 2.10: Results from a generalised linear model assessing the relationship of the ratio of the concentration of pathogen target to host DNA in larvae naturally infected with *Melissococcus plutonius,* with samples treated with different host depletion treatments (PMA and NEBNext® enrichment).The family used for the model was Gaussian with a log link. (*) represents a significant difference.

Variable	t-value	p-value	sig
(Intercept)	3.777	0.001	*
Enriched	0.256	0.800	
PMA	9.474	0.000	*

The concentration of *M. plutonius* appeared to decrease in the treated samples (Figure 2.13). A generalised linear model (GLM) of the *M. plutonius* concentration showed there was a significant decrease in the concentration between both the NEB enrichment treatment and the PMA treatment compared to the untreated samples (Table 2.11). For

the GLM it was shown using a normality test that the residuals did not deviate from a normal distribution (Figure A.8).

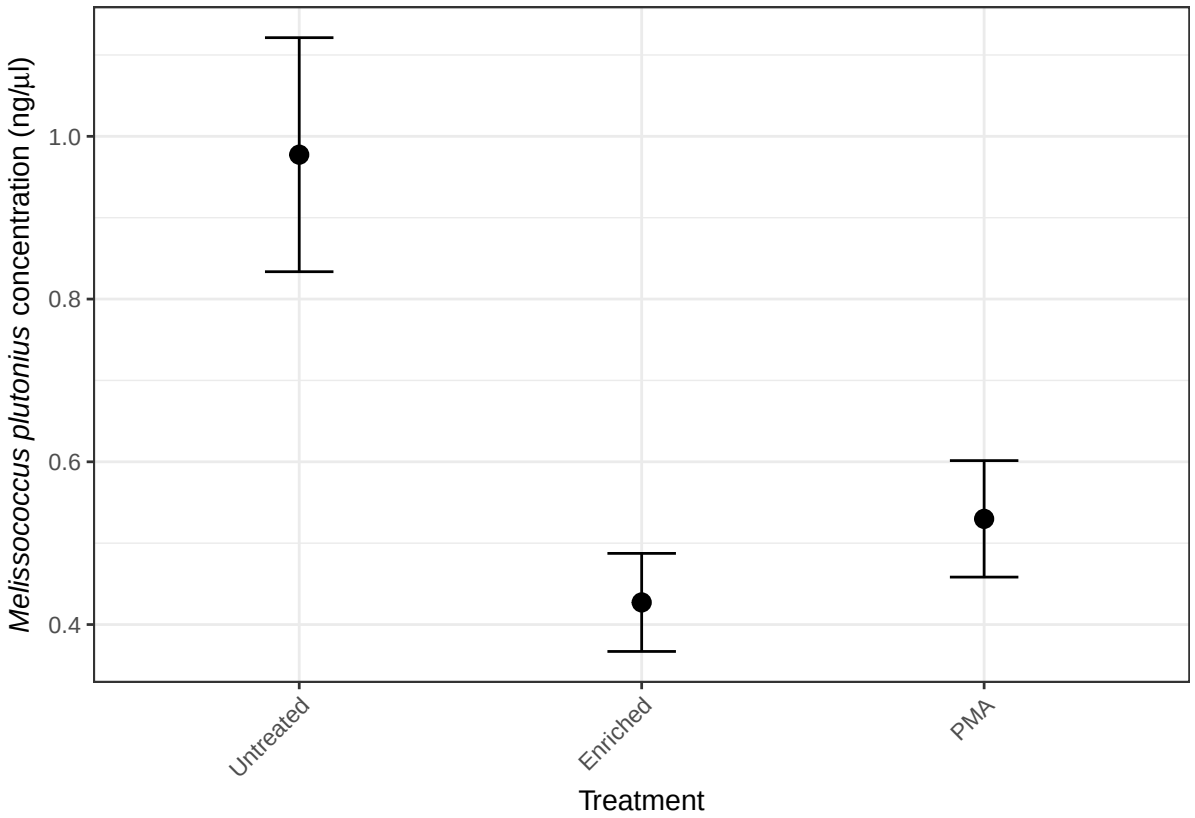


Figure 2.13: Effect of host depletion treatments (PMA and NEBNext® enrichment) on the the concentration of *Melissococcus plutonius* in larvae naturally infected with *Melissococcus plutonius*

Table 2.11: Results from a generalised linear model assessing the relationship of the concentration of *Melissococcus plutonius* DNA in larvae naturally infected with *Melissococcus plutonius,* with samples treated with host depletion treatments (PMA and NEB enrichment). The family used for the model was Gaussian. (*) represents a significant difference.

Variable	t-value	p-value	sig
(Intercept)	9.867	0.000	*
Enriched	-3.928	0.001	*
PMA	-3.195	0.004	*

2.3.1.5 Testing a reduced volume of PMA treatment

Host depletion was visualised as a change in the ratio of *M. plutonius* to *A. mellifera* DNA for each percentage of PMA treatment on naturally infected samples (Figure 2.14). A generalised linear model (GLM) was performed to assess the relationship between the ratio of untreated samples and the PMA treated samples (Table 2.12). The GLM showed that the the ratio of the concentration of pathogen to host with the 100% PMA treatment was significantly higher than the untreated samples. There was no significant difference in the ratio of the concentration of pathogen to host between the untreated samples and both the 1% PMA treatment and the 10% PMA treatment. For the GLM it was shown using a normality test that the residuals did not deviate from a normal distribution (Figure A.9).

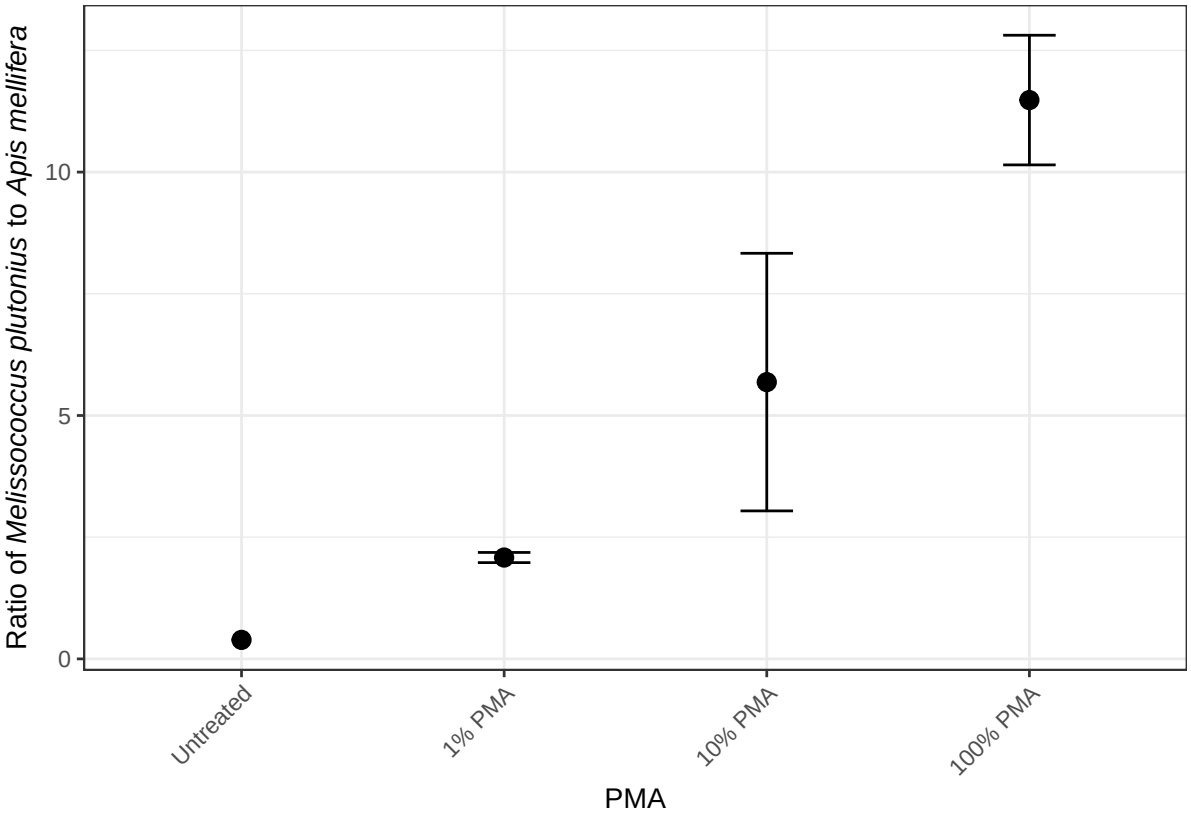


Figure 2.14: Effect of different percentages of PMA host depletion treatments compared to the recommended volume (1%, 10% and 100%) on the ratio of the concentration of pathogen target to host DNA in larvae naturally infected with *Melissococcus plutonius*

Table 2.12: Results from a generalised linear model assessing the relationship of the ratio of the concentration of pathogen target to host DNA in larvae naturally infected with *Melissococcus plutonius,* with samples treated with different percentages of PMA host depletion treatments compared to the recommended volume (1%, 10% and 100%). The family used for the model was Gaussian. (*) represents a significant difference.

Variable	t-value	p-value	sig
(Intercept)	0.132	0.899	
1% PMA	0.494	0.639	
10% PMA	1.547	0.173	
100% PMA	3.240	0.018	*

The concentration of *M. plutonius* appeared to remain unchanged between PMA treated and untreated samples (Figure 2.15). A generalised linear model (GLM) of the *M. plutonius* concentration showed there was no significant difference in concentration between all the PMA treated and the untreated samples (Table 2.13). For the GLM it was shown using a normality test that the residuals did not deviate from a normal distribution (Figure A.10).

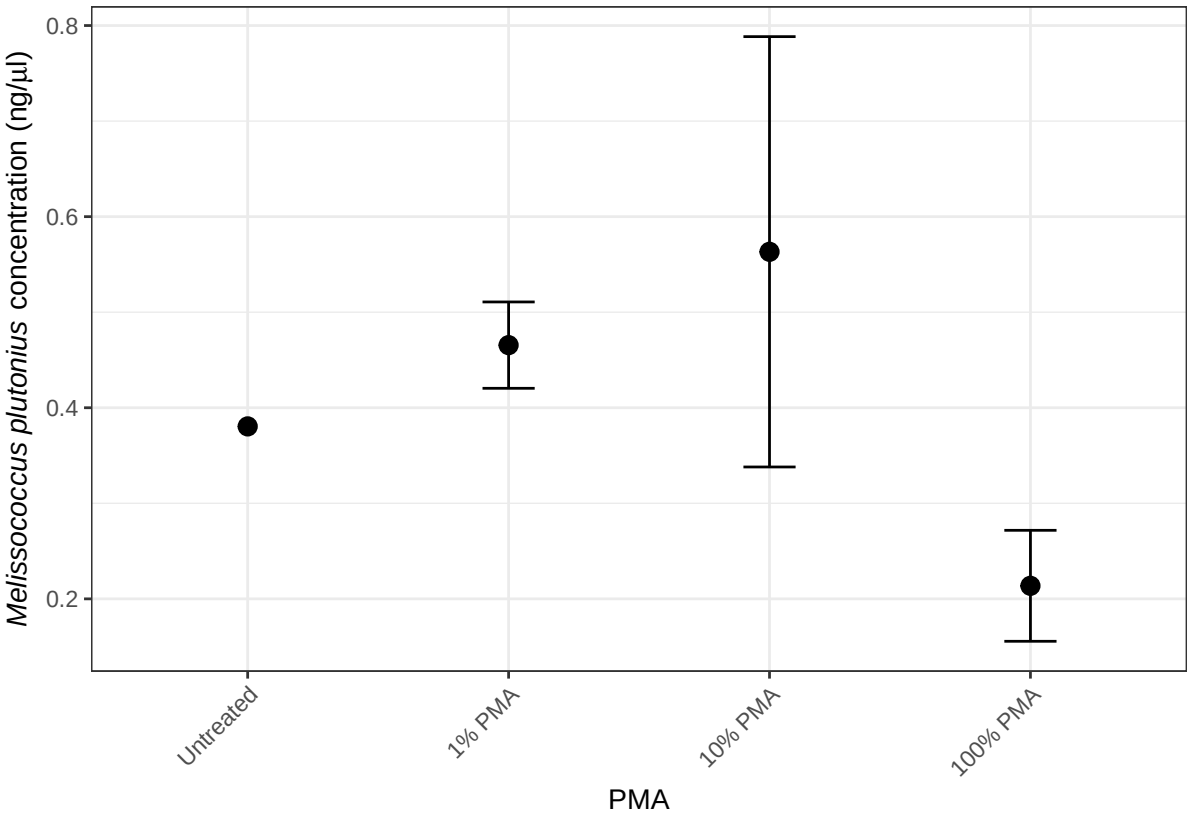


Figure 2.15: Effect of different percentages of PMA host depletion treatments compared to the recommended volume (1%, 10% and 100%) on the concentration of *Melissococccus plutonius* DNA in larvae naturally infected with *Melissococcus plutonius*.

Table 2.13: Results from a generalised linear model assessing the relationship of the ratio the concentration of *Melissococcus plutonius* DNA in larvae naturally infected with *Melissococcus plutonius,* with samples treated with different percentages of PMA host depletion treatments compared to the recommended volume (1%, 10% and 100%). The family used for the model was Gaussian. (*) represents a significant difference.

Variable	t-value	p-value	sig
(Intercept)	1.606	0.159	
1% PMA	0.311	0.766	
10% PMA	0.668	0.529	
100% PMA	-0.610	0.564	

2.3.2 DNA extraction development

The costs for commercial kits were almost double the cost from the home-made methods, and and the commercial costs do not include added consumables (Table 2.14). The cheapest method for DNA extraction was the home-made bead extraction.

Table 2.14: Cost per sample for DNA extraction methods both home-made and commercial. The cost per sample was calculated from the retail price of kits and consumables at the time of writing. The cost for the commercial kits does not include the additional consumbales cost.

Extraction Method	Type	Cost per sample (£)
Qiagen DNeasy blood and tissue kit	Commercial	4.05
Promega® Food kit	Commercial	4.53
Promega® HMW kit	Commercial	5.11
Home-made column	Homemade	2.22
Home-made bead	Homemade	1.56
Caterpillar method	Homemade	2.22

2.3.2.1 Testing different column DNA extractions

The efficacy of the extraction of bacterial DNA was visualised as a measure of the concentration of *M. plutonius* (Figure 2.16). The concentration of *M. plutonius* was nearly 0 ng/μl for the caterpillar method. A generalised linear model (GLM) of the *M. plutonius* concentration showed there was a significant decrease in concentration between both the caterpillar method and home-made column extraction compared to the commercially available Qiagen DNeasy kit (Table 2.15). For the GLM it was shown using a normality test that the residuals did not deviate from a normal distribution (Figure A.11).

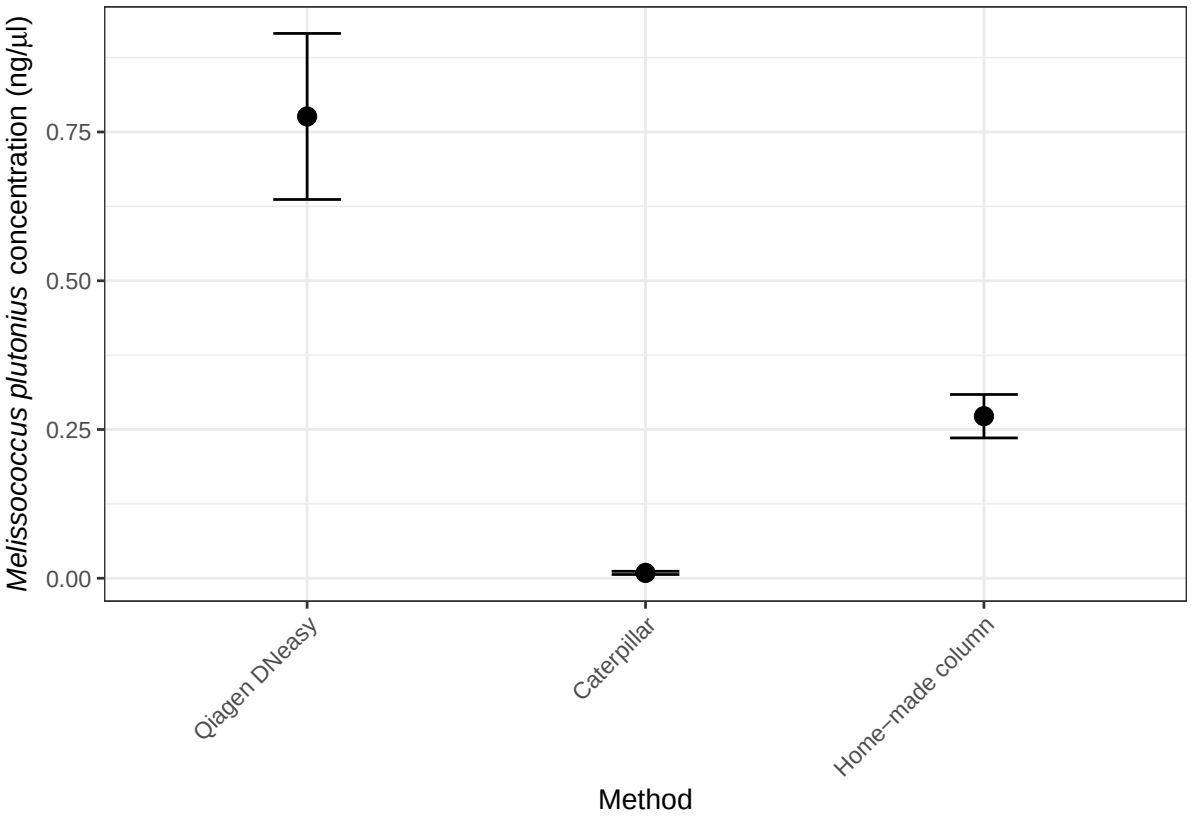


Figure 2.16: Effect of DNA extraction methods (caterpillar, home-made column and Qiagen DNeasy) on the concentration of *Melissococcus plutonius* DNA in larvae naturally infected with *Melissococcus plutonius*

Table 2.15: Results from a generalised linear model assessing the relationship of the concentration of *Melissococcus plutonius* DNA in larvae naturally infected with *Melissococcus plutonius* with different DNA extraction methods being used (caterpillar, home-made column and Qiagen DNeasy). The family used for the model was Gaussian. (*) represents a significant difference.

Variable	t-value	p-value	sig
(Intercept)	9.319	0.000	*
Caterpillar	-6.513	0.000	*
Home-made column	-4.277	0.001	*

Host depletion was visualised as a change in the ratio of *M. plutonius* to *A. mellifera* DNA for each extraction method on naturally infected samples (Figure 2.17). A generalised linear model (GLM) was performed to assess the relationship between the ratio of the different extraction methods, (Table 2.16). GLM showed that there was a significant

decrease in the ratio of the concentration of pathogen to host with both the home-made column extraction and the caterpillar method compared to the commercially available Qiagen DNeasy kit. For the GLM it was shown using a normality test that the residuals did not deviate from a normal distribution (Figure A.12).

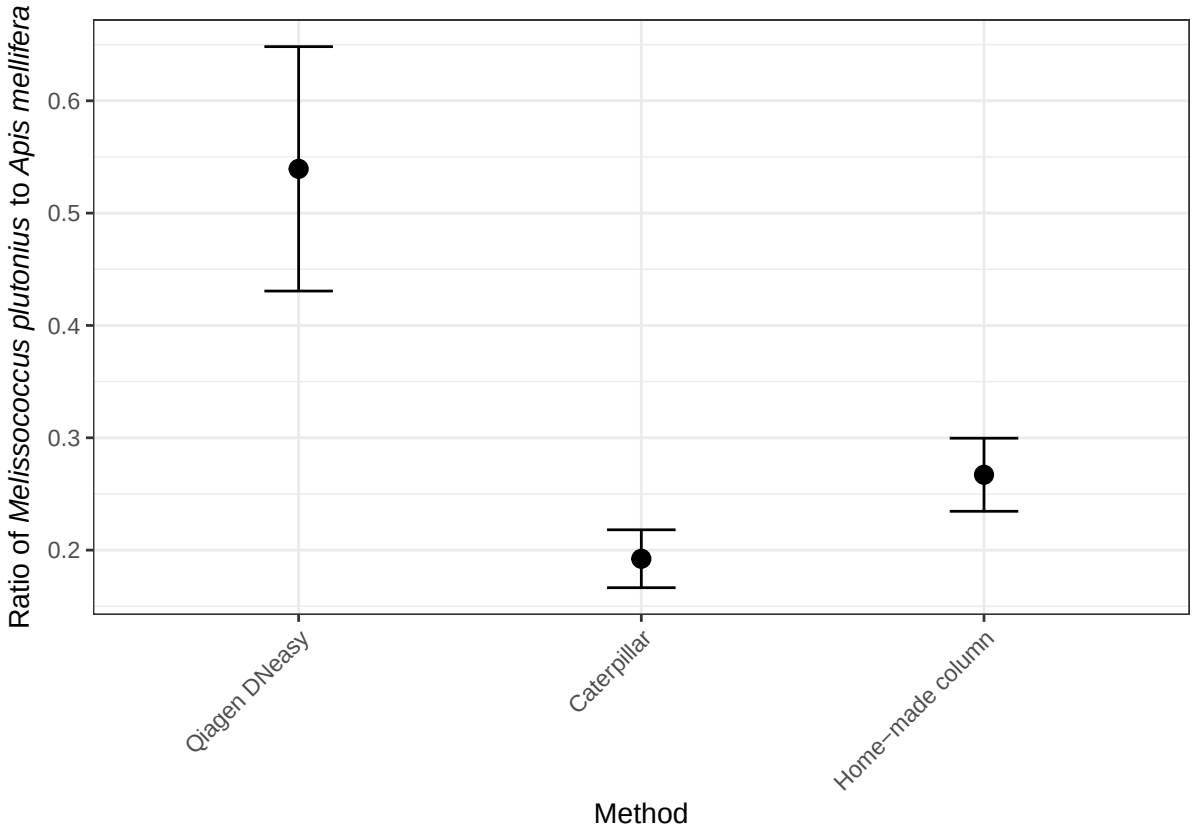


Figure 2.17: Effect of DNA extraction methods (caterpillar, home-made column and Qiagen DNeasy) on the ratio of the concentration of pathogen target to host DNA in larvae naturally infected with *Melissococcus plutonius*

Table 2.16: Results from a generalised linear model assessing the relationship of the ratio of the concentration of pathogen target to host DNA in larvae naturally infected with *Melissococcus plutonius,* with different DNA extraction methods being used (caterpillar, home-made column and Qiagen DNeasy). The family used for the model was Gaussian. (*) represents a significant difference.

Variable	t-value	p-value	sig
(Intercept)	8.024	0.000	*
Caterpillar	-3.651	0.002	*
Home-made column	-2.865	0.012	*

2.3.2.2 Comparing column methods with a bead based method

The efficacy of the extraction of bacterial DNA was visualised as a measure of the concentration of *M. plutonius* (Figure 2.18). A generalised linear model (GLM) of the *M. plutonius* concentration showed there was a significant decrease in concentration between both the home-made bead method and home-made column extraction compared to the commercially available Qiagen DNeasy kit (Table 2.17). For the GLM it was shown using a normality test that the residuals did not deviate from a normal distribution (Figure A.13).

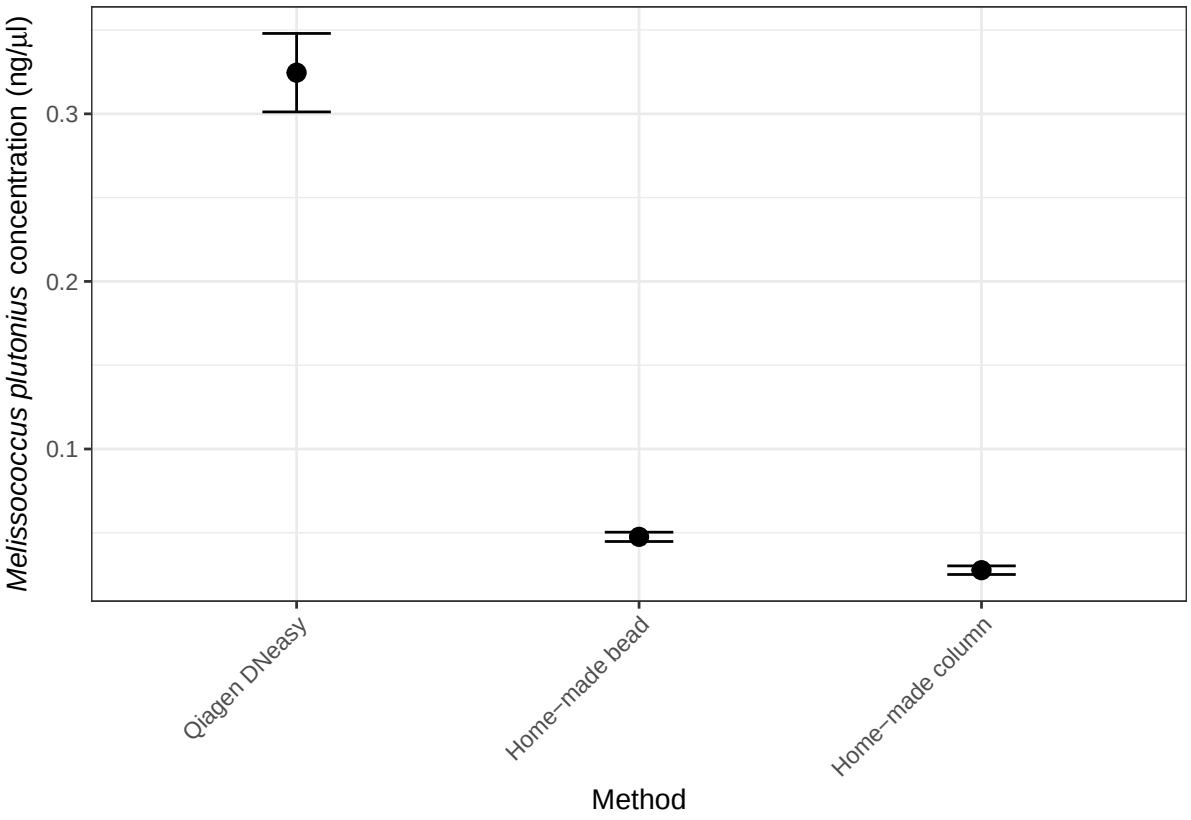


Figure 2.18: Effect of DNA extraction methods (home-made bead, home-made column and Qiagen DNeasy) on the concentration of *Melissococcus plutonius* DNA in larvae naturally infected with *Melissococcus plutonius*

Table 2.17: Results from a generalised linear model assessing the relationship of the concentration of *Melissococcus plutonius* DNA in larvae naturally infected with *Melissococcus plutonius,* with different DNA extraction methods being used (home-made bead, home-made column and Qiagen DNeasy). The family used for the model was Gaussian. (*) represents a significant difference.

Variable	t-value	p-value	sig
(Intercept)	23.703	0	*
Home-made bead	-14.304	0	*
Home-made column	-15.332	0	*

Host depletion was visualised as a change in the ratio of *M. plutonius* to *A. mellifera* DNA for each extraction method on naturally infected samples (Figure 2.19). A generalised linear model (GLM) was performed to assess the relationship between the ratio of the different extraction methods (Table 2.18). GLM showed that there was a significant decrease in the ratio between the home-made bead method and the Qiagen DNeasy method. There was a significant increase in the ratio between the home-made column method and the Qiagen DNeasy kit. For the GLM it was shown using a normality test that the residuals did not deviate from a normal distribution (Figure A.14).

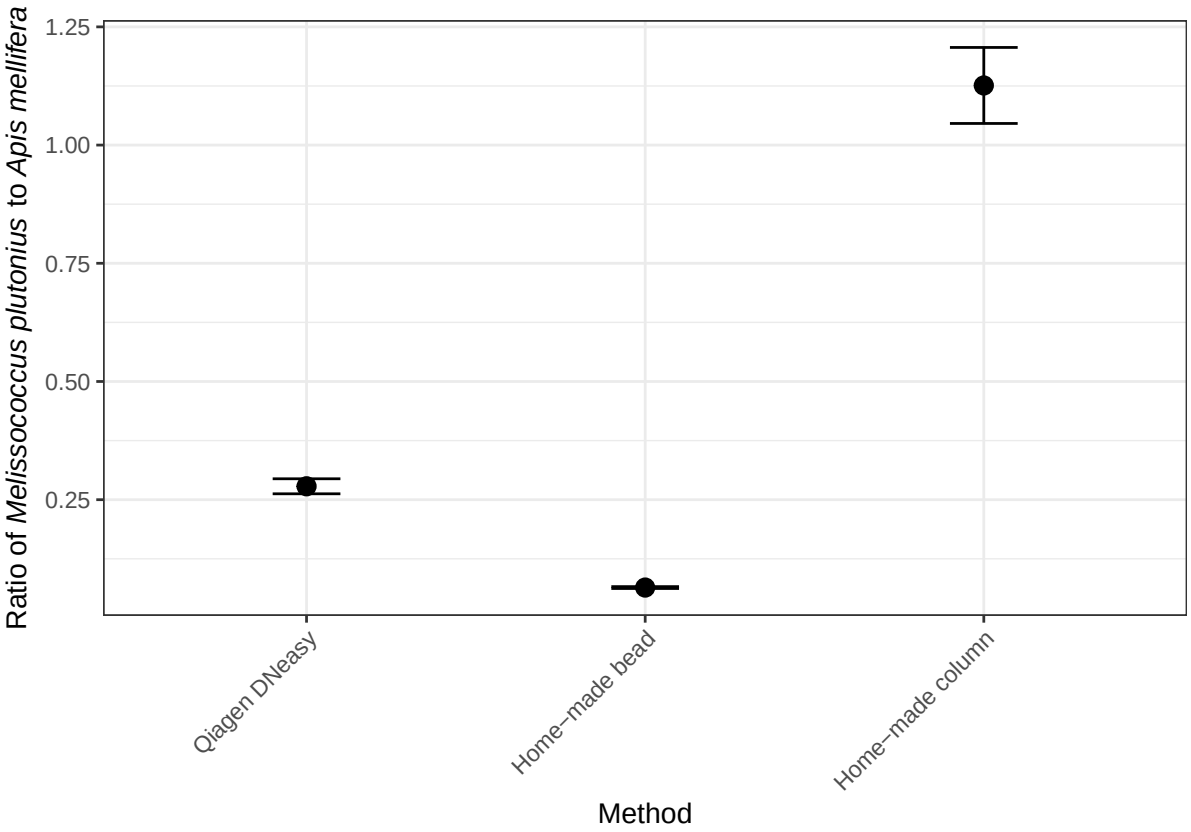


Figure 2.19: Effect of DNA extraction methods (home-made bead, home-made column and Qiagen DNeasy) on the ratio of the concentration of pathogen target to host DNA in larvae naturally infected with *Melissococcus plutonius*

Table 2.18: Results from a generalised linear model assessing the relationship of the ratio of the concentration of pathogen target to host DNA in larvae naturally infected with *Melissococcus plutonius,* with different DNA extraction methods being used (home-made bead, home-made column and Qiagen DNeasy). The family used for the model was Gaussian. (*) represents a significant difference.

Variable	t-value	p-value	sig
(Intercept)	5.884	0.001	*
Home-made bead	-3.201	0.019	*
Home-made column	12.670	0.000	*

2.3.2.3 Testing different elution solutions with a chosen DNA extraction method

The efficacy of the extraction of bacterial DNA was visualised as a measure of the concentration of *M. plutonius* (Figure 2.20). A generalised linear model (GLM) of the *M. plutonius* concentration showed there was a significant decrease in concentration between the elution using the caterpillar buffer and water buffer with the home-made column extraction method compared to the commercially available Qiagen DNeasy kit (Table 2.19). There was no significant difference between using elution buffers AE and TE with the home-made column extraction method compared to the commercially available Qiagen DNeasy kit. For the GLM it was shown using a normality test that the residuals did not deviate from a normal distribution (Figure A.15).

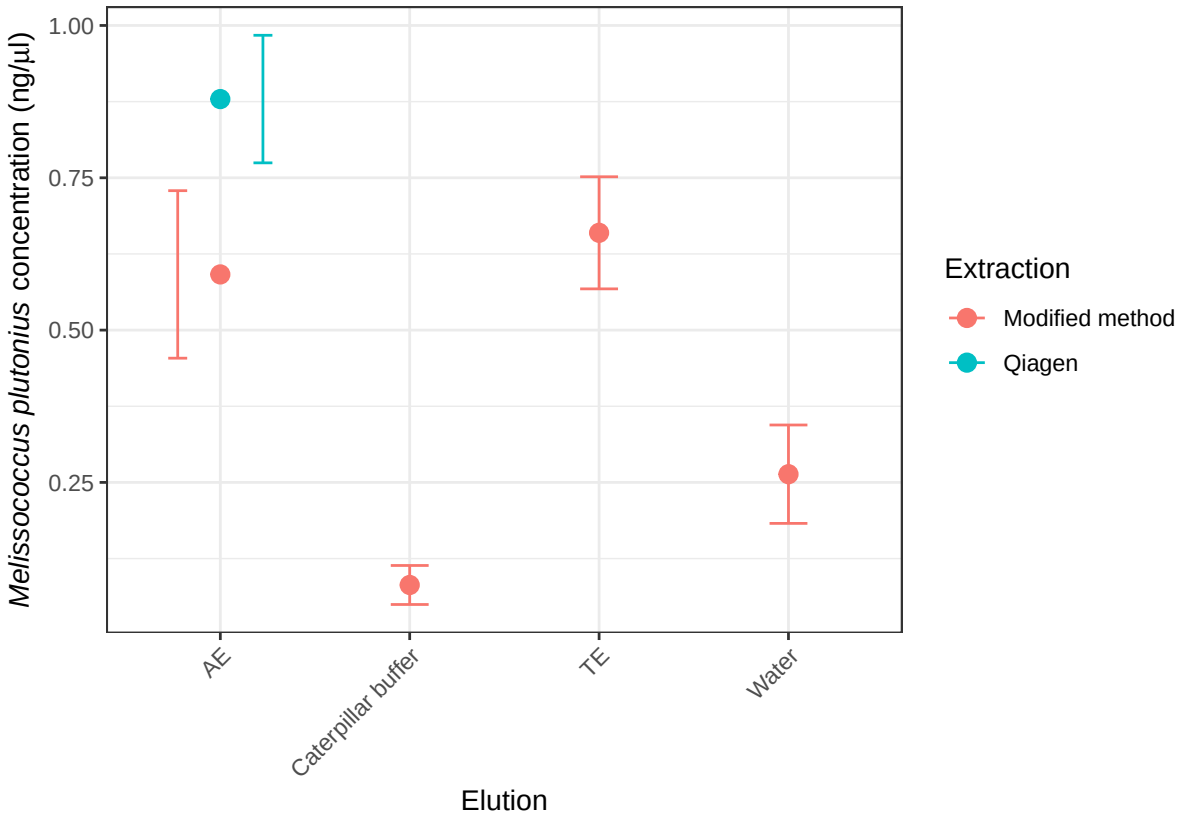


Figure 2.20: Effect of using different elution buffers (AE, TE, caterpillar and water) with the home-made column extraction method, as well as the standard Qiagen DNeasy kit with AE buffer, on the concentration of *Melissococcus plutonius* DNA in larvae naturally infected with *Melissococcus plutonius*

Table 2.19: Results from a generalised linear model assessing the relationship of the concentration of *Melissococcus plutonius* DNA in larvae naturally infected with *Melissococcus plutonius,* with different elution buffers (AE, TE, caterpillar and water), with the home-made column extraction method, as well as the standard Qiagen DNeasy kit with AE buffer. The family used for the model was Gaussian. (*) represents a significant difference.

Variable	t-value	p-value	sig
(Intercept)	9.179	0.000	*
AE	-2.125	0.060	
Caterpillar buffer	-5.888	0.000	*
TE	-1.620	0.136	
Water	-4.545	0.001	*

Host depletion was visualised as a change in the ratio of *M. plutonius* to *A. mellifera* DNA for each elution method on naturally infected samples (Figure 2.21). A generalised linear model (GLM) was performed to assess the relationship between the ratio of the different elution methods with the home-made column extraction compared with the standard Qiagen DNeasy kit with AE elution buffer (Table 2.20). GLM showed that there was a significant decrease in the ratio between the elution using the caterpillar buffer and the Qiagen DNeasy method. There was no significant difference on the ratio of using elution buffer AE, TE or water compared to Qiagen DNeasy method. For the GLM it was shown using a normality test that the residuals did not deviate from a normal distribution (Figure A.16).

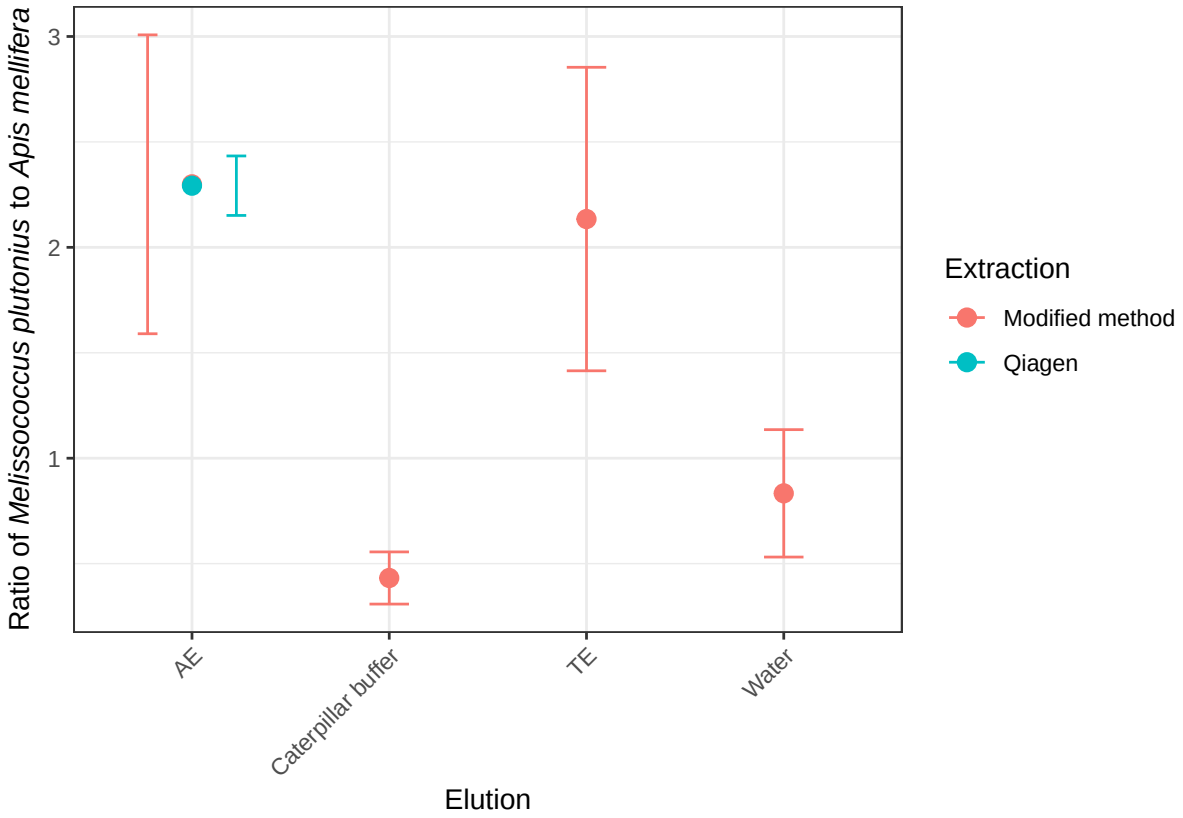


Figure 2.21: Effect of using different elution buffers (AE, TE, caterpillar and water) with the home-made column extraction method, as well as the standard Qiagen DNeasy kit with AE buffer, on the ratio of the concentration of pathogen target to host DNA in larvae naturally infected with *Melissococcus plutonius*

Table 2.20: Results from a generalised linear model assessing the relationship of the ratio of the concentration of pathogen target to host DNA in larvae naturally infected with *Melissococcus plutonius,* with different elution buffers (AE, TE, caterpillar and water), using the home-made column extraction method, compared to the standard Qiagen DNeasy kit with AE buffer for eluting DNA. The family used for the model was Gaussian. (*) represents a significant difference.

Variable	t-value	p-value	sig
(Intercept)	4.787	0.001	*
AE	0.010	0.992	
Caterpillar Buffer	-2.747	0.021	*
TE	-0.234	0.820	
Water	-2.154	0.057	

2.3.3 Combining host depletion and extraction methods to generate a final sequencing pipeline

Using the methods established from the results of the previous experiments in section 2.3.1 and 2.3.2 a final method was tested using qPCR and whole genome sequencing methods.

2.3.3.1 Combining the selected DNA extraction method with host depletion at different volumes

Host depletion was visualised as a change in the ratio of *M. plutonius* to *A. mellifera* DNA for each host depletion treatment with the selected extraction method (home-made column with a TE elution), and the sample input volume on naturally infected samples (Figure 2.22). A generalised linear model (GLM) was performed to assess the relationship between the ratio of the different host depletion methods and smple input volumes (Table 2.21). GLM showed that there was a significant increase in the ratio of the concentration of pathogen to host with the DNase with water lysis treatment compared to the untreated samples. There was no significant difference between the ratios for the water lysis only and all the PMA treatments compared to the untreated samples. There was no significant difference in the sample input volume between both 1 ml and 2 ml compared to 200 μl. For the GLM it was shown using a normality test that the residuals did not deviate from a normal distribution (Figure A.17).

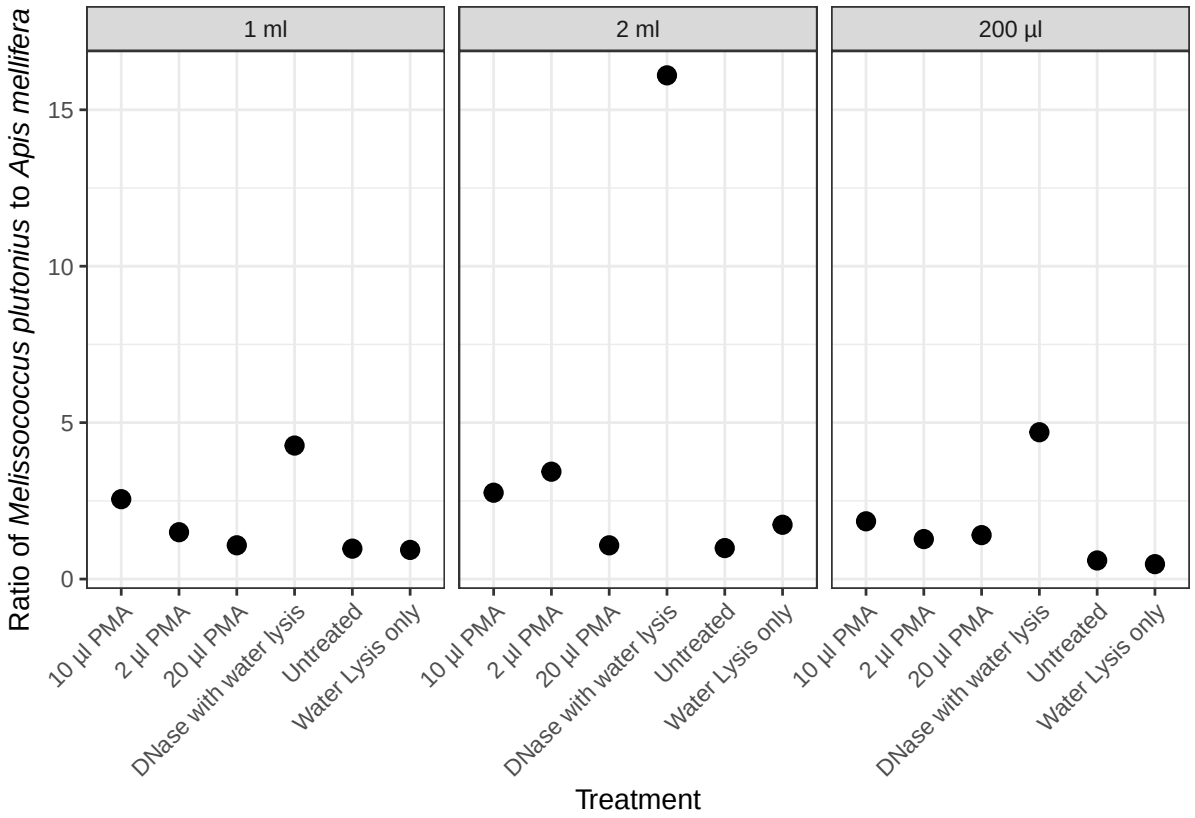


Figure 2.22: Effect of host depletion treatments (water lysis only, DNase with water lysis, 2 μl PMA, 10 μl PMA and 20 μl PMA) and sample input volume (200 μl, 1 ml and 2 ml) on the ratio of the concentration of pathogen target to host DNA in larvae naturally infected with *Melissococcus plutonius*

Table 2.21: Results from a generalised linear model assessing the relationship of the ratio of the concentration of pathogen target to host DNA in larvae naturally infected with *Melissococcus plutonius,* with samples treated with different host depletion treatments (water lysis only, DNase with water lysis, 2 μl PMA, 10 μl PMA and 20 μl PMA) and sample input volume (200 μl, 1 ml and 2 ml). The family used for the model was Gaussian. (*) represents a significant difference.

Variable	t-value	p-value	sig
(Intercept)	-0.047	0.963	
Water Lysis	0.092	0.929	
DNase with water lysis	3.516	0.006	*
2 µl PMA	0.570	0.581	
10 µl PMA	0.719	0.489	
20 µl PMA	0.157	0.879	
1 ml Volume	0.111	0.914	
2 ml Volume	1.746	0.111	

The efficacy of the extraction of bacterial DNA was visualised as a measure of the concentration of *M. plutonius* (Figure 2.23). A generalised linear model (GLM) of the *M. plutonius* concentration showed there was no significant difference in concentration between all the host depletion treatments compared to the untreated samples (Table 2.22). There was a significant increase in the concentration for the 1 ml and 2 ml input volumes compared to 200 μl. For the GLM it was shown using a normality test that the residuals did not deviate from a normal distribution (Figure A.18).

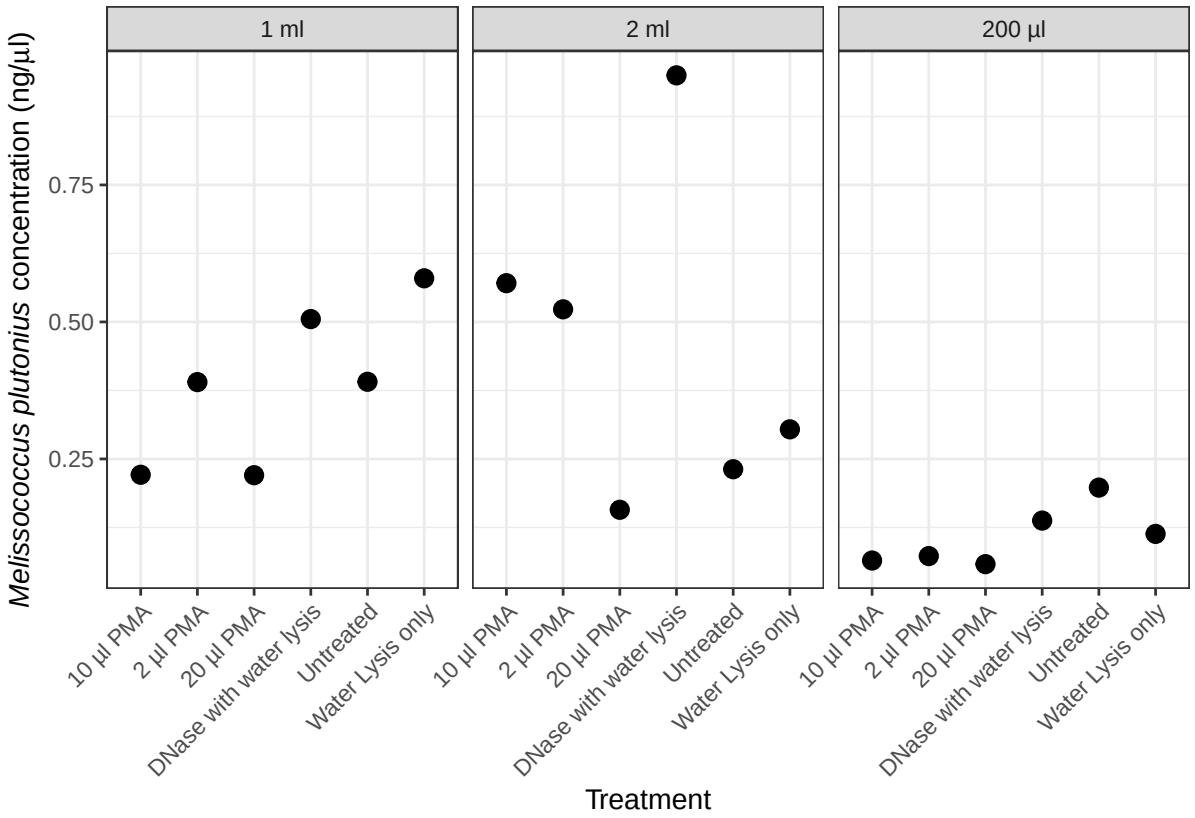


Figure 2.23: Effect of host depletion treatments (water lysis only, DNase with water lysis, 2 μl PMA, 10 μl PMA and 20 μl PMA) and sample input volume (200 μl, 1 ml and 2 ml) on the concentration of *Melissococcus plutonius* DNA in larvae naturally infected with *Melissococcus plutonius*

Table 2.22: Results from a generalised linear model assessing the relationship of the concentration of *Melissococcus plutonius* DNA in larvae naturally infected with *Melissococcus plutonius,* with samples treated with different host depletion treatments (water lysis only, DNase with water lysis, 2 μl PMA, 10 μl PMA and 20 μl PMA) and sample input volume (200 μl, 1 ml and 2 ml). The family used for the model was Gaussian. (*) represents a significant difference.

Variable	t-value	p-value	sig
(Intercept)	0.550	0.594	
Water Lysis only	0.411	0.690	
DNase with water lysis	1.793	0.103	
2 µl PMA	0.385	0.708	
10 µl PMA	0.085	0.934	
20 µl PMA	-0.891	0.394	
1 ml Volume	2.728	0.021	*
2 ml Volume	3.431	0.006	*

2.3.3.2 Testing DNase Treatment with and without water lysis using the Flongle

Host depletion was visualised as a change in the ratio of *M. plutonius* to *A. mellifera* DNA sequencing reads for both DNase treatment with and without a water lysis step, using the selected DNA extraction method (home-made column extraction with TE elution and 2 ml sample input volume) (Figure 2.24). Upon a visual inspection the ratio increased in all samples but to different extents. For E2683 there was a difference in ratio by approximately 30 compared with the untreated with no lysis sample. E2672 there was an increase in ratio by approximately 10 and for E2666 only a small increase of approximately 1 compared with the untreated with no lysis sample.

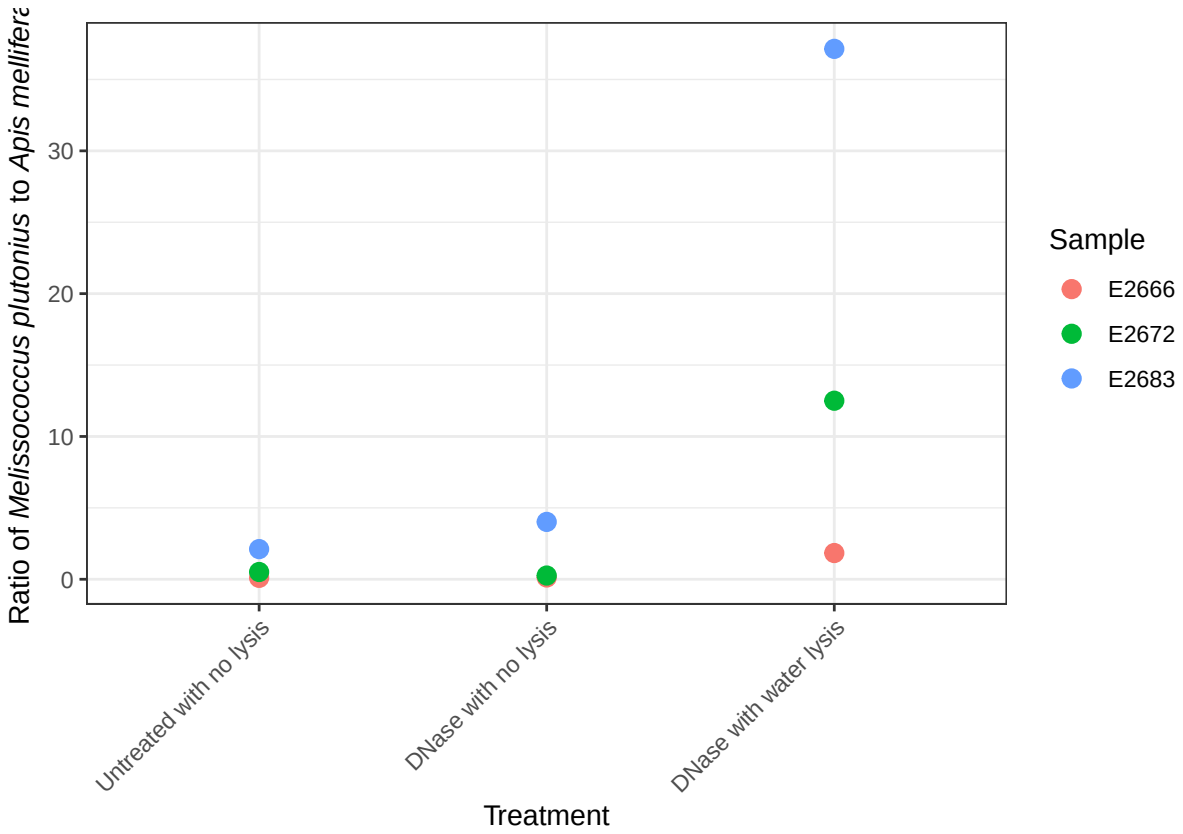


Figure 2.24: Effect of both DNase treatment with and without a water lysis on the ratio of the sequencing reads of pathogen target to host DNA in larvae naturally infected with *Melissococcus plutonius*

2.3.3.3 Testing final method using qPCR and Flongle

Host depletion was visualised as a change in the ratio of *M. plutonius* to *A. mellifera* DNA for the selected host depletion method (DNase with a water lysis) and the selected extraction method (home-made colum extraction with TE elution and a 2 ml sample input volume) (Figure 2.25). The ratio increased for all samples with the DNase with water lysis treatment but to different extents. The difference between DNase treated and untreated appeared to be the greatest in E2559. Sample E2540 appeared to have a very small increase between DNase treatment and no treatment.

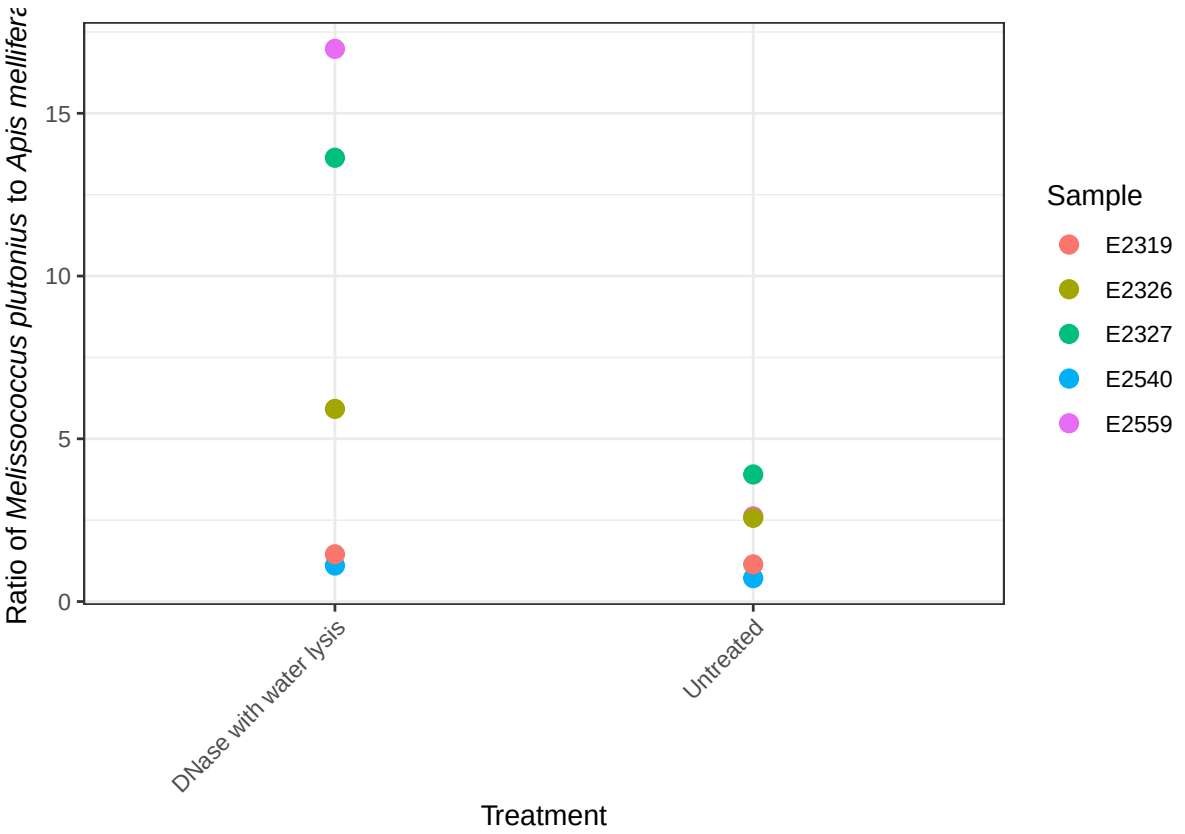


Figure 2.25: Effect of DNase treatment with a water lysis on the ratio of the concentration of pathogen target to host DNA in larvae naturally infected with *Melissococcus plutonius*

The efficacy of the extraction of bacterial DNA was visualised as a measure of the concentration of *M. plutonius* (Figure 2.26). For sample E2540 there was an increase in the *M. plutonius* concentration in the DNase treated sample compared to untreated sample. For the rest of the samples there was an increase in *M. plutonius* concentration in the untreated sample compared to the DNase treated one.

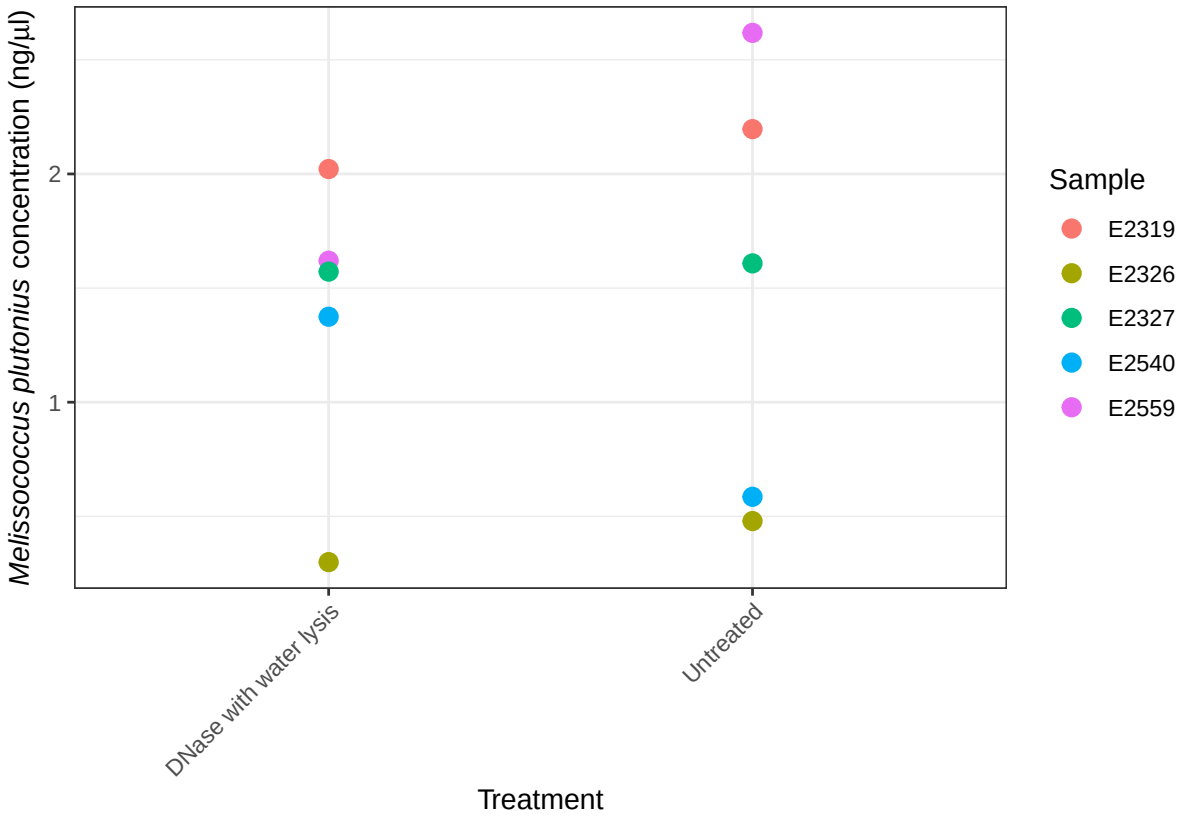


Figure 2.26: Effect of DNase treatment with a water lysis on the concentration of *Melissococcus plutonius* DNA in larvae naturally infected with *Melissococcus plutonius*

For the two samples that were taken forward for Oxford Nanopore flongle sequencing the host depletion was visualised as a change in the ratio of *M. plutonius* to *A. mellifera* percentage sequencing reads for the selected host depletion method (DNase with a water lysis) and the selected extraction method (home-made column extraction with TE elution and a 2 ml sample input volume) (Figure 2.27). The DNase with water lysis treatment increased the ratio of *M. plutonius* to *A. mellifera* percentage reads for both samples.

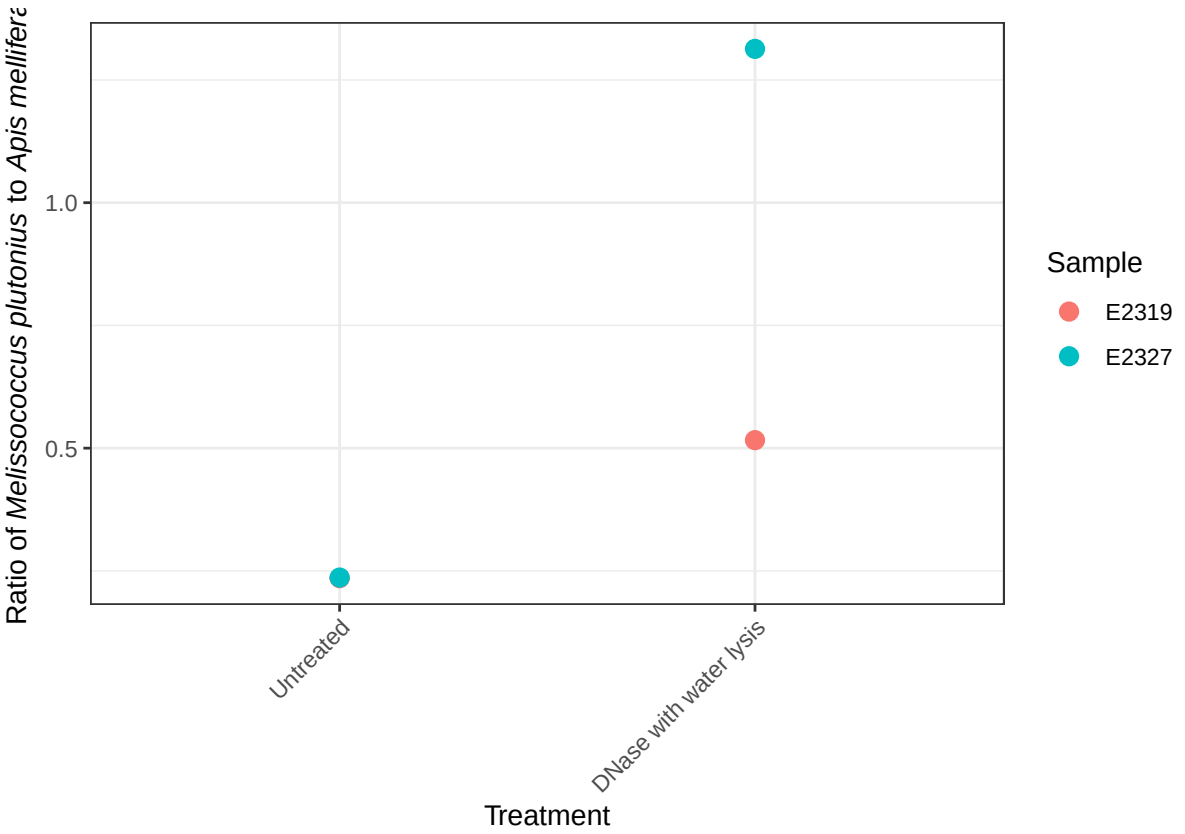


Figure 2.27: Effect of DNase treatment with a water lysis on the ratio of the percentage reads of pathogen target to host DNA in larvae naturally infected with *Melissococcus plutonius*

2.3.3.4 Cost of final pipeline

The cost of each of the developed pipeline steps was assessed using costs of each process at the time of writing, including all consumables required (Figure 2.28).

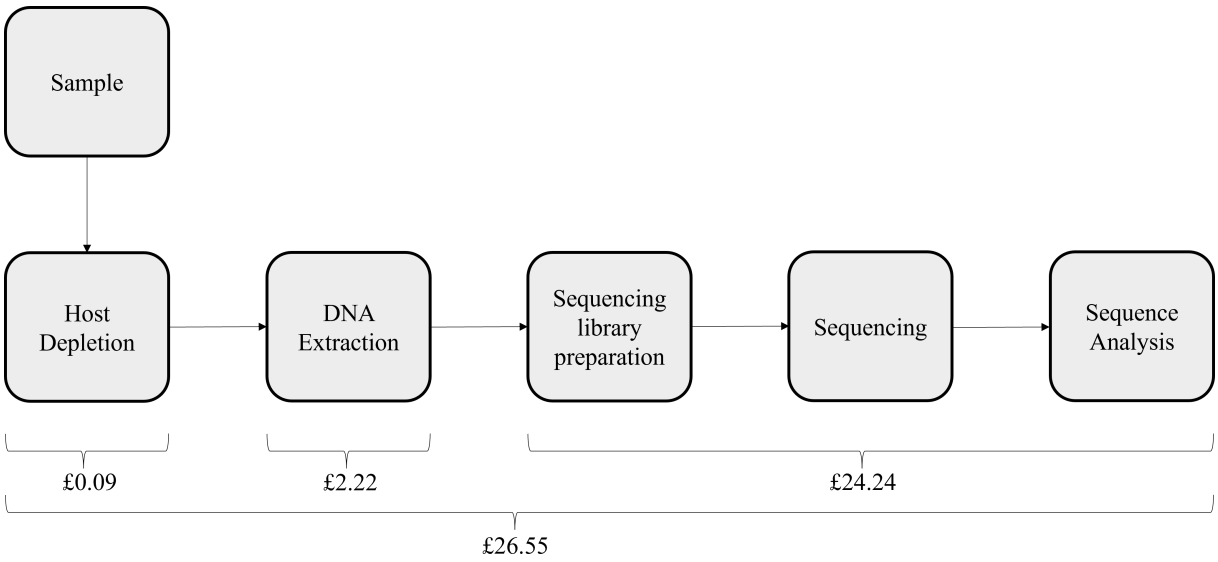


Figure 2.28: The cost of each stage involved in whole genome sequencing using the developed methods.

2.4 Discussion

Here I present a novel, cost-effective pipeline for whole genome sequencing from naturally infected larvae stored in existing LFD buffer bottles containing sodium azide using Oxford Nanopore sequencing. This method can be used on samples already collected in the field with no requirement for additional sampling. Different host depletion methods and extraction methods were assessed to produce a DNA extract that has a high concentration of *M. plutonius* to *A. mellifera* DNA, and is affordable enough to be deployed at scale.

Other studies that have generated whole genome sequencing data from *M. plutonius* culture the bacterium prior to sequencing (Djukic et al., 2018; Okumura et al., 2018). Such a workflow has costs and benefits. Firstly, the culturing step adds complexity with considerable time costs, and given that not all strains grow on media with equal vigor, could select against sequencing some strains that are reluctant to grow or get out competed by those that grow well. One clear benefit is that all sequencing data generated are for *M. plutonius*, resulting in high depths of sequence. My methods sought to use samples collected in the field that contained larval material and other microbes that would dilute any sequencing efforts. I therefore needed to deplete *A. mellifera* DNA to achieve high depths of *M. plutonius* sequencing. High sequencing depth is an important quality marker when generating whole genome sequencing data to infer pathogen epidemiology (Jiang et al., 2019). I used an increasing ratio of *M. plutonius* to *A. mellifera* as an indicator

of success when depleting samples of host DNA measured using two methods, qPCR and Flongle sequencing. The qPCR method was used for a majority of the method development as it is much cheaper than running each experiment on the Flongle. For the first experiment, in 2.2.2.3, to test the 16 samples after extraction using the Flongle it would have cost approximately £400 for sequencing alone, and with qPCR it cost approximately £15. This cost saving allowed more methods to be tested and optimised. When the ratio was assessed using both the flongle and qPCR (Figure 2.25 and Figure 2.27), the results matched demonstrating that calculating the ratios using qPCR data gave a clear indication of host depletion. Increased ratios of *M. plutonius* to *A. mellifera* indicated that several methods were effective at depleting the samples of host DNA including PMA with blood, water and no lysis and DNase with a water lysis and a blood lysis. When PMA was used as a host depletion technique with saliva samples a water lysis step was required, unlike my study that found PMA with the water lysis and no lysis to be equally effective at depleting the host in larvae (Marotz et al., 2018). The difference could mean that the host DNA was already exposed to the PMA in my study hence not requiring a lysis step. The DNase treatment with a saponin selective lysis was found to be the most effective at depleting human tissue host samples (Bruggeling et al., 2021), my study showed that a lysis was also required for effective host depletion using DNase. The host depletion efficiency is important but also is cost.

Reducing costs is important as the whole genome pipeline needs to be cost effective to be adopted as a routine method, at scale and within tight funding budgets. For the host depletion step, from the effective PMA treatments the PMA with no lysis method was chosen to take forward, as without the lysis step there was a reduced consumable cost. From the effective DNase treatments, the DNase with water lysis was taken forward as it did not require a buffer from commercially available kit like the blood buffer, and water is cheap and readily available. Saving costs at this stage is important in reducing the cost of the whole protocol. An attempt to save costs with the PMA was acheived by reducing the volume added but this, in turn, negatively impacted efficiency. The home-made column extraction was taken forward to be optimised as it was half the price of the Qiagen extraction kit (Table 2.14). Due to the cost saving and that the ratio of *M. plutonius* to *A. mellifera* was higher compared to the Qiagen extraction, section 2.3.2.2, the home-made column extraction was taken forward, although the *M. plutonius* concentration was lower. The *A. mellifera* genome is over 100 x the size of the *M. plutonius*

genome (Okumura et al., 2011; Wallberg et al., 2019). This size difference means that even in small quantities the host DNA can still have a large influence on the number of host reads, therefore a higher ratio of *M. plutonius* to *A. mellifera* is more important than the *M. plutonius* concentration at ensuring the most target reads are produced (Ganda et al., 2021). The choice of elution buffer was influenced by cost, the most effective elution buffers were TE and AE. AE buffer was part of the Qiagen DNeasy kit and could have been purchased only through the company, whereas the TE buffer was a common chemical found in many molecular labs, could have been purchased from many different companies and in multiple forms, making it a cheaper option. The cost of whole genome sequencing has decreased over time, and it was predicted that this decrease in cost will continue (Pollie, 2023). Outsourcing sequencing always costs more than in-house sequencing as commerical companies need to make profit, so my method will be cheaper. At £26.55 a sample, the developed pipeline was cost-effective, but time is also money.

Time was also an important factor for development of this pipeline. When dealing with outbreak situations transmission events need to be identified quickly so further transmission can be prevented and field management can be put into place in problematic areas. Whole genome sequencing can be used to identify cluster outbreaks allowing transmission events to be investigated (Gilchrist et al., 2015). For my pipeline the cluster information would be sent to the bee inspector and with their knowledge of the beekeepers movements and practices, transmission events can be identified and appropriate field management put into place. The time to generate whole genome sequencing data has decreased, making this method more compatible to outbreak analysis than in the past, this developed pipeline took approximately a week from sample receipt to sequence output (Gilchrist et al., 2015). A quick turnover of samples to whole genome sequence data enabled cluster outbreaks, and subsequently transmission events to be identified in both a local hospital scale and a country wide scale of SARS-CoV-2 (Francis et al., 2022). The sequencing process was quick, it took one day to prepare the library from DNA extracts following the quick and simple Oxford Nanopore library preparation protocol, section 2.2.4.2, 24 hours to run the flongle and a day for data analysis. The plan was to use this pipeline on all 2020 samples, as this was a larger scale the samples would be run on the Oxford Nanopore MinION which runs for an extra 24 hours, but generating a lot more samples worth of data. The developed methods can be performed in 96 well plates so can be easily scaled up for this additional throughput, requiring no additional time per

sample than in individual tubes. Other parts of the pipeline were designed to save time in particularly the sample choice. Using samples that are already collected saves time, as no additional sampling is required from the bee inspector. Eliminating the need to culture *M. plutonius* from the infected larvae as done previously for *M. plutonius* sequencing saved a lot of time (Djukic et al., 2018; Okumura et al., 2018). To culture from the larval samples it would have taken at least two weeks from receiving the sample to generation of sequencing data, for the pipeline I developed this would take five days. For the host depletion, cost was not the only reason the PMA with no lysis was taken forward for further analysis, it also was to save time as there was one less step in the process. The time that could not be controlled was the time between collection and lab storage.

The time and storage conditions between the sample being taken and receipt at the lab was unknown, as there were no storage guidelines and no time limit between the two, so they could have been left in a car or stuck in the post. The sampling was also performed by the apiary's local bee inspector, so although there is a protocol in place, there may be a difference between the handler, meaning human error could have occured possibly by putting too much or too little material into the bottle. The possible variation in sampling and sample handling could explain why there was a large variation in the ratio of *M. plutonius* to *A. mellifera* in some samples after host depletion treatment, section 2.3.3.3. If more larval material was collected in some samples this could have impaired the host depletion efficiency, as there was more material to deplete, or the DNA extraction, as the silica membrane can be overloaded. There also could have been more dead larval material making the host depletion method more effective in some samples. There was also a lot of variation in the starting *M. plutonius* concentration between each of the samples, and in some cases when treated the *M. plutonius* concentration decreased. The host depletion method used, DNase, is also used to detect cell viability (Reyneke et al., 2017). The decrease of *M. plutonius* DNA concentration in some of the samples could have been due to some of the *M. plutonius* cells being dead, possibly due to sampling but also just naturally. If the bacterial cells were dead, they would have had damaged membranes and therefore would have been more susceptible to osmotic lysis, or the dead cells have lysed and some of the DNA detected in the non depleted samples was free DNA. In both cases the DNA would be susceptible to DNase digestion during host depletion, therefore reducing *M. plutonius* levels (Cangelosi & Meschke, 2014). This was not a huge issue as the ratio still increased in the decreased *M. plutonius* samples. The difference

in *M. plutonius* concentration could also be due to sampling methods but also levels of infection will likely have varied between samples naturally. Storage was assessed and samples stored in the fridge and freezer made no difference to the quality of the DNA in the spiked samples, however in naturally infected samples storage could have an effect, as the spiked samples appeared to react differently to the naturally infected samples.

Initial experiments on host depletion methods used larval samples spiked with cultured *M. plutonius*. When these larvae were tested with different lysis and host depletion treatments, the ratio was significantly higher in both PMA with blood lysis and PMA with water lysis than in the untreated samples (Table 2.6), suggesting that these methods were having the desired effect of depleting host DNA. However, when looking at the *M. plutonius* DNA concentration for these same samples, there was significantly less DNA in treated samples compared to untreated samples. These results suggest that the treatment was destroying most of the *M. plutonius* DNA, so although the increase in pathogen to host ratio looked promising, the low DNA concentrations were unacceptable for sequencing. When the same experiment was performed on naturally infected samples all methods but the DNase treatment with no lysis significantly increased the *M. plutonius* to *A. mellifera* ratio, again suggesting the methods are depleting the host DNA. However, the *M. plutonius* DNA concentration remained similar to the untreated controls, suggesting pathogen DNA was being retained. Taken together, these data suggest that cells of cultured *M. plutonius* are not as robust as *M. plutonius* from naturally infected samples. Both lysis and host depletion steps destroyed all the bacterial DNA in cultured bacteria indicating the structural integrity of these were not as robust as bacteria in naturally infected samples. *M. plutonius* has been difficult to culture (Govan et al., 1998), especially the atypical strains (Arai et al., 2012). This difficulty in growing could suggest that the media is not optimal for this bacteria, if certain nutrients are lacking it could influence the structure of the membranes (Malouin et al., 1991), therefore causing the cell wall to be more sensitive to lysis treatments. Peptidoglycan make up the cell wall in Gram-positive bacteria (Pasquina-Lemonche et al., 2020), certain media was shown to encourage the binding of cell wall lytic enzymes to the peptidoglycan in Gram-positive bacteria (Bhagwat et al., 2021). It is possible that these lytic enzymes were digesting the cell walls making them exposed to any host depletion treatments. Given the importance of ensuring future methods were appropriate for field samples, cultures of *M. plutonius* were no longer used and future method development concentrated on naturally infected samples.

To conclude, the method was developed with consideration of cost, time, ratio of *M. plutonius* to *A. mellifera* and *M. plutonius* concentration, to be effective and costs £26.55 per sample and from samples that are already collected routinley through inspections. This method can be used on a large scale to generate whole genome sequencing data of all European Foulbrood positive buffer bottles from each year. The next steps would be to develop either a SNP profile or cgMLST to be able to use data generated from this method to distinguish outbreaks from within the current sequencing types. To further develop the method an *in silico* method could be tested as a host depletion method, such as adaptive sequencing. Adaptive sequencing works by either rejecting the unwanted reads or accepting the wanted reads as the DNA strands come through the Nanopore. For example, if *A. mellifera* DNA is to be rejected a file is uploaded with the *A. mellifera* genome information at the sequencing set up stage using MinKnow. With this information, if the DNA fed through the pore matches the uploaded file it stops recording the data and pushes the strand out of the pore, leaving it free to sequence another next strand (Payne et al., 2020). Adaptive sequencing is a live method so will save more time and cost than lab-based host depletion methods. Storage conditions of naturally infected samples could also be investigated with access to an infected apiary.

Chapter 3

A whole genome sequencing pipeline to investigate European Foulbrood outbreak sites across the UK

3.1 Introduction

The European honey bee, *Apis Mellifera*, is an important pollinator in the UK making it a valuable asset to the economy and agricultural systems (Breeze et al., 2011; Smith et al., 2011). Pollination is not the only way honeybees contribute to the economy, products they produce such as honey and beeswax are exported and in 2022 approximately £13 million was made exporting these products (ITC, 2023). Colony loss is a major problem threatening honeybee populations (Gray et al., 2023), some causes of this are believed to be climate change, pesticide use and disease (Christen et al., 2016; Conte & Navajas, 2008; Potts et al., 2016). Disease is a major problem for honeybees, a common bacterial disease is European Foulbrood (EFB). EFB is a larval infection, caused by *Melissococcus plutonius*, a Gram-positive bacteria. The bacteria colonise in the gut of the larvae, competing for food, leaving the larvae to die after approximately three to five days after infection. The diease is distinguished by an irregular brood pattern and larvae twisted with opaque or creamy white guts (White, 1912). The disease is thought to be spread by the nurse bees after cleaning out the cells with the infected dead larvae, through their mouth parts (Forsgren, 2010; Bailey, 1983). In the UK EFB is a notifiable diseases, so all suspected cases must be reported to government funded honey bee health inspectors

of the National Bee Unit (NBU), and appropriate action will be taken if a positive case is confirmed (Government, 2006). Recent years have seen a switch from antibiotic usage, to a combination of destruction or shook swarm, where only the infected combs are destroyed and the bees are shaken off into new foundation, and clean equipment (Wilkins et al., 2007; Waite et al., 2007). NBU bee health inspectors have used lateral flow devices (LFD) to confirm the visual diagnosis and allow immediate action to be taken.

To test for EFB using an LFD a sample is taken from an infected cell, and put into a bottle of sodium azide buffer containing ball bearings. The bottle is shaken and the sample is added to a device like those used for COVID-19 testing. As the sample moves through the membrane, if positive, it reacts with specific antibodies to produce a control line and a second blue line (Tomkies et al., 2009). The positive samples are tested at the National Bee Unit (NBU) for the sequence type using multi-locus sequence typing (MLST) (Haynes et al., 2013). MLST typing is a common method used to detect different strain types of a pathogen using fragments, loci, within a selection of coding genes (Maiden et al., 1998), there are over 130 MLST schemes (Jolley et al., 2018). For *M. plutonius* the MLST scheme is referred to as an MLST-like scheme as it uses some loci from non coding regions, it is also unusual as it only uses four loci unlike *Salmonella* and *E. coli* that uses seven (Jolley et al., 2018). Loci within two coding regions are used in the EFB MLST scheme, these are present within the acety-lornithine deacetylase encoding gene, argE and a putative secreted antigen, gbpB. The other two loci in the scheme are within intergenic regions, these are regions between the coding genes. The purR locus can be found upstream of the purine operon repressor and galK is found between two fragments of galaktokinase. Using primers specific for regions of each of the four loci and Sanger sequencing each sample is assigned an allele type, the combination of the allele types across all four genes, an allelic profile, assigns the sample type, for example argE-2, galK-3, gbpB-2 and purR-4 is sequence type 3 (ST3). This scheme was developed in 2013 and currently there are 46 sequence types of EFB across the world (Haynes et al., 2013; Jolley et al., 2018). The sequence types are divided into clonal complexes.

Clonal complexes (CCs) group related sequence types that share a specific similarity in their allelic profiles defined by the GoeBurst algorithm (Francisco et al., 2009). Usually, clonal complexes are created based on sequence types (STs) sharing five identical loci out of the seven present (Madera et al., 2023). For each clonal complex, an allelic profile is calculated to be the ancestral or most common type after which the CC the is named.

For example, if ST7 was the ancestral strain the clonal complex would be defined as CC7 (Budge et al., 2014). As there are only four genes in the MLST scheme for EFB the algorithm starts by creating groups based on three out of four identical loci rather than the traditional five out of seven. There are currently three clonal complexes of EFB, CC3, CC12 and CC13. When new allelic combinations are discovered or profiles with new alleles such as a new galK allele a new sequence type is assigned. The allelic profiles of any new sequence types are passed through the goeburst algorithm to assign it a clonal complex (Grossar et al., 2023). Currently there are 16 sequence types in CC3, 10 in CC12, 18 in CC13 and 2 that have not been assigned yet (Table 3.1) (Jolley et al., 2018). The spread of the sequence types vary across the globe, for example studies show there are 12 different sequence types present in Switzerland and 15 in the UK (Budge et al., 2014; Grossar et al., 2023). Reports show that the most prevalent clonal complex in the UK is CC3, and the most widespread sequence types are ST3, ST5 and ST23 making up 72% of positive cases (Budge et al., 2014). These widespread STs create large clusters, or multiple smaller clusters within the landscape (Budge et al., 2014). A higher resolution of genetic information is required to track transmission events within these foci. Single nucleotide polymorphisms (SNPs) generated from the whole genome sequencing can provide such increased resolution for epidemiological studies.

Table 3.1: The sequence types present in each of the clonal complexes.

Clonal Complex	Sequence types
3	2,3,5,6,7,11,22,23,24,29,30,31,35,39,46,47
12	10,12,16,19,21,25,27,33,34,36
13	4,8,9,13,14,15,17,18,20,26,28,32,40,41,42,43,44,45
Unclassified	48,49

A SNP is a mutation of one nucleotide to another one at a single point in the genome. SNPs can occur in all regions of the genome and could have a functional effect if present in regions that code for genes, for example loss or gain of function (Schork et al., 2000). In medicine SNPs are used frequently as biomarkers for certain diseases in humans or for assessing how patients may react to certain medication (Schork et al., 2000). For

investigating the epidemiology of pathogens SNP profiles across multiple samples can be used. With the SNP profile the phylogeny of individuals can be assessed (Faison et al., 2014). A study used a SNP profile to investigate how different species of the zoonotic disease *Brucella* are related, and the different species were very clearly resolved through this analysis (Foster et al., 2009). SNP profiles have not only been able to distinguish different species, they can also be used for studying the phylogeny of the same species, for example in *Bacillus anthracis* (Pearson et al., 2004). Using SNPs to investigate the phylogeny of different *M. plutonius* samples would be a useful tool to deeper understand the epidemiology, expanding on the the MLST scheme by using whole genome sequence data.

A method to produce whole genome data of *M. plutonius* from the buffer bottle samples already collected was developed in chapter 2, using Oxford Nanopore sequencing. Oxford Nanopore was associated with high error rates, but with each new release the error rate is improving (Cuber et al., 2023). With the error rate improving the technology is appearing in more studies particularly in healthcare settings. With antimicrobial resistant bacteria being a major concern, rapid ways to survey outbreaks are being studied. Oxford Nanopore sequencing is a rapid sequencing technology, so many studies have been investigating ways to implement it into epidemiological studies. A study developed a method to rapidly sequence isolates of *Staphylococcus aureus* using Oxford Nanopore sequencing and SNP profiles, validated with Illumina MiSeq (Ferreira et al., 2021). Illumina sequencing is a different sequencing platform that is commonly used due to low error rates (Stoler & Nekrutenko, 2021). The method was used to test methicillin sensitive *S. aureus* on real outbreak samples in a hospital, and identified 2 separate outbreaks in less than 31 hours (Ferreira et al., 2021). In multiple bacterial species, the SNP profiles generated from Oxford Nanopore sequencing were compared to those generated from Illumina sequencing. For some bacterial species the SNP profiles between the technologies were nearly identical, and some had differing numbers of SNPs, but the phylogeny matched known outbreaks (Linde et al., 2023). This shows that using Oxford Nanopore technology has the potential to generate useful SNP profiles to study phylogeny.

An assembled genome is required to generate a SNP profile because each SNP is spatially anchored to a location within the genome. Sequencing data is generated as fragments called contigs, which need to be linked together like pieces of a jigsaw to construct a whole

genome. This process is called assembly. To assemble a genome there are two common methods, reference guided and *de novo*. A reference guided assembly is where contigs are mapped to a reference genome of the desired output to guide the assembly. The *de novo* approach creates a novel genome without being guided by a reference, this method has been frequently used to discover new pathogens (Ekblom & Wolf, 2014). To use reference guided assembly genome data has to be available. There are currently 27 published whole genome records from different strains of *M. pluonius* on the National Center for Biotechnology Information (NCBI) database, ranging from 2012 to 2022 (NCBI, 2021). One record is identified as the reference genome, NCBI refSeq: GCF_003966875.1, which is the most complete genome and seen to be representative of the species (Okumura et al., 2011). Looking at the reference genome the whole genome is 2.1 Mb long, and this includes two plasmids pMP19 which are approximately 20,000 bp and pMP1 which is approximately 200,000 bp (Okumura et al., 2011). SNPs occur on both plasmid and chromosomal DNA, so a SNP profile can be generated from all three sources.

The overall aim of this chapter is to produce and analyse whole genome sequencing data from all positive buffer bottles from 2020 to generate higher resolution epidemiological data to better elucidate movements. Sequencing data will be generated using the pipeline developed in chapter 2. Data analysis will begin by recreating the MLST typing provided from the traditional Sanger sequencing. A bioinformatics pipeline will be created to generate a SNP profile on the chromosomal DNA and both plasmids from samples that pass the criteria developed. From the SNP profile phylogenetic trees will be created to assess relationship between samples. Tracing the source of the disease with the SNP data will also be tested.

3.2 Methods

3.2.1 Sample preparation

3.2.1.1 Sample Collection

LFD buffer bottles each containing a single larva from infected sites in 2020 were received from Fera Science Ltd via DEFRA. MLST typing for each sample was performed by Victoria Tomkies at the National Bee Unit (NBU) before being sent to Newcastle University. All samples came with MLST typing information, location, county, and

a beekeeper number to help link ownership of outbreaks. The samples that were successfully sequence typed were selected for downstream analysis, and some samples which failed the MLST were also included providing a total of 570 samples.

3.2.1.2 Host Depletion and DNA Extraction

For each selected sample 2 ml was transferred into an empty 2 ml tube. The aliquoted samples were centrifuged at 7,500 g for 10 minutes, the supernatant was removed, and pellets re-suspended in 20 µl DNase buffer I, 1 µl DNase I (10649890, Fisher Scientific), 180 µl distilled water, and incubated at 37°C for 60 minutes. To inactivate the DNase, 20 µl 50 mM EDTA was added, and the samples incubated at 65°C for 10 minutes. The samples were centrifuged at 10,000 g for 5 minutes and the pellet was re-suspended in 180 µl Gram-positive lysis buffer. The suspended samples were transferred into 96 well plates to increase throughput. The plates were incubated at 37°C for 30 minutes, then centrifuged at 20,000 g for 5 minutes and the supernatant was discarded. The pellets were re-suspended in 200 µl of PDB (5 mM guanidine HCl) and 25 µl proteinase K (Qiagen) was added and the plates were incubated at 56°C for 30 minutes. Then 200 µl 100% ethanol was added and the sample transferred to a cEZ-10 RNA Mini Spin Column and spun down at 6000 g for 1 minute. The flow through was discarded and 500 µl of wash buffer 1 (7 M Guanidine HCl, 56% ethanol) was added and the column spun down at 6000 g for 1 minute the flow through discarded, 500 µl of wash buffer 2 (70% ethanol, 10 mM Tris HCl) was added and spun down at 6000 g for 3 minutes the flow through was removed and a 1 minute spin was performed. The columns were transferred to a clean 96 well plate, 50 µl TE buffer was added directly to the column and then centrifuged at 10,000 g for 1 minute, this was repeated once. The DNA concentration was measured for each sample using the dsDNA quantification using Sybr Green I V.2 protocol (dx.doi.org/10.17504/protocols.io.b34gqqtw).

3.2.2 Oxford Nanopore MinION sequencing

The samples were run across six Minion runs and prepared for sequencing using the SQK-LSK110 (Oxford Nanopore) kit with the EXP-PBC096 (Oxford Nanopore) PCR barcoding kit. A modified Ligation sequencing gDNA - PCR barcoding protocol from Oxford Nanopore was used. Half the volume of all components was used to cut costs, up

until the end prep step. For the PCR step conditions, the number of cycles was 18 and the extension time was 8 minutes, with a final extension of 10 minutes. The samples were pooled in equal volumes, 2 µl of each sample and then 48 µl of the pool taken forward. The short fragment buffer was used for the final clean up. Based on the final pool concentration being low from the first 3 plates, for the last three plates after pooling the PCR products, a bead clean was performed to concentrate the samples. For the bead clean 400 µl of AMPure XP beads (A63881) were added to the pooled PCR products. The mixture was incubated for 5 minutes at room temperature, and the tube placed on a magnet to pellet the beads. Whilst on the magnet the supernatant was pipetted off and 1 ml 80% ethanol was added and then removed by pipetting, this step was repeated. The pellet was left to dry for 30 seconds, the tube removed from the magnet and the beads were resuspended in 50 µl nuclease free water. The beads were incubated for 2 minutes, and the tube was placed back onto the magnet. After the beads were pelleted the 50 µl of supernatant was used for further library preparation according to the protocol. The pooled library was loaded onto a R9.4.1 Minion flow cell (Oxford Nanopore, FLO-MIN106D) and run for 72 hrs using the MinKNOW Software (Oxford Nanopore). Within the MinKNOW software live, high accuracy basecalling was performed, converting electrical signal to nucleotides. Each sample was barcoded with a unique sequence and the samples were separated using live barcoding through the software, as well as demultiplexing, removing all the adaptors. The MinKNOW software also has a built-in quality control (QC) step that anything with a Phred score of less than 7 is placed in a failed folder. The final output was raw reads in FASTQ format for each sample.

3.2.3 Bioinformatics

3.2.3.1 Initial analysis

All packages used for bioinformatics were through the Bioconda package manager (Grüning et al., 2018). Barcoding was performed live during sequencing and the reads placed into folders matching the barcode, with those that do not match any of the barcodes are placed into an unclassified folder. Initially porechop (Version:0.2.4) (Wick et al., 2017) was used to detect missed barcodes in the unclassified files. A reference guided assembly was carried out using Minimap2 (Version:2.24) the raw reads for each sample were mapped to the reference genome, and two plasmids individually, Table 3.2

(Li, 2018). A reference guided assembly was performed opposed to a *de novo* assembly because the samples were not pure *M. plutonius* and the average sequencing depth was too low. Using Samtools (Version:1.16.1) the mapped reads were converted to binary alignment map files (BAM) required for downstream analysis.

Table 3.2: Details of the samples used in the reference guided assembly for each sample.

Genome	Genbank ID	Size (Mb)
Melissococcus plutonius DAT561 DNA	AP018492.1	1.80
pMP1	AP018493.1	0.20
pMP19	AP018494.1	0.02

3.2.3.2 Overview of sample selection and SNP analysis

To ensure as many samples are included as possible in the final analysis a pipeline was developed (Figure 3.1). This pipeline ensures that only the samples that I can be confident with are included.

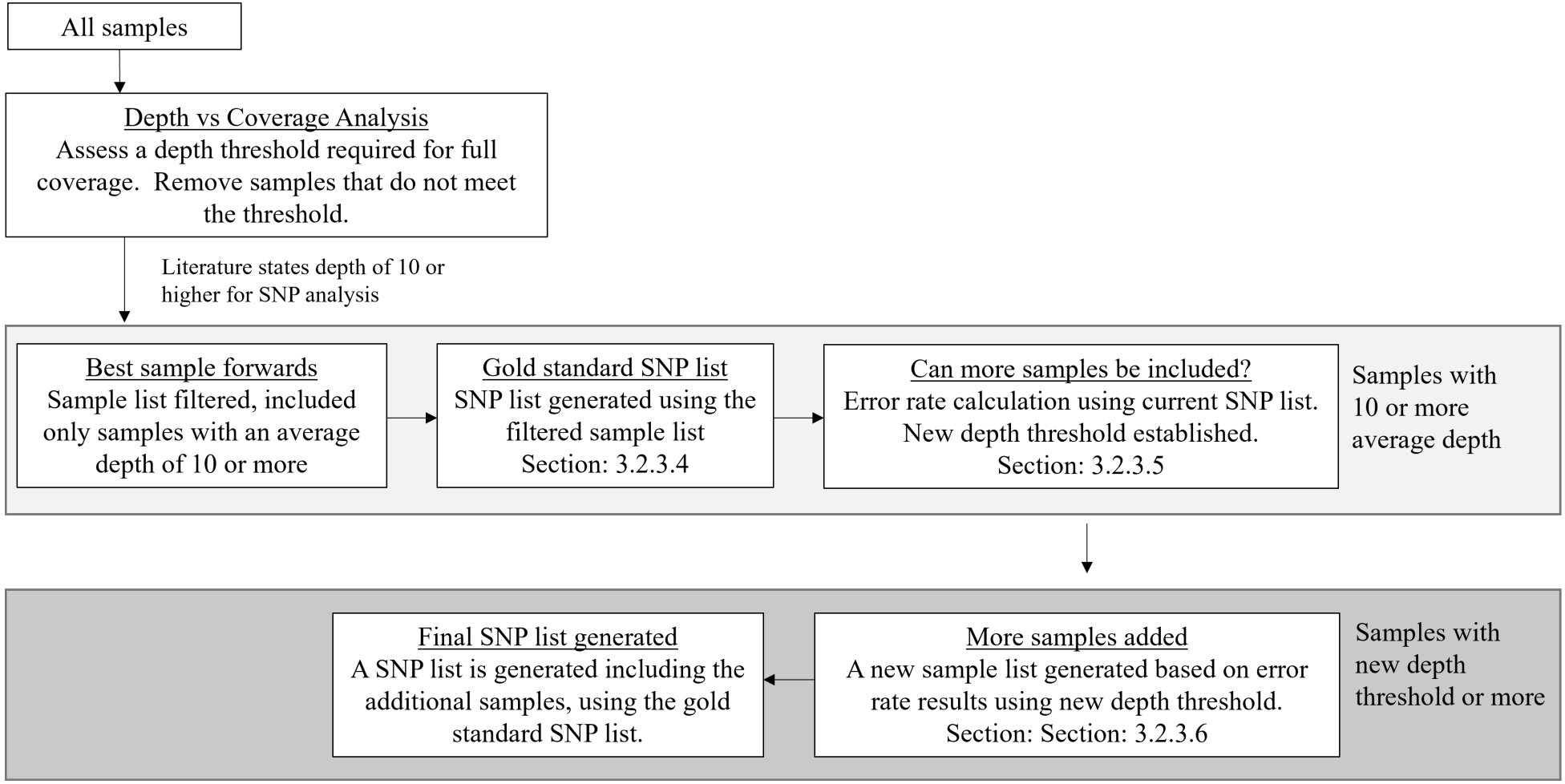


Figure 3.1: Overview of the workflow used to select samples and generate SNPs

3.2.3.3 Depth vs coverage analysis

Using the BAM files generated in 3.2.3.1 the average genome coverage of the reads to the references, and the average depth of the sequencing was calculated using Samtools coverage (Version:1.16.1) (Danecek et al., 2021). These results were plotted against each other to assess the depth required for full genome coverage. The inflection point was calculated to determine a depth threshold that enables sufficient coverage and samples that did not reach this threshold were removed from further analysis to avoid poor quality data from pervading the SNP data.

3.2.3.4 SNP Generation

Nanocaller (Version:3.0.1) was used to generate SNPs for each of the selected samples against the reference genome for *M. plutonius* (Ahsan et al., 2021), using the setting for all, covering SNPs, insertions and deletions. Only the SNP file was used, as I was only concentrating on SNPs for this study. The software was initially designed for diploid genomes therefore SNPs were assigned either homopolymers or heteropolymers, using R the heteropolymers were filtered out leaving only homopolymer SNPs. A list of SNPs was generated for each sample. From the literature, SNPs can be accurately called at a depth of 10, so the sample list was filtered to include only samples with an average depth of 10 or higher (Jiang et al., 2019). The SNP files for the chosen samples were combined using bcftools (Version:1.9) (Danecek et al., 2021) to generate a core SNP list. Using the BAM files generated in 3.2.3.1 the depth of each nucleotide across the whole genome was calculated using Samtools depth for each sample (Danecek et al., 2021). Using the depth information, each position in the core SNP list was assigned a depth for each sample, to ensure sufficient depth across all samples, as the sample list was selected on average depth. The SNPs were filtered at different depths from 1 to 10, so if any sample had a depth of less than the filter at any specific SNP the SNP would be removed from the core list. The number of SNPs at each filtered depth was recorded. The number of SNPs and minimum depth were plotted on a graph to calculate the optimum depth. The optimum depth was calculated from the inflection point of the plot. The depth selected was used for further analysis. The SNP analysis was performed on the two plasmids, pMP19 and pMP1 using Nanocaller and the chosen depth filter from the *M. plutonius* was used, the SNPs at this depth generated the core SNP list.

3.2.3.5 Error rate calculation

Using the *M. plutonius* BAM files generated in 3.2.3.1 for the chosen samples that average above depth 10 and a list of the positions of the chosen SNPs generated in 3.2.3.4 a mpileup file was generated using Samtools mpileup. The mpileup file generated contains all the reads at each SNP position. In R (Version: R 4.2.1) using the mpileup files for each sample the number of each nucleotide at each SNP position was calculated. The probability of correctly calling a SNP was calculated with increasing number of reads, and the percentage correctly called, and number of reads was plotted. This was also performed per nucleotide to see if any of them are more error prone with this technology. From this a depth cut off was decided based on the results, this was used for sample selection further down the pipeline.

3.2.3.6 Sample Selection

Based on the results from the error rate calculation a lower SNP depth was assessed to try and include more samples. All samples, excluding those that did not meet the depth threshold for sufficient coverage in 3.2.3.3. The depth at each of genome positions present in the core SNP list for each sample were obtained from the previous Samtools results. The positions with a depth of the determined cut off or more were marked with a 1 and those with less than the cut off a 0. The sum across all SNP positions for each sample was calculated, and in a loop the sample with the lowest sum was removed, the SNP number recorded until there was only one sample left. The number of samples and SNPS were plotted, and the inflection point gave the chosen sample list to use.

3.2.3.7 Assessment of genomes

The whole pipeline was also performed on the two plasmids pmp1 and pmp19 (Figure 3.1). Using the core SNP list generated for *M. plutonius*, a binary matrix was created showing absence and presence of each SNP across all chosen samples. Using the matrix, a phylogenetic tree was created using the Neighbourhood-Joining Tree Estimation method and bootstrapping in the Ape package (Version:5.7) (Paradis et al., 2023) with graphics through ggtree (Version:3.6.2) (Yu, 2020). This was repeated for the two plasmids. The trees were used as a preliminary assessment to assess how informative the groups were, and which ones to take forward to final analysis.

3.2.3.8 MLST Typing

The MLST files for each of the genes in the scheme, purR, argE, galK and gbpB, were downloaded from pubMLST (Jolley et al., 2018). The files contain the sequence for each variant within each gene in fasta format. The sequencing reads from each of the samples on the final sample list after sample selection and SNP analysis were mapped to each of the fasta files using Samtools (Version:1.16.1). The average genome coverage of the reads for each gene, and the average depth were calculated using Samtools coverage (Danecek et al., 2021). The allele for each locus was compared to the results obtained from the traditional MLST typing, and the percentage matched was calculated.

3.2.3.9 Map generation

A map of England and Wales was generated to include all the positive EFB cases in 2020 and symbolised by ST type. The map was generated in R (Version: R 4.2.1) using the sf package (Version:1.0-15) (Pebesma, 2023) and ggrepel package (Version:0.9.4) (Slowikowski, 2023) . A second map was generated showing same information but using only the selected samples from the pipeline to see how representative the analysis is of disease cover.

3.2.3.10 Tree generation

Initially a neighbourhood joining tree was create using the R package Ape (Version:5.7) (Paradis et al., 2023) for the traditional MLST method performed at Fera Science Ltd, using the four genes. Using the core SNP list generated for the chosen sample list for *M. plutonius* a SNP matrix was created showing each base for each sample at each SNP position. Using the matrix, a maximum likelihood tree was created using phyml (Version:3.3.20220408) with bootstrapping set at 100. The tree was visualised using ggtree (Yu, 2020; Guindon et al., 2005). Another tree was generated using the same method but removing samples that formed an outgroup. Individual phylogenetic trees were generated for ST2, ST3, ST6, ST7, ST23 and ST39 to assess in further detail if individual groups form within the sequence type. To generate these trees for each sequence type the list of SNPs were filtered to remove redundant SNPs and the total number recorded.

3.2.3.11 Update of Nanocaller

A haploid option was introduced for Nanocaller in a later update (Version:3.3.0) (Ahsan et al., 2021), the Nanocaller step was repeated using this new option, the filtering of homopolymer step was not required. The new SNP data was run through the same steps in 3.2.3.4, 3.2.3.5 and 3.2.3.6. A maximum likelihood tree was created using the same method as 3.2.3.10.

3.2.3.12 Grapetree

The SNP matrix was used to create two minimum spanning trees using Grapetree. Grapetree uses a novel algorithm to generate a minimum spanning tree to analyse genetic relationships between multiple samples (Zhou et al., 2018). The two trees were identical in genetics, but one was labelled using sequence type and the other by county to assess groupings. Inkscape was used to edit the tree so the ST colours match previous work (Project, 2020).

3.3 Results

3.3.1 Initial analysis

To assess the average depth required for a high genome covered the average depth and the percentage covered was plotted (Figure 3.2) and the inflection point measured. The inflection point was calculated to be 2, this was the depth required to obtain a high coverage of the *M. plutonius* genome using this sequencing pipeline. The total number of samples that passed this threshold was 348, the rest were removed from further analysis.

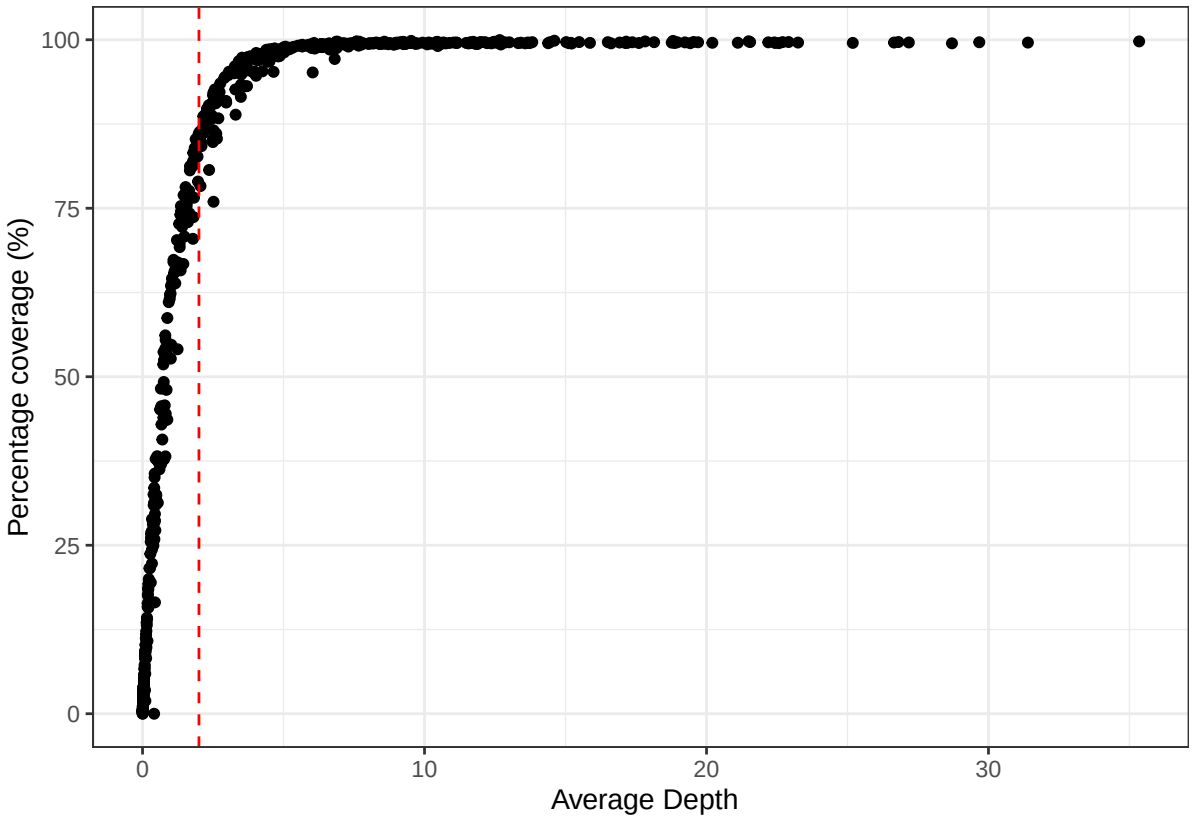


Figure 3.2: The average depth of the sequencing reads mapped to the *Melissococcus plutonius* genome across all nucleotides for each sample, compared to the percentage coverage of the *Melissococcus plutonius* genome. The red dotted line represents the inflection point of the plot, the depth threshold for a high coverage.

3.3.2 Depth analysis

The number of SNPs found at each depth was recorded and compared (Figure 3.3). The inflection point was calculated to choose the depth when the SNPs stabilise to reduce noise this was depth 5 with a SNP number of 1775.

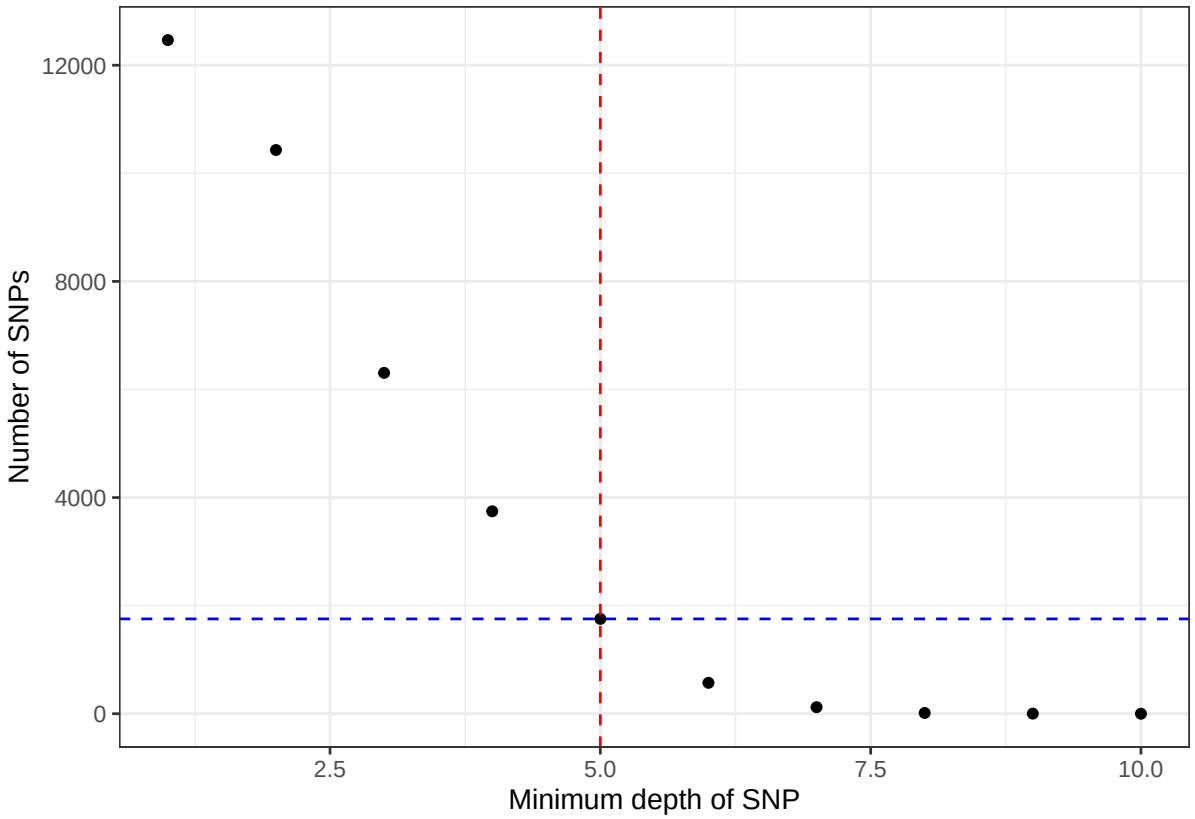


Figure 3.3: The number of SNPs compared to the minimum depth of all the SNPs called. The blue and red dashed lines represent the point of inflection, therefore the chosen SNP number and depth used for future analysis.

3.3.3 Error calculation

The proportion of SNPs being called correctly and the number of reads were compared for all nucleotides combinations and for individual nucleotides (Figure 3.4). For all nucleotides the proportion correctly called never dropped below 0.985 even at the low number of reads. For the individual nucleotides all of them had a correct classification rate of more than 0.98, showing no difference across each nucleotide. The error rate was low allowing the depth to be lowered without a reduction in data quality.

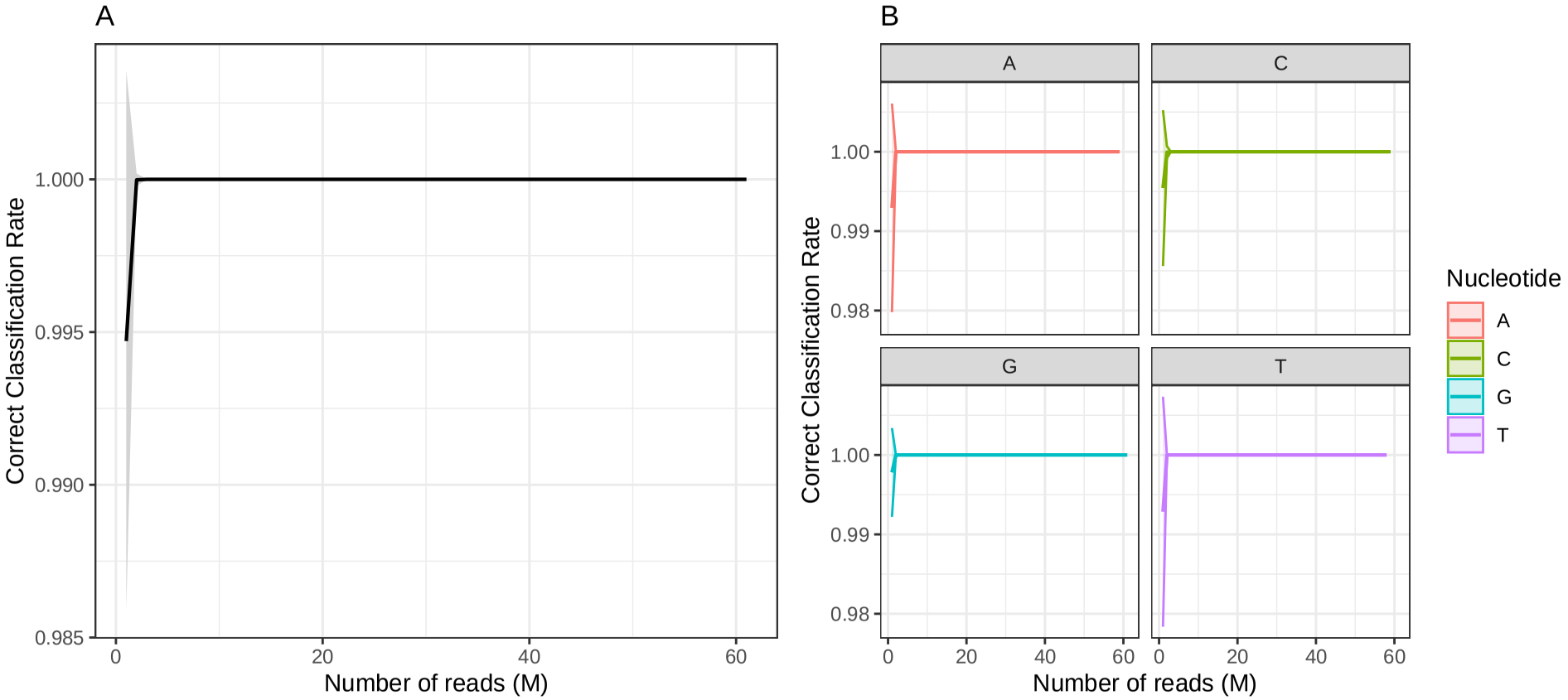


Figure 3.4: (A): The percentage successful call across all nucleotides (B): The percentage successful call across each nucleotide

3.3.4 Sample selection

Using the 1775 SNPs selected in 3.3.2 and the knowledge of a low error rate in 3.3.3, SNPs with a depth of 3 or more were included. The sample list has only included samples with an average depth of 10 higher up until this point, but the selected SNPs may reach the minimum depth of 3 in some of the other samples. All samples with an average depth of 2 or higher as shown in 3.3.1 that have a good coverage of genome were selected for further analysis. The number of SNP positions that had a depth of 3 or higher were recorded everytime the sample with the least matches was removed and were plotted (Figure 3.5). The data shows an inflection point at 136 samples and 1407 SNPs, this optimises the balance of SNPs and number of samples.

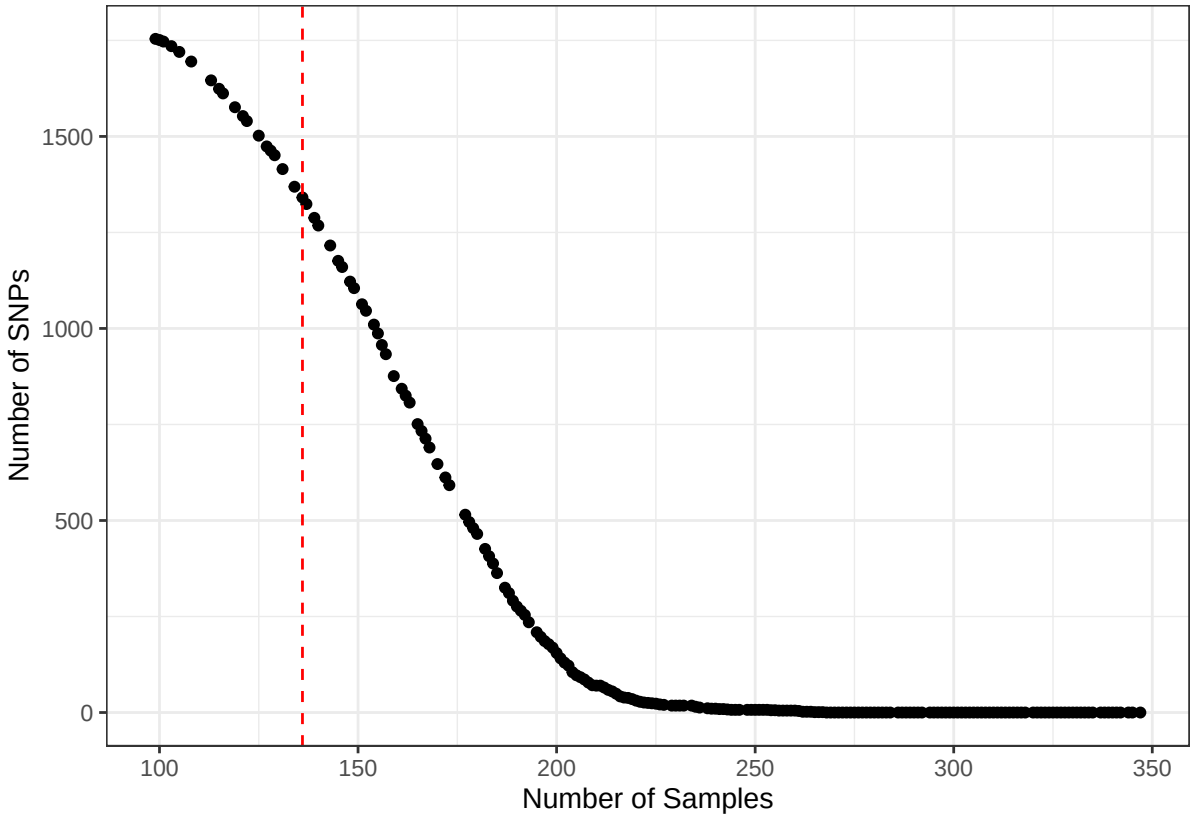


Figure 3.5: The number of SNPs with a depth of 3 or higher compared to the number of samples. The red line represents the point of inflection, 136 samples and 1407 SNPs.

3.3.5 Plasmid assessment

A neighbourhood joining tree was generated for the *M. plutonius* genome and the two plasmids using the SNP matrix. For pMP19 and pMP1 no clear groups apart from an outgroup at ST39 were shown in the trees, suggesting the SNP profile was not informative.

For *M. plutonius* groups were forming outside of the clear ST39 group (Figure B.1 and B.2). Only the *M. plutonius* SNP profile was used for final analysis.

3.3.6 MLST typing

Using the MLST scheme each sample was assigned a variant for each of the four genes. These results were compared to the results obtained from the traditional MLST scheme testing performed at Fera Science Ltd and were reported (Table 3.3). For three out of four of the genes over 90% of the variants matched the traditional testing, with around 10% of these presenting more than one variant with a 100% match. For galK only 50% matched with the traditional MLST scheme, and 34% of these had more than one variant with 100% match. The biggest difference for the galK call was that variant 3 in the traditional MLST scheme was called as variant 8 using whole genome sequencing for most of the unmatched samples. For all downstream analysis the traditional MLST scheme data was used to assign sequence types.

Table 3.3: The percentage of samples that matched each gene in the MLST scheme with the traditional MLST scheme result performed at Fera Science Ltd, and those that did not.

Gene	Percentage samples matched (%)	Percentage of matched with muliple calls	Percentage of samples not matched	Number of failed/unknown samples with the tradtional MLST
argE	96.21	10.60	3.79	4
galK	17.70	32.28	82.30	6
gbpB	100.00	7.58	0.00	4
purR	98.47	6.11	1.53	5

3.3.7 Maps

The maps of all the 2020 samples compared to the selected samples (Figure 3.6), showed that there was still country wide spread represented, as most of the same counties

were covered. There are some patches in the selected samples, the ST13 outbreak in Nottinghamshire and an ST2 cluster in Dorset is not represented.

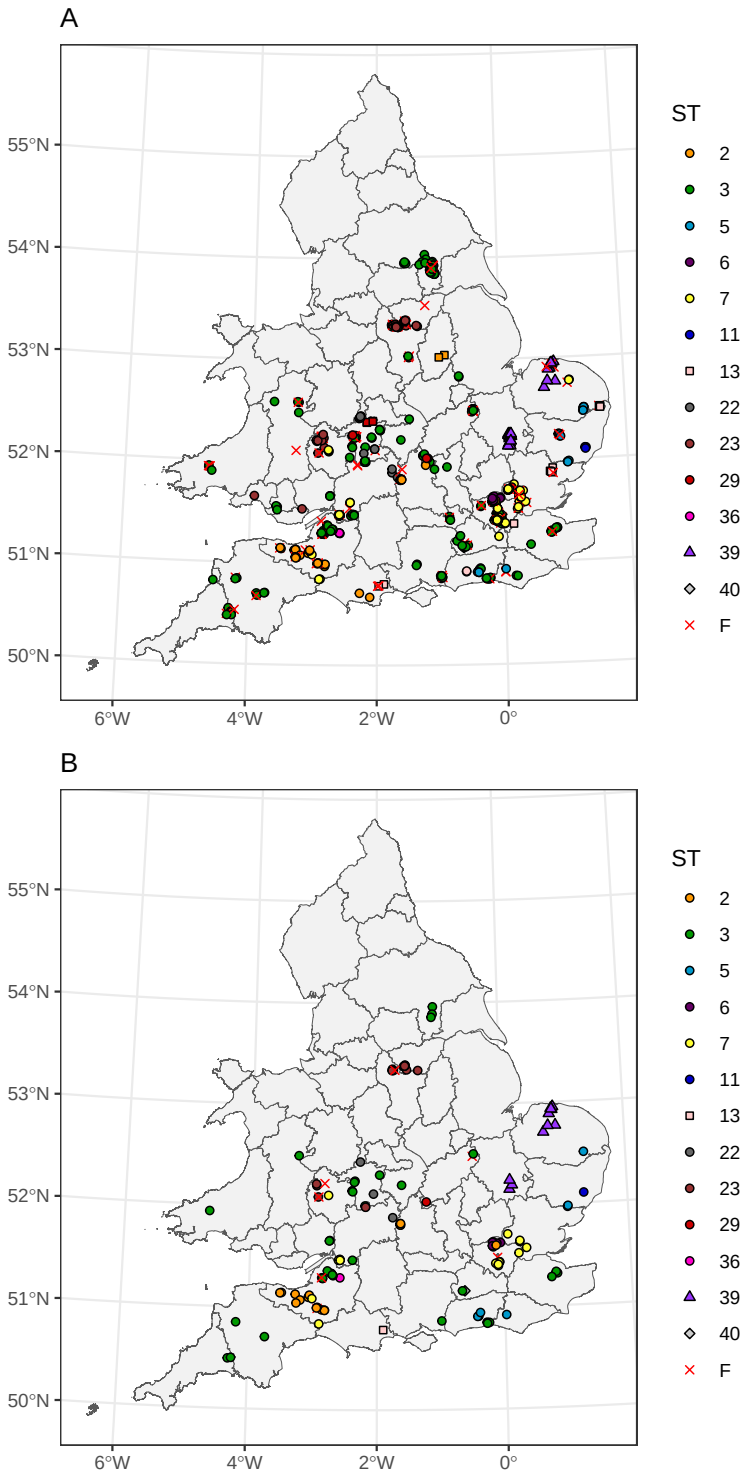


Figure 3.6: A) Map of England and Wales with all the positive EFB samples in 2020, separated by sequence type. B) Map of England and Wales with the positive EFB samples in 2020 that were included in the final analysis, separated by sequence type. The shape of the symbols represent the clonal complex. Circle: Clonal Complex 3. Diamond: Clonal Complex 12. Square: Clonal Complex 13. Triangle: Predicted as a new Clonal Complex

3.3.8 Tree generation

3.3.8.1 Tree using MLST data

A Neighbourhood joining tree was created of the traditional MLST method using the four genes in the MLST scheme (Figure 3.7). The tree shows the samples separated into 13 groups. The biggest group is ST3 and that contains samples from more than 10 counties, showing that geographical outbreaks were not identified. For all trees produced the BK number is a randomly assigned number to each beekeeper for anonymity.

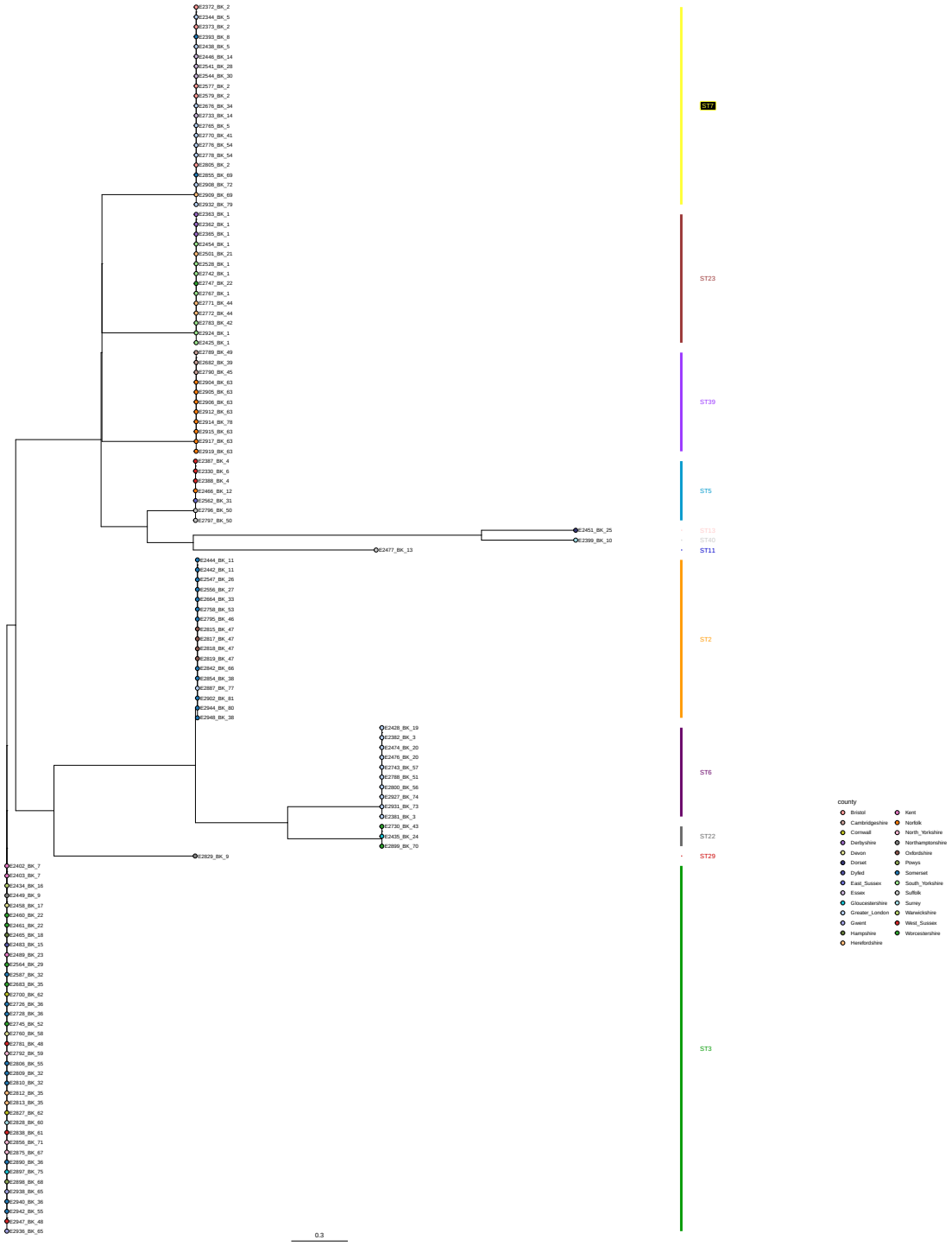


Figure 3.7: A neighbourhood joining tree created using the four gene MLST scheme results for the samples chosen for analysis minus the failed and unknown samples as full MLST data was not avaliable. The E number is the unique identifier for the samples, BK represents beekeeper number. The scale bar represents the phylogenetic distance.

3.3.8.2 Tree using SNP data

A maximum likelihood tree was created of the chosen 136 samples and 1407 SNPs using the Nanocaller without the new haploid function (Figure 3.8). The tree shows that clear groups have formed based on the SNP profile chosen throughout this analysis. Groups have occurred within the sequence types to separte them out. ST3 appeared to have 8 distinct groups, with 2 separate clusters for Somerset. An outgroup has formed of ST39 showing it was genetically different from the other samples. The ST13 and ST40 sample were also very genetically different from the other STs. For ST7 there appeared to be a separate cluster for Essex, Bristol and Greater London. The samples for ST23 appeared to form 2 distinct clusters, one for Herefordshire and another for South Yorkshire with Derbyshire. For ST5 there were 2 clear clusters, a West Sussex cluster and a Suffolk with Norfolk cluster. It appeared that there were three ST2 clusters in Somerset and one in Oxfordshire. There only appeared to be one cluster of ST6 with all samples from Greater London. The outgroups were removed to give clarity to the other groups (Figure 3.9).

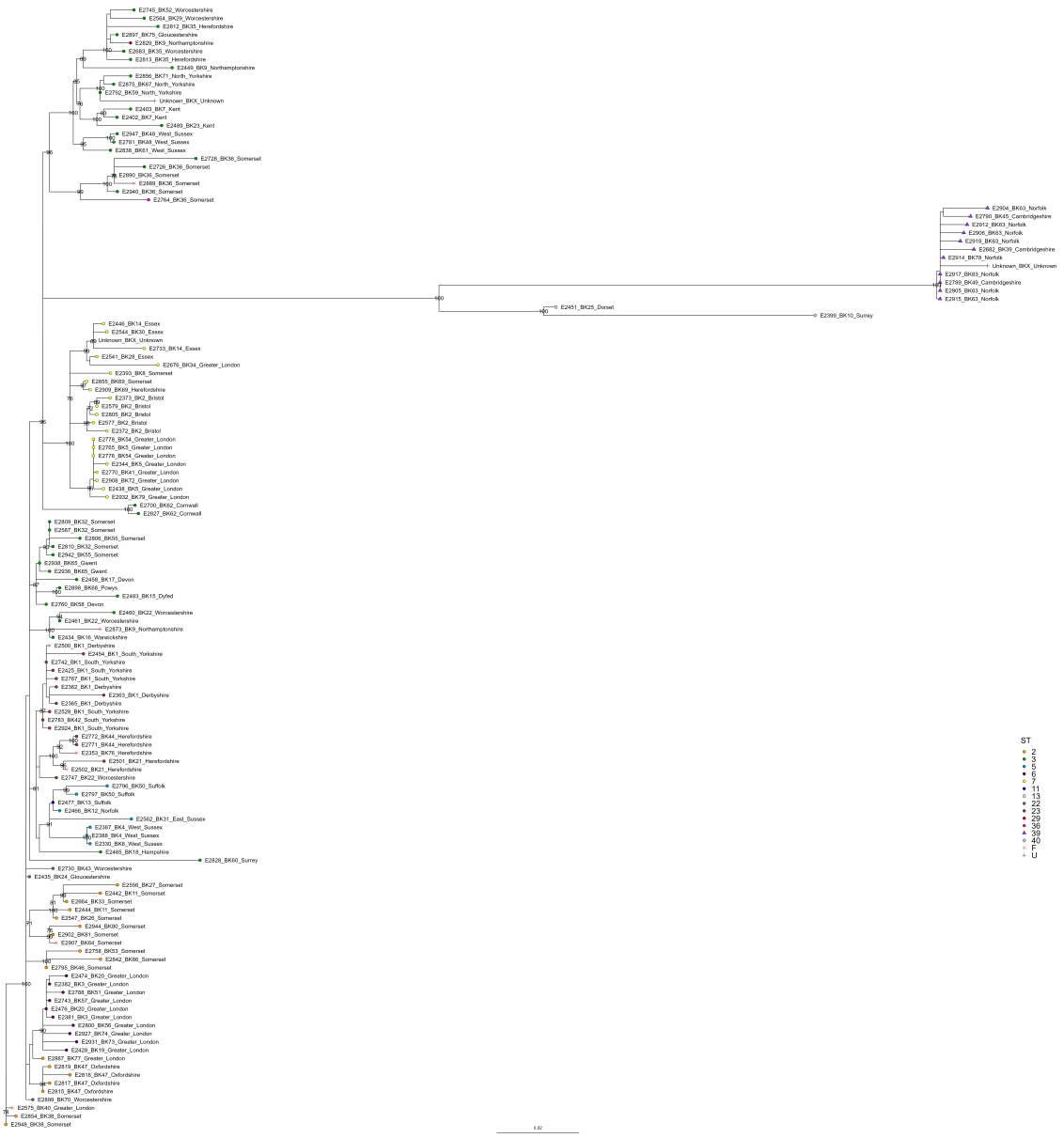


Figure 3.8: A maximum likelihood tree representing all the chosen 136 samples and using the 1407 SNP profile generated using Nanocaller with no haploid option. The E number is the unique identifier for the samples, BK represents beekeeper number. The scale bar represents the phylogenetic distance. The numbers on the branches represent the bootstrap percentage after 100 replications. The shape of the symbols represent the clonal complex. Circle: Clonal Complex 3. Diamond: Clonal Complex 12. Square: Clonal Complex 13. Triangle: Predicted as a new Clonal Complex

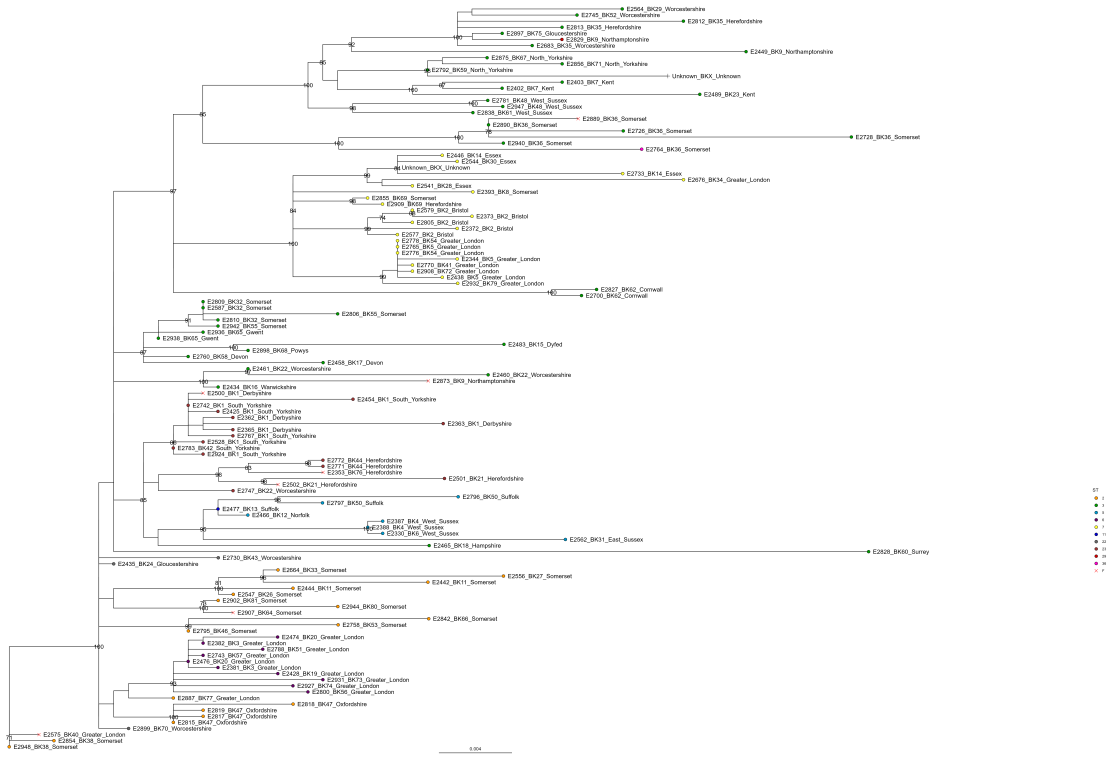


Figure 3.9: A maximum likelihood tree representing all the chosen 136 samples and using the 1407 SNP profile generated using Nanocaller with no haploid option, with the ST39, ST13 and ST40 removed. The E number is the unique identifier for the samples, BK represents beekeeper number. The scale bar represents the phylogenetic distance. The numbers on the branches represent the bootstrap percentage after 100 replications. The shape of the symbols represent the clonal complex. Circle: Clonal Complex 3. Diamond: Clonal Complex 12. Square: Clonal Complex 13. Triangle: Predicted as a new Clonal Complex

A maximum likelihood tree was created of the chosen 138 samples and 1227 SNPs using the Nanocaller with the new haploid function (Figure 3.10). There was no clear clustering of sequence types, and a lot of chaining of branches. This analysis was not taken forward.

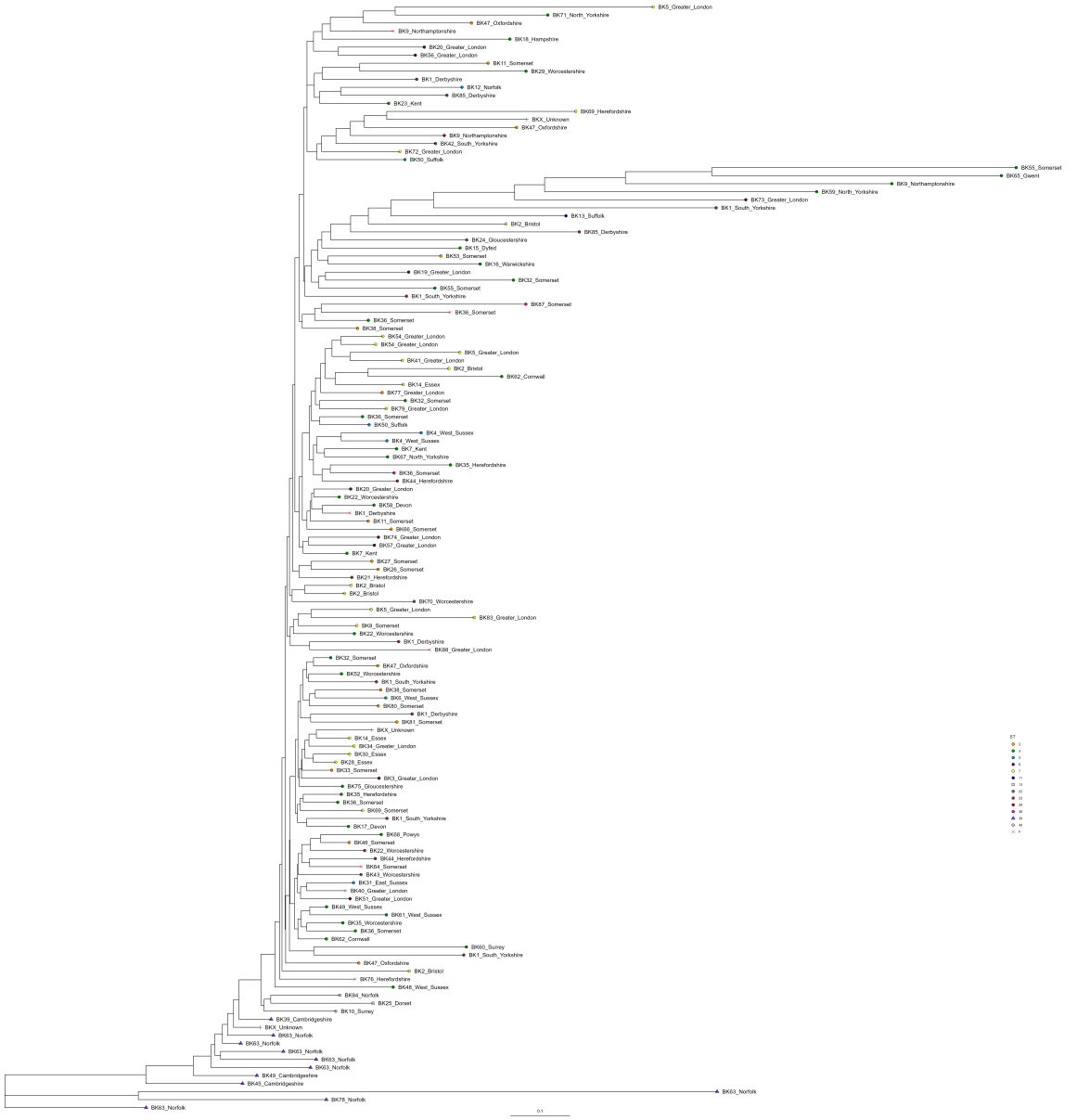


Figure 3.10: A maximum likelihood tree representing all the chosen 138 samples and using the 1227 SNP profile generated using Nanocaller with the haploid option. The E number is the unique identifier for the samples, BK represents beekeeper number. The scale bar represents the phylogenetic distance. The shape of the symbols represent the clonal complex. Circle: Clonal Complex 3. Diamond: Clonal Complex 12. Square: Clonal Complex 13. Triangle: Predicted as a new Clonal Complex

3.3.8.3 Sequence type 2

A tree was created for just the samples that were assigned sequence type 2 in the traditional MLST scheme (Figure 3.11). This showed clear clades within the ST with bootstrap of 90 or higher. There was a clear Oxfordshire cluster, that shares the same ancestor as the Greater London sample. There were three distinct Somerset clusters.

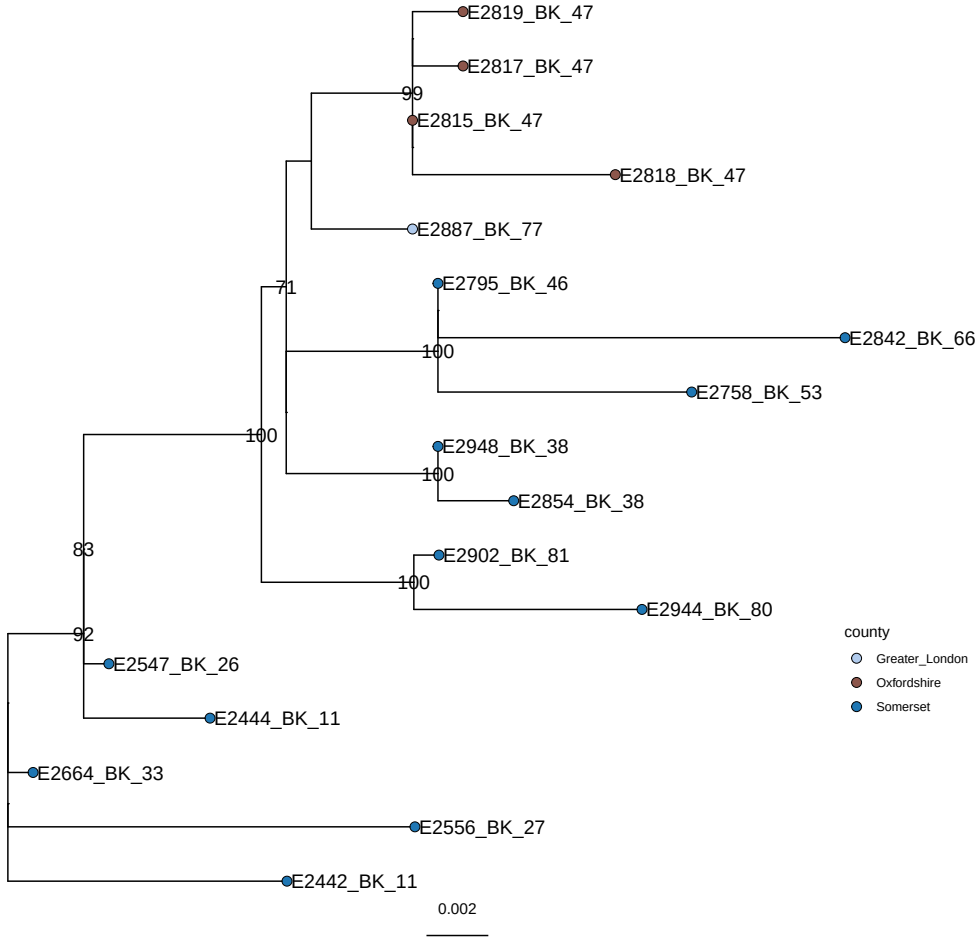


Figure 3.11: A maximum likelihood tree containing only samples that were assigned ST2 from traditional MLST typing using the 1407 SNP profile generated using Nanocaller with no haploid option. The E number is the unique identifier for the samples, BK represents beekeeper number. The scale bar represents the phylogenetic distance. The numbers on the branches represent the bootstrap percentage after 100 replications.

3.3.8.4 Sequence type 3

A tree was created for just the samples that were assigned ST3, the most common, in the traditional MLST scheme (Figure 3.12). This showed five clear clades within the ST with bootstrap of 90 or higher. The appeared to be two separate Somerset groups. The

Somerset group at the top shares a common ancestor with samples from Gwent, as well as a group formed of 2 Welsh counties Dyfed and Powys, and two samples from Devon. The samples from Cornwall formed their own clade. All the samples from beekeeper 36 have formed a clade, that share a common ancestor with all the samples above it on the tree. Samples from Worcestershire, Herefordshire and Gloucestershire have grouped together, with a sample from Northamptonshire sharing a common ancestor. Samples from North Yorkshire grouped together as well as those from Kent.

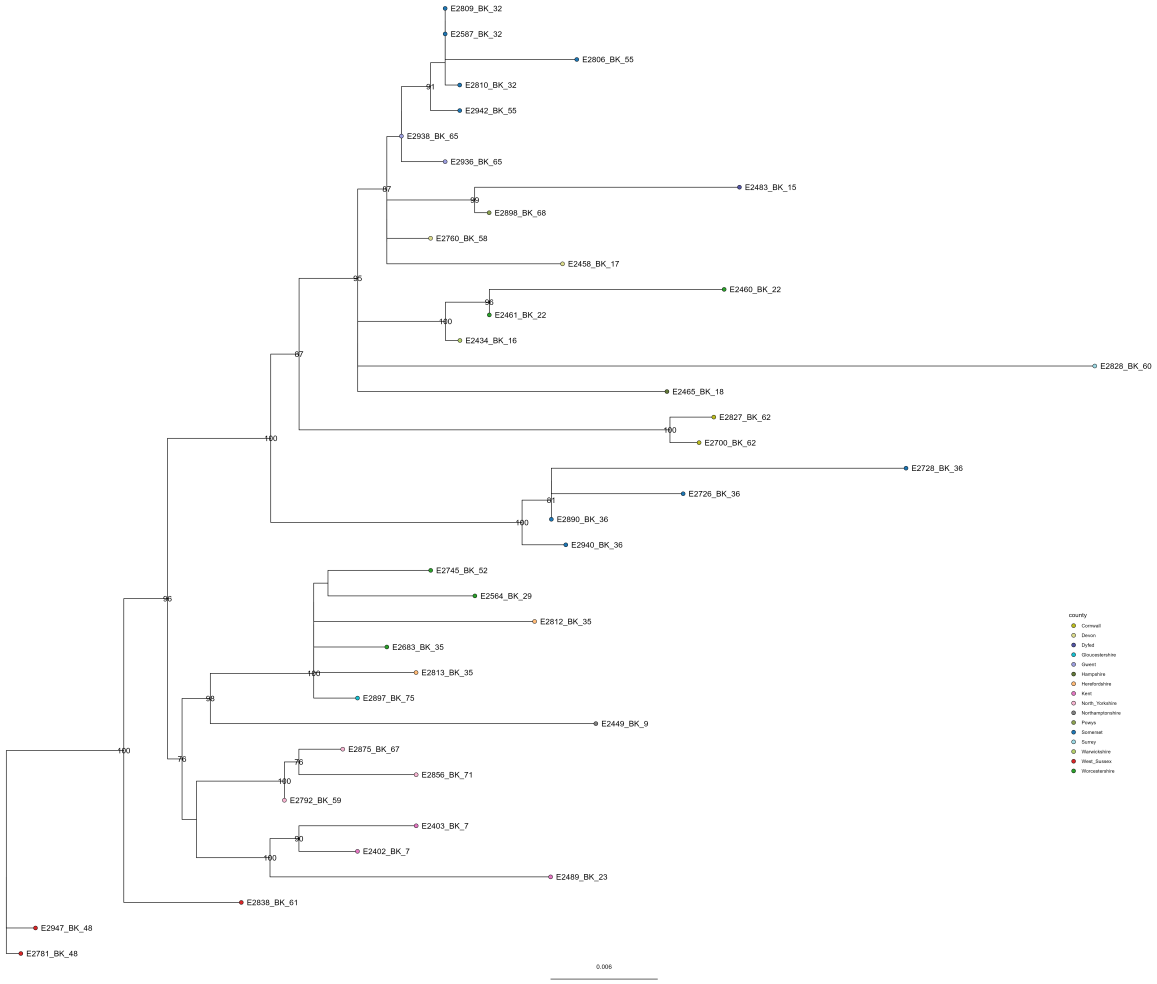


Figure 3.12: A maximum likelihood tree containing only samples that were assigned ST3 from traditional MLST typing using the 1407 SNP profile generated using Nanocaller with no haploid option. The E number is the unique identifier for the samples, BK represents beekeeper number. The scale bar represents the phylogenetic distance. The numbers on the branches represent the bootstrap percentage after 100 replications.

3.3.8.5 Sequence type 6

A tree was created for just the samples that were assigned ST 6 in the traditional MLST scheme (Figure 3.13). There were no clear clades within ST 6.

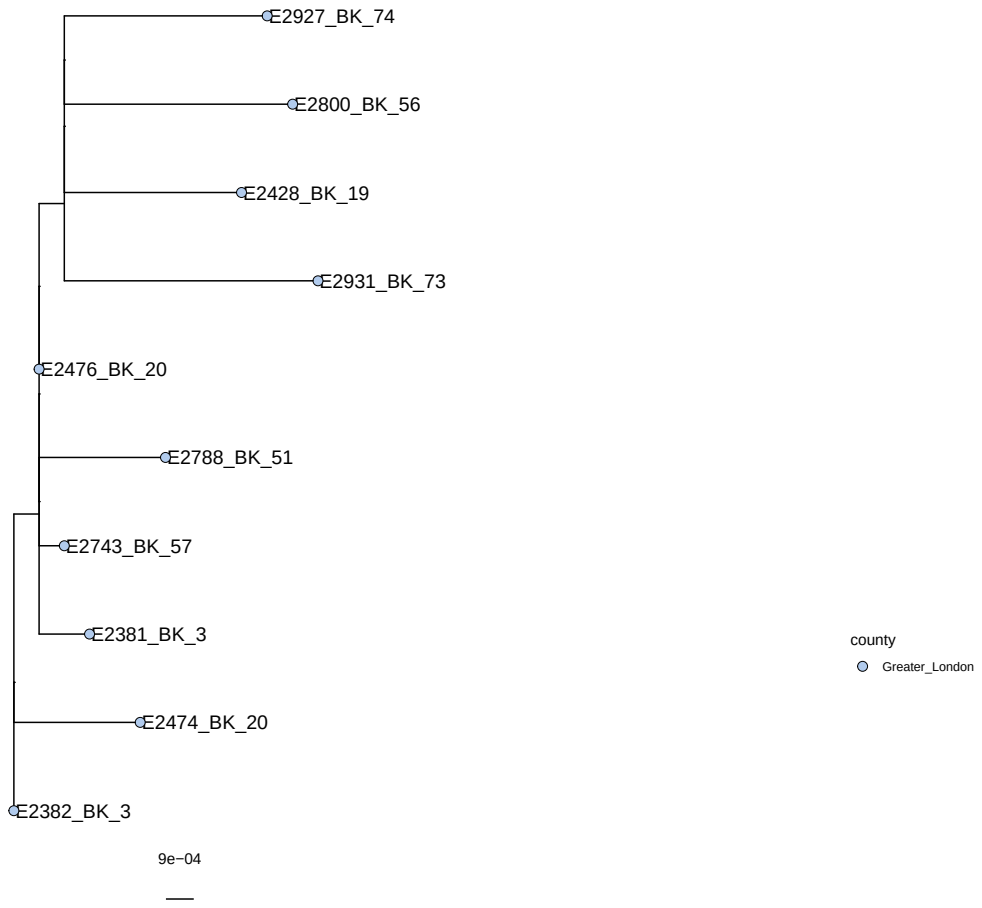


Figure 3.13: A maximum likelihood tree containing only samples that were assigned ST6 from traditional MLST typing using the 1407 SNP profile generated using Nanocaller with no haploid option. The E number is the unique identifier for the samples, BK represents beekeeper number. The scale bar represents the phylogenetic distance. The numbers on the branches represent the bootstrap percentage after 100 replications.

3.3.8.6 Sequence type 7

A tree was created for just the samples that were assigned ST7 in the traditional MLST scheme (Figure 3.14). There are clear groupings within the ST7 samples. There appeared to be an Essex, Greater London and Bristol cluster. There is also a cluster for beekeeper 69, with samples from both Somerset and Herefordshire.

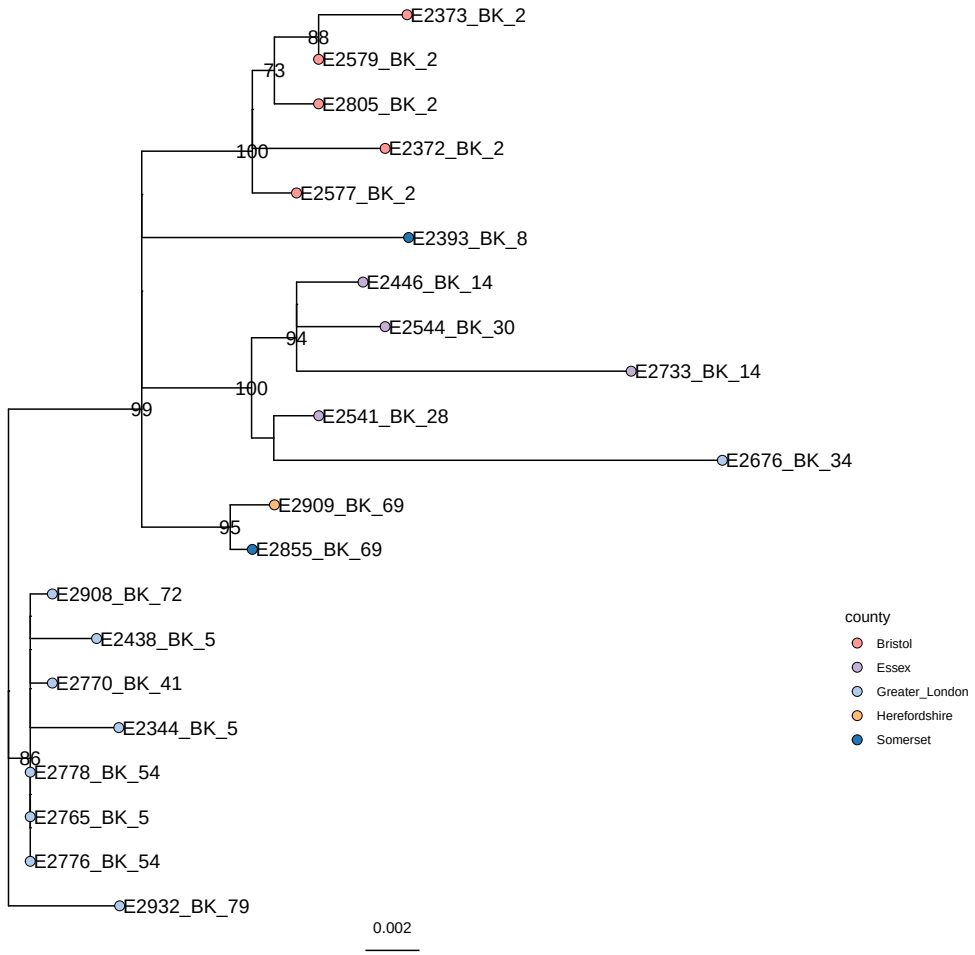


Figure 3.14: A maximum likelihood tree containing only samples that were assigned ST7 from traditional MLST typing using the 1407 SNP profile generated using Nanocaller with no haploid option. The E number is the unique identifier for the samples, BK represents beekeeper number. The scale bar represents the phylogenetic distance. The numbers on the branches represent the bootstrap percentage after 100 replications.

3.3.8.7 Sequence type 23

A tree was created for just the samples that were assigned ST23 in the traditional MLST scheme (Figure 3.15). The samples from beekeeper 1 share a common ancestor with beekeeper 21 and 22.

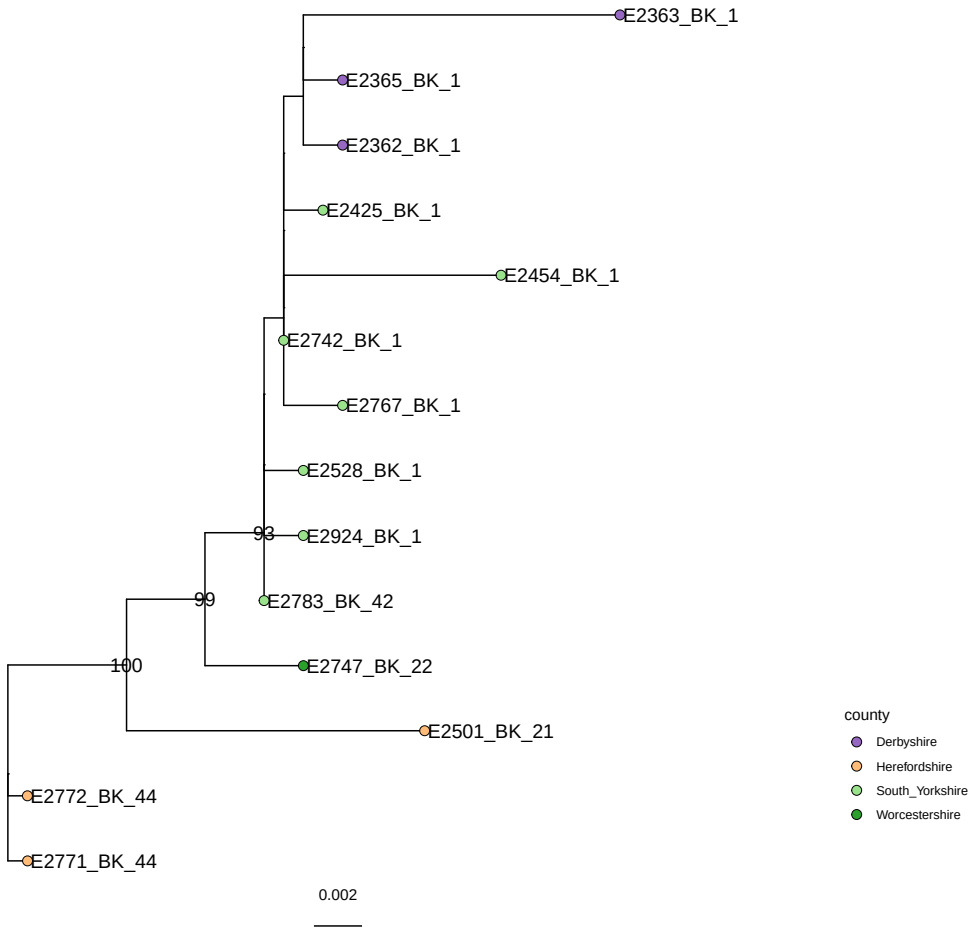


Figure 3.15: A maximum likelihood tree containing only samples that were assigned ST23 from traditional MLST typing using the 1407 SNP profile generated using Nanocaller with no haploid option. The E number is the unique identifier for the samples, BK represents beekeeper number. The scale bar represents the phylogenetic distance. The numbers on the branches represent the bootstrap percentage after 100 replications.

3.3.8.8 Sequence type 39

A tree was created for just the samples that were assigned ST39 in the traditional MLST scheme (Figure 3.16). There are no clear clades within ST39.

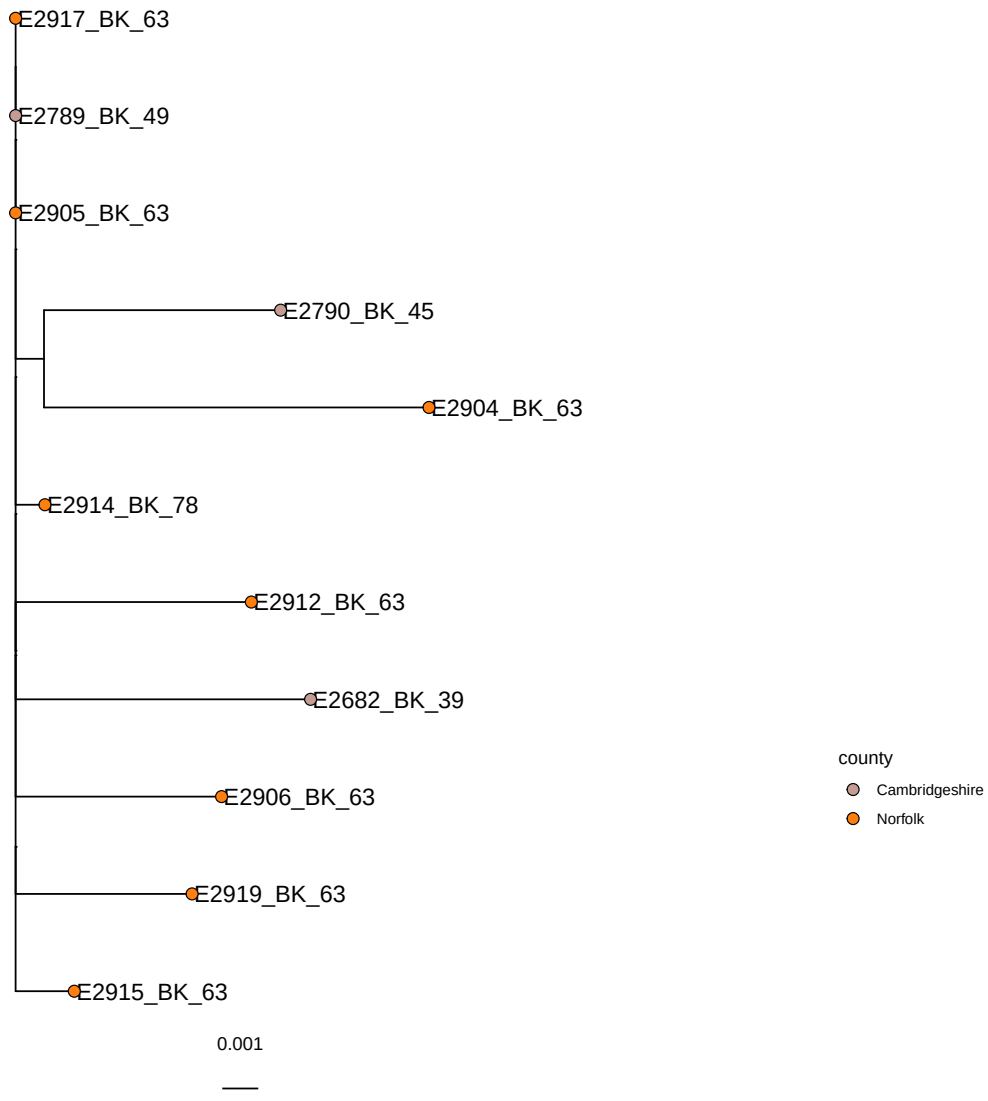


Figure 3.16: A maximum likelihood tree containing only samples that were assigned ST39 from traditional MLST typing using the 1407 SNP profile generated using Nanocaller with no haploid option. The E number is the unique identifier for the samples, BK represents beekeeper number. The scale bar represents the phylogenetic distance. The numbers on the branches represent the bootstrap percentage after 100 replications.

3.3.9 Minimum spanning tree

A minimum spanning tree was created using Grapetree for the chosen 136 samples using the 1407 SNPs using ST to colour (Figure 3.17). The same tree was created but using

county to colour the points (Figure 3.18). There appeared to be three clear clusters, one for ST39, including samples from both Cambridge and Norfolk, one with a central point of ST22 and another with the centre being ST23. In the the ST22 cluster, samples from ST3, ST6 and ST2 were present. In the ST23 samples from ST3, ST11, ST5, ST7, ST36 and ST29 were present. From the tree it appeared that ST5 evolved from ST11 and ST29 from ST3. The ST13 sample had 146 differences in SNPs to the ST22 cluster and from that the ST40 sample had 81 differences. Some samples were grouped by county, some Somerset clusters and London clusters.

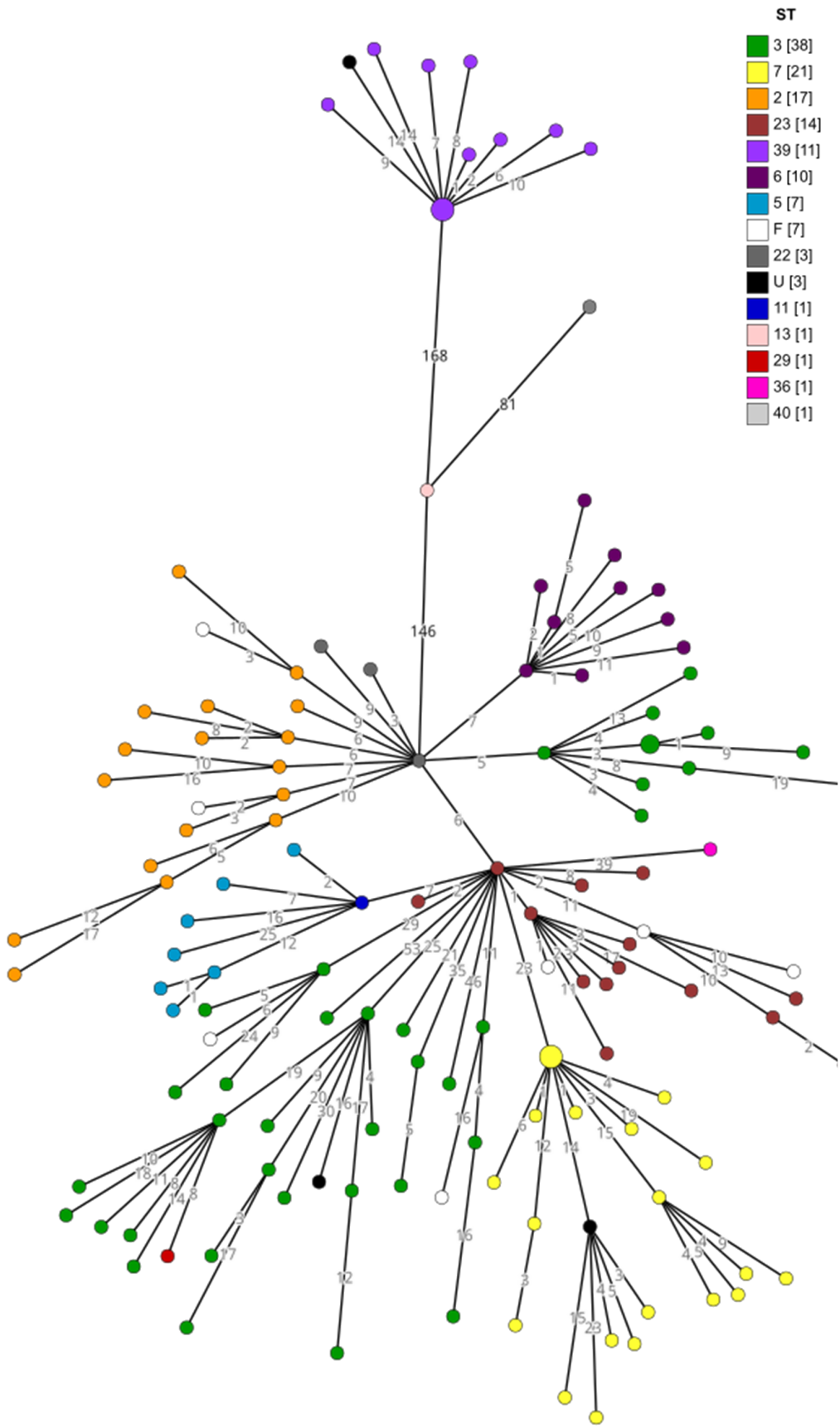


Figure 3.17: A minimum spanning tree created by GrapeTree software. Coloured by sequence type. The tree is on a log scale. The numbers on the branches represent the number of SNP differences.

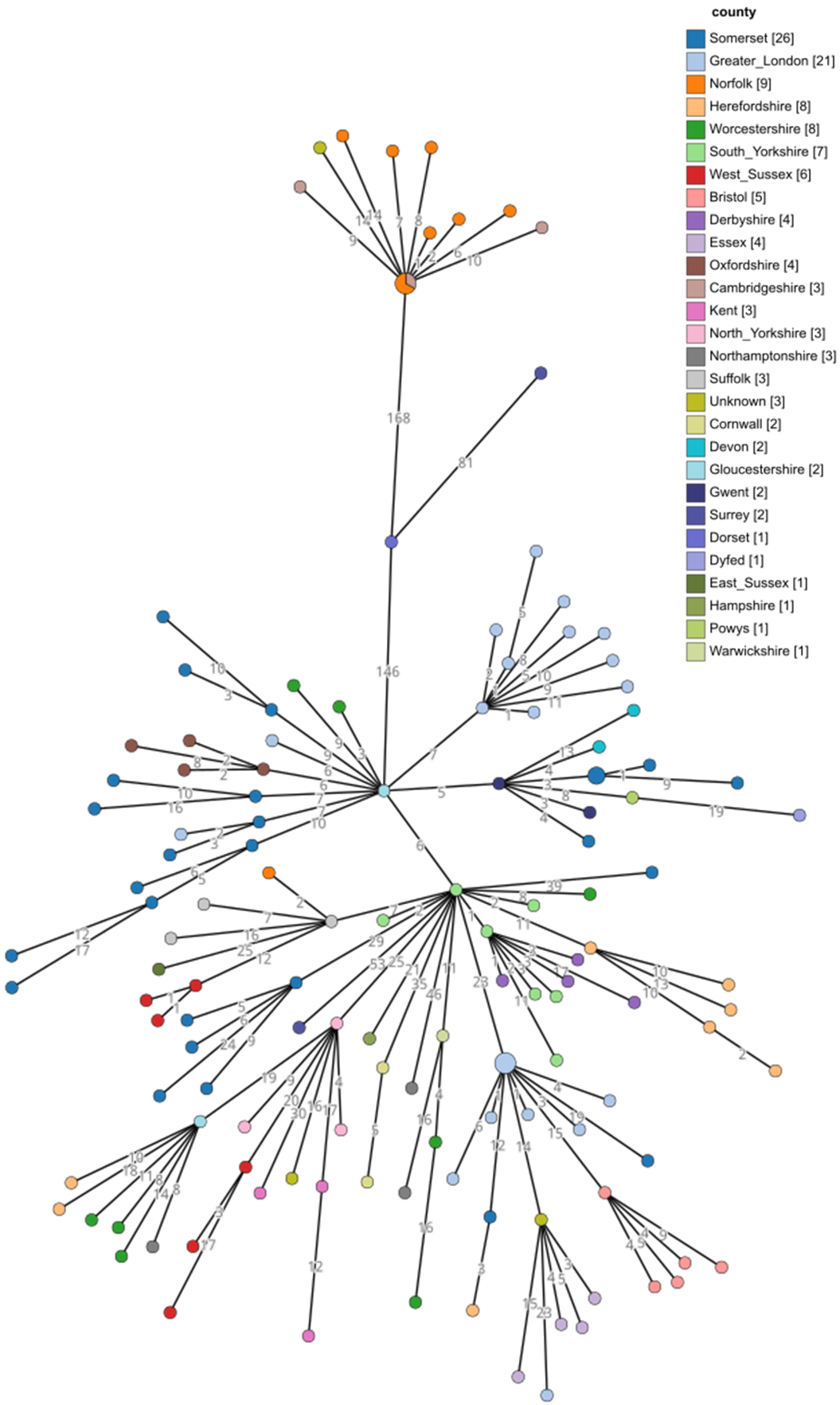


Figure 3.18: A minimum spanning tree created by GrapeTree software. Coloured by county. The tree is on a log scale. The numbers on the branches represent the number of SNP differences.

3.3.10 Case studies

3.3.10.1 Sequence type 3 in Somerset

There were two clearly genetically distinct clusters of ST3 in Somerset identified using bootstrap values to support these groupings, as presented in the phylogenetic tree for ST3 (Figure 3.19). The two clusters were from three different beekeepers. Cluster 1 were geographically clustered, and cluster 2 was only associated with beekeeper 36.

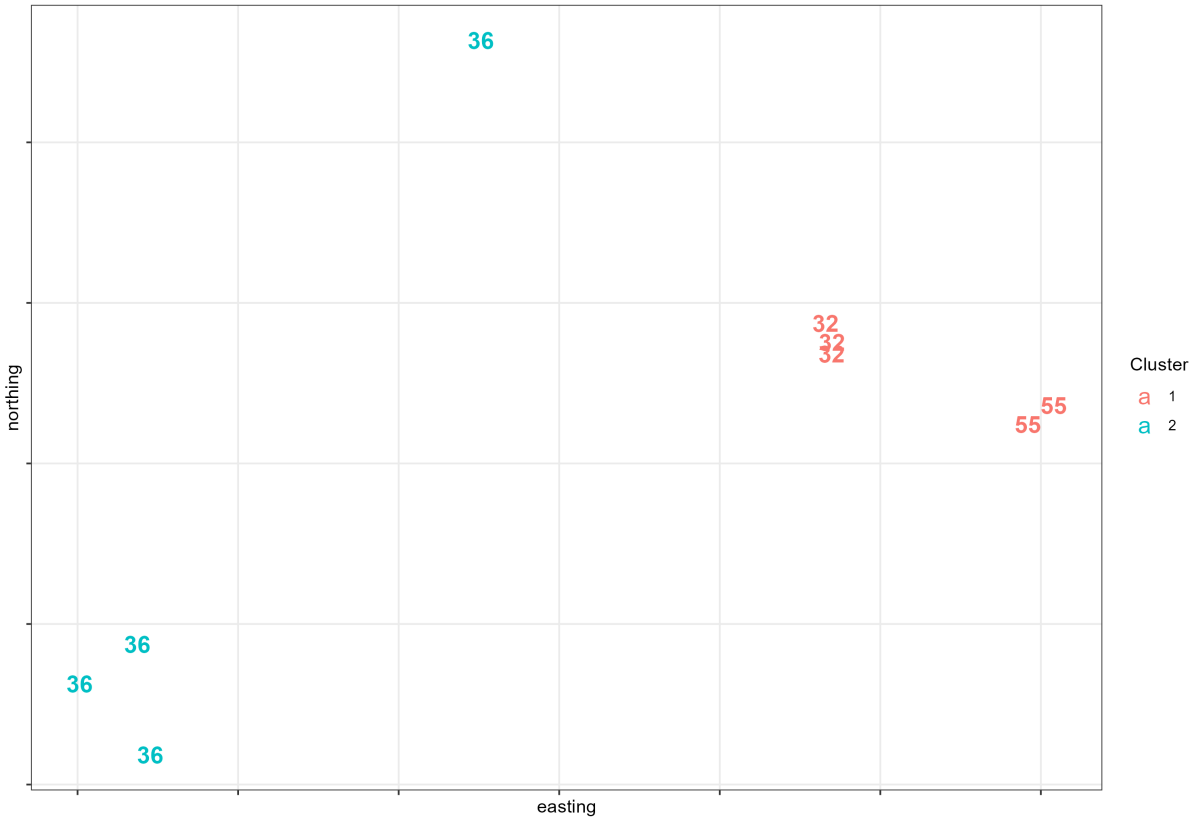


Figure 3.19: The Geographic location of the ST3 samples from Somerset, the values are hidden for anonymity, and points jittered to be able to see overlapping points. The cluster represents an outbreak cluster identified by eye and bootstrap values from the ST3 specific phylogenetic tree. The number is the beekeeper number, matching those presented in the phylogenetic tree.

3.3.10.2 Sequence type 2 in Somerset

There were eight clusters identified from the phylogenetic tree in, Section 3.3.8.3, for ST2 in Somerset (Figure 3.20). Beekeeper 11 had two cluster types of ST2. Cluster 1 and cluster 3 were geographically clustered. Beekeeper 38 had the same cluster type.

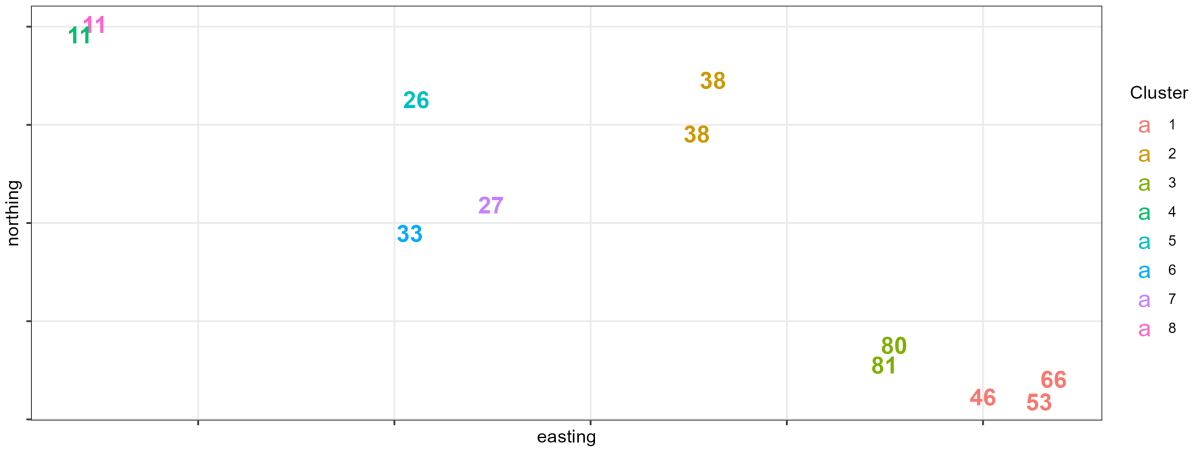


Figure 3.20: The Geographic location of the ST2 samples from Somerset, the values are hidden for anonymity, and points jittered to be able to see overlapping points. The cluster represents an outbreak cluster identified by eye and bootstrap values from the ST2 specific phylogenetic tree. The number is the beekeeper number, matching those presented in the phylogenetic tree.

3.3.10.3 Sequence type 7 in London and Essex

Three clusters were identified from the phylogenetic tree in section 3.3.8.6 for ST7 across Essex and Greater London (Figure 3.21). Both cluster 1 and 3 were geographically clustered. Cluster 3 was present in both Essex and Greater London but were clustered on the county line. Cluster 2 only had one sample and was geographically clustered with cluster 1.

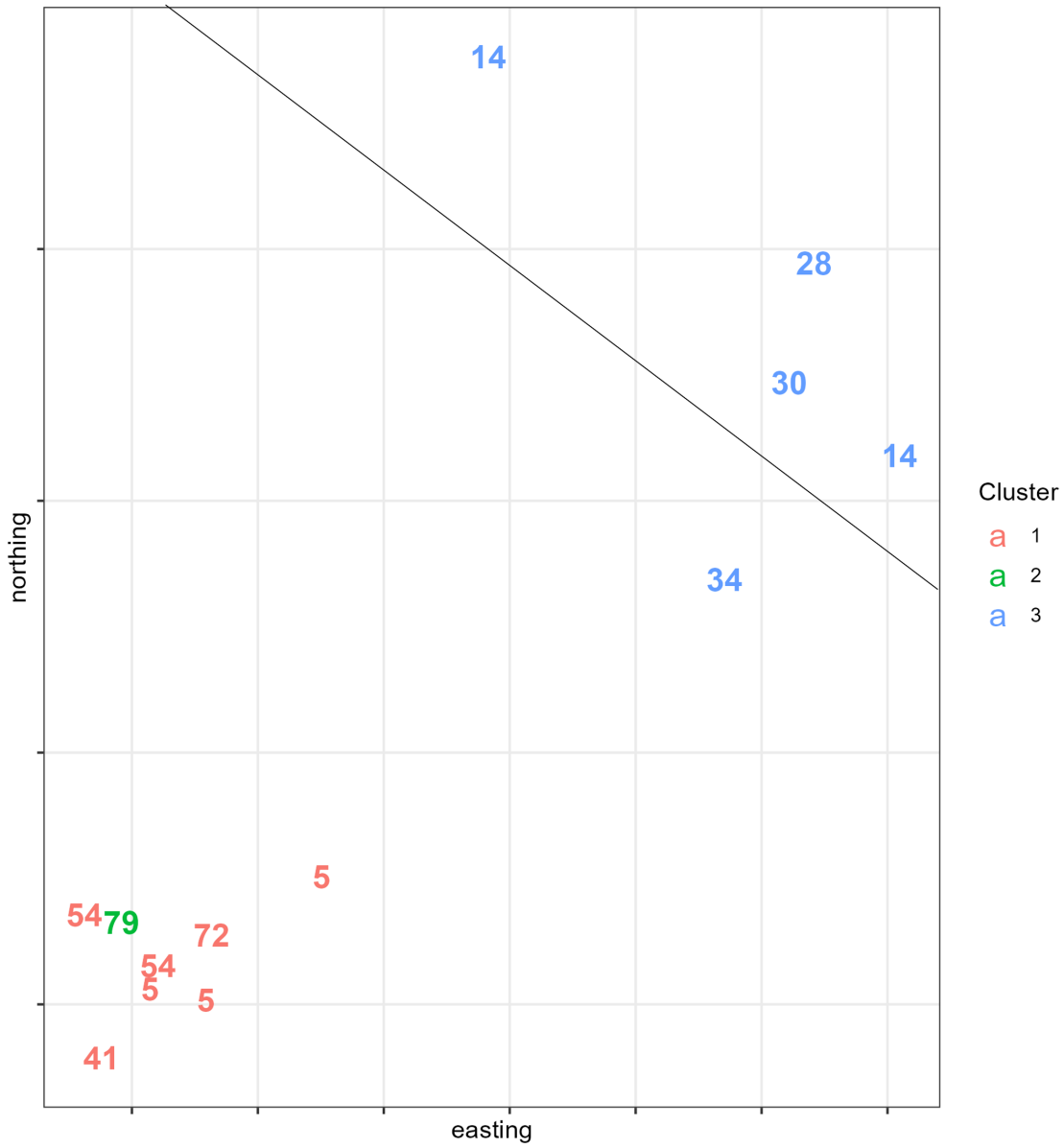


Figure 3.21: The Geographic location of the ST7 samples from London and Essex, the values are hidden for anonymity, and points jittered to be able to see overlapping points. The line represents the county line between Greater London and Essex, London being below it. The cluster represents an outbreak cluster identified by eye and bootstrap values from the ST7 specific phylogenetic tree. The number is the beekeeper number, matching those presented in the phylogenetic tree.

3.3.10.4 Sequence type 23 across all counties

Four clusters were identified from the phylogenetic trees in Section 3.3.8.7, for ST23. Beekeeper 1 had the same cluster type across all of their colonies, that was geographically clustered with beekeeper 42. Beekeeper 44 has the same cluster type (Figure 3.22).

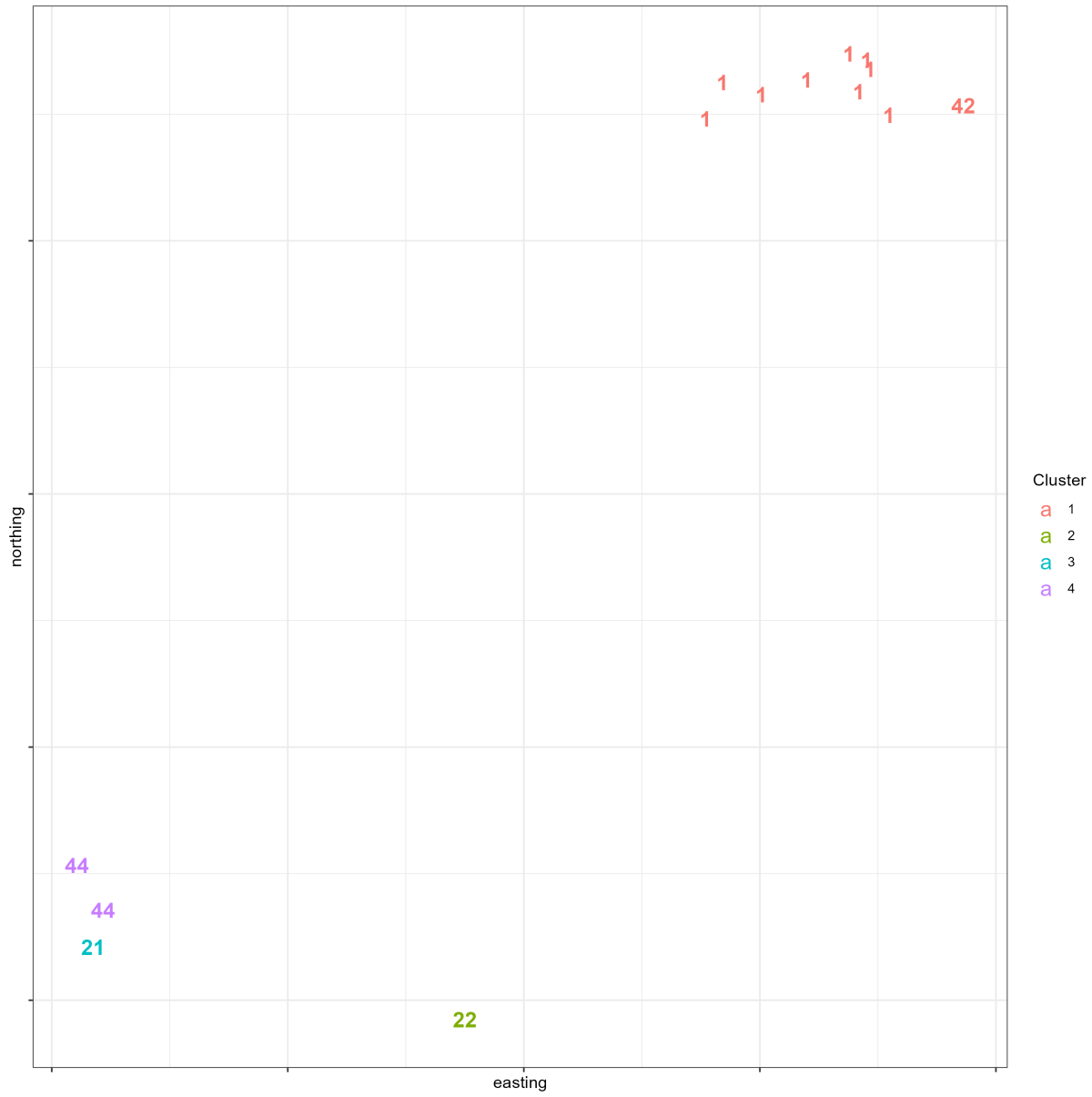


Figure 3.22: The Geographic location of the ST23 samples, the values are hidden for anonymity, and points jittered to be able to see overlapping points. The cluster represents an outbreak cluster identified by eye and bootstrap values from the ST23 specific phylogenetic tree. The number is the beekeeper number, matching those presented in the phylogenetic tree.

3.4 Discussion

Here for the first time, I demonstrated that informative whole genome data can be obtained from buffer bottles containing a single infected larva, using a sequencing pipeline that uses Oxford Nanopore Technologies. Whole genome data of *M. plutonius* has been generated in previous studies but this was using bacterial cultures (Arai et al., 2012), which takes time and resources, especially when producing multiple genomes. Previous

studies have used different sequencing technologies such as MiSeq and PacBio, but this is the first time using handheld Oxford Nanopore Technology. Using the genomic data, I have created a SNP profile to provide more geographical clarity than the MLST method currently used, even with samples with a lower depth. This data may give more of an insight into localised outbreaks of European Foulbrood, to then help improve the management of the disease.

Using whole genome data gave more clarity into the outbreaks than the traditional MLST typing. The traditional MLST scheme divided the samples into 12 outbreak groups (Haynes et al., 2013; Budge et al., 2014). The geographical distribution within each of the groups was vast for example for ST3, samples were present from North Yorkshire and also Cornwall which are approximately 400 miles apart. With the SNP profile generated from the whole genome data geographical clusters appeared to form within the MLST groupings, the samples from Cornwall and North Yorkshire were present in separate clusters, suggesting that they are from different outbreaks as they were genetically diverse. Interestingly, most of the geographic clusters were made up of a single sequence type suggesting that the MLST types are linked to the SNP clusters. Identifying these smaller outbreak regions is a valuable tool when it comes to managing the spread of European foulbrood. Possible transmission events may be identified between beekeepers with apiaries in the same outbreak cluster, and the possible cause of an outbreak in a particular area. There are samples that appeared to not be clustered with any other sample like the Surrey sample, this is likely due to not having full sampling coverage. Only 136 out of the 570 samples sequenced were included in this analysis which means there are gaps in the full picture of the disease in 2020, so some of the samples removed from the analysis may be genetically clustered with the samples that appear to be clusters made up of only one sample. Even if all of the samples were included there are still feral bees, or beekeepers not registered on beebase that are missing in the analysis (Manning et al., 2007; NBU, 2024). Some clusters included samples from more than one sequence type. An ST36 sample was clustered with samples from the same beekeeper but they were all ST3. ST 36 is a new strain (Thebeau et al., 2022), so this could have evolved from ST3. This shows that the whole genome data could also show how *M. plutonius* has evolved. The whole genome data has split up some sequence types into more than one outbreak cluster, for example ST3.

Individual outbreaks of ST3 can be identified using whole genome sequencing. ST3 is

the most common sequence type in England and Wales (Budge et al., 2014). In this analysis approximately 28% of samples included were ST3. There are clear independent outbreaks of ST3 across England and Wales shown when using whole genome sequencing data. Somerset appeared to have two independent outbreaks of ST3, with high bootstrap values to support the groupings. Geographically the samples in each outbreak cluster were also clustered (Figure 3.19). There was a sample that was geographically separate in cluster 2, but all of cluster 2 were from the same beekeeper that has multiple apiary sites. This separated site also appeared in a separate clade on the phylogenetic tree, whilst sharing a common ancestor with the samples from the other site. Cluster 2 could be an isolated case with that beekeeper but due to the missing data points this cannot be confirmed. This case study demonstrated that with the whole genome data along with geographic data and other information a more detailed picture of the outbreaks can be formed than with the MLST data. These two outbreaks, from the data analysed, appear to be spread locally. Somerset cluster 1 shared a common ancestor with samples from Gwent, these two counties are not bordering counties, but it is likely they have *M. plutonius* from the same source. For this case some samples that link the samples geographically could be missing, or it could be that there were incidents like sharing of beekeeping equipment between apiaries in each of the counties that could have been a transmission route. Professional beekeepers could be the source of transmission too, as they tend to sell materials or move colonies between areas. These samples also share a common ancestor with samples from Devon and other Welsh counties, so this could also be a link. Both of these case studies showed that if extra information is gathered a deeper understanding of causes of transmission could be made. Whilst ST3 is high in Somerset there were also cases of ST2.

Outbreak clusters of ST2 identified from the genetics were geographically clustered in Somerset. Somerset had cases of ST2 between 2011 and 2012 (Budge et al., 2014), so the sequence type had been present at least 8 years prior to 2020. In Somerset there were 3 clear outbreak clusters that shared the same ancestor, for each cluster the samples were also clustered geographically (Figure 3.20). Each of the three clusters were towards the East of the county. This information suggested that the bacteria evolved between these three outbreaks, as they were geographically close together spread could have happened through close proximity. Forager bees in the apiaries in close proximity could have shared foraging sites, flowers had been shown to be a hub of disease between pollinators so spread

of *M. plutonius* could have occurred this way in Somerset (Graystock et al., 2015). There could also have been missing samples present between them that did not make the final analysis. Beekeeper 11 appeared to have two different outbreak types in the same apiary, this could have been an error in the pipeline, like a mix up of samples, or this beekeeper could have infection from two different sources. The other clusters only had one sample in them so not much could be inferred from this. Finding out more about the beekeepers in these clusters could explain how the disease spread, for example do they share equipment or collect swarms. There was an Oxford cluster within ST2 however this came from the same beekeeper, without full coverage of the samples it is difficult to tell if this was a county wide outbreak or an isolated outbreak. The Oxford cluster shares an ancestor with the sample from Greater London, so it is likely that a transmission event occurred between the two. Again, this demonstrates that more clarity is gained within the sequence types.

ST7 has a high prevalence in Greater London. There appeared to be 4 outbreak clusters and a single isolated sample within ST7. The samples from both Essex and Greater London formed two clusters and one isolated sample (Figure 3.21). It appeared that the isolated sample, labelled as cluster 2, evolved from cluster 1, as they are geographically linked and closely related genetically. The similarity in the sequences present within geographical clusters suggest historic movement to each area, followed by subsequent evolutionary changes. This suggests that the strains have evolved from a common ancestor as opposed to multiple introductions. Spread within this cluster could have been the same as ST2 as they are in close proximity, and with more information a bee inspector could investigate transmission routes to try and manage spread. Cluster 3 has spread between Essex and Greater London. This cluster has not grouped geographically as the sites appear far apart, however there could be missing samples in between as only 18% of the ST7 samples from Essex and Greater London were included. It was interesting that beekeepers 30 and 28 had sites in between the two sites of beekeeper 14, some movement of equipment or bees could have occurred between them. There were two other ST7 clusters outside of London and Essex. One cluster contained samples all from Bristol, but from the same beekeeper, so it was hard to infer anything from this. The second cluster contained samples from both Herefordshire and Somerset, but again from the same beekeeper. This London, Essex case study for ST7 and all the other case studies have shown that many of the clusters identified from the whole genome sequencing are also geographically clustered.

All the information of these genetic clusters is useful as this could be passed onto the bee inspector who is more aware of movements and contacts within their area to try and manage the spread. Not all sequence types had clear clusters.

ST23 had some clustering but not as clear as ST2 and 3. There was one big cluster and then single samples all with supporting bootstrap values (Figure 3.22). The samples included for ST23 represented 29% of cases, which is why some of the clusters included only one sample. Cluster 1 is geographically clustered, mainly all from one beekeeper, but it is likely transmission has occurred between beekeepers 1 and 42. Sequence types that showed less clustering were ST39 and ST6. Samples from ST6 all came from Greater London, even those that were not included in this analysis. ST6 was present in Greater London between 2011 and 2012 (Haynes et al., 2013), to my knowledge it has not been found anywhere other than England. There was no clustering observed for ST6 suggesting all occurrences are from the same source, even though they are not geographically clustered, the samples were spread across all of Greater London. This could suggest this sequence type is less virulent and could have possibly been out competed by more virulent strains as Greater London also has the presence of ST2,3,6,7,13 and 40. The sequence type could also have evolved into another one and the colonies with ST6 is a reoccurring disease. The whole genome analysis for ST39 showed no clusters with high bootstrap values. ST39 had only recently entered England in 2020 it is likely that it has not had time to evolve yet. ST39 is present in two geographic clusters, one in Norfolk and one in Cambridge, approximately 52 miles apart. The distance between the two counties suggested that some human instigated transfer had occurred between the two sites. ST39 is genetically diverse from all the other sequence types in England and Wales.

ST39 is a new clonal complex. The phylogenetic tree and the minimum spanning tree created from the whole genome data showed ST39 to be genetically distinct from the samples from clonal complex 3 and distinct but not to the same extent from clonal complex 12 and 13. On pubMLST ST39 is recorded as belonging to clonal complex 3 (Jolley et al., 2018). ST39 came up in different clonal complexes when Grossar et al performed different analyses with the traditional MLST data, in GoeBurst it was CC13 and on a phylogenetic tree it grouped more with CC3 (Grossar et al., 2023). The argE-2 and galK-3 combination match to a lot of the other CC3 sequence types, nearly all of CC13 share the purR-4 variant with ST39, however the gbpB gene for ST39 is unique to that sequence type with gbpB-20 (Jolley et al., 2018). As ST39 failed to follow patterns to

place it in a current clonal complex, showed genetic divergence in this study and has unique combinations this supported that ST39 is a new clonal complex. This new clonal complex was likely to have come from another country as it is so genetically diverse from all the other commonly found sequence types in the UK. There is limited sequence type data published for EFB outside of Europe and Japan, ST39 has not been recorded in anywhere other than the UK (Grossar et al., 2023; Okumura et al., 2019). If the strain has come from outside the UK, it has likely come in through an import. In 2018 the UK was the fourth biggest importer of honey in the world, with China being the main source (García, 2018). *M. plutonius* can survive in honey for a long time, so this could be a possible route for this strain to enter the UK, no EFB typing data for China is publicly available for this be looked into (Hornitzky & Smith, 1998). Another common import is of honey bee queens, between 2007 and 2017 the main countries of origin for UK imports were Greece, Slovenia, Italy, Hawaii, Denmark, New Zealand and Cyprus (Budge et al., 2020). *M. plutonius* can be present in adult bees from infected colonies but also from colonies with low levels of the bacteria where the disease is not presenting yet (McKee et al., 2004). It could be possible that bees with low levels of *M. plutonius* from what appear to be healthy colonies are included in the imports, and bringing new strains into the UK, however queen imports are more regulated and have to be tested. This demonstrated how new strains could be detected, but still comparing to the traditional MLST scheme.

The new whole genome method cannot be used to compare with historic samples using the traditional MLST scheme. For argE, gbpB and purR there was a high match between the results from the traditional MLST method and the whole genome method, but not for galK. For the galK gene only a small percentage of samples matched between the methods. The predominant miscall in the whole genome data was that galK-8 called whereas in the traditional MLST method it was galK-3. The genetic difference between the two galK variants is a single base deletion, however this deletion occurs in a polyA region, known as a homopolymer (Figure 3.23). The galK-3 variant has a repeating sequence of 8 As and galK-8 has a repeating sequence of 7 (Jolley et al., 2018). The mismatch is likely to have been caused by an error in the Oxford Nanopore technology rather than an error in the Sanger sequencing used in the traditional MLST typing. The MinIon flow cell R9.4 flow cell used in this study has a known issue with homopolymers (Xian et al., 2022). As the number of repeated bases increase the chance of missing a read becomes higher (Delahaye & Nicolas, 2021). The homopolymer issues appear to have been improved with

the development of the Minion flow cell 10.4 (Sereika et al., 2022). The technology now has a head that reads the DNA strand twice on its way through the pore, as well as a longer head. Some methods have been developed to correct these errors during the analysis of the sequencing data such as homopolish, which uses data from homologous sequences to try and correct the errors present (Huang et al., 2021). If the sequencing was repeated with the new technology or the homopolish analysis, it may improve the accuracy of using the traditional MLST scheme, but with this current data and analysis the whole genome sequencing data cannot be used to MLST type using the traditional method.

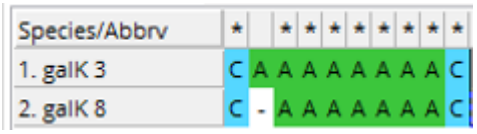


Figure 3.23: A sequence alignment of a section of galK-3 and galK-8 using MEGA showing the absence of an A in galK-8 compared to galK-3.

The limit to this study was the missing data, clear groups have formed showing the potential of using SNP profiles to study outbreaks, but only 136 out of the 570 sequenced were included. Potential reasons for many samples failing sequencing could be sampling error at the field, all these samples were taken by many different bee inspectors, although a protocol exists there could be some samples with more material in or too little. Too much larval material could include inhibitors for the PCR step or too little may not be enough bacteria left to sequence (Scheu et al., 1998). There are no instructions for storage or handling, these samples are posted to the lab so there is no control over the environment they have been stored in or for how long. A way to control these issues is to generate a more specific protocol or have a second sample from the buffer bottle collected soley for sequencing. Other issues could have been limitations in the lab processing methods, the method may bias certain levels of bacteria or cleaner samples. Further development and testing of the sequencing pipeline need to be performed to include more samples in the analysis. Even if all the data from the 2020 would have been included it does not give a full picture of the country overall, as wild bees could be transmitting the disease and unregistered beekeepers may exist.

Future work from this study would be to gather more sequencing data particularly from other years to compare with the 2020 data. Sequencing the 2021 samples would give an idea if the outbreaks identified are spreading, evolving or even stopped over time.

Gathering more information, possibly through a survey, about the possible cause of disease transmission and opinions of beekeepers could generate an insight into how EFB is spreading in certain areas. Redoing the 2020 samples with the new sequencing flow cell could see an improvement in the MLST typing, possibly eliminating the galK-3 and 8 confusion. ST6 is very interesting, as it does not appear to be evolving or clustering, this would be an interesting case study. Gathering functionality data from all sequence types either from this sequencing data or even RNA sequencing could give more of an insight into what may be causing this. Functionality data would then allow ST6 to be compared to the other sequence type to assess if there are genes that are making it potentially less virulent.

This study has demonstrated that affordable whole genome sequencing can be used to give more clarity on outbreaks within the UK from infected larvae samples, even at low sequencing depths. Along with information about beekeeper activity from the bee inspector this could be a useful tool to manage and monitor outbreaks building on from the traditional MLST scheme. Even with only a fraction of the samples used some useful information can be gained, and it also highlights the potential whole genome sequencing has on monitoring and managing EFB.

Chapter 4

Using whole genome sequencing from infected larave to look beyond the phylogeny of *Melissococcus plutonius*

4.1 Introduction

European Foulbrood (EFB) is a honey bee disease that is found internationally. EFB is caused by the Gram-positive, non sporulating bacteria *Melissococcus plutonius* (White, 1912). The bacteria grows in the larval mid gut, and causes larval death three to five days after infection (Bailey, 1956). The cause of larval death is thought to be starvation, as the *M. plutonius* outcompetes the larvae for food (Forsgren, 2010). Spread of the disease occurs when an infected larva survives to the pupal stage and excretes their gut contents into the cell, and nurse bees then contact the bacteria during cell cleaning. After cleaning, the bacteria could be transferred to the mouth parts of the nurse bee leading to a potential spread risk (Forsgren, 2010; Bailey, 1983). *M. plutonius* has been found in adult honey bees, and also honey bee products such as honey (Mckee et al., 2003). Evidence of an EFB infection include larvae that have become twisted and have either opaque or creamy white guts and an irregular brood pattern (Bailey, 1983). Due to EFB being a subject to statutory control measures any suspicion that a colony has EFB must be reported to the local bee inspector (Government, 1982). To confirm an EFB infection a bee inspector

carries out a serological test using a lateral flow device (LFD), similar to those used during the COVID-19 pandemic (Tomkies et al., 2009; Wang et al., 2020). In positive cases action is taken, shook swarm is the most common practice performed. Shook swarm involves removing the infected combs, scorching the box and onto new foundation shaking off the bees present on the infected comb. Using an antibiotic, oxytetracycline (OTC), was the routine method between 1967 and 1984 (Waite et al., 2007). OTC is a broad-spectrum antibiotic commonly used in livestock (Li et al., 2008). OTC falls within the group of natural tetracyclines and binds to the 30S ribosomal subunit to inhibit protein synthesis of the bacteria, preventing replication (Pickens & Tang, 2010). OTC is now only used in a limited number of cases, one reason being because shook swarm is more effective (Budge et al., 2010) and another is the environmental risks of using antibiotics.

One environmental risk of using OTC is that residues can be present in food products produced by honey bees such as honey, which if consumed could be a health risk (Thompson et al., 2006). These residues could also lead to other bacteria in the environment as well as the target pathogen developing antimicrobial resistance (Kirchhelle, 2018). Antimicrobial resistance (AMR) is a major global health concern particularly with the overuse and exposure of antibiotics across all aspects of life (Ruckert et al., 2024). One way AMR can occur in bacteria is through random mutation and selection. Bacteria multiply at a fast rate and each division poses the risk of random mutations occuring. These mutations could lead to a gene coding for different proteins with alternative functions, potentially giving bacteria the ability to resist the effect of antibiotics, genes with this ability are called antibiotic resistance genes (ARGs). If exposed to antibiotics the resistant bacteria survive through selection and carry on multiplying (Davies, 1996). Another mechanism for bacteria to acquire ARGs is through horizontal gene transfer, usually assisted by plasmids. Horizontal gene transfer is the passing of genetic information, usually beneficial to the recipient, like ARGs, to other organisms that are not their offspring (Sprague, 1991; Barlow, 2009). This transfer can occur through different mechanisms, such as conjugation and transformation. Transfer through conjugation occurs when the two species are in physical contact, and genetic information is transferred across the membrane. Transfer through transformation occurs when the genes are present in the environment, and they can pass to the genome of the bacteria (Lorenz & Wackernagel, 1994; Sprague, 1991). There are a few different mechanisms through which AMR genes can provide the bacteria with the ability to

prevent the action of antibiotics. The AMR gene may code for an efflux pump that rejects the incoming antibiotics, this prevents the antibiotic from making it to the target and providing antibiotic resistance to the bacteria. Alternatively, an AMR gene might produce a protein mimic to the antibiotics target, so it binds to that instead of the intended target (Munita & Arias, 2016). Regardless of the mechanism or mode of attainment, the presence of ARGs in *M. plutonius* would likely reduce OTC efficacy and would highlight the frailty of antibiotic usages. As well as the direct effects ARGs may have on *M. plutonius* control, ARGs pose a wider global threat to human health, when genes pass to bacteria that are pathogenic to humans, making treatment of human disease more difficult. This is especially important given the limited suite of effective antibiotics now available (Prestinaci et al., 2015). *M. plutonius* DNA has been found to be present in honey (Mckee et al., 2003). ARGs in honey give exposure of these genes to the environment, increasing the chances of contact with bacteria that is pathogenic to humans. Identifying ARGs in infected cases of *M. plutonius* could influence the decision of the bee inspector as to whether to use antibiotics or not, for example in extreme cases if there are OTC resistant genes present then OTC may not be effective. ARGs identification could also be used to investigate the risk of honey bees and products on general antimicrobial resistance. Although ARGs can improve the survival of *M. plutonius*, the bacteria require virulence genes to survive in the larval host.

Virulence genes determine the expression of virulence factors in pathogenic bacteria. Virulence factors are molecules that give the pathogenic bacteria the tools to enter, survive and cause disease within the host. Virulence factors can assist with many different aspects of virulence and so fall into hundreds of different categories. For example, adherence factors can help the pathogenic bacteria attach to the cell wall of the host, increasing survival assisting with the formation of biofilms. Exotoxins are virulence factors, that attack the core processes of the host cell, resulting in cellular damage. Other virulence factors assist with the evasion of the immune system of the host by forming protective structures such as capsules (Johnson, 2018). Databases like VFDB represent a repository of many virulence factors from a collection of pathogenic bacteria (Chen et al., 2016). Identifying virulence factors is of particular interest in drug development, where virulence factor inhibition could prevent infection and spread of a particular pathogen (Wu et al., 2008). Little has been studied about the process of invasion and infection of *M. plutonius* in the honey bee larvae, identifying and understanding the virulence factors involved could

be useful to understand the mechanism and give more clarity to the virulence process and potentially provide targets for future treatments. Several virulence genes and gene clusters have been identified in *M. plutonius*; one particular gene is found on the pMP19 plasmid.

Plasmids are naturally present in bacteria and independent from the chromosomal DNA a plasmid is small circular sequence of double stranded DNA. Plasmids replicate independently and usually contain genes that provide beneficial traits to the bacteria, for example virulence factors or ARGs (Rodríguez-Beltrán et al., 2021). *M. plutonius* has two plasmids, pMP1 and pMP19 (Okumura et al., 2011), although Nakamura et al showed that the pMP19 plasmid was not present in all samples (Nakamura et al., 2020). The virulence gene Melissotoxin A (MtxA) has been identified in *M. plutonius* and is found on pMP19 (Djukic et al., 2018; Grossar et al., 2020). The exact mechanism of action of MtxA is unclear but based on sequence similarity to factors in other bacteria, it has been hypothesised that MtxA could work by damaging the gut cells of the larvae (Grossar et al., 2020). Using molecular techniques *M. plutonius* samples are given a sequence type based on their genetic sequences, and these are grouped into one of 3 groups called clonal complexes (CC3, CC12 and CC13). Typically, in culture *M. plutonius* loses virulence, but atypical strains present in CC13 remain virulent even after multiple rounds of sub culturing (Arai et al., 2012). Gene knock out studies showed that in CC3 the pMP19 plasmid was essential for virulence, but possessing the MtxA gene was not. The pMP19 plasmid is not required for virulence of CC12 and the atypical CC13 strains (Nakamura et al., 2020). A study of MtxA using Swiss isolates of *M. plutonius* found that MtxA was present in all CC12 samples, half of the CC3 samples and a small percentage of the CC13. Some other potential virulence factors have been identified in *M. plutonius*, such as cell surface adhesion proteins and capsule forming proteins (Djukic et al., 2018). Identifying genes that code for virulence factors could be useful for understanding disease aetiology and ultimately developing treatments that prevent the spread of EFB. Virulence genes or ARGs usually do not make up the core genome, and so could be used as an additional indicator of intra-species diversity.

Identifying the genes that do not make up the core genome can be achieved by studying the pangenome. Pangenome analysis allows diversity across the species or genus to be analysed, as non-core genes are usually not essential to the core functions of the organism but may provide an advantage, like the virulent genes and ARGs (Gong et al., 2023). A pangenome analysis categorises the genes present across all the samples in a dataset. Core

genes are those present in all of the samples within a dataset, soft core genes are those present between 95% and 99% of samples, shell genes present in between 94% and 15% and cloud genes in less than 15% of samples (Tonkin-Hill et al., 2020). Pangenome analysis has helped identify novel species of *Myxococcaceae* due to genetic differences (Chambers et al., 2020). This method could be used to identify genes that may contribute to the genetic diversity of *M. plutonius*. Importantly, the infection rate and pangenome can be influenced by the presence of other bacteria such as gut microbes.

The gut microbiome of the honey bee larva forms an important symbiotic relationship that can convey many attributes to the host, including protection from disease (Maes et al., 2016). The gut microbiome of a healthy larva is predominately made up of *Apilactobacillus kunkeei* and *Bombella apis* (Smith et al., 2021). As a larva grows the diversity of bacteria present in the gut microbiome increases. There are some bacterial species that are found in all adult honey bees including *Snodgrassella alvi*, *Lactobacillus apis* and *Gilliamella apicola* (Anderson et al., 2023; Raymann & Moran, 2018). The bacterial composition of the gut microbiome changes when a larva is infected with *M. plutonius*, and also new bacteria known as secondary invaders take advantage of this can colonise. Common secondary invaders found in infected larvae are *Paenibacillus dendritiformis*, *Paenibacillus alvei*, *Enterococcus faecalis*, *Brevibacillus laterosporus*, *Bacillus pumilus* and *Achromobacter eurydice* and their presence may impact disease development (Forsgren, 2010). For example, the presence of *P. alvei* appeared to increase mortality and infection rates, and in other studies it appeared to have the opposite effect (Lewkowski & Erler, 2019; Giersch et al., 2010). The composition of the larval microbiome in infected larvae, both symptomatic and asymptomatic, not only varies depending on apiary site but it also varies between individuals. A study of the composition of the micobiome of infected larvae showed that the across many infected larvae levels of common gut bacteria, *F. perrara*, *G. apicola*, *S. alvi* and *L. apis* increased, as well as *Bombella apis* and other *Bombella spp.* The levels of the secondary invader *E. faecalis* also increased, along with three other bacterial species *F. fructosus*, *L. iberica* and *Bifidobacterium* (Anderson et al., 2023). Understanding the impact *M. plutonius* has on the microbiome composition can be an important tool for understanding infection processes.

The overall aim of this chapter was to investigate what other information can be gained from a sequencing pipeline initially designed to investigate the phylogeny of *M. plutonius* in the UK. This extra information included virulence factors, ARGs, shell genes and

other microbes. After generating the sequencing data using the method developed in chapter 2, the first step was to predict the genes present and look for ARGs and virulence genes using information from current databases. The second step was to perform a pangenome analysis to assess the shell genes and identify any clustering of samples from this information. The final step was to assess if any of the common larval gut bacteria and secondary invaders were present in any of the 136 samples and see if any clustering within samples occur based on bacteria composition.

4.2 Methods

4.2.1 Generating sequencing data

4.2.1.1 Sample Collection

LFD buffer bottles each containing a single larva from infected sites in 2020 were received from Fera Science Ltd via DEFRA. MLST typing for each sample was performed by Victoria Tomkies at the National Bee Unit (NBU) before being sent to Newcastle University. All samples came with MLST typing information, location, county, and a beekeeper number to help link ownership of outbreaks. The samples that were successfully sequence typed were selected for downstream analysis, and some samples which failed the MLST were also included providing a total of 570 samples.

4.2.1.2 Host Depletion and DNA Extraction

For each selected sample 2 ml was transferred into an empty 2 ml tube. The aliquoted samples were centrifuged at 7,500 g for 10 minutes, the supernatant was removed, and pellets re-suspended in 20 µl DNase buffer I, 1 µl DNase I (10649890, Fisher Scientific), 180 µl distilled water, and incubated at 37°C for 60 minutes. To inactivate the DNase, 20 µl 50 mM EDTA was added, and the samples incubated at 65°C for 10 minutes. The samples were centrifuged at 10,000 g for 5 minutes and the pellet was re-suspended in 180 µl Gram-positive lysis buffer. The suspended samples were transferred into 96 well plates to increase throughput. The plates were incubated at 37°C for 30 minutes, then centrifuged at 20,000 g for 5 minutes and the supernatant was discarded. The pellets were re-suspended in 200 µl of PDB (5 mM guanidine HCl) and 25 µl proteinase K

(Qiagen) was added and the plates were incubated at 56°C for 30 minutes. Then 200 µl 100% ethanol was added and the sample transferred to a cEZ-10 RNA Mini Spin Column and spun down at 6000 g for 1 minute. The flow through was discarded and 500 µl of wash buffer 1 (7 M Guanidine HCl, 56% ethanol) was added and the column spun down at 6000 g for 1 minute the flow through discarded, 500 µl of wash buffer 2 (70% ethanol, 10 mM Tris HCl) was added and spun down at 6000 g for 3 minutes the flow through was removed and a 1 minute spin was performed. The columns were transferred to a clean 96 well plate, 50 µl TE buffer was added directly to the column and then centrifuged at 10,000 g for 1 minute, this was repeated once. The DNA concentration was measured for each sample using the dsDNA quantification using Sybr Green I V.2 protocol (dx.doi.org/10.17504/protocols.io.b34gqqtw).

4.2.2 Oxford Nanopore MinION sequencing

The samples were run across six Minion runs and prepared for sequencing using the SQK-LSK110 (Oxford Nanopore) kit with the EXP-PBC096 (Oxford Nanopore) PCR barcoding kit. A modified Ligation sequencing gDNA - PCR barcoding protocol from Oxford Nanopore was used. Half the volume of all components was used to cut costs, up until the end prep step. For the PCR step conditions the number of cycles was 18 and the extension time was 8 minutes, with a final extension of 10 minutes. The samples were pooled in equal volumes, 2 µl of each sample and then 48 µl of the pool taken forward. The short fragment buffer was used for the final clean up. Based on the final pool concentration being low from the first 3 plates, for the last three plates after pooling the PCR products, a bead clean was performed to concentrate the samples. For the bead clean 400 µl of AMPure XP beads (A63881) were added to the pooled PCR products. The mixture was incubated for 5 minutes at room temperature, and the tube placed on a magnet to pellet the beads. Whilst on the magnet the supernatant was pipetted off and 1 ml 80% ethanol was added and then removed by pipetting, this step was repeated. The pellet was left to dry for 30 seconds, the tube removed from the magnet and the beads were resuspended in 50 µl nuclease free water. The beads were incubated for 2 minutes, and the tube was placed back onto the magnet. After the beads were pelleted the 50 µl of supernatant was used for further library preparation according to the protocol. The pooled library was loaded onto a R9.4.1 Minion flow cell (Oxford Nanopore, FLO-MIN106D) and run for 72 hrs using the MinKNOW Software (Oxford Nanopore). Within the MinKNOW software

live, high accuracy basecalling was performed, converting electrical signal to nucleotides. Each sample was barcoded with a unique sequence and the samples were separated using live barcoding through the software, as well as demultiplexing, removing all the adaptors. The MinKNOW software also has a built-in quality control (QC) step that anything with a Phred score of less than 7 is placed in a failed folder. The final output was raw reads in FASTQ format for each sample.

4.2.3 Initial bioinformatics

All packages used for bioinformatics were through the Bioconda package manager (Grüning et al., 2018). Barcoding was performed live during sequencing and the reads placed into folders matching the barcode, with those that do not match any of the barcodes are placed into an unclassified folder. Initially porechop (Version:0.2.4) (Wick et al., 2017) was used to detect missed barcodes in the unclassified files. A reference guided assembly was carried out using Minimap2 (Version:2.24) the raw reads for each sample were mapped to the reference genome, and two plasmids individually (Table 4.1) (Li, 2018). A reference guided assembly was performed opposed to a *de novo* assembly because the samples were not pure *M. plutonius* and the average sequencing depth was too low. Using Samtools (Version:1.16.1) the mapped reads were converted to binary alignment map files (BAM) required for downstream analysis.

Table 4.1: Details of the samples used in the reference guided assembly for each sample.

Genome	Genbank ID	Size (Mb)
Melissococcus plutonius DAT561 DNA	AP018492.1	1.80
pMP1	AP018493.1	0.20
pMP19	AP018494.1	0.02

4.2.4 Gene assignment analysis

4.2.4.1 Overview

An overview was shown to highlight the bioconda packages used and outputs in this methods section (Figure 4.1).

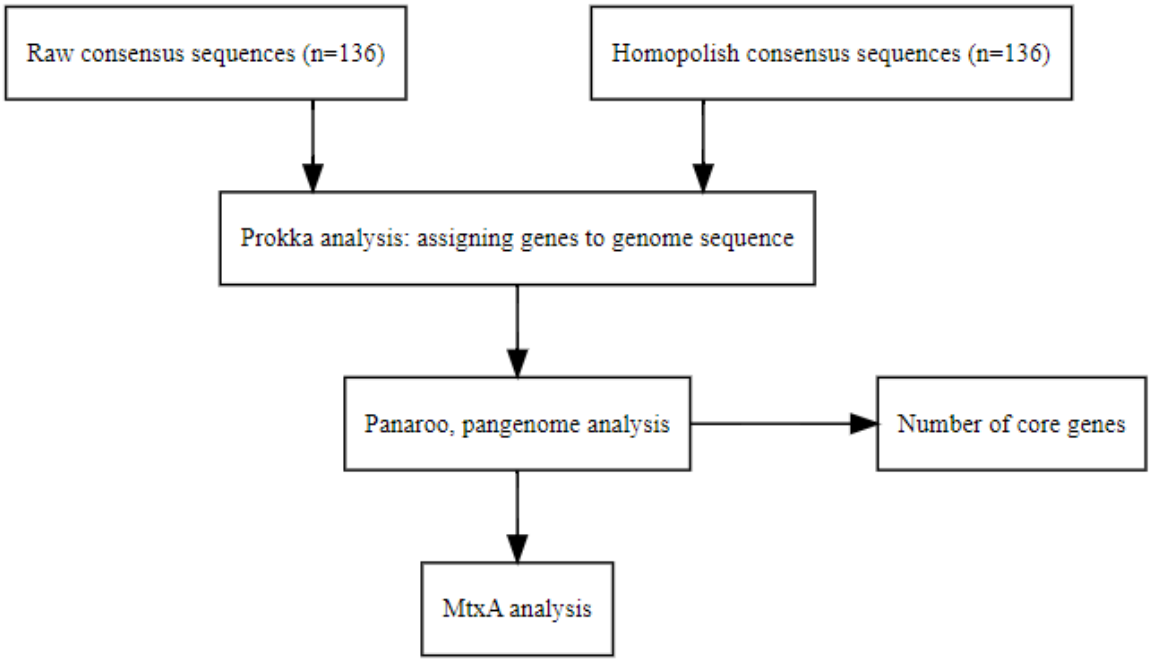


Figure 4.1: An overview of the bioinformatics process for both homopolished and raw consensus sequences used for the analysis in this section.

4.2.4.2 Prokka analysis

A consensus sequence was assembled from the *M. plutonius* BAM files generated in 4.2.3 for each of the 136 samples that were chosen in chapter 2. The consensus sequence was generated using the Samtools consensus function in the Samtools package (Version:1.16.1) (Danecek et al., 2021). To predict the genes present in each sample, the consensus sequence was run through Prokka (Version:1.14.6) using the *M. plutonius* reference genome (Genbank:AP018492.1) as guidance (Seemann, 2014; Clark et al., 2016). The prokka output included a gff file, that contained all the genetic features found on the input genome, and a fasta file that contained individual nucleotide sequences for each genomic feature found. Both the consensus sequence generation and the Prokka

analysis were repeated for both the plasmids, pMP1 and pMP19, using (AP018493.1 and AP018494.1 respectively).

4.2.4.3 Panaroo analysis

Panaroo (Version:1.1.2) (Tonkin-Hill et al., 2020) was used to conduct a pangenome analysis that identified the core genes that were present in every sample, and those that were present in only a selection of the samples. More specifically, this process identified core genes, shell genes, that were present in most samples, and cloud genes, that were only present in a minority of the samples. The gff files produced in the prokka analysis in 4.2.4.2 were used as input into the Panaroo function. A cluster analysis was performed on the shell genes identified in the panaroo analysis. The Hopkins' statistic was calculated to assess if the samples could be clustered, this was calculated using the factoextra and package in R (Version:1.07) and the fanny clustering in the cluster package (Version:2.1.3) (Kassambara, 2016; Maechler, 2018). The cluster results were presented.

4.2.4.4 MtxA gene analysis

For each sample, the presence and number of fragments of the MtxA gene were manually recorded using the tsv file from the Prokka analysis peformed in section 4.2.4.2. The MtxA assigned sequences were extracted and visualised in MEGA11 (Tamura et al., 2021). Existing MtxA primers tox_MEPL_for:5 -GCTCAAGCAGCAACTTTTACG-3 and tox_MEPL_rev:5 -TTCCCCTGGTATTACTTGTAGATG-3 were aligned to the MtxA fragments to assess which part of the genes the published primers were amplifying (Grossar et al., 2023).

4.2.4.5 Homopolish

The consensus sequences generated in 4.2.4.2, for *M. plutonius*, pMP19 and pMP1, were polished to correct common Nanopore homopolymer issues using Homopolish (Version:0.4.1) (Huang et al., 2021). Homopolish uses information from reference genomes to correct the homopolymer regions. The polish function was used on the consensus sequences to generate a new polished consensus sequences for each sample.

4.2.4.6 Prokka, panaroo and MtxA analysis on homopolished sequences

The prokka analysis and the panaroo anlysis in 4.2.4.2 and 4.2.4.3 were repeated using the homopolished sequences generated in 4.2.4.5. A comparison of the output genes for both the polished and unpolished results was recorded. The MtxA sequences from the prokka output for pMP19 performed in 4.2.4.5 were aligned using MEGA-11 to look for differences in the gene sequence across samples, and to assess if the genes were still fragmented. The individual nucleotide differences between the samples were recorded.

4.2.4.7 Searching for AMR and Virulence genes

A search for potential AMR genes and virulence genes was performed on the polished consensus sequences generated in 4.2.4.5. Abricate was used to search for these genes of interest. The database used needed to be specified so for AMR genes I used CARD, NCBI AMRfinderPlus, Resfinder, ARG-ANNOT, MEGARES 2.00 and EcOH (Doster et al., 2020; Gupta et al., 2014; Ingle et al., 2016; Zankari et al., 2012; Feldgarden et al., 2019; Jia et al., 2017). For the virulence genes I used the VFDB database (Chen et al., 2016).

4.2.5 Analysis of secondary invaders and gut microbiome bacteria

4.2.5.1 Searching for secondary invaders and gut microbiome bacteria

Further bioinformatics were performed to assess the presence of secondary invaders of *M. plutonius* and common gut microbes in each sample sequenced in section 4.2.1. A reference guided assembly was performed for each of the bacteria outlined (Table 4.2) using Minimap2 (Version:2.24) (Li, 2018). Each of the mapped read files were converted to BAM files using Samtools (Version:1.16.1) (Danecek et al., 2021). Samtools was used to map the reads to the reference genomes and report the coverage. The percentage genome coverage was calculated using the base pair coverage value for each bacterial species. The number of reads and percentage reads was also calculated for *A. mellifera*, using the genome (Genbank:GCA_003254395.2) (Wallberg et al., 2019). Initially the percentage reads of *M. plutonius* were compared to the percentage of disease in the colony the sample was found, assessed by the bee inspector. This comparison was performed to see if there

was a correlation between presentation of disease and the sequencing data, and if this differed between sequence types.

Table 4.2: Details of the microbiome and secondary invader bacterial species analysed for each sample and the reference of the genome sequence used in the guided assembly.

Bacteria	GenBank reference	Genome size (Mb)
Paenibacillus alvei	GCA_026797385.1	6.7
Apilactobacillus kunkeei	GCA_019575995.1	1.5
Bacillus plumilus	GCA_003020795.1	3.6
Bifidobacterium asteroides	GCA_000304215.1	2.2
Bombella apis	GCA_025289935.1	1.9
Brevibacillus lacterosporus	GCA_002706795.1	5.4
Enterococcus faecalis	GCA_000393015.1	2.9
Frishella perrara	GCF_025291255.1	2.3
Fructobacillus fructosus	GCA_014489725.1	1.4
Gilliamella apicola	GCA_000599985.1	3.1
Lactobacillus apis	GCA_003150935.1	1.7
Lonsdalea iberica	GCA_002111585.1	3.8
Snodgrassella alvi	GCA_000600005.1	2.5
Paenibacillus larvae	GCA_002951935.1	4.4
Paenibacillus dendritiformis	GCA_021654795.1	6.5

4.2.5.2 Analysis of secondary and gut microbiome bacteria

The percentage genome coverage for each of the bacterial species, along with *M. plutonius* were compared by visualising the data with a box and whisker plot, created using ggplot2

(Version: 3.4.4) (Wickham, 2016). The percentage reads for each species were also plotted in the same way to assess the composition of the samples. A cluster analysis was performed with the bacterial reads proportion for each species per sample, to see if the compositions could be explained by ST, geography or collection date. The bacterial read proportion was the proportion of the summed reads of all the chosen bacterial species for each individual species. As a supervised clustering was used the first step was to calculate the number of clusters. The Hopkins' statistic was calculated to assess if the samples could be clustered, this was calculated using the factoextra package in R (Version:1.0.7) (Kassambara, 2016). The cluster tendency was visualised, and the optimum number of clusters was determined using fanny clustering, with the elbow method. The elbow method calculated the within cluster sum of squares for each number of clusters, and when plotted the elbow of the plot indicated the optimum cluster number. The clusters were visualised in a PCA using ggplot2. Once the optimum cluster number was obtained, the cluster analysis was performed using fanny clustering within the R package cluster (Version:2.1.3) (Maechler, 2018). Each sample was assigned a cluster group based on the species composition. The bacterial composition for samples in each cluster group were assessed to see which bacteria was driving the clusters. This was achieved by plotting the bacterial reads proportion for each sample in each cluster group. The same plot was generated with percentage of total reads for each bacterium across samples, rather than just proportion of the selected bacterial reads. This was performed to highlight there were unknown species also present. The assigned cluster groups were also plotted on a map of England and Wales. The map was generated using the sf (Version:1.0-15) (Pebesma, 2023) and ggrepel (Version:0.9.4) (Slowikowski, 2023) package in R (Version: R 4.2.1). This map was used to see if the bacterial clusters were grouped geographically. A summary of the ST types in each cluster was also produced to assess if there was a specific composition that favoured specific sequence types. Whether the compositions of bacterium changed over time was investigated. The average bacterial reads proportion for each bacteria across the samples was calculated for each month of collection and visualised on a plot.

4.2.5.3 AMR gene assessment in bacterial species

To identify the AMR genes present in the gut microbiome and the secondary invader bacteria, consensus sequences were made for species that had more than 75% coverage,

using Samtools consensus (Version:1.16.1). To the consensus sequences generated the same analysis as 4.2.4.7 was perfomred to identify potential AMR genes. The AMR genes were compared to those identified in the *M. plutonius* genome within each sample.

4.3 Results

4.3.1 Gene assignment analysis

4.3.1.1 The MtxA gene

The pMP19 and evidence of the MtxA gene was found in all 136 samples. From the original prokka analysis of pMP19 the MtxA gene was fragmented in some samples. The fragment number varied, 46% of samples were fragmented into 2 fragments, 29% into 3 fragments and the the remaining 25% of samples were not fragmented and covered the whole gene. When the two fragments were aligned to the full gene it became clear that a premature TGA stop codon, is causing the fragmentation (Figure 4.2). The difference in the whole gene and the fragment sequences was a different number of Ts in a homopolymer further up the sequence. The primers used by Gossar et al, aligned into the first fragment of all samples (Grossar et al., 2020). After using the homopolish the MtxA gene in each sample was the full 897 bps and no fragments were present. When the homopolished MtxA sequences were aligned three mutations were found. The 304C>T mutation was found in all the ST39, ST13 and ST40 samples, but this was a silent mutation not changing the amino acid. The 343G>A mutation was found in all the ST39 sequences, this mutation changed valine to isoleucine. The 704A>G mutation was also found in all ST39 samples, this changed the amino acid from glutamic acid to glycine.

Figure 4.2: The sequence alignment of the whole MtxA gene, with the two fragments generated in the prokka analysis of one of the samples analysed.

4.3.1.2 The pangenome analysis

The number of each type of gene was compared between the homopolished prokka output and the raw consensus prokka output, for pMP19, pMP1 and the main genome sequence. For the main genome, pMP1 and pMP19 the total gene number, soft core genes, shell genes and cloud genes all decreased after homopolishing (Figure 4.3). The number of core genes increased after polishing across the main genome, pMP1 and pMP19. The Hopkins' statistic for the shell genes of pMP1 was 0.5440197 and for the main genome 0.5820848, both suggesting that there was very weak clustering of the samples using the shell genes. A Hopkins statistic of >0.5 is classed as clusterable data, but the closer to 1 the value is the more highly clustered the data is (Banerjee & Davé, 2004). No Hopkins' statistic for the shell genes of pMP19 was reported as too few genes were identified.

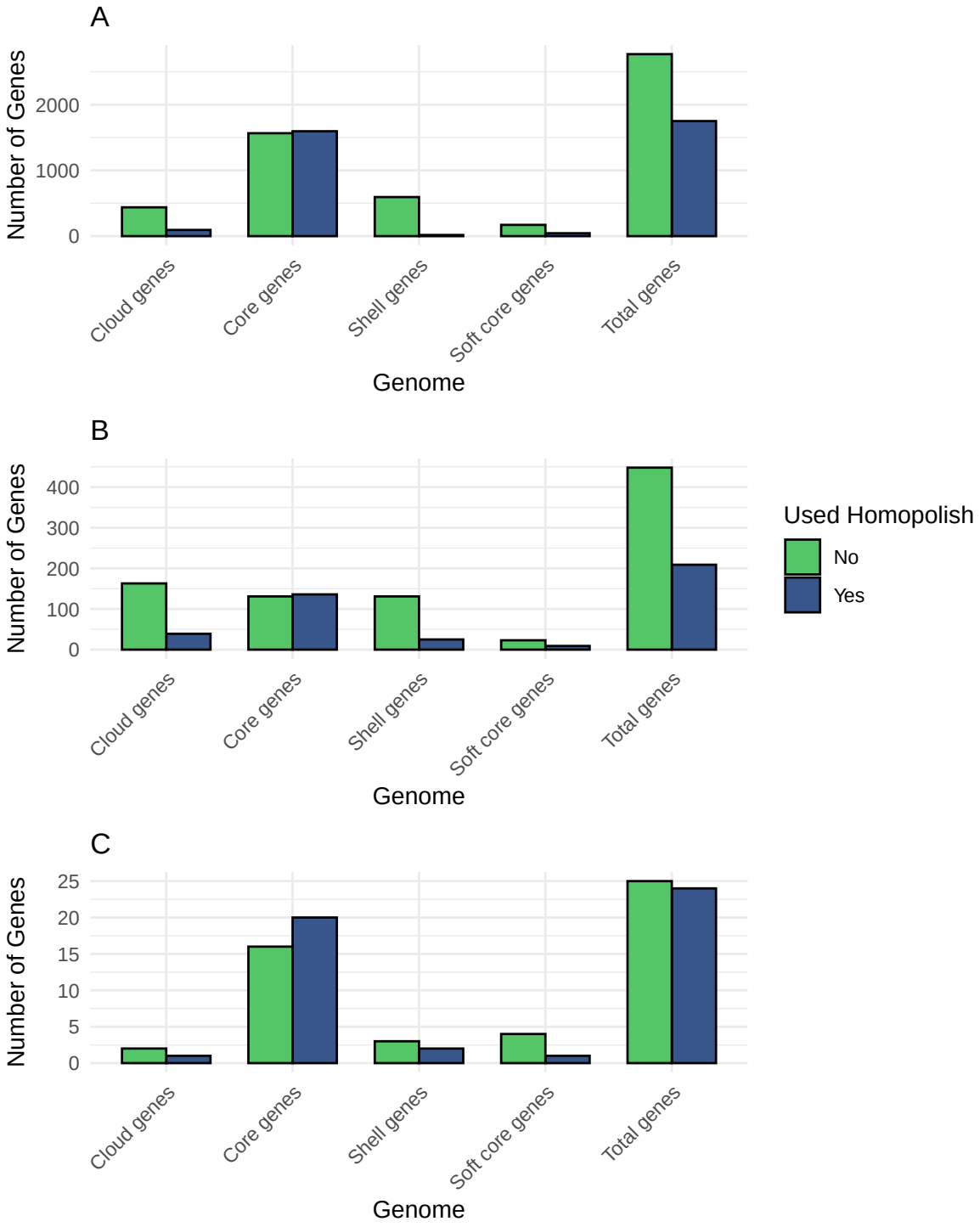


Figure 4.3: The number of genes for each gene type calculated using the panaroo analysis for both raw consensus as an input, no, and for consensus sequences after homopolishing, yes. A) Genes present in pMP1 B) Genes present in pMP19 C) Genes present in the chromosomal DNA

4.3.1.3 Virulence genes and AMR genes

Potential AMR and virulence genes were found in the main genome (Chromosome) and a potential AMR gene was found in pMP1 (Table 4.3). The majority of putative virulence genes were detected in all samples (Table 4.3). E2381, an ST7 sample from Greater London, was the only sample to contain hasC. The absence of VanRL on the pMP1 plasmid was more variable from different ST 2,3,5,6,7,13,23 samples, and some samples that failed traditional MLST typing, from fifteen different counties spread across the UK. When looking at the genetic sequence for each sample there was no genetic difference between the potential virulence genes clpP, cspA, cspB, hasC and psaA. For cap8E there was a mutation present, 618 T>C, in the samples that were ST39, ST40 and ST13, this was a silent mutation.

Table 4.3: The AMR and virulence genes found through the Abricate analysis across the main *M. plutonius* genome (Chromosome), pMP1 and pMP19

Found in	Gene	Type	Average Coverage (%)	Average ID (%)	Database found	Samples found in (%)
Chromosome	efrA	AMR	98	76	Megares and CARD	100
Chromosome	efrB	AMR	95	78	Megares and CARD	100
Chromosome	(MLS)mph(D)	AMR	98	77	Argannot	100
pMP1	VanRL	AMR	77	70	Meagres, Argannot, CARD and NCBI	79
Chromosome	Cap8E	Virulence	85	75	vfdb	100
Chromosome	clpP	Virulence	94	78	vfdb	100
Chromosome	cspA	Virulence	96	74	vfdb	100
Chromosome	cspB	Virulence	98	70	vfdb	100
Chromosome	hasC	Virulence	91	71	vfdb	99
Chromosome	psaA	Virulence	88	72	vfdb	100

4.3.2 Analysis of secondary invaders and gut microbiome bacteria

4.3.2.1 Percentage reads in sequencing compared to percentage disease in the colony in field

There was no correlation between the percentage of disease observed in the colony to the percentage reads of *M. plutonius* across all sequence types (Figure 4.4).

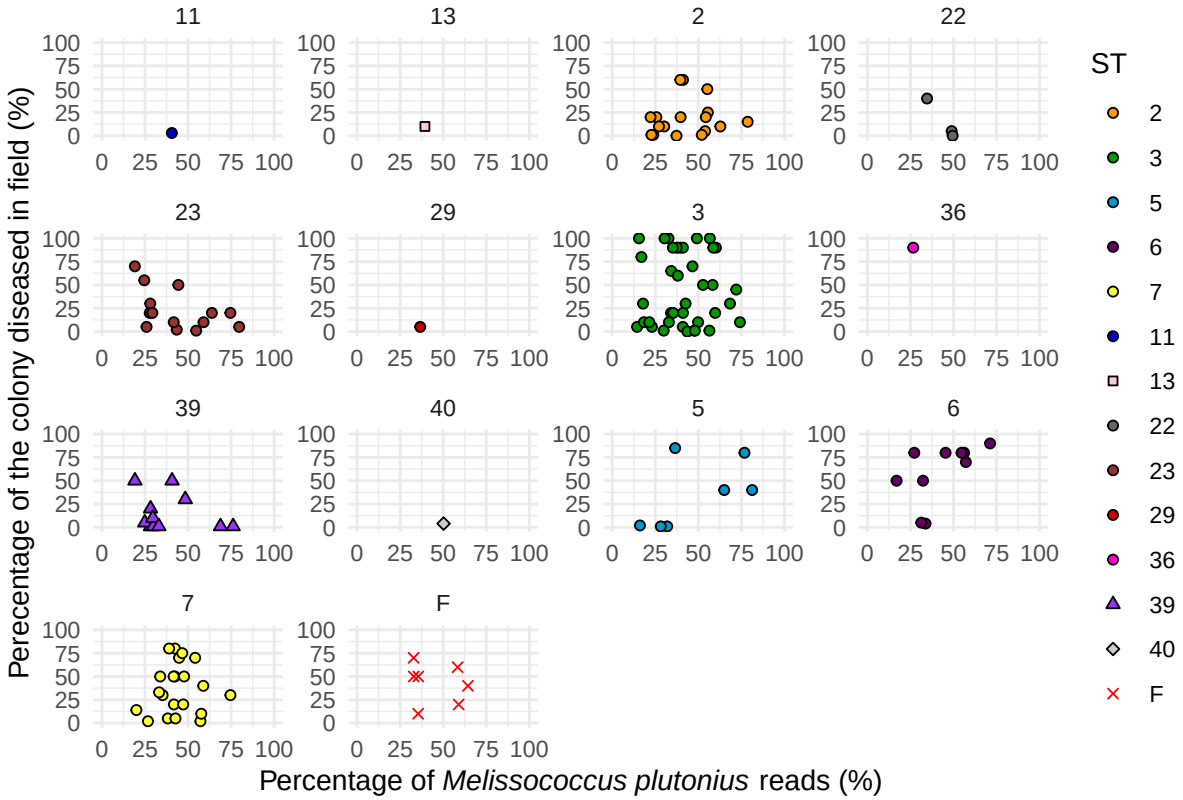


Figure 4.4: The percentage of *Melissococcus plutonius* reads from the whole genome sequencing compared to the percentage of colony diseased when observed in the field.

4.3.2.2 Assessment of genome coverage of each bacteria

The presence of each bacterial species differed across samples based on the percentage of the genome coverage for each species (Figure 4.5). The secondary invader bacterial species, *Bacillus plumilus* and *Brevibacillus lacterosporus* were not found in any of the 136 samples. The bacterial species associated with the adult honey bee gut microbiome, *Frishella perrara*, *Gilliamella apicola*, *Lonsdalea iberica* and *Snodgrassella alvi* were not found in any of the samples. Small traces of *Bombella apis* and *Paenibacillus dendritiformis* were found in a small selection of samples. For *Apilactobacillus kunkeei* the percentage genome

coverage was very varied across all the samples, the median value was approximately 85%, suggesting this bacteria was present in many of the samples. Most samples had no or small traces of *Bifidobacterium asteroides*, but one sample had more than 75% genome coverage. A few samples had a high genome coverage for *Enterococcus faecalis* with a majority having a small trace with little variation. There was a large variation of percentage genome coverage for *Fructobacillus fructosus* with the median being quite low suggesting most samples only had traces. A majority of samples only has small traces of *Lactobacillus apis* but there was a selection of samples that had a high percentage genome coverage. For *Paenibacillus alvei* there was not much variation in percentage genome coverage, but some samples had a high percentage. None of the samples had a trace of *Paenibacillus larvae*, the causative agent of American foulbrood (AFB).

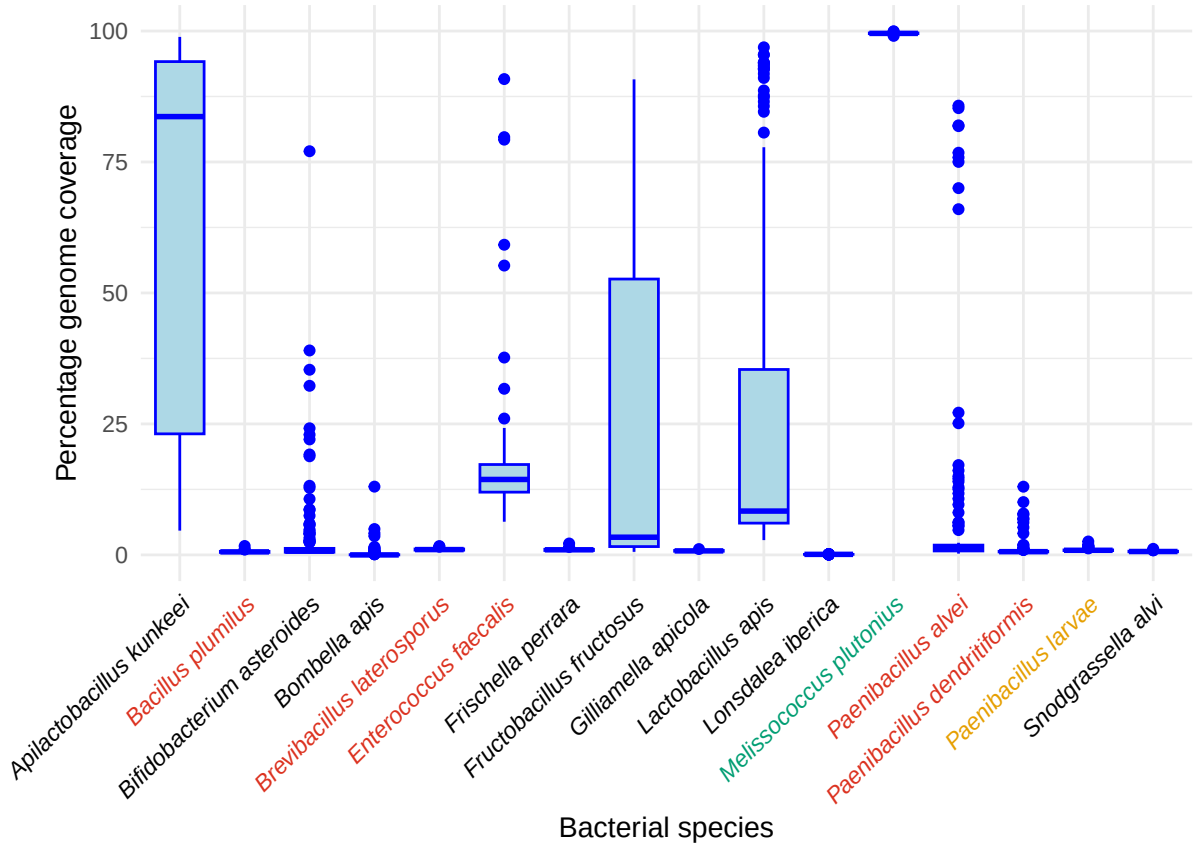


Figure 4.5: A box plot showing the percentage genome coverage of each bacterial species across all 136 samples analysed. The species in black are naturally found in the honey bee gut microbiome, the red are secondary invaders that follow *Melissococcus plutonius* infection, the green is *M. plutonius*, the species of interest, and the yellow is the causitive agent of another foulbrood disease, AFB.

4.3.2.3 Percentage composition of each bacterial species sequencing reads

The percentage reads for each bacterial species and *A. mellifera* showed to vary in levels across samples (Figure 4.6). The secondary invader bacterial species, *B. plumilus*, *P. dendritiformis* and *B. lacterosporus* were either absent or present in low percentages across the 136 samples. The bacterial species associated with the honey bee gut microbiome, *B. apis*, *F. perrara*, *G. apicola*, *L. iberica* and *S. alvi* were also found to be absent or present in low percentages. For a majority of the samples *A. kunkeei* was only present in a low percentage of reads, but some of the samples were made up of more than 40% of *A. kunkeei* reads. All samples had low percentage reads of *B. asteroides* apart from one sample that had over 10% of total reads being identified as *B. asteroides*. Most samples had a low trace of *E. faecalis*, but some samples had over 10% of the reads identified as *E. faecalis*. A majority of the samples had low percentage reads of *F. fructosus*, but a selection of samples had between 5% and 35% of the total reads identified as *F. fructosus*. The percentage reads for *L. apis* was low in a high number of the samples, but some did have percentage reads that varied between 5% and 45%. The percentage reads for *P. alvei* was the same as for *F. fructose* and *L. apis* that most samples were low, but some were present in a high percentage. The variation of percentage reads for *M. plutonius* varied on average between 30% and 55%, with a small number of samples present in higher than this thershold and some lower. A majority of the samples had *A. mellifera* reads below 20% of the total reads, but there were samples that had as high as 55%. None of the samples had reads of *P. larvae*. The composition of each sample (Figure 4.7) showed that there was a huge variation across samples.

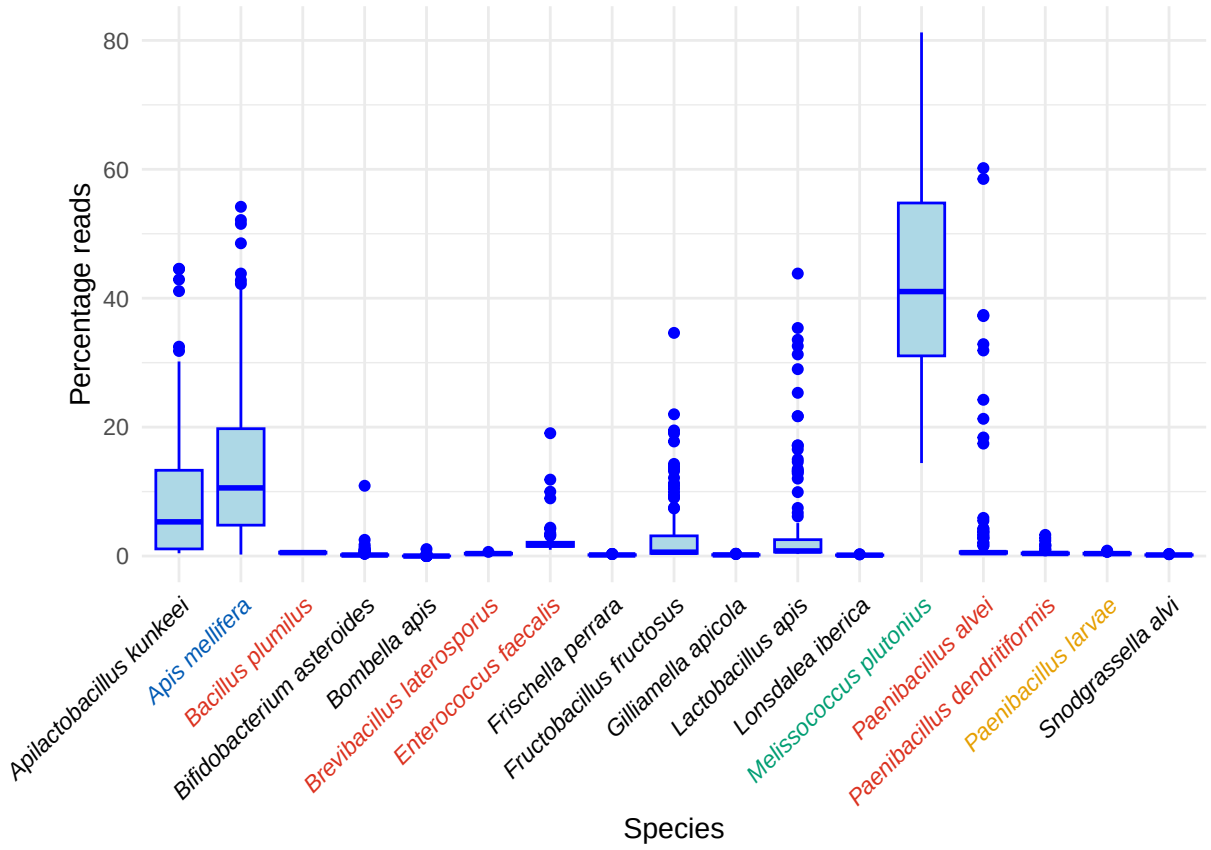


Figure 4.6: A box plot showing the percentage of the total reads of each bacterial species across all 136 samples analysed. The species in black are naturally found in the honey bee gut microbiome, the red are secondary invaders of *Melissococcus plutonius*, the green is *M. plutonius*, the species of interest, the yellow is the causitive agent of another foulbrood disease, AFB and the blue is honey bee reads

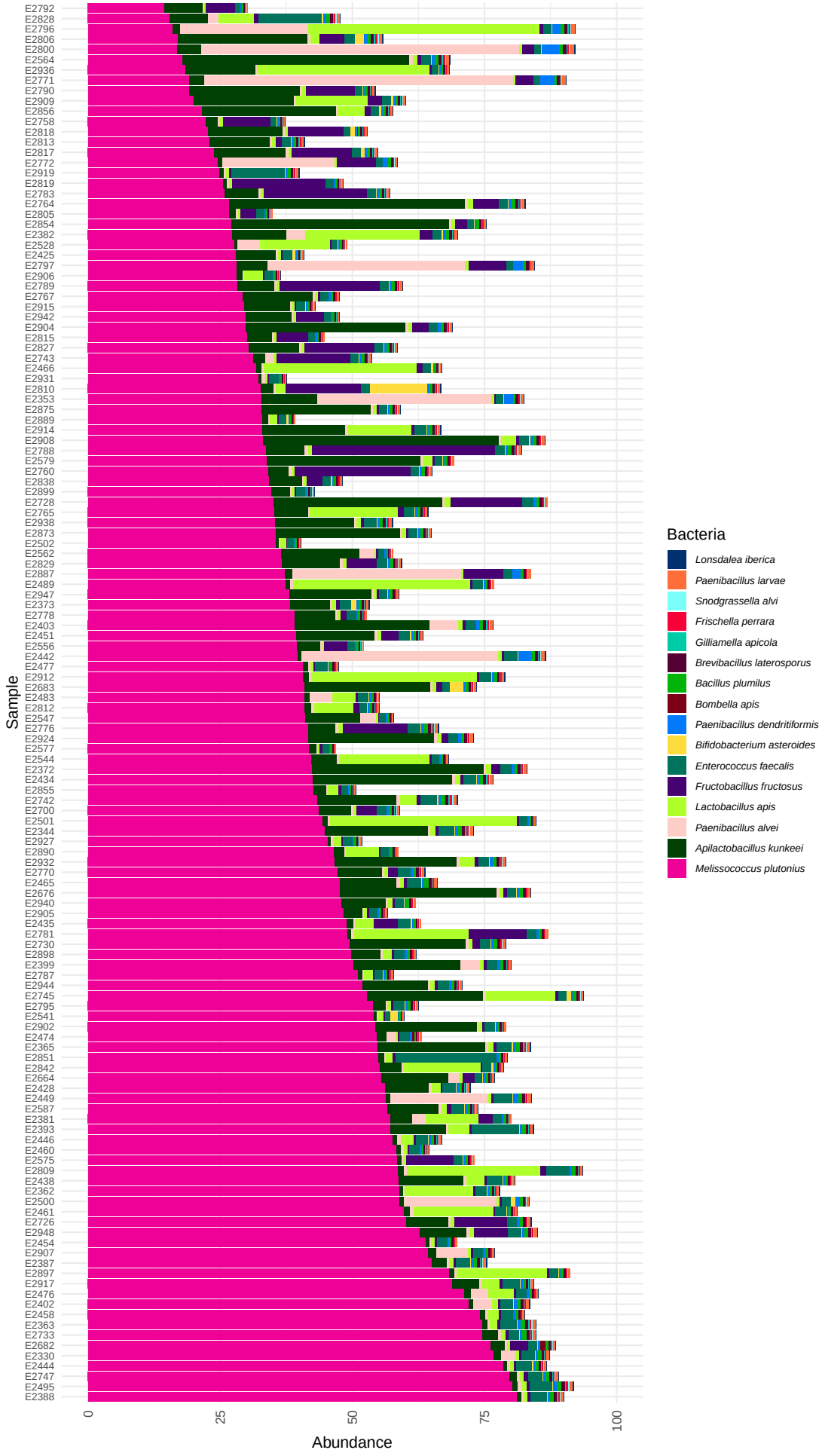


Figure 4.7: A stacked boxplot highlighting the composition of bacterial DNA for each of the 136 samples.

4.3.2.4 Clustering of the bacterial composition across all samples

The Hopkins' statistic for the clustering was 0.904912, showing there is a cluster tendency. The visual cluster analysis suggested that there were three boxes (Figure 4.8). The elbow method showed the elbow to be present at 3 clusters (Figure 4.9). The PCA analysis suggested that one of the samples overlapped between cluster 1 and 2, and that approximately 77% of the variation was explained by these three clusters (Figure 4.10).

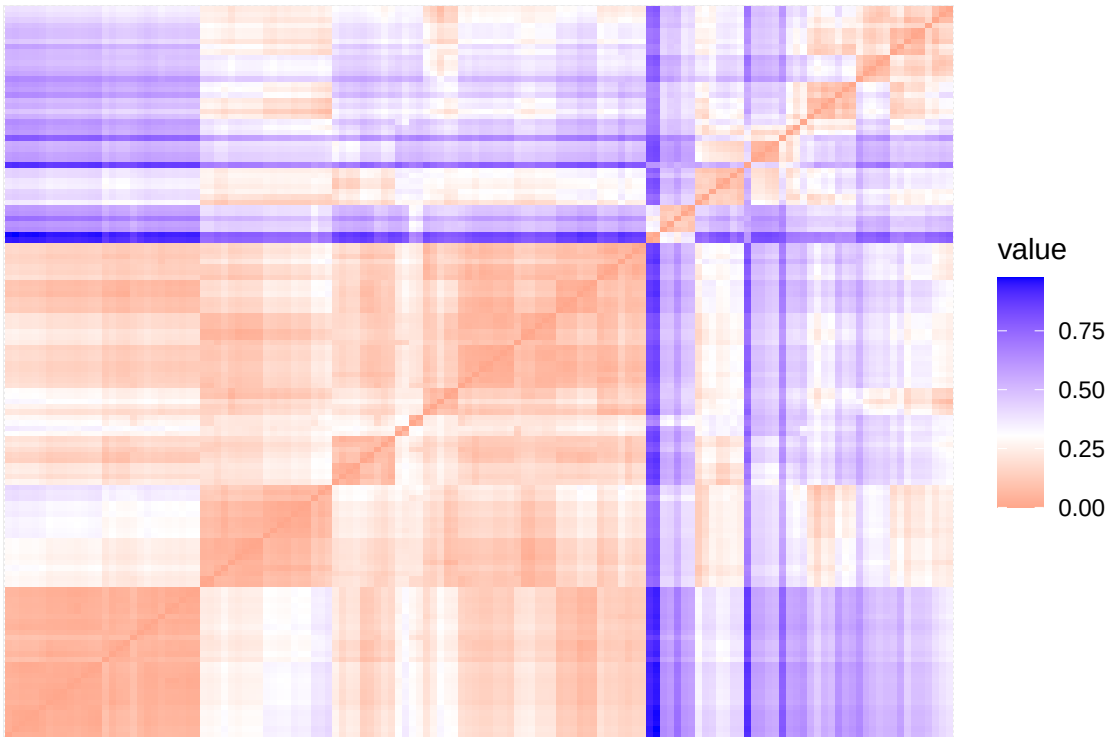


Figure 4.8: A visual assessment of cluster tendancy of bacterial compostion across the 136 samples.

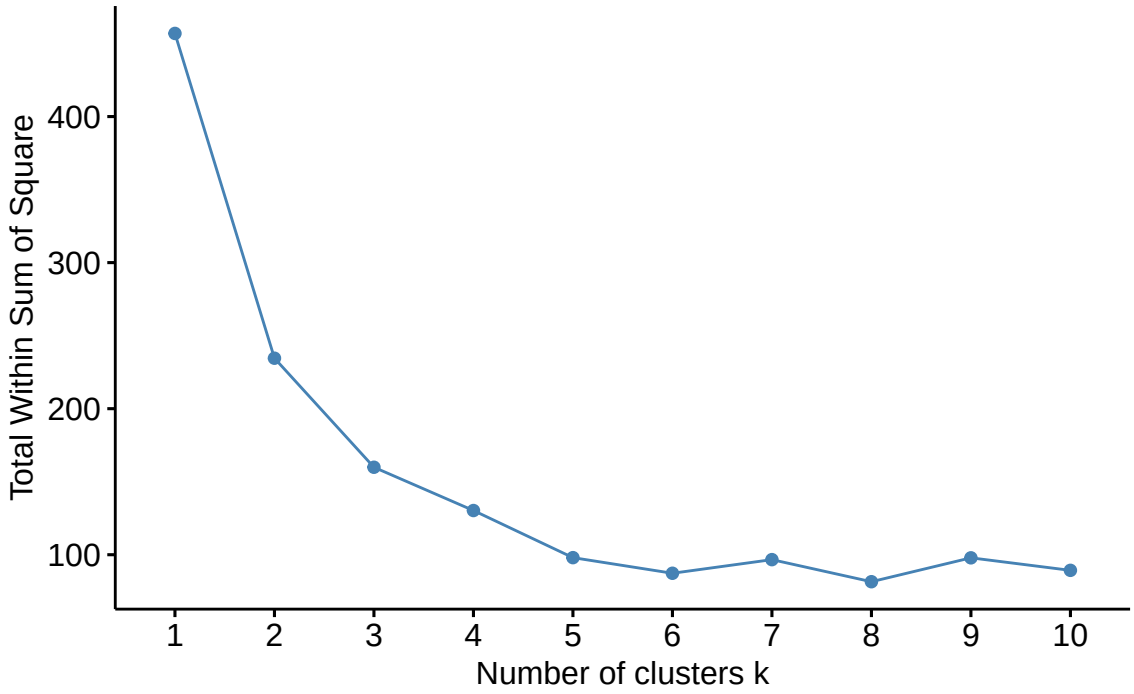


Figure 4.9: The optimal number of clusters comparing the total within-cluster sum of square for bacterial composition clusters.

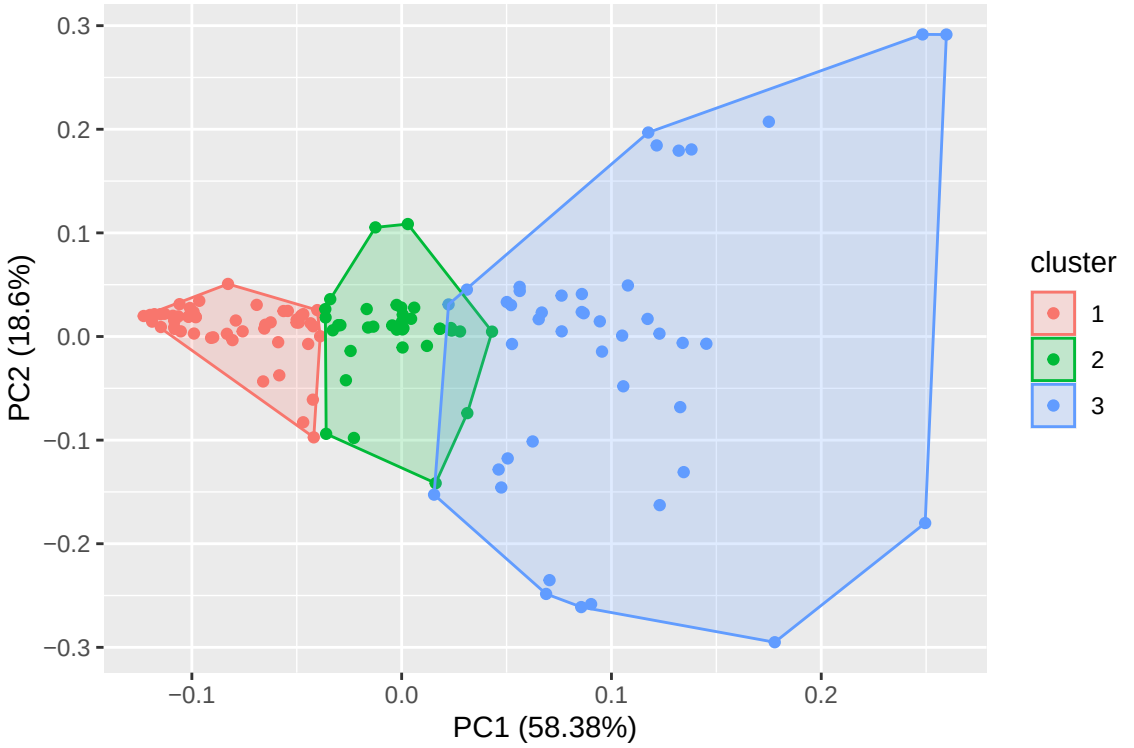


Figure 4.10: A PCA visualising the bacterial composition clusters

4.3.2.5 Bacteria composition anlaysis of the clusters

There was a higher percentage of *M. plutonius* in cluster 1 and the lowest percentage in cluster 3 (Figure 4.11).

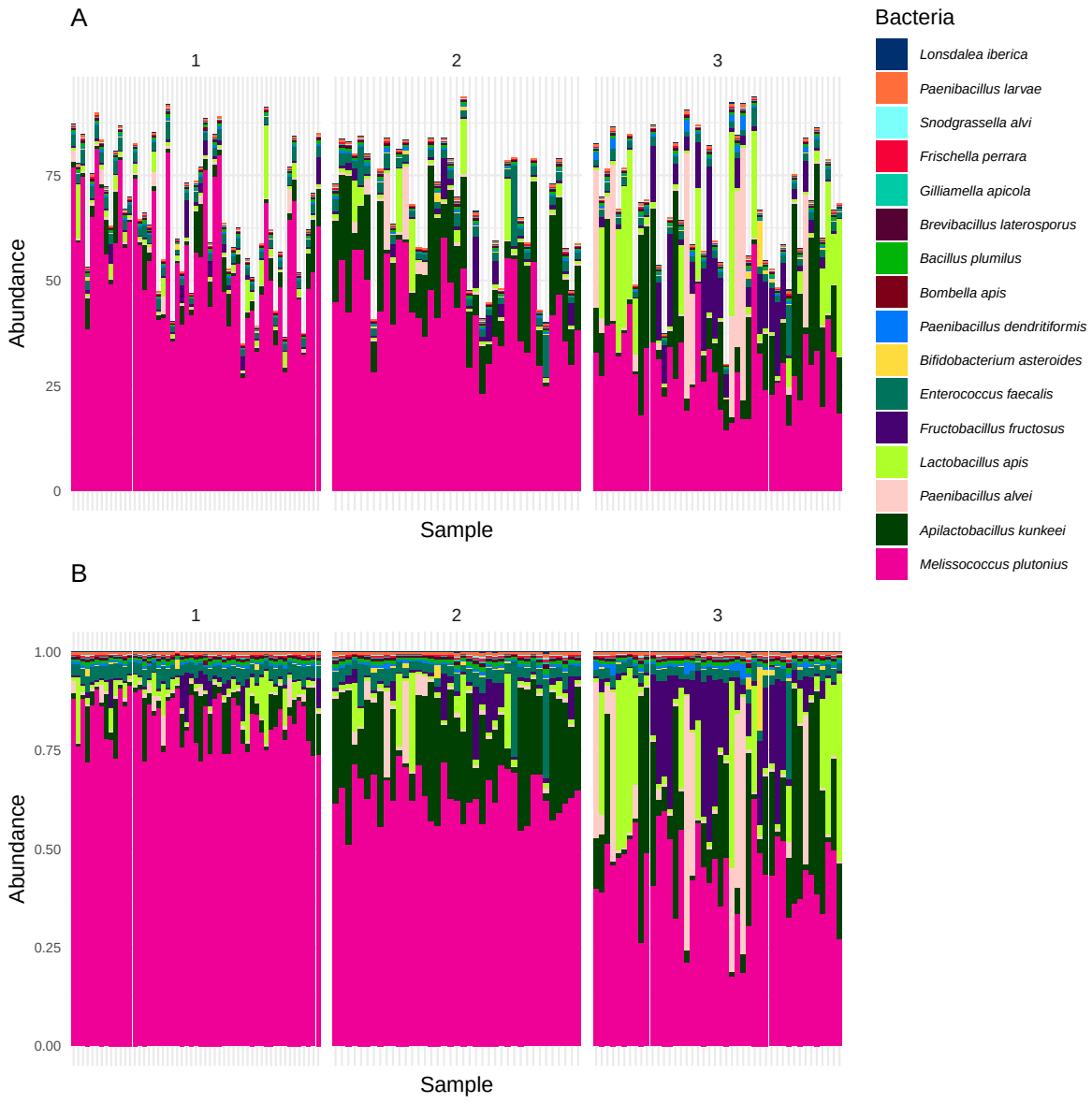


Figure 4.11: Stacked boxplots highlighting the composition of bacterial species for each of the 136 samples. The samples are grouped by the clusters assigned in the cluster analysis. A) The composition as a percentage of the total sequencing reads. B) The composition as proportion of the sum of the reads for the selected bacterial species.

There was no clear geographical clustering of the three species composition clusters (Figure 4.12). There was some clustering within counties, in North Yorkshire there are only samples in cluster 3 present and in Kent only cluster 2. There was no clear grouping of ST types within the three species composition clusters (Table 4.4).

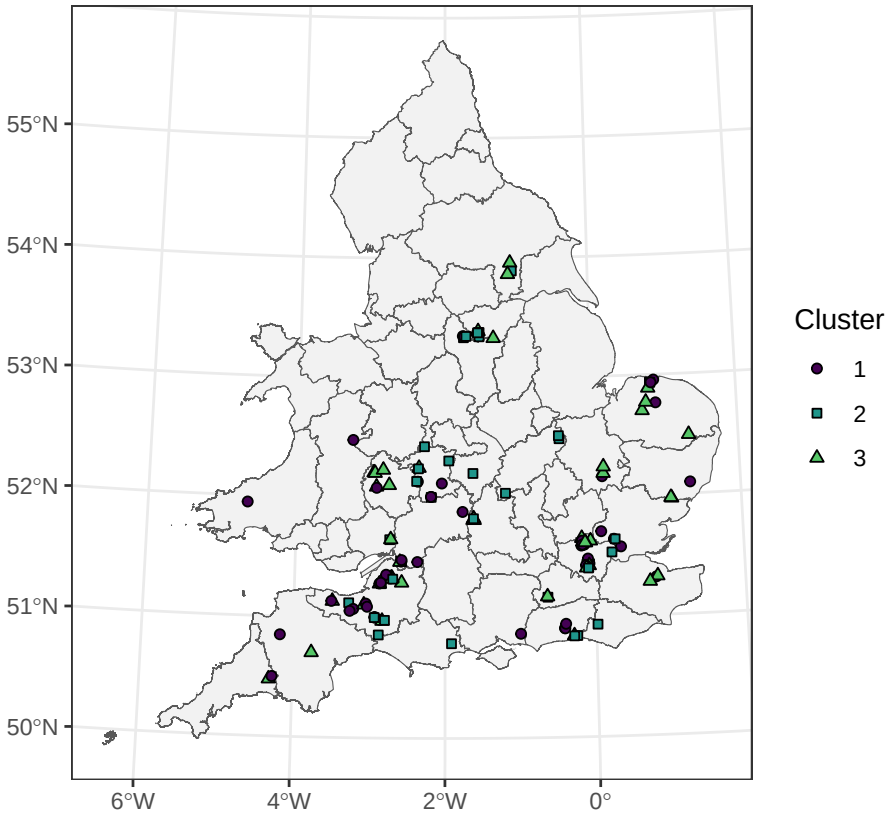


Figure 4.12: A map of England and Wales, highlighting the geographic location of each cluster of bacterial compostion.

Table 4.4: A table highlighting the ST types present in each of the bacterial compostion clusters

ST	Cluster 1 samples (n)	Cluster 2 samples (n)	Cluster 3 samples (n)
2	6	4	7
3	12	12	14
5	3	1	3
6	5	1	4
7	10	7	4
11	1	0	0
13	0	1	0
22	2	1	0
23	4	5	5
29	0	1	0
36	0	0	1
39	4	2	5
40	1	1	0
F	4	2	1
U	2	1	0
Total	54	39	44

4.3.2.6 Bacterial compostion for each month of inspection

The average percentage reads of *M. plutonius* peaked decreased between May and August and increased in September (Figure 4.13 and Figure 4.14). The percentage reads of *A. kunkeei* was highest in June, and *P. alevei* peaked in August.

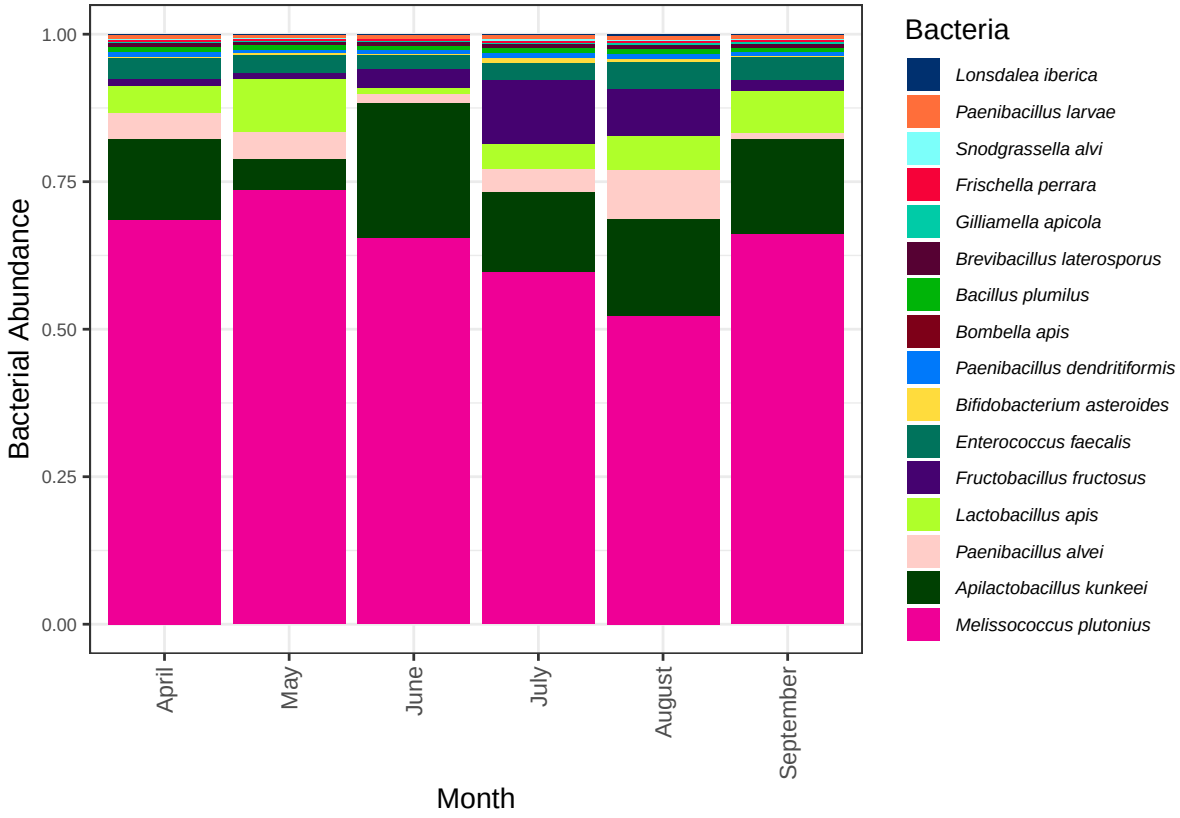


Figure 4.13: A stacked boxplot highlighting the average composition of bacterial speices DNA for each month of inspection.

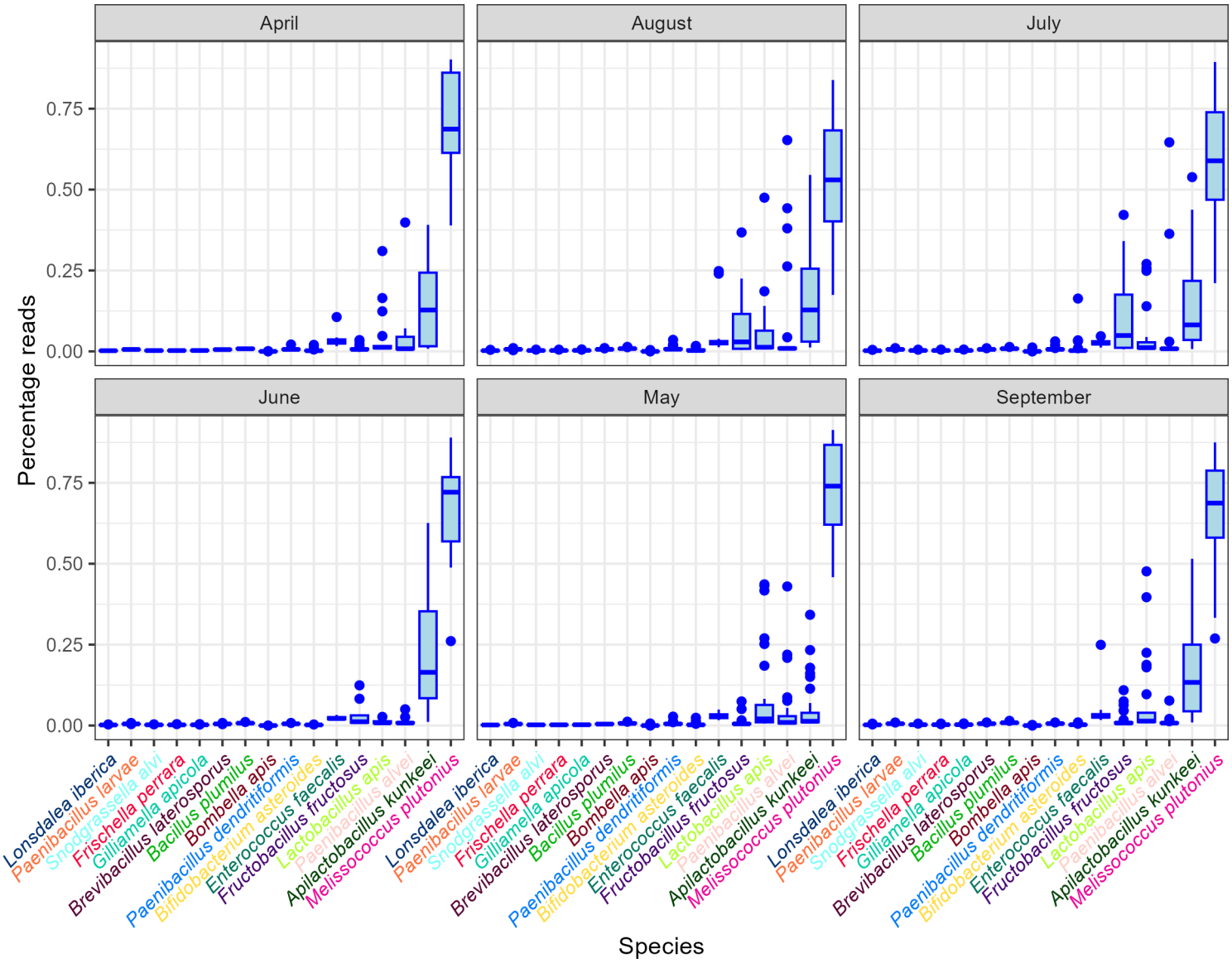


Figure 4.14: A box plot showing the percentage of the total reads of each bacterial species across all 136 samples analysed for each inspection month. The colours coordinate with the bar plots.

4.3.2.7 AMR gene anlaysis in bacteria found alongside *Melissococcus plutonius*

The samples with more than 75% of the other bacterial species were taken forward for AMR analysis. The samples chosen were 74 samples containing *A. kunkeei*, 1 sample containing *B. asteroides*, 3 samples containing *E. faecalis*, 27 samples containing *F. fructosus*, 23 samples containing *L. apis* and 7 samples containing *P. alvei*. No AMR genes were identified in *F. fructosus*, *A. kunkeei* and *L. apis*. Potential AMR genes were found in *P. alvei* (Figure 4.15), and one in *B. asteroides*, when compared to the AMR genes found in *M. plutonius* none of the genes were the same. AMR genes were also found in the samples containing *E. faecalis* (Figure 4.16). For *E. faecalis*, all three samples contained the (MLS)mph(D) and the efrB genes also found in *M. plutonius*, section 4.3.1.3. Two of the samples containing *E. faecalis* also had the efrA gene, the same as *M. plutonius*.

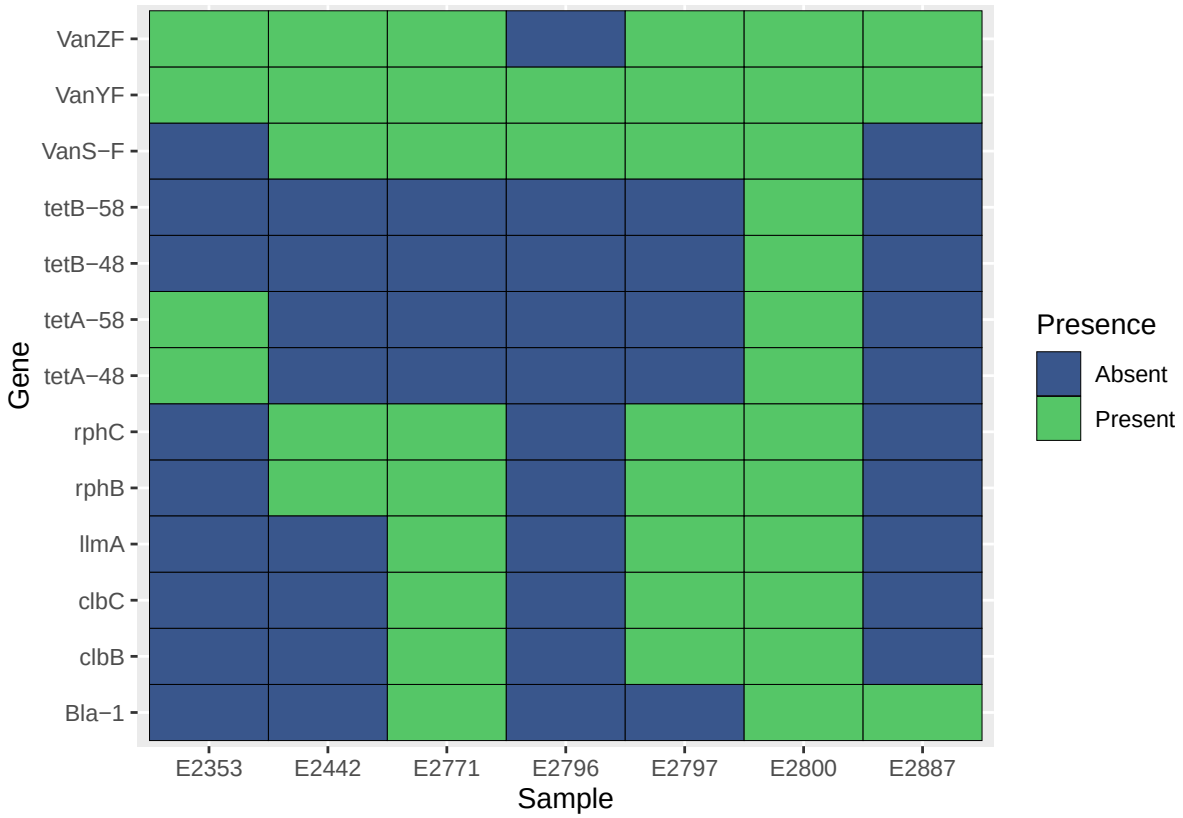


Figure 4.15: A presence and absence matrix showing the antimicrobial resistant genes found in the secondary invader *P. alvei*, that was found in the presented samples alongside *Melissococcus plutonius*.

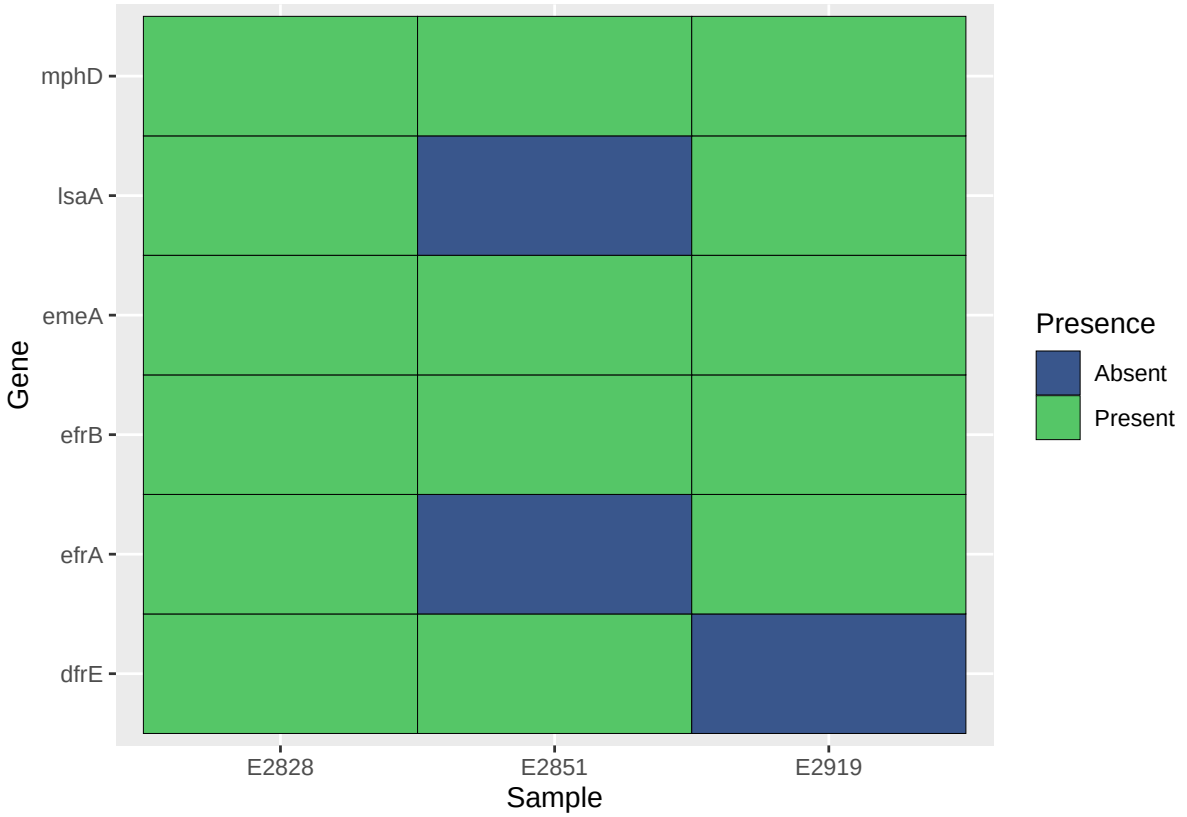


Figure 4.16: A presence and absence matrix showing the antimicrobial resistant genes found in the secondary invader *E. faecalis*, that was found in the presented samples alongside *Melissococcus plutonius*.

4.4 Discussion

Here I present potential virulence genes and ARGs found using a sequencing pipeline already developed for the purpose of studying EFB epidemiology. I have also gathered more information about the secondary invaders present in real EFB positive larval samples using whole genome sequencing (WGS). These findings show the potential of WGS for understanding more about the aetiology of EFB.

The Homopolish package appeared to improve the gene annotations for the *M. plutonius* main genome and the plasmids. When inputting the raw consensus sequences into the Panaroo analysis there were more total genes found. This is likely to be a result of fragments of genes caused by premature stop codons, like those highlighted in the MtxA gene. Premature stop codons do exist in bacteria and can contribute to the evolution of the organism by changing a genes function (Wong et al., 2008). A homopolymer of Ts were present slightly upstream from the MtxA gene. As there is a known issue of the MinIon R9.4 flow cell with inaccurate reading of homopolymers (Xian et al., 2022;

Delahaye & Nicolas, 2021), it is likely in this case that misreading of homopolymers is the cause of the premature stop codons. A study by De Vivo et al using Nanopore sequencing also had the same issue there were a lot more stop codons in the minion data than other sequencing techniques (Vivo et al., 2022). As this is the likely cause Homopolish was used to try and minimise this issue. There were fewer total genes after Homopolish suggesting the fragments have been resolved, there were also more core genes and fewer shell core, shell and cloud genes, suggesting a lot of the noise was removed. Homopolish also completed all the MtxA genes leaving zero fragments present suggesting excellent data cleansing. Although there was a potential issue with Homopolish when it comes to access to representative genomes. Homopolish uses similarity to known sequences, so in this case it was published sequences of *M. plutonius* which there are 27 published records. However, there are 46 sequence types listed on pubMLST, so not all STs are represented in the database (NCBI, 2021; Jolley et al., 2018). Partial representation of STs means there is a risk that correct homopolymers may be 'corrected', even though they are genuine deletions or insertions simply because they were absent in the original database. An example is that the difference between ST29 and ST3 when performing sequence typing is a deletion in a polyA region, using Homopolish this is likely to call them as the same sequence type (Jolley et al., 2018).

The pMP19 plasmid and the MtxA gene was found to be present in all samples. Previous studies in Nakamura et al did not detect pMP19 in all samples across the three clonal complexes. Previous work suggested that the presence of pMP19 increased the virulence of sequence types belonging to CC3, a majority of the samples in this study were CC3 (87%) (Nakamura et al., 2020). These results may suggest that the UK isolates have a high virulence. The MtxA gene was also not found on all samples, unlike this study. The presence of MtxA was tested using conventional PCR on cultured bacteria from infected sites (Grossar et al., 2023). The primers used were aligned to the MtxA regions of the samples in this study to assess if they annealed across the possible premature stop codon regions, they all joined to the first fragment. As the primers joined to the first fragment, even if the premature stop codons were real the PCR would still be positive for my samples, so there was no false negatives. Although Grossar et al found that MtxA was absent in some samples in Switzerland, as samples were taken between 2005 and 2006, and also samples taken in 2010, they did find an increase in the number of samples with MtxA between the 2005 and 2006 samples and the 2010 ones (Grossar et al., 2023). As

my samples were collected in 2020, 10 years after these samples the MtxA gene could be driving spread and therefore could explain why MtxA is present in all the samples in this study across England and Wales. To date, studies on MtxA have only covered some sequence types, the Swiss study only included ST3 and 7 that overlapped with my samples and Nakamura et al only included a few samples for each clonal complex, CC12, CC13 and CC3 (Grossar et al., 2023; Nakamura et al., 2020). All these studies were also carried out on cultured *M. plutonius*, and it has been shown that the plasmid can be lost through multiple passages (Arai et al., 2012), whereas the samples in this study were all straight from the infected larvae. As well as MtxA some other virulence genes were found.

There were six potential virulence genes found, Cap8E, clpP, cspA, cspB, hasC and psaA, all apart from hasC were detected across all samples. Cap8E is a gene that codes for a type 8 capsular polysaccaride synthesis protein and was found in *Staphylococcus aureus*, as part of a capsule forming operon (Takeuchi et al., 2005). As it is a capsule forming protein this would help the *M. plutonius* evade the immune system, but only one of the genes has been found in this sample but could indicate a similar operon is present in *M. plutonius* (Johnson, 2018). ClpP is a gene that codes for a serine protease and was found in *Listeria monocytogenes*. The protease allows the pathogen to continue growing under stress therefore it is only switched on under stress (Gaillot et al., 2000). *Listeria* is a Gram-positive pathogen that colonises in the gut, so could share similar genes to *M. plutonius* (Barbuddhe & Chakraborty, 2009). CspA and CspB are also capsule forming virulence genes, found as a virulence gene in *Enterococcus faecalis* a bacteria that is in the same family as *M. plutonius*. CspA and CpsB are not essential for the capsule formation, but CpsC-K are and have not been detected in this study (Hancock & Gilmore, 2002). This again could suggest a similar pathway exists in *M. plutonius*. HasC has been found to be a virulence gene in both *Bacillus* and *Streptococcus* species, and produces UTP–glucose-1-phosphate uridylyltransferase, to create a hyaluronic capsule to evade the immune system of the host (Ferretti et al., 2001; Okinaka et al., 1999). In *Bacillus* it is associated with the plasmid region (Okinaka et al., 1999). The gene hasC is the last gene in an operon for capsule formation, but was shown that hasC was not essential (Ashbaugh et al., 1998; Crater et al., 1995). Without the other genes *M. plutonius* would not be able to produce a capsule, but genes similar to hasA and B could exist to form the operon. The final gene detected, psaA, is a gene that along with psaB and psaC make up an ABC transport system for metals, protecting the bacteria from oxidative stress within the host,

found in *Streptococcus* species (Novak et al., 1998). The other genes psaB and C were not found in *M. plutonius*, but similar genes could be to make up this operon. It appears that all the virulence genes found were part of operons, with counterparts not found in *M. plutonius*. The absent genes could suggest that the operons could have existed in *M. plutonius* but were more divergent than the species they were initially detected in, further work could involve using other methods to explore these genes. The sequences of the virulence genes were compared across the samples.

The sequence comparison of the virulence genes across all the samples showed a mutation in cap8E although this was silent it was present in all ST13, ST39, and ST40 samples. There was also a silent mutation in MtxA that was only present in the same three ST types. These mutations could suggest that evolutionary these three sequence types are related (Barraclough et al., 2012). The other two mutations detected in MtxA were only present in ST39, one was silent but the other mutation, 704A>G, changed the amino acid from glutamic acid to glycine. Glutamic acid is a negatively charged, polar amino acid and can play a major role in protein structure, for example it can contribute to salt bridges, stabilising the protein (Brosnan & Brosnan, 2013; Hendsch & Tidor, 1994). Glycine is the most basic of the amino acids, it is non-polar, and it lacks a side chain making it small in size (Hall, 1998). The change from a polar to a non-polar animo acid could affect the structure of the toxin produced by the MtxA gene, and therefore affect its function. This mutation information also suggests that ST39 is genetically distinct from the other sequence types, but closer related to ST13 and ST40. As well as virulence genes ARGs were detected too.

Three ARGs detected on the chromosome, efrA, efrB and (MLS)mph(D) and one of pMP1, VanRL. The genes efrA and efrB were detected in all samples and make up a EfrAB efflux pump, which function by pumping out the antibiotics. Initially detected in *Enterococcus sp*, it was shown that in the presence of certain antibiotics, expression of EfrAB increased, suggesting resistance to these (Lerma et al., 2014). The antibiotics were gentamicin, streptomycin and tochloramphenicol. Gentamicin is usually used to treat Gram-negative infections in humans, but is commonly used in pigs (Chaves & Tadi, 2023; Tams et al., 2023). Streptomycin is a broad-spectrum antibiotic targeting the 30s region of the ribosome like OTC, and has been used as a pesticide in the past (Lewis et al., 2016). Tochloramphenicol is also a broad-spectrum antibiotic and has been banned from use in food production (Hanekamp & Bast, 2015). The EfrAB genes were also detected in the *E.*

faecalis present in two of the samples, suggesting a possible gene transfer between the two species. The ARG, (MLS)mph(D) is an acquired macrolide–lincosamide–streptogramin B gene, and codes for an enzyme that inactivates the antibiotics (Hoek et al., 2011). Macrolide and lincosamides are commonly used in pigs and cattle (Pyörälä et al., 2014). As these antibiotics have been used in agriculture exposure of *M. plutonius* to these could have occurred through the environment. The (MLS)mph(D) gene was detected in all three of the samples containing *E. faecalis*, again suggesting a transfer event. The ARG detected on pMP1, VanRL, is a VanR gene found in the VanL cluster, and makes the pathogen resistant to vancomycin (Boyd et al., 2008). Vancomycin is naturally made by a bacterium found in soil and used to treat various human diseases (Levine, 2006). Other vancomycin resistance genes were found in *P. alvei* but not the same ones as *M. plutonius*. As most of these antibiotics are used in human medicine, it is concerning that these genes may get into the environment. There were other ARGs detected in both *P. alvei* and *E. faecalis*. For *P. alvei* this included genes that cause resistance to vancomycin, rifamycin, licosamide and penicillins, again all used in human medicine (Fraimow et al., 2005; Materon et al., 2003; Pawlowski et al., 2016). Interestingly, a couple of the genes detected in *P. alevi*, tetA and B, cause resistance to tetracyclins, the group OTC belongs to (Pawlowski et al., 2016). There is no evidence to show that OTC was used on the apiaries the samples containing the tet genes came from, so resistance may have come from the environment, however historic exposure to OTC cannot be ruled out. As well as ARGs the bacterial composition may be important for studying EFB.

There was no clustering of bacterial composition between counties or ST types, suggesting that this differs between apiaries and individuals. It has been previously reported that there was a significant difference between the bacterial species found between apiaries infected with EFB, like in this study, including the percentage of *M. plutonius* (Anderson et al., 2023). Over time the percentage of *M. plutonius* was at the lowest in June, July and August, this could suggest that the larva were more susceptible to disease during these months, as all samples presented with the disease. This study did not find any traces of the natural gut bacteria, *F. perrara*, *L. iberica*, *G. apicola* and *S. alvi*, but *F. fructosus* was present in some of the samples. In some infected samples it was shown that the levels of these natural gut bacteria increased when in the presence of *M. plutonius* (Anderson et al., 2023). The apparent absence of these four bacteria may be due to methodological bias, because the sequencing pipeline was primarily designed to sequence *M. plutonius*

a Gram-positive bacteria. The gut bacteria not detected were all Gram-negative, but the *F. fructosus* was Gram-positive (Table 4.5). The DNA extraction method used a lysis buffer particularly targeting the cell wall of Gram-positive bacteria, and the host depletion method may act differently with Gram-negative bacteria, as found by Ganda et al when host depletion between Gram-positive and negative was compared (Ganda et al., 2021). As a result of the bias of the method, only Gram-positive bacteria can be assessed using this method which includes all the reported secondary invaders. One secondary invaders that has not been included was *Achromobacter eurydice* this was because a reference genome was not found, however there was evidence to suggest that it could possibly be *A. kunkeei*, which is usually found in the honey bee gut and was found in most of the samples in this study (Lewkowski & Erler, 2019; Erler et al., 2018). The secondary invaders *P. alvei*, *E. faecalis* and *L. apis* were found in some of the samples, as found historically in samples with *M. plutonius*. None of the samples contained the other reported secondary invaders, *B. plumilus* and *B. laterosporus* (Forsgren, 2010). It is still unclear how and if these secondary invaders contribute to the development of EFB (Stanisavljević et al., 2023). The percentage reads of *P. larvae* were investigated as this causes another foulbrood disease, AFB, none of the samples had traces of this pathogen, and according to records none of the samples tested had reports of an AFB infection, so it is unknown from this study if AFB can also be detected and studied.

Table 4.5: Details of the Gram stain of each of the gut microbiome and secondary invader bacterial species

Bacteria	Gram stain
Paenibacillus alvei	Gram-positive
Apilactobacillus kunkeei	Gram-positive
Bacillus plumilus	Gram-positive
Bifidobacterium asteroides	Gram-positive
Bombella apis	Gram-negative
Brevibacillus laterosporus	Gram-positive
Enterococcus faecalis	Gram-positive
Frischella perrara	Gram-negative
Fructobacillus fructosus	Gram-positive
Gilliamella apicola	Gram-negative
Lactobacillus apis	Gram-positive
Lonsdalea iberica	Gram-negative
Snodgrassella alvi	Gram-negative
Paenibacillus larvae	Gram-positive
Paenibacillus dendritiformis	Gram-positive

Future work could include modifying the sequencing pipeline to incorporate Gram-negative bacteria to give a clearer picture of bacterial composition. What would also be interesting is to investigate whether the *M. plutonius* composition is correlated with the disease severity. During inspections a number is given for disease progression, this could be assessed against the percentage reads for *M. plutonius*. Testing samples with known AFB infection would investigate whether this sequencing pipeline can be used to study this disease as well as EFB. Looking at the correlation between the reads from the other bacteria and percentage of disease in the field could also give an indication if any of the other pathogens are driving or potential causing presentation of EFB symptoms. To conclude, more information was gathered from the whole genome pipeline designed for investigating phylogeny of *M. plutonius*. Potential virulence genes were identified in the infected samples, suggesting similar virulence operons to other pathogens may exist in *M. plutonius*, as well as the previously reported MtxA gene. Some AMR genes were identified in *M. plutonius* but none of them caused resistance to OTC, so they are likely to have come from either exposure of antibiotics in the environment, horizontal gene transfer from other organisms or historic exposure to antibiotics from other sites. Secondary invader Gram-positive pathogens were identified in some of the samples, expanding the capabilities of whole genome sequencing.

Chapter 5

Regulation and management to improve the control of EFB

5.1 Introduction

European foulbrood (EFB) is a bacterial honey bee disease caused by *Melissococcus plutonius*, that infects the honey bee at the larval stage (Bailey, 1983). After three to five days of infection the larvae tend to die most never reaching pupation (Forsgren, 2010). The clinical signs of this disease include an irregular brood pattern, healthy brood has uniform cells that have been capped, whereas a diseased colony has a scattering of cells that are both capped and uncapped, know as a pepper pot patterning (Bailey, 1983). The appearance of an infected larva depends on infection stage and age. When a young larva is infected, it becomes transparent, whereas older larvae become twisted. The colour changes throughout the infection, healthy larvae are shiny white, an infected larva transitions to a light yellow and eventually a brown, dark scale in the cell (Forsgren et al., 2013). In the UK EFB is a notifiable disease, meaning that if a bee keeper has any suspicion of an infection in their apiary, they are legally bound to report it to the government local bee inspector (Government, 1982). Once disease has been reported, a bee inspector will inspect the colony for symptoms and use a lateral flow device (LFD) that contains a monoclonal antibody that is specific to *M. plutonius*, to confirm EFB. Should field diagnostics remain undetermined, a sample is sent to the laboratory to test for the presence of the causative bacteria using microscopic and real-time PCR tests (Tomkies et al., 2009). There are three methods of control available for EFB control in the UK, the

use of Oxytetracyclin (OTC), a husbandry method called shook swarm, where adult bees are transferred to a clean hive, and destruction which is only used in the most severe cases of disease. Shook swarm involves shaking the bees into a new colony and then scorching the infected hive (Waite et al., 2007). This action is disruptive to the beekeeper and in worst case scenarios the entire colony could be destroyed, an estimated replacement costs for a hive after EFB infection was reported by a beekeeper as £571.99 (NBU, 2013). A risk-based inspection programme uses a traffic light system to assign risk of EFB to each apiary registered on a national database known as Beebase (NBU, 2024). Green (low risk) is over 10 km away from an infected colony, amber (medium risk) is between 3 km and 10 km, and red (high) is below 3 km. Inspectors from the NBU can also be called out by the beekeeper to visit colonies because of suspicion of disease for additional training and advice (Budge et al., 2012). Usually when an apiary tests positive for EFB, apiaries within 3 km are inspected (NBU, 2013). Promoting bee health is important to reduce the impact of honey bee diseases like EFB on the honey bee population.

The National Bee Unit (NBU) are responsible for delivering programmes to promote bee health on behalf of the government in England and Wales, to manage and control bee diseases (Wilkins et al., 2007). The NBU run the apiary inspection program, this program trains bee inspectors to test for EFB and another foulbrood disease, American foulbrood (AFB) and deploy relevant control. These are the bee inspectors called out when EFB is suspected (Thompson & Brown, 1999). The NBU run a website (www.nationalbeeunit.com) which contains training resources for beekeepers as well as information about legislation, and live disease reports for EFB for each 10 km square (NBU, 2024). There are also historic disease records available. These resources allow the beekeepers to monitor disease in their areas. Beekeepers can also register for disease alerts if EFB and AFB are found in the area surrounding their apiaries. Being registered on BeeBase allows the NBU to have a record of beekeepers in England, Wales and Scotland, allowing disease to be controlled and monitored (Thompson & Brown, 1999). The BeeBase website contributes towards the Healthy Bee Plan 2030 which is a 10-year £23 million flagship government policy to improve honey bee health (Defra & Government, 2021).

Published in 2020, the healthy bee plan 2030 is an implementation plan to improve honey bee health created by the Department of Environment Food and Rural Affairs (DEFRA) and the Welsh government. The plan aims to provide more effective biosecurity, good

husbandry standards, enhance the beekeepers and farmers skill sets, provide scientific evidence for actions taken on supporting bee health and to create more opportunities for knowledge exchange. Some of the initiatives in the plan involve encouraging beekeepers to partake in a basic assessment for beekeeping, to encourage all beekeepers to sign up to beebase and a review of the NBU literature provided on their website (Defra & Government, 2021). An extensive library of educational resources for beekeepers are provided by the NBU through their website, in forms such as leaflets and videos (NBU, 2024).

Educational resources targeted to the beekeepers are important tools in the management of EFB. The detection of EFB relies on the recognition of EFB by the beekeeper to report it to the local inspectors, so they need to have the ability to recognise EFB. It was shown that beekeepers with more experience and more education in beekeeping are less likely to get disease in their apiaries, therefore training and educational resources are a valuable prevention tool (Jacques et al., 2017). Through the British beekeepers association (BBKA) beekeepers can take written examinations in different areas of beekeeping, called modules. There are seven modules that can be taken (Table 5.1). The BBKA also carry out practical assessments that give the beekeeper qualifications when passed, these include basic assessment covering the basics of beekeeping, honey bee health, general husbandry and advanced husbandry (BBKA, 2024). For each module and assessment there is a link to recommended reading containing resources from the BBKA themselves as well as a magazine called BeeCraft and the NBU (NBU, 2024; BeeCraft, 2024). BeeCraft has articles that can be purchased, on topics such as swarming and American Foulbrood recognition (BeeCraft, 2024). BeeBase contains a page on diseases and pests, including information about how to spot Foulbrood and links to other resources such as an advisory leaflet on Foulbrood disease but also more generic guides on best practice and hive cleaning (NBU, 2011, 2017, 2018). A survey investigating the reason for colony loss across Europe highlighted that as well as the education level of beekeepers, beekeeping practice, played a key role in colony loss (Jacques et al., 2017).

Table 5.1: A list of the modules that can be examined by the BBKA.

Module Number	Module
1	Honey Bee Mangement
2	Honey Bee Products and Forage
3	Honey Bee Pests, Diseases and Poisioning
5	Honey Bee Biology
6	Honey Bee Behaviour
7	Selection and Breeding of Honey Bees
8	Beekeeping and Honey Bee Management

Good husbandry practices such as high hygiene within the apiary is thought to prevent disease spread both between colonies and apiaries. Disease can be spread through the equipment like the hive tools, so ensuring tools are washed with washing soda either by scrubbing or leaving to soak between colonies is important. Wax or propolis can contain pathogens such as bacteria, so making sure any residues are removed from brood boxes using a flame torch or scraping them off is good practice, as well as ensuring no residues remain on the smoker. Disease can also spread through clothing, when starting beekeeping for the day beekeepers should ensure that their bee suit and boots are clean and washed regularly (NBU, 2018). Ensuring gloves are clean before commencement of beekeeping is also important to minimise disease spread, and disposable gloves worn over the reusable gloves is recommended as those can be disposed of immediately, decreasing risk further. *M.plutonius* has been found in the hive debris of EFB positive hives, so ensuring that hives are properly cleaned and disinfected after use prevents disease spreading to the new colonies (Mckee et al., 2003; NBU, 2011).

Honey bee swarming is a common occurrence, particularly during the summer months. Swarms occur when a queen along with some worker bees leave the colony in order to establish another colony elsewhere. There are many factors that can cause swarming, the age distribution of worker bees in the colony, congestion, colony size or disease (Winston,

1987; Zacepins et al., 2021). The BBKA have a webpage for members of the public to report swarms and find a local volunteer that will collect them (BBKA, 2024). When a swarm is collected it is usually from an unknown source so could be carrying diseases or pests, *M. plutonius* has been found on adult honey bees and is known to survive in honey. In Australia, nearly 70% of all bulk honey samples contained detectable levels of *M. plutonius* and the pathogen was found to be viable by culture in nearly 30% of samples (Mckee et al., 2003). Infected honey poses a risk of disease spread as honey bees will rob honey if they have access to it, for example from honey barrels or jars that have not been sufficiently cleaned and sent to be disposed of or leaks in shipping containers (Mutinelli, 2011). Prior to swarming, honey bees gorge on honey until their crops are full and use this energy resource to help build their new nest (Winston, 1987). It was found that beekeepers who bought swarms were more likely to experience colony loss, disease may have had an impact on this (Jacques et al., 2017). How swarms are housed is key in reducing the risk of any carried diseases passing on to neighbouring apiaries or colonies. It was shown that infected swarms housed on fresh foundation eliminated EFB, but on drawn comb the disease emerged (Russel et al., 1937). This is because husbandry practices which encourage the honey bee to use their honey resources before brood is present are less likely to spread disease to the newly developed larvae. As such, the addition of drawn frames would increase risk because the time between the swarm arriving in the new equipment and producing larvae is reduced (Winston, 1987).

There is a complex interaction between the epidemiology of EFB, population density of bees and honey bee management. Management is associated with manipulating colonies, and this is dependent on human behaviour as well as the regulatory framework of disease management. The interaction of these processes likely drives disease risk, but behaviours of beekeepers have not been quantified thus making it difficult to articulate the epidemiology for EFB. Beliefs have an impact on behaviour, for example during the COVID-19 pandemic people who believed in conspiracies theories were less likely to get the vaccine, and comply with the recommendations from the government (Earnshaw et al., 2020). We need to investigate human behaviour, beliefs and disease within the regulatory framework to provide a more complete understanding of EFB epidemiology. Here I create a conceptual model of EFB epidemiology, that includes beekeepers' beliefs and behaviours, using survey data and modelling methodologies, that can account for the complexity and relative significance of contrasting factors that contribute to

disease. I included survey data that requested information about beekeepers' beekeeping experience, education, husbandry practise and their direct experience with EFB. The survey was sent to two areas with a high EFB risk, Cambridgeshire and Somerset. A conceptual model was challenged with the survey data, using structural equation modelling (SEM). SEM is a method that quantifies the relationship between multiple different observed variables or latent variables, through multiple structural equations (Kline, 2005). Latent variables are not measured or observed directly, but are interpreted by multiple observed variables, for example education of a beekeeper is a latent variable that could represent number of BBKA modules taken and related literature read (Thakkar, 2020). The survey data were used to challenge the conceptual model using SEM to better understand the contribution of beekeeping behaviours and beliefs on EFB epidemiology.

5.2 Methods

5.2.1 Cambridgeshire and Somerset disease incidence

To assess the disease occurrence over time for each county inspection data was obtained from Beebase (NBU, 2024). The number of apiaries inspected between 2006 and 2023 for each county was presented, as well as the number of EFB positive cases found during those inspections. The number of cases per 100 inspections was calculated and presented.

5.2.2 Survey

A survey was produced to ascertain the views of beekeepers about EFB, to understand their experience and education to investigate their behaviours around biosecurity and swarm collection, and to highlight any shift in response to contact from the NBU. The survey was prepared on the JISC online survey system (https://newcastle.onlinesurveys.ac.uk/european-foulbrood-in-somerset) and the link for completion sent to membership secretaries of local beekeeping associations around Somerset and Cambridgeshire for distribution to their membership. These regions were chosen because of their high disease risk for EFB. Key summary statistics were presented.

5.2.3 Conceptual model for SEM

A conceptual model was generated to hypothesise how beekeeper experience, behaviours and beliefs interacted with the epidemiology of EFB across the two counties, Somerset and Cambridgeshire (Figure 5.1). The conceptual model was populated with three latent variables, swarm practices, education and experience and biosecurity practices. The swarm practices latent variable was predicted to be explained by the risk of house swarming methods, whether the beekeeper collected swarms and the distance travelled to collect the swarms. It was predicted that the education and experience latent variable was explained by the number of beekeeping qualifications, the years of experience, the number of colonies owned and whether the beekepeer had read appropriate literature. It was predicted that the biosecurity latent variable was explained by the use of disposable gloves, whether disinfecting of hive tools occured and whether sharing of honey extracting equipment occured. It was hypothesised that the more education and experience a beekeeper has the more likely they were to practise low risk swarm and biosecurity practices, have been inspected by the NBU and perceived EFB as a problem. If a beekeeper perceived EFB as a problem or had seen foulbrood it was expected that they were more likely to practise low risk swarm and husbandry practices. When a beekeeper had been inspected by the NBU, I expected that they were more likely to had seen EFB, and therefore more likely to practise low risk swarm and husbandry practices, and to have perceived EFB as a problem. I predict that if a beekeeper had received an alert to say EFB is present in the area they were more likely to see EFB as a problem and exhibit improved biosecurity and swarm practices.

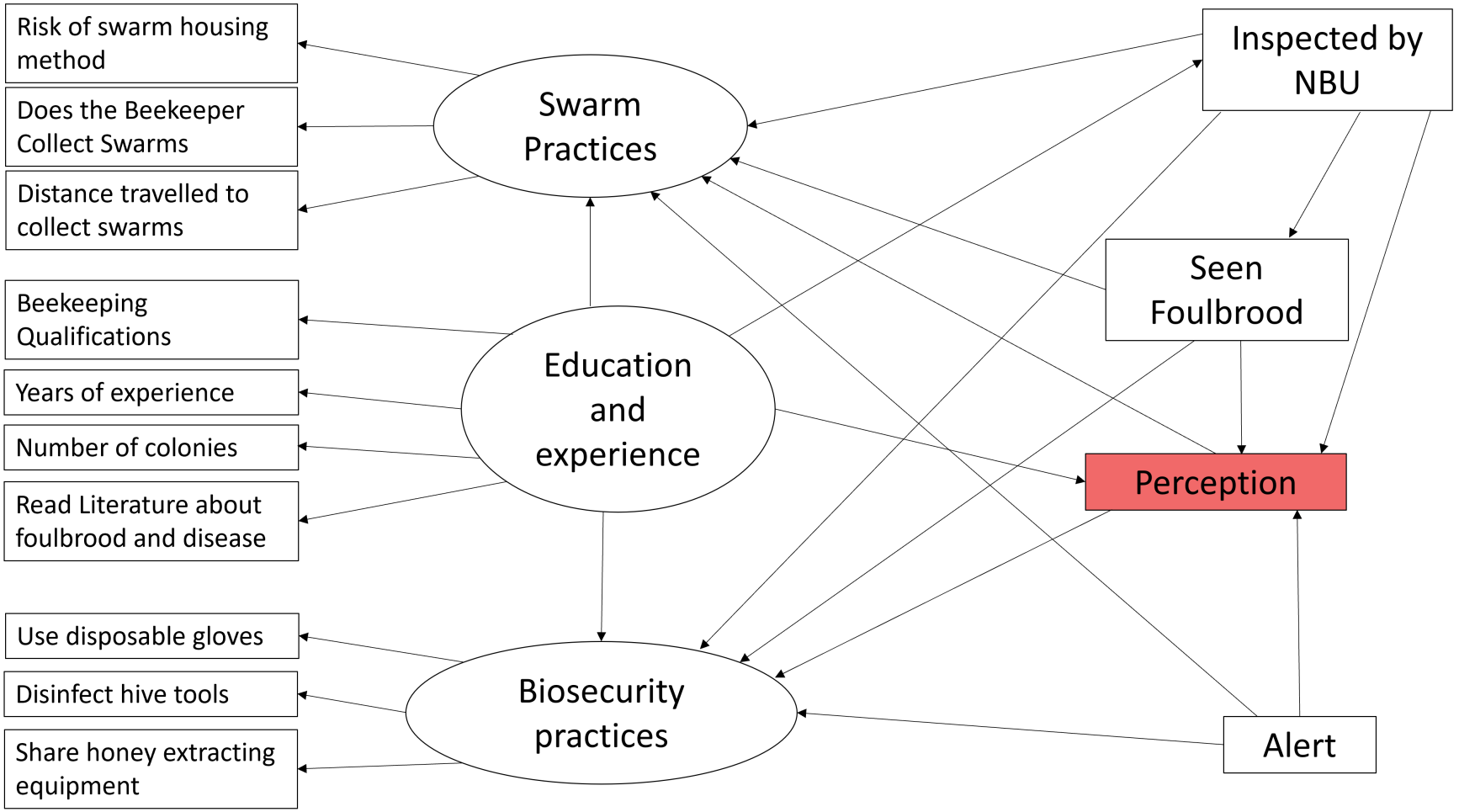


Figure 5.1: A Conceptual model of the variables impacting the perception of EFB and the behaviour of beekeepers.

5.2.4 Correlation assessment

To ensure all the variables in the conceptual model were not redundant a heterogeneous correlation matrix was generated in R (Version:4.2.1) using the hetcor function in the polycor package (Version:0.8-1) (Fox, 2022). A correlation matrix was produced and assessed for redundant variables. This was visualised using the corrr package (Version:0.4.4)(Kuhn et al., 2022) and the ggcorrplot (Version:0.1.4.1) using a correlation matrix and a correlation network (Kuhn et al., 2022). The matrix shows the correlation coefficient on a gradient to highlight the strength of the correlation between each variable. The network highlights relationships between the variables, the variables that are closer together in space are more correlated than those further apart.

5.2.5 Preparing data for SEM input

The data for each of the latent and observed variables was formatted for input to the SEM. The individual variables specified in the conceptual model (Figure 5.1) were from the answers to the questions specified in the survey (Table 5.2).

Table 5.2: The variables used in the models specifying which questions in the survey they came from

Survey section	Question number	Question	Variable in model
Beekeeping information	1	How many production colonies do you own?	Number of colonies
Beekeeping information	6	How many years have you kept bees?	Years of experience
Beekeeping information	8	Do you have any beekeeping qualifications?	Beekeeping Qualifications
Beekeeping information	10	Have you used any NBU literature for training/education?	Read Literature about foulbrood and disease

Table 5.2: The variables used in the models specifying which questions in the survey they came from

Survey section	Question number	Question	Variable in model
Questions about your husbandry methods	1	Do you collect swarms?	Does the Beekeeper Collect Swarms
Questions about your husbandry methods	1a	How do you house swarms?	Risk of swarm housing method
Questions about your husbandry methods	1b	In kilometres, what is the maximum distance you would travel from your apiary to collect swarms?	Distance to travelled collect swarms
Questions about your husbandry methods	6	Do you share any honey extraction equipment (honey spinner/uncapping tray/settling tanks)?	Share honey extracting equipment
Questions about your husbandry methods	8	Have your bees ever been inspected by a bee inspector form the National Bee Unit?	Inspected by NBU
Questions about your husbandry methods	9a	Have you had notification of EFB in the last 2 years?	Alert
Questions about your husbandry methods	10a	Do you use disposable gloves?	Use disposable gloves
Questions about your husbandry methods	11	Do you soak your hive tool in disinfectant in between apiary visits?	Disinfect hive tools
Questions about your experience with Foulbrood	1	Do you consider that EFB is a problem for beekeepers in your area?	Perception
Questions about your experience with Foulbrood	4	Have you ever seen European foulbrood in a honey bee colony?	Seen Foulbrood

Table 5.2: The variables used in the models specifying which questions in the survey they came from

Survey section	Question number	Question	Variable in model

5.2.5.1 Swarm practices input data

The Swarm practices latent variable (Figure 5.1) was produced by assigning high or low risk to each question. For example, the housing methods specified in the survey, *questions about husbandry-Q1a*, were given a risk assignment to generate a binary variable, where 0 was low and 1 was high (Table 5.3). The distance travelled variable was used as reported by the respondent in Km in *questions about husbandry-Q1b*.

Table 5.3: The risk assessment of the different swarm housing methods performed by beekeepers.

Housing Method	Risk Assignment
In a box containing only foundation	Low
In a box with drawn comb	High
In a box with honey and pollen	High
Initially in a quarantine apiary on their own	Low
Deliberately starve for 3 days	Low
Moved into a hive	High
Does not collect swarms	Low

5.2.5.2 Education and experience

For the education and experience variable (Figure 5.1) the number of beekeeping qualifications were summed, using the answers to the question *Beekeeping information- Q8*. The years of experience, *Beekeeping information- Q6*, and the number of colonies, *Beekeeping experience-Q1*, were used as reported. The read literature about foulbrood and disease variable was whether the respondent had read disease relevant literature, yes (1) or no (0), specifically the NBU documents: Foulbrood Disease of Honey Bees and Common Pests and Diseases and Disorders of the Adult Honey Bee, *Beekeeping information- Q10*.

5.2.5.3 Biosecurity practices

For the biosecurity practices (Figure 5.1), the use of disposable gloves, *Questions about your husbandry methods-Q10a*, was assigned as yes, 1 and no, 0, and the same was assigned to the disinfect hive tools, *Questions about your husbandry methods-Q11*. For the share honey equipment, *Questions about your husbandry methods-Q6*, the yes was assigned 0 and the no 1, this was to make this variable align with the other two. If the sharing equipment response was yes that would indicate a higher disease risk, but a yes response to wearing disposable gloves would be a lower risk practise.

5.2.5.4 Perception and experience with EFB

There were four individual variables (Figure 5.1), inspected by the NBU, seen foulbrood, perception and alert. Both inspected by the NBU, *Questions about your husbandry methods-Q8*, and seen foulbrood, *Questions about your experience with Foulbrood-Q4*, were kept as the reported yes and no, 1 and 0. The alert variable used was received an alert in the last two years, *Questions about your husbandry methods-Q9a*, and again was yes and no, 1 and 0. The perception variable is whether the respondent sees EFB as a problem, *Questions about your experience with Foulbrood-Q1*, yes or no (1 and 0). For all the above variables if the answer was unsure it was reported as no.

5.2.6 Running the SEM

The conceptual model (Figure 5.1) was challenged with the formatted data from the survey, using the R package Laavan (Version:0.6-16)(Rosseel, 2012). Two SEMs were produced, one for each county to assess the relationship between general practice, education, perception and experience with EFB. The standardised coefficients were reported on the conceptual model for each county. The Comparative Fit Index value was presented, this is a value between 0 and 1 that assesses the fit of the model, the closer to 1 the value the better the fit (Laar et al., 2021). The Root Mean Square Error of Approximation (RMSEA) was also presented, this measures the difference between the predicted values in the model to the actual values, ideally should be <0.05 for a good fit, and anything above 0.1 is a poor fit (Kim et al., 2016).

5.2.7 Literature search for swarm housing information

Based on the survey and SEM results a literature search was carried out, both using the internet and beekeeping books, to scope out what information is available to the beekeeper regarding swarm housing and linking swarm collecting with EFB.

5.2.8 Literature search for biosecurity practices linked to EFB

A literature search of the material available on the NBU website for beekeepers was performed. The fact sheets and information that links disease to biosecurity practices were analysed.

5.3 Results

5.3.1 Cambridgeshire and Somerset disease incidence

EFB was an emerging disease in Cambridgeshire, as there were very few cases per 100 visits by the NBU in Cambridgeshire between 2006 and 2015, from 2016 onwards EFB cases increased (Figure 5.2). In Somerset EFB was established, as there were cases of EFB per 100 visits across the whole time frame. The number of visits increased in 2009 across both counties and in Somerset the number of EFB cases in 2010 reached a trough.

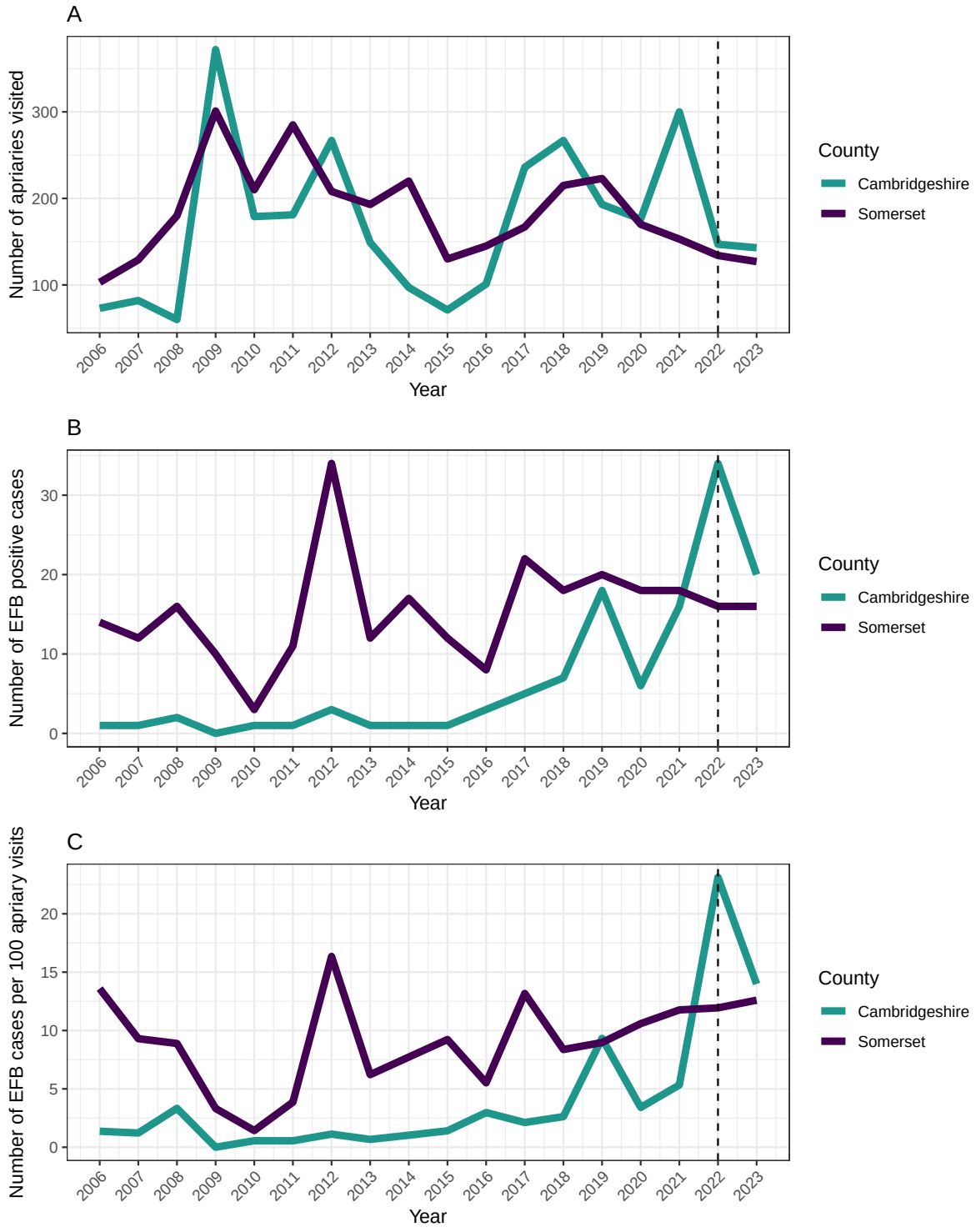


Figure 5.2: The EFB incidence and number of visits by the NBU between 2006 and 2023 for both Cambridgeshire and Somerset. The dashed line on all the three plots represents the year the survey was carried out A) The number of apiaries visited by the NBU per year for both counties. B) The number of EFB positive cases per year for both counties. C) The number of EFB cases per 100 visits by the NBU per year for both counties.

5.3.2 Survey

In total, 258 respondents returned a complete survey for Somerset, but only 254 were used for the analysis. Three responses were removed as the number of apiaries was over 100 and the number of colonies were less than 5, so appeared to be an error in filling out the survey so the results are not reliable. One response was removed as they stated they travelled more than 600 km to collect swarms, which is one third of the length of the UK, and so this appeared incorrect. In total there were 109 respondents for Cambridgeshire all surveys were used for subsequent analyses. Overall, 94% and 85% of respondents were registered with the NBU respectively from Somerset and Cambridgeshire. The full survey with all the responses for Somerset can be found in C.1.1. The full survey with all the responses for Cambridgeshire can be found in C.1.2.

5.3.2.1 Responses related to swarm practices

The majority of respondents across both counties collected swarms, and there was no significant difference in the proportion of yes and no between both counties (Figure 5.3). The most common swarm housing practises for both counties was using a box containing only foundation, which was assigned low risk for EFB (Figure 5.3). The second most common practice for both counties was housing in a box with drawn comb, assigned as high risk for EFB. Travelling between 5 km and 10 km to collect swarms was the most frequent response for both counties and not many respondents travelled less than 5 km for swarms (Figure 5.3).

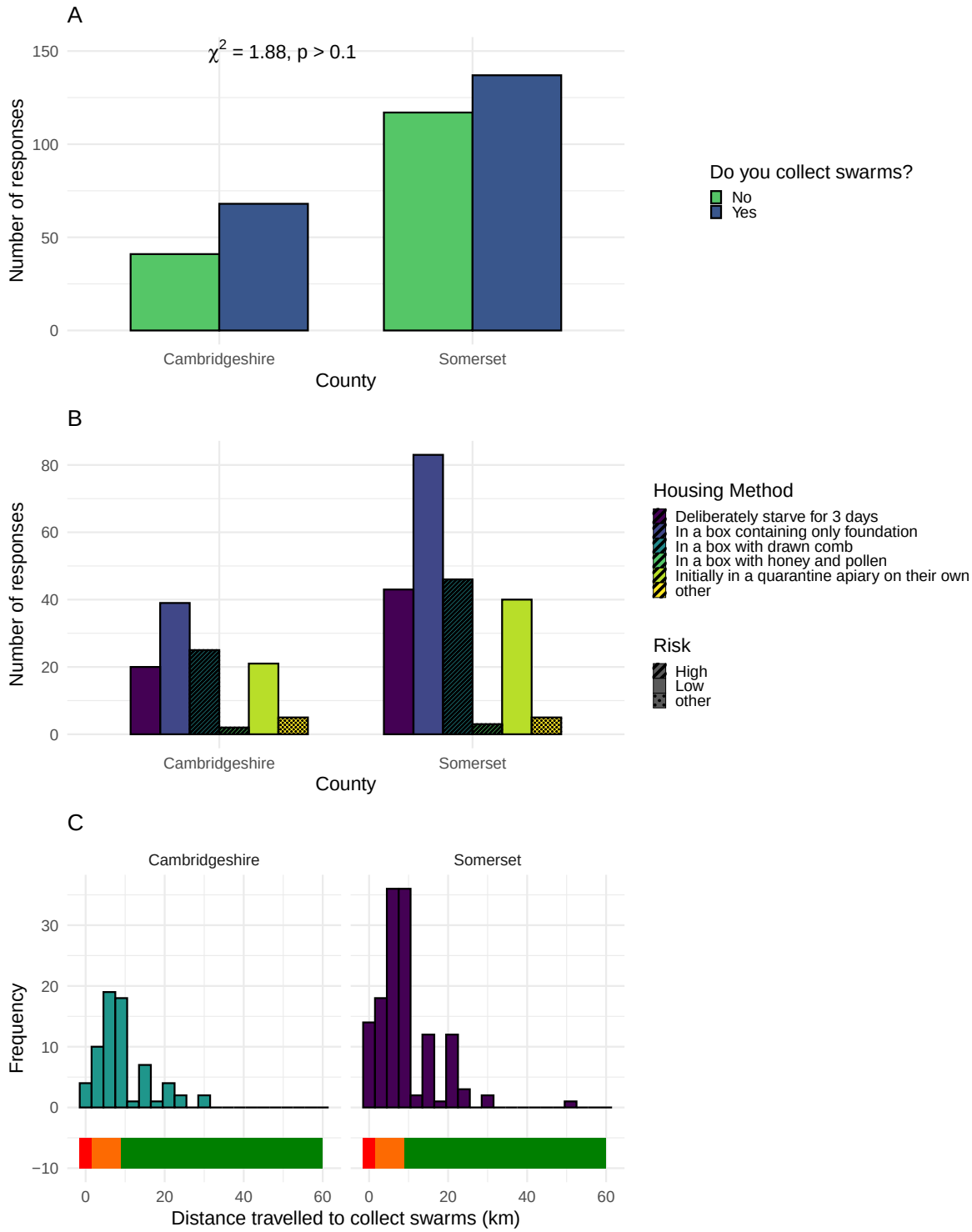


Figure 5.3: The responses to the survey investigating EFB spread from both Cambridgeshire and Somerset related to swarm practices. A) A summary of the responses to the question of whether they collect swarms by county, $chi^2 = 1.88$, p-value >0.1. B) A summary of the responses to which housing method they use if they collect swarms by county and the risk assessment highlighted. C) A histogram showing the distribution of the distance travelled by respondents to collect swarms for both counties. The bar at the bottom represents the EFB risk traffic light system, red represents high risk zone, orange medium risk and green low.

5.3.2.2 Responses related to education and experience

A high number of respondents from Somerset and Cambridgeshire had no beekeeping qualifications, 62% of respondents in Somerset and in Cambridgeshire 50% (Figure 5.4). Very few respondents had more than 1 qualification across both counties. The average number of years active for both counties was between 5 and 10 years. For Somerset, 70 years was the highest response for number of years active, far higher than the maximum for Cambridgeshire which was 44 years. Most beekeepers had between 5 and 10 active colonies in both Cambridgeshire and Somerset, with very few respondents having more than 20 colonies. There was no significant difference between the proportion of those who read the EFB relevant literature and those that did not across the two counties.

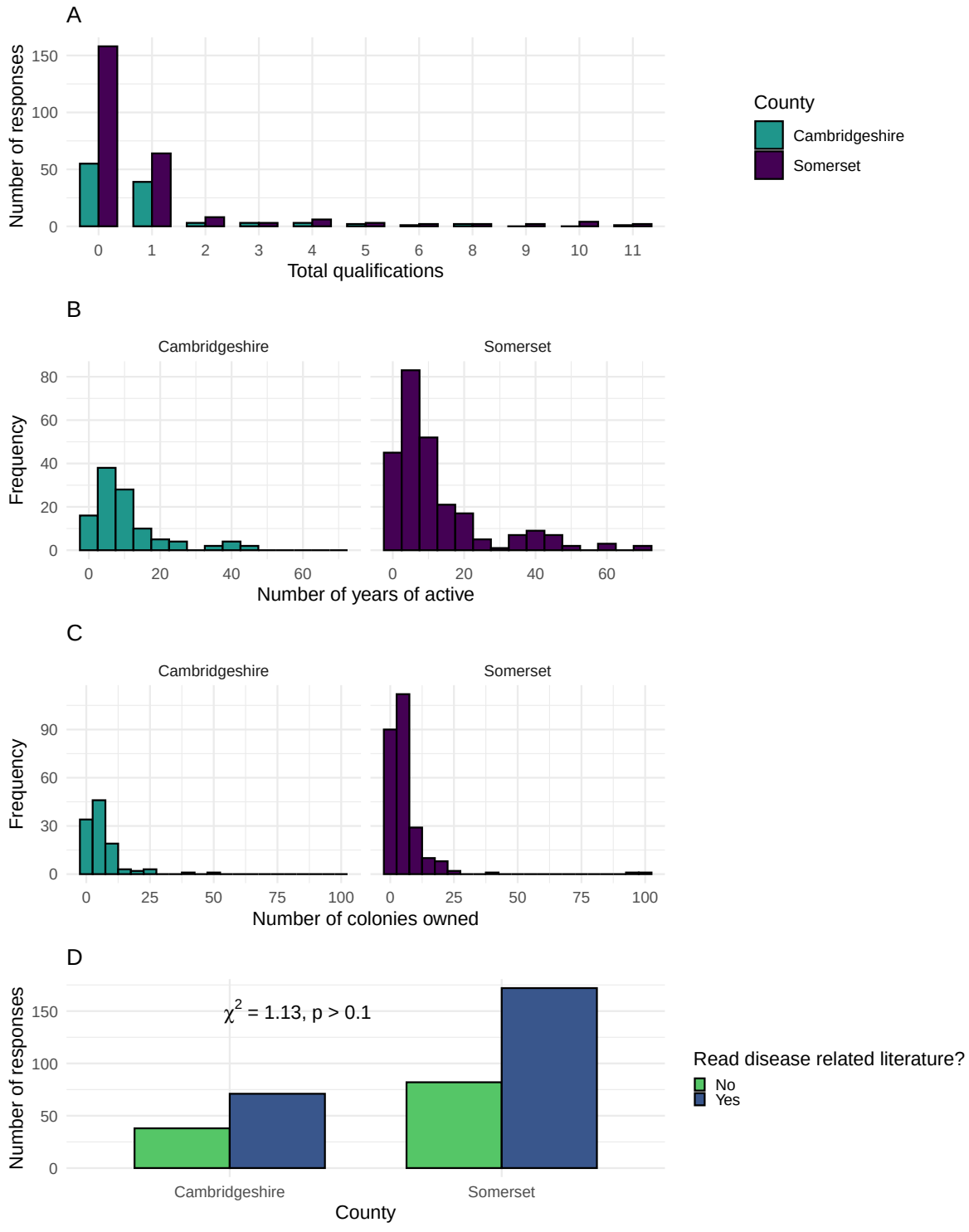


Figure 5.4: The responses to the survey investigating EFB spread from both Cambridgeshire and Somerset related to education and experience. A) A plot showing the number of qualifications the respondents have. B) A histogram showing the distribution of number of years the respondent had been active. C) A histogram showing the distribution of number of active colonies the respondents owned. D) A summary of whether the respondents had read disease relavent literature,$chi^2 = 0.13$, p>0.1.

5.3.2.3 Responses related to biosecurity

Across both Cambridgeshire and Somerset more respondents wear disposable gloves and disinfect their tools than do not (Figure 5.5). There was no significant difference between the proportion of those who responded yes and no for disinfecting tools when comparing the two counties. This was the same for wearing disposable gloves. For both counties more respondents did not share equipment, and there was no significant difference between the proportion of yes and no between the two counties.

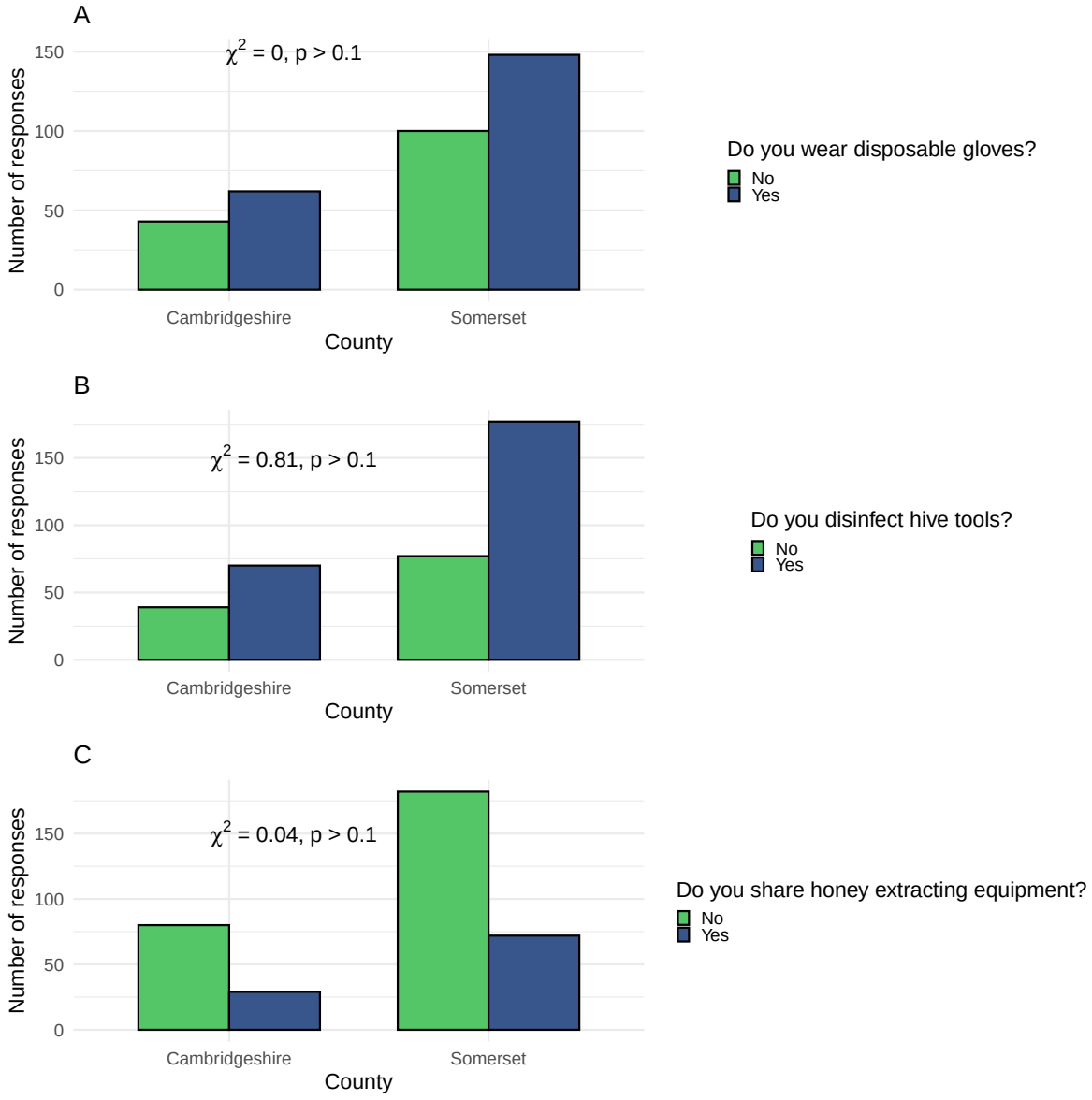


Figure 5.5: The responses to the survey investigating EFB spread from both Cambridgeshire and Somerset related to biosecurity practices. A) A summary of whether the respondent wears gloves, $chi^2 = 0$, p>0.1. B) A summary of whether the respondent disinfects their hive tools, $chi^2 = 0.81$, p>0.1. C) A summary of whether the respondent shares honey extracting equipment, $chi^2 = 0.04$, p>0.1.

5.3.2.4 Responses related to perception and experience with EFB

In Cambridgeshire 73% of the respondents had been inspected by the NBU, whereas in Somerset a lower proportion had been inspected, 53%, this was a significant difference (Figure 5.6). Very few of respondents in both counties had seen Foulbrood, and a small number were unsure, most had not seen foulbrood, the proportion of answers was not significantly different between the two counties. More respondents see EFB as a problem than those who do not for both counties, approximately 65%, no significant difference between the two counties. The number of respondents that have signed up for alerts in Cambridgeshire were 82, 75% of the total respondents. The number of respondents that have signed up for alerts in Somerset were 154, 60% of the total respondents. For both counties there were beekeepers that were unsure if they were signed up for alerts. In Cambridgeshire more respondents had received an alert, 40, than had not, 35. Less respondents had received an alert than had not in Somerset.

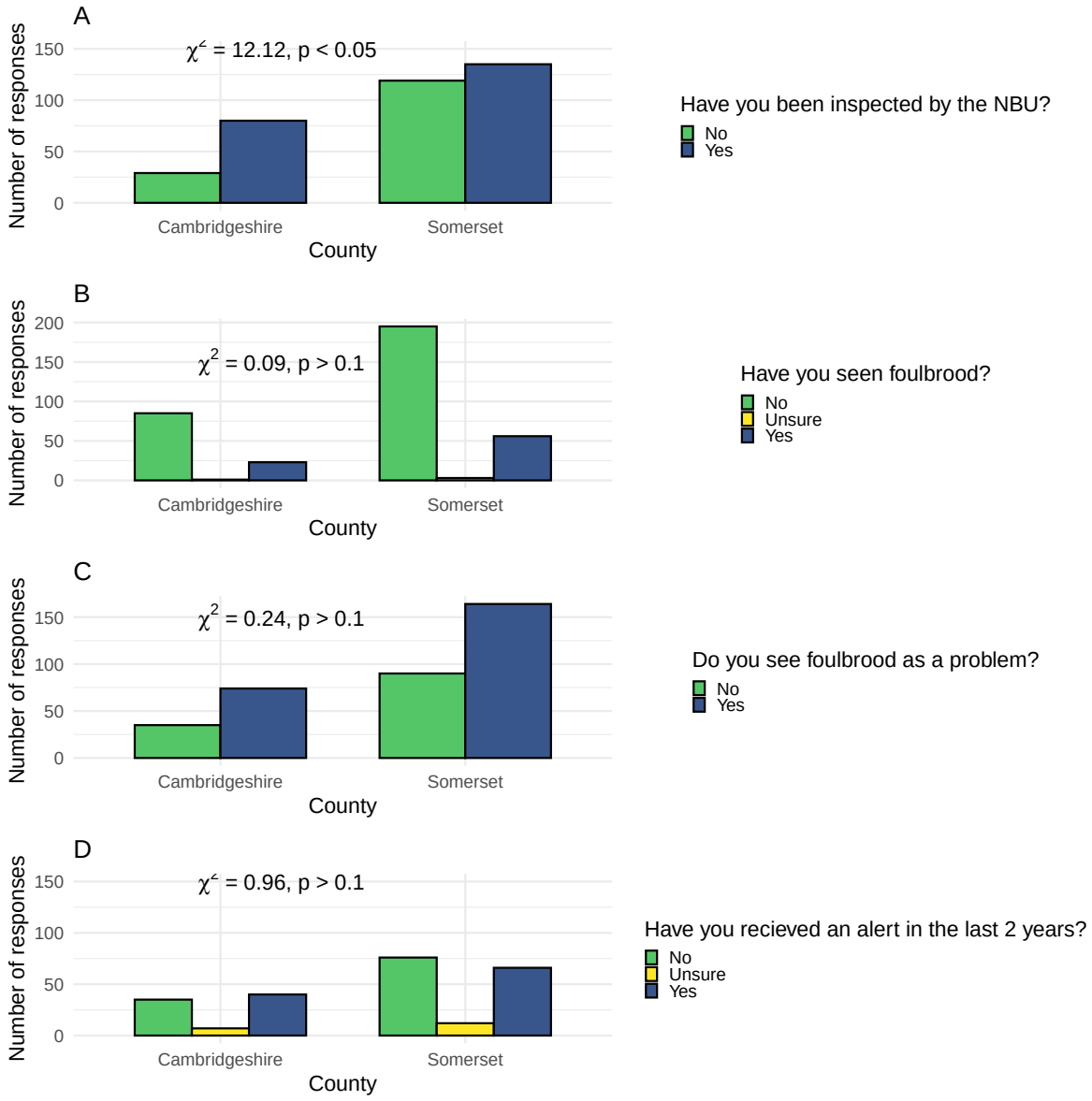


Figure 5.6: The responses to the survey investigating EFB spread from both Cambridgeshire and Somerset related to perception and experience with EFB. A) A summary of whether the respondent has been inspected by the NBU, $chi^2 = 12.12$, p<0.05. B) A summary of whether the respondent has seen foulbrood, $chi^2 = 0.09$, p>0.1. C) A summary of whether the respondent sees foulbrood as a problem, $chi^2 = 0.24$, p>0.1. D) A summary of whether the respondent has receieved an alert in the last 2 years, $chi^2 = 0.96$, p>0.1.

5.3.2.5 Opinions of Beekeepers

The most common response for the reason behind EFB spread for both counties was a feeling that respondents with disease cannot recognise disease and engage in poor beekeeping practices at diseased sites (Figure 5.7). A high number of respondents were unsure about why the disease is spreading across both counties.

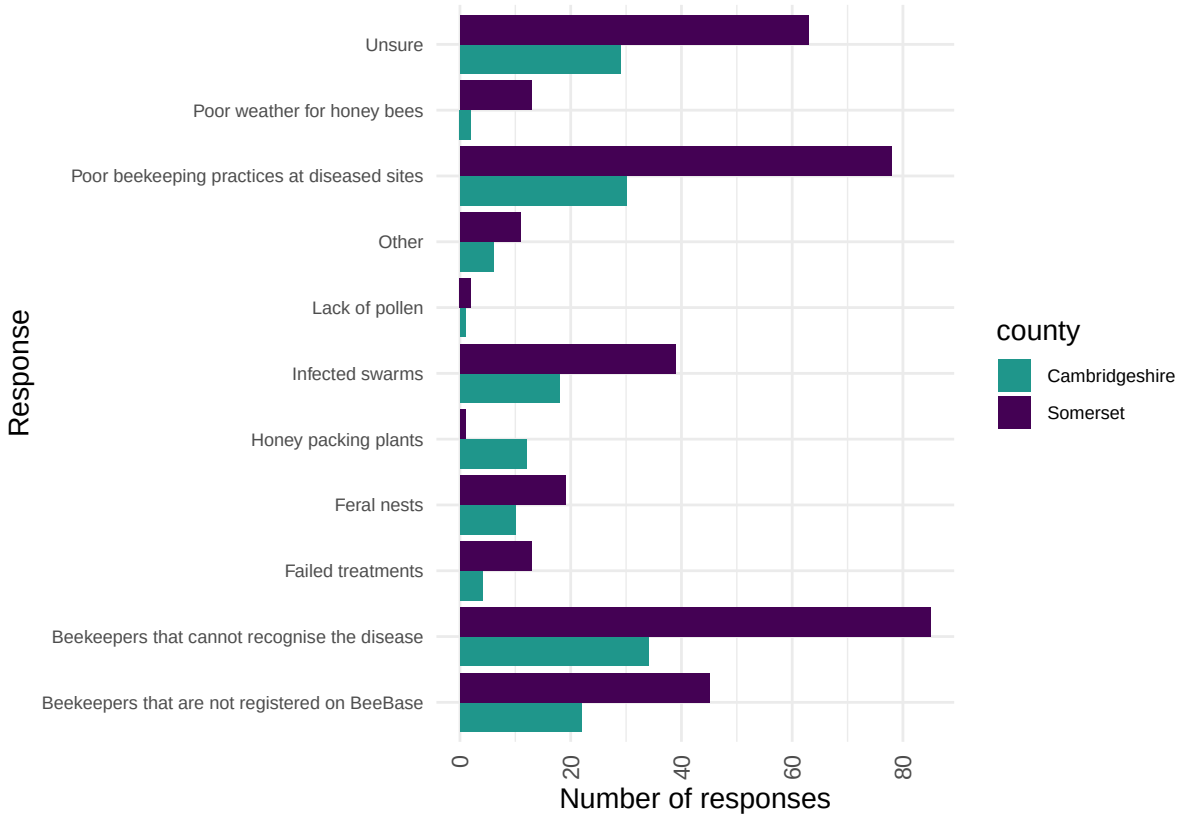


Figure 5.7: A summary of the beekeepers responses to the question: *What do you think are the main causes of EFB in your area?*, for each county

5.3.3 Correlation analysis

5.3.3.1 Somerset

For Somerset there were 254 respondents so the critical value for significance, $p<0.05$, is 0.12 according to Pearson's correlation table. This means that the Pearsons's correlation coefficient values above 0.12 were considered significant, both negative and positive. For Somerset there was a strong positive correlation between collect swarms and house method, and also between collect swarms and distance to collect, both with a correlation coefficient of nearly 1 (Figure 5.8). There were groups formed within the correlations, particularly between the swarming variables (Figure 5.9).

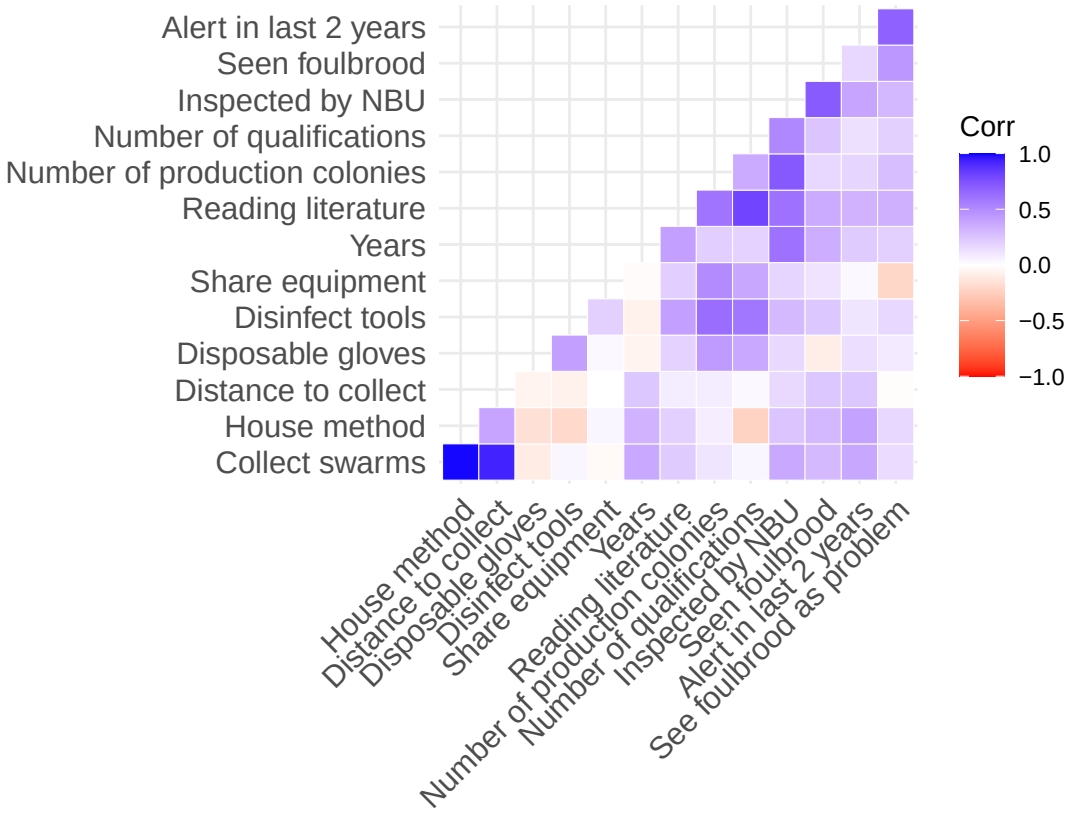


Figure 5.8: A correlation matrix, representing the Pearson correlation coefficient values between each of the variables generated from the responses to the EFB survey carried out in Somerset.

Figure 5.9: A correlation network, representing the Pearson correlation coefficient values between each of the variables generated from the responses to the EFB survey carried out in Somerset. Only the significant correlations are shown, those with a corrlation coefficient of >0.12.

5.3.3.2 Cambridgeshire

For Cambridgeshire there were 109 respondents so the critical value for significance, $p<0.05$, is 0.19. This means that the Pearsons's correlation coefficient values above 0.19 were considered significant, both negative and positive. There was also a strong positive correlation between collect swarms and house method, and also between collect swarms and distance to collect in Cambridgeshire (Figure 5.10). There were groups formed within the correlations, particularly between the swarming variables (Figure 5.11).

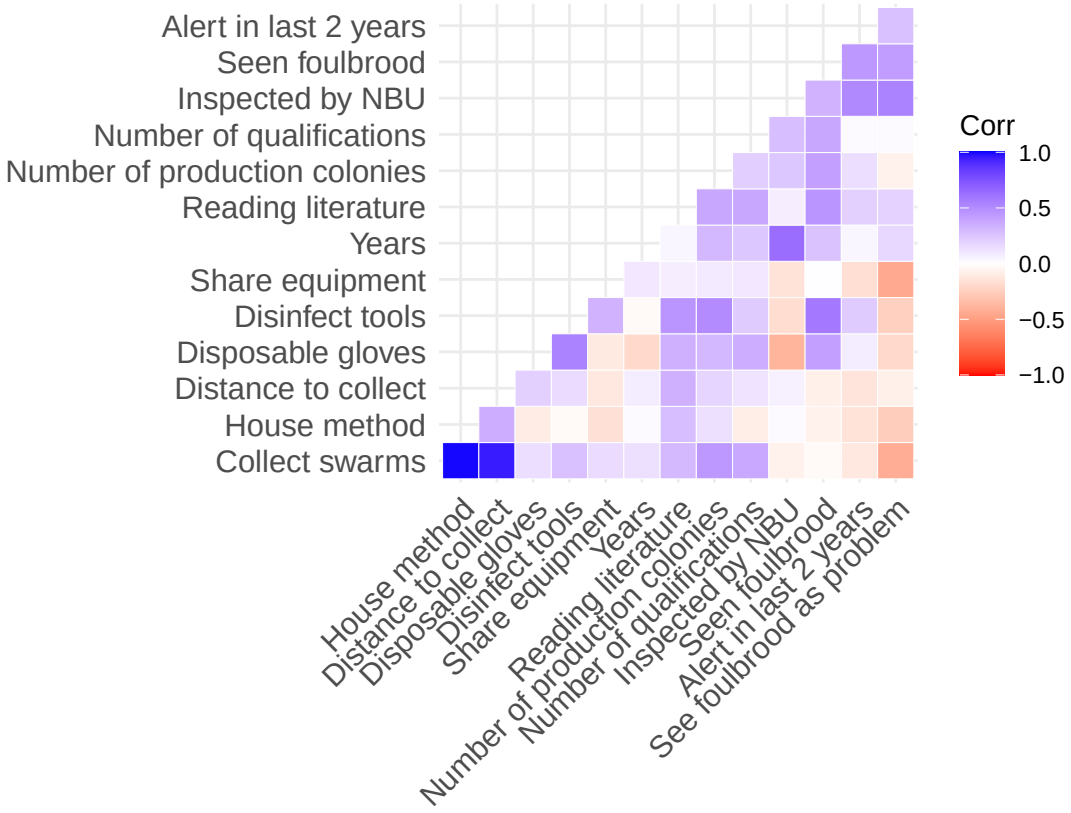


Figure 5.10: A correlation matrix, representing the Pearson correlation coefficient values between each of the variables generated from the responses to the EFB survey carried out in Cambridgeshire.

Figure 5.11: A correlation network, representing the Pearson correlation coefficient values between each of the variables generated from the responses to the EFB survey carried out in Cambridgeshire. Only the significant correlations are shown, those with a corrlation coefficient of >0.19.

5.3.4 SEM

5.3.4.1 Somerset

The SEM generated for Somerset had a CFI value of 0.847, indicating a good fit and a RMSEA value of 0.074 indicating an acceptable fit. In the Somerset SEM, swarm practices were significantly influenced by all three variables, risk of swarm housing method, does the beekeeper collect swarms and distance travelled to collect swarms (Figure 5.12). Education and experience were significantly influenced by all four variables approximately equally, beekeeping qualifications, years of experience, number of colonies and read literature about foulbrood and disease. Biosecurity practices were significantly influenced by all three variables, use of disposable gloves, disinfect hive tools and share honey extracting equipment. Education and experience had a significant positive effect on inspected by the NBU and biosecurity practices, but no significant effect on perception and swarm practices. Alert had a significantly positive effect on perception and on swarm practices but no significant effect on biosecurity practices. Seeing foulbrood had a significantly positive effect on perception but no significant effect on swarm practices and biosecurity practices. Inspected by the NBU had a significant positive effect on seen foulbrood and no significant effect on perception, swarm practices and biosecurity practices. Perception had no significant effect on swarm practices and biosecurity practices. The full SEM output can be found in C.2.1.

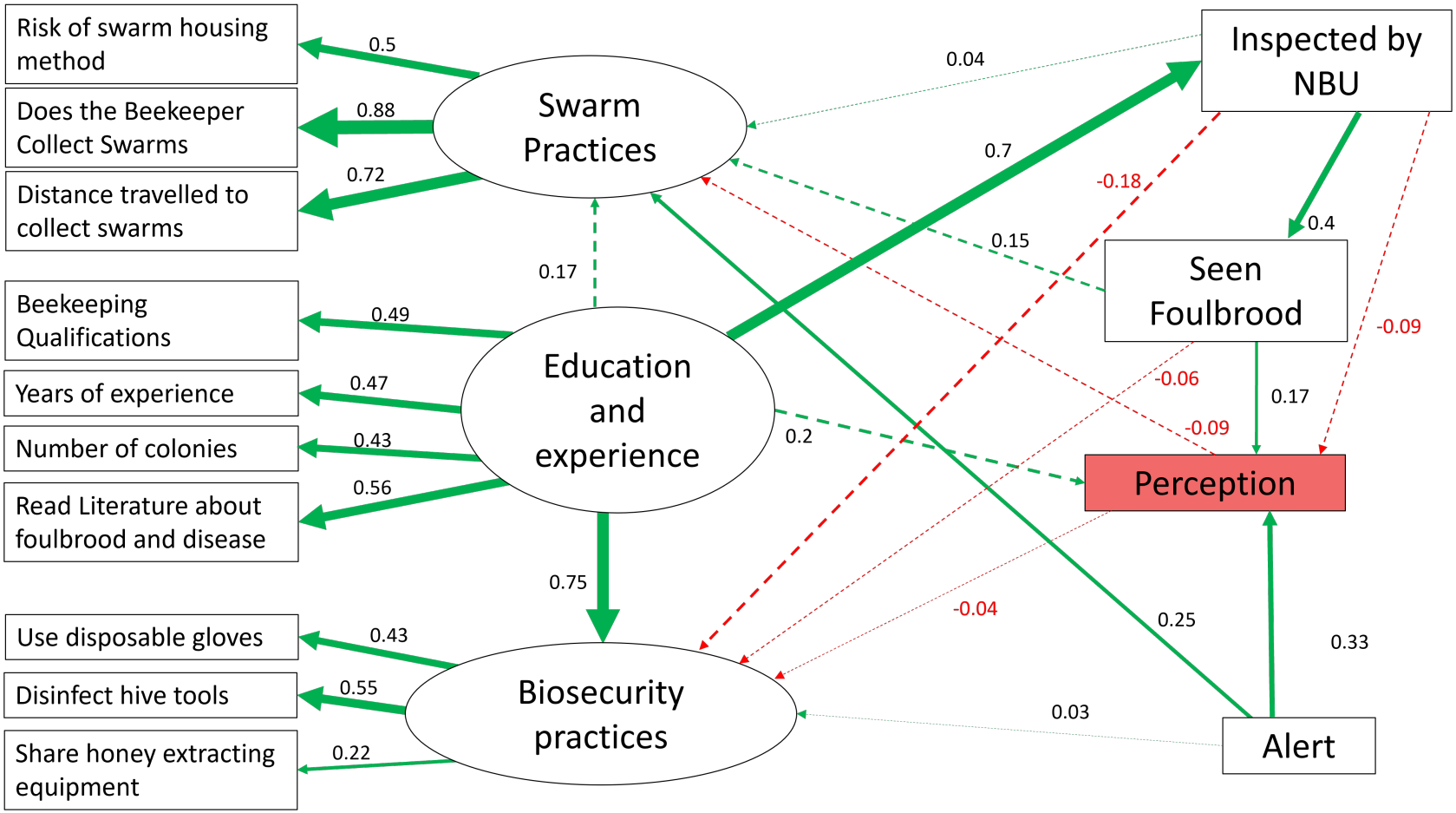


Figure 5.12: The results of the structural equation model investigating the impact of education, experience and exposure to EFB on beekeeping practices and perception, in the county of Somerset. The green lines represent a positive effect and red a negative effect. The solid lines represent significant effects and the dotted lines non significant.

5.3.4.2 Cambridgeshire

The SEM generated for Cambridgeshire had a CFI value of 0.782, indicating a good fit and a RMSEA value of 0.084 indicating an acceptable fit. In the Cambridgeshire SEM, swarm practices were significantly influenced by all three variables, risk of swarm housing method, does the beekeeper collect swarms and distance travelled to collect swarms (Figure 5.13). Education and experience were significantly influenced by all four variables approximately equally, beekeeping qualifications, years of experience, number of colonies and read literature about foulbrood and disease. Biosecurity practices were significantly influenced by use of disposable gloves and disinfect hive tools, but not by share honey extracting equipment. Education and experience had a significant positive effect on inspected by the NBU, swarm practices and biosecurity practices. There was no significant effect of education and experience on perception. Inspected by the NBU has a significant positive effect on perception, a significantly negative effect on biosecurity practices and no significant effect on seen foulbrood and swarm practices. Seen foulbrood had a significantly positive effect on perceptions and had no significant effect on the biosecurity practices or swarm practices. Perception had no significant effect on biosecurity practices or swarm practices. Alert had no significant effect on perception, swarm practices or biosecurity practices. The full SEM output can be found in C.2.2.

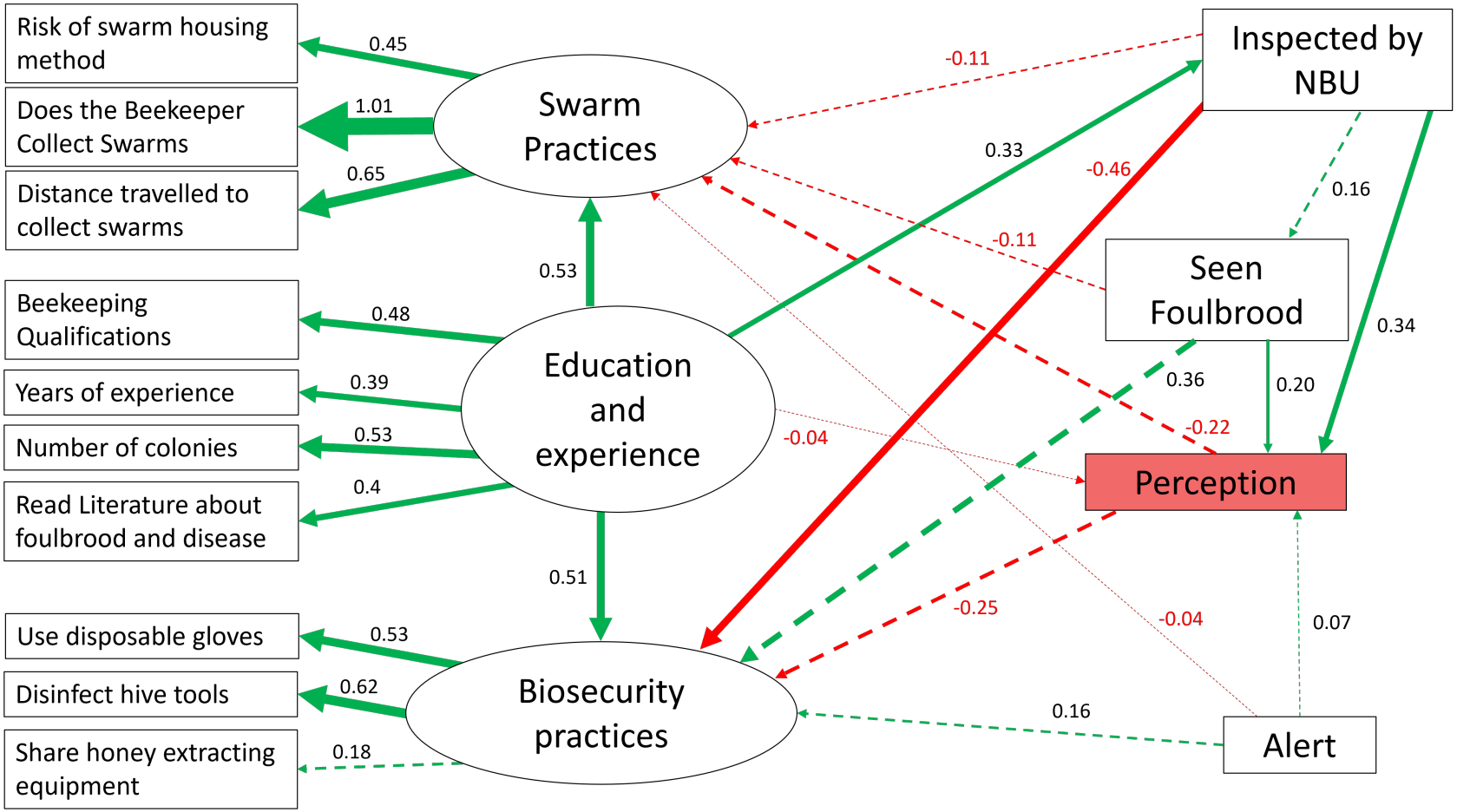


Figure 5.13: The results of the structural equation model investigating the impact of education, experience and exposure to EFB on beekeeping practices and perception, in the county of Cambridgeshire. The green lines represent a positive effect and red a negative effect. The solid lines represent significant effects and the dotted lines non significant.

5.3.5 Literature search on swarm housing methods and linking swarms with EFB

Most of the literature found on swarm hiving suggested housing swarms on drawn comb (Table 5.4). Only one source mentioned disease risk with swarms and that was from 1943. Very little information was found in the literature for linking swarms to disease, but the leaflet from the NBU, Foulbrood disease of honey bees states, in their 10 key rules of EFB, to be wary of swarms and hive them on foundation not drawn comb (NBU, 2017). This information leaflet was one of those listed in the survey and selected in the variable for reading EFB literature in the SEM.

Table 5.4: A list of the literature available to beekeepers stating hiving swarm methods.

Title	Disease Mentioned	Suggested swarm housing method	Reference
Beekeeping for all	Yes	On foundation	Edawardes (1943)
Beekeeping up-to-date	No	Brood chamber and frames with foundation	Tinsley (1945)
Beekeeping for profit and pleasure	No	Worker comb or foundation	Webb (1945)
Honey Farming	No	On foundation	Manley (1946)
Beekeping practice	No	On foundation preferably also with drawn combs with some containing honey	Stuart (1947)
The Beekeepers encyclopedia	No	On foundation	Deans (1949)
Beekeeping	No	Drawn comb and foundation	Smith (1963)
Background to beekeeping	No	On foundation, advised to have a brood frame	Waine (1975)
Bees and Beekeeping	No	Hive with drawn comb	Morse (1975)
Amateur beekeeping	No	On foundation, advised to have a brood frame	Sechirst (1976)

Table 5.4: A list of the literature available to beekeepers stating hiving swarm methods.

Title	Disease Mentioned	Suggested swarm housing method	Reference
The Bee book: The history and Natural History of the Honeybee	No	Prepared Hive, no mention on drawn or foundation	More (1976)
Keeping Bees	No	Drawn brood comb and foundation	Beckly (1977)
Mastering the art of Beekeeping	No	Drawn brood comb, with eggs, honey, larvae and pollen	Ormond Harry (1982)
Principles of practical beekeeping	No	With stores	Couston (1990)
Taking and Hiving a Swarm	Yes	Foundation and if late in the season sterilised drawn comb	NBU (2012)
The BBKA Guide to beekeeping	No	Wax foundation and one frame of drawn comb preferably clean	Davis & Cullum-Kenyon (2015)
The beecraft apiary guide to swarming and swarm control 5a	No	On frames of comb or foundation	BeeCraft (2022)

5.3.6 Literature search on biosecurity practices linking to EFB

A factsheet from the NBU, Apiary and Hive Hygiene, was found on BeeBase that mentioned having good apiary and hive hygiene reduced the risk of transmission of diseases (NBU, 2011). An information leaflet from the NBU, hive cleaning and sterilisation, mentioned EFB specifically and how it is important to keep the hive clean to prevent it (NBU, 2018). In the same information leaflet mentioned in 5.3.5, Foulbrood disease of honey bees, it mentioned good hive hygiene and disinfecting used equipment to prevent disease spread. In the NBU starting right with bees it mentions the word disease 45 times, it covered legislation, individual diseases and hive hygiene in relation to disease (NBU, 2016). It appeared that disease linked to biosecurity practices is covered in a lot of the material available on the NBU.

5.4 Discussion

Here I present a structural equation model generated using survey data from beekeepers in counties with a high prevalence of EFB. This model links beliefs with behaviours to increase the understanding of the epidemiology of EFB in the UK. The model has highlighted key behaviours that are misaligned with the perception of EFB risk. These findings could be used to advise the sector on producing information leaflets and training materials to improve the uptake of low-risk activities by beekeepers, but also highlights some evidence gaps around disease and risk from swarms.

The risk perception of a disease is important for tackling its management and prevention (Agrebi et al., 2022). Interestingly, in Somerset having an inspection by the NBU did not appear to influence the perception, but seeing the disease did. This indicated that if colonies were disease free, EFB was not seen as a risk, however in Cambridgeshire inspections did influence the perception, as well as seeing EFB, and EFB was seen as a problem. In Somerset if an inspection by the NBU had taken place it was more likely that EFB had been seen, however in Cambridgeshire being inspected did not impact on whether EFB had been seen. This could be that because cases were higher in Somerset, and so the prioritised inspection regime would mean that visits would preferentially be to areas of high disease risk. As such fewer visits in Cambridgeshire would have been conducted in high risk (red) regions (Budge et al., 2012). As bee inspections are run regionally, as opposed to the literature and training provided nationally (NBU inspection programme), there may be a county difference between the information given at the inspections when the colony was negative for EFB. Creating a leaflet that highlights the risk of EFB to hand to beekeepers, when the colony is negative for EFB could help to unify the perception of EFB risk by beekeepers post inspection. Education and experience did not appear to impact the perception in both counties, suggesting a disconnect between accessing educational materials and appreciating EFB risk. This observation was supported by reading the wider literature, which found a constant lack of explicitly stating EFB risk. Perception of EFB risk was influenced by whether an alert had been received in Somerset, but not in Cambridgeshire. In Somerset EFB was more likely to have been seen as a problem if an alert had been received, this could be because of EFB was an emerging disease in Cambridgeshire and an established disease in Somerset. In Somerset more exposure to the disease historically could have influenced the reaction when EFB was nearby, whereas in Cambridgeshire the disease remains a new entity, so potentially an

unknown threat. Also, the sampling power could have had an impact on this as there were more responses in Somerset. There were other differences but also similarities between the responses of the two counties.

When comparing the individual survey responses between the two counties, Somerset and Cambridgeshire, there appeared to be little difference when it came to swarm collection practices, biosecurity practices and education. This is likely to be because the information given out about general beekeeping in forms such as leaflets and information booklets are mainly provided by the NBU and BBKA at a national level (BBKA, 2024; NBU, 2024). As such, the same information and advice is available to both counties, perhaps explaining why the same behaviours were exhibited. Across both counties with more education and experience the bee inspector was more likely to be called out, this could infer that the literature is successful at teaching beekeepers to recognise EFB. Work by Mezher et al supported this as it was found that most beekeepers could recognise EFB (Mezher et al., 2021). There was also a strong correlation between years and inspections. This suggests that the longer a beekeeper has been around the more chance they would have been inspected. The one clear difference between the two counties based on the survey responses was the NBU inspections, a higher proportion of colonies in Cambridgeshire were inspected than in Somerset. There could be a number of reasons for this, one being that because EFB is an emerging disease in Cambridgeshire beekeepers are being more cautious and calling out the beekeeper with any suspicions. Another reason could be that again because cases were higher in Somerset, this could have resulted in less available bee inspectors, resulting in only red areas being inspected whereas in Cambridgeshire also amber areas (Budge et al., 2012). This high percentage, 75%, of Cambridgeshire respondents being inspected by the NBU is also reflected in the number of inspections, it peaked in 2021 the year they survey was carried out, 300 that season as opposed to 150 in Somerset the same year. However, being inspected by the NBU did not appear to have an impact on swarm practices.

Swarm collection and housing practices were not seen as an EFB risk by respondents from either Somerset or Cambridgeshire. The SEM output for both counties indicated that being inspected by the NBU or seeing foulbrood had no effect on swarm collection and housing practices. Even the perception of EFB as a problem did not affect the swarm collecting and housing practices. These data suggest that the respondents do not link EFB disease risk with swarm collection and housing practices. It was interesting to note that

the education and experience did not influence swarm collection and housing practices in Somerset, again this suggesting that swarm collection was not seen as a risk. When searching the literature available to beekeepers about hiving swarms, very few mentioned risks from disease, and in particular foulbrood. This suggested that there may not be enough literature explicitly stating the risk of EFB when swarm collecting or hiving. In Somerset more risky swarm practices were more likely performed if a disease alert had been received. However, it was also interesting to note that many travelled more than 10 km to collect swarms, which would be beyond the range of any disease alert and would mean that the disease risk local to the collected swarm was unknown. Indeed, there was a positive correlation between the distance travelled to collect swarms and receiving a disease alert. In Cambridgeshire the education and experience had a positive effect on the swarm practices, meaning that the more experience and education gained the more likely that high risk swarm practices were carried out, further supporting that swarms were not seen as a disease risk. There was a positive correlation between the years and collecting swarms, suggesting that new beekeepers were less likely to collect swarms.

The most common housing of swarm practice was on foundation, the low-risk method, and it is interesting to note that the NBU advise that this method offers a lower disease risk (NBU, 2012). The theory is that when honey bees swarm, they carry with them honey which can contain *M. plutonius*. If placed on foundation, the honey is used to produce wax required to create brood comb within which the queen can lay her eggs and begin the new generation. Converting the honey to wax theoretically reduces the risk of exposing young bees to *M. plutonius*. The second most common method of housing a swarm was by providing drawn brood comb. In this case the queen is allowed to lay eggs immediately, resulting in any *M. plutonius* contaminated honey being stored or coming into direct contact with the growing larvae, thus increasing theoretical risk (Winston, 1987; Fries et al., 2006; Mckee et al., 2003). My work highlighted a gap in the evidence for this theory. I could only find one 90-year-old study comparing the housing methods from swarm colonies suffering EFB to observe disease transmission to the daughter colonies. In this case three swarms were housed from three diseased colonies, two were provided foundation and one was provided drawn comb. Neither daughter colonies housed on foundation developed EFB, but the colony housed on drawn comb did suffer from the disease (Russel et al., 1937). We sought information from other EFB researchers on the topic and found some unpublished data from Switzerland. Swiss researchers monitored

mother colonies and subsequent swarms for the presence of *M. plutonius* using qPCR. Swarms from parent colonies with known presence of *M. plutonius* were found to have between 10 and 50x lower levels of *M. plutonius* than the parent colony using real-time PCR (Jean Daniel Charriere, Pers, Comm.). It was noteworthy that 10 out of 11 swarms remained positive for *M. plutonius*, highlighting the risk of *M. plutonius* spread, albeit at lower levels. The husbandry methods for housing swarms were not monitored in this experiment which highlights a significant evidence gap. The advice in the beekeepeing literature found during the search was very mixed as to whether to use foundation, brood frames or even honey stores. Although my results suggested that swarm collection and housing practices were not linked with perception of disease risk. Two respondents from each county mentioned that they believed swarms to be represent a risk. Some respondents did state infected swarms as the reason why EFB was a problem, ~40 in Somerset and ~20 in Cambridgeshire, but most believed it was poor beekeeping practices and not recognising the disease as the main reason.

Respondents believed that poor beekeeping practices were the most common reason for EFB spread, along with beekeepers not being registered and those who cannot recognise the disease. These reasons all suggested other beekeepers actions, hence blaming others. An interesting quote was made by Morland, "In assessing the value of reports it is necessary to bear in mind that some beekeepers consider that the admission of the existence of foulbrood carries a stigma. It should be made clear that unless foulbrood is neglected, there is no slur on the beekeeper" (Morland, 1934), this suggests that the beekeeper should not feel blamed for having EFB, however blaming for disease is human nature. Blame and disease have been companions throughout history scapegoats of disease in Greek mythology were blamed, where many stories included sacrificing humans to end plagues, as they believed certain individuals that had angered the gods were the cause. During the black death massacres of particular religious groups occurred, as they were blamed for the outbreak. In the 1700s a yellow fever outbreak occurred in Philadelphia, and the victims were blamed. When the reasons for spread are unknown, then humans consistently seek to blame others (Cohn, 2018). Stepping forward to more recent times, during the COVID-19 pandemic, blame was placed on various governments and nationalities. This was believed to be because when people are experiencing uncertain times, they are more susceptible to heuristics cues (Sharma et al., 2022). This natural human behaviour was also reflected in this study as the opinions of how EFB spread were

primarily behaviours of others. It is also common human behaviour to blame outsiders for an issue, for example when an outbreak of a crop disease, Cassava Mosaic Disease, hit a community in Tanzania, a survey showed that a majority of the locals blamed another country for the disease (Rugalema et al., 2009). In data not shown from this study a few respondents in the survey blamed importing of bee products to local industry, highlighting again a common human behaviour in disease outbreaks.

This study has highlighted an evidence gap in swarm practices, as only one study has compared hive housing practices. Future work could be to test the different swarm hiving methods, to gain evidence. This would provide more concrete evidence of swarm housing risk, as the risk for each method in this study has primarily calculated from the biology of swarming and disease spread not evidence. It has shown that the alerts funded by government money, had an effect on perception of disease in one county studied but not the other. This shows the potential the alerts can have on perception, and it is a cheap way to inform beekeepers, so maybe the alerts need to provide more information about biosecuirty practices and swarm collections. This study only reflects two counties with high prevalence of EFB, one with emerging disease and the other with an established disease. To provide more evidence it would be helpful to get a countrywide perspective, especially from more counties with both established and emerging disease. The study has also shown there is a disconnect between biosecurity practices and swarm practices with perspective of disease. This study could be used to advise the NBU to include information stating more clearly that certain behaviours increase the risk of foulbrood.

Chapter 6

General discussion

M. plutonius was first identified in 1908 (White, 1912) and causes a statutory honey bee brood disease in the UK (Government, 1982), however, unlike the other foulbrood disease American Foulbrood (AFB) which appears to be controlled, cases have increased in recent years (Giles Budge, as per comms) (Figure 6.1). The failure to control EFB suggests that more work is required to better understand disease movements and control failures. This study presents two different, equally important, tools that can be used to try and improve management and control of European Foulbrood. The first was an affordable molecular pipeline that can be used to study individual, localised outbreaks of EFB, as well as gain further information about what genes, and other bacteria that may be present in each sample. The second was using a structural equation model as a tool to understand beekeeper perceptions and behaviours by analysing survey results that highlight the misalignments between beekeepers perception, behaviours and beliefs.

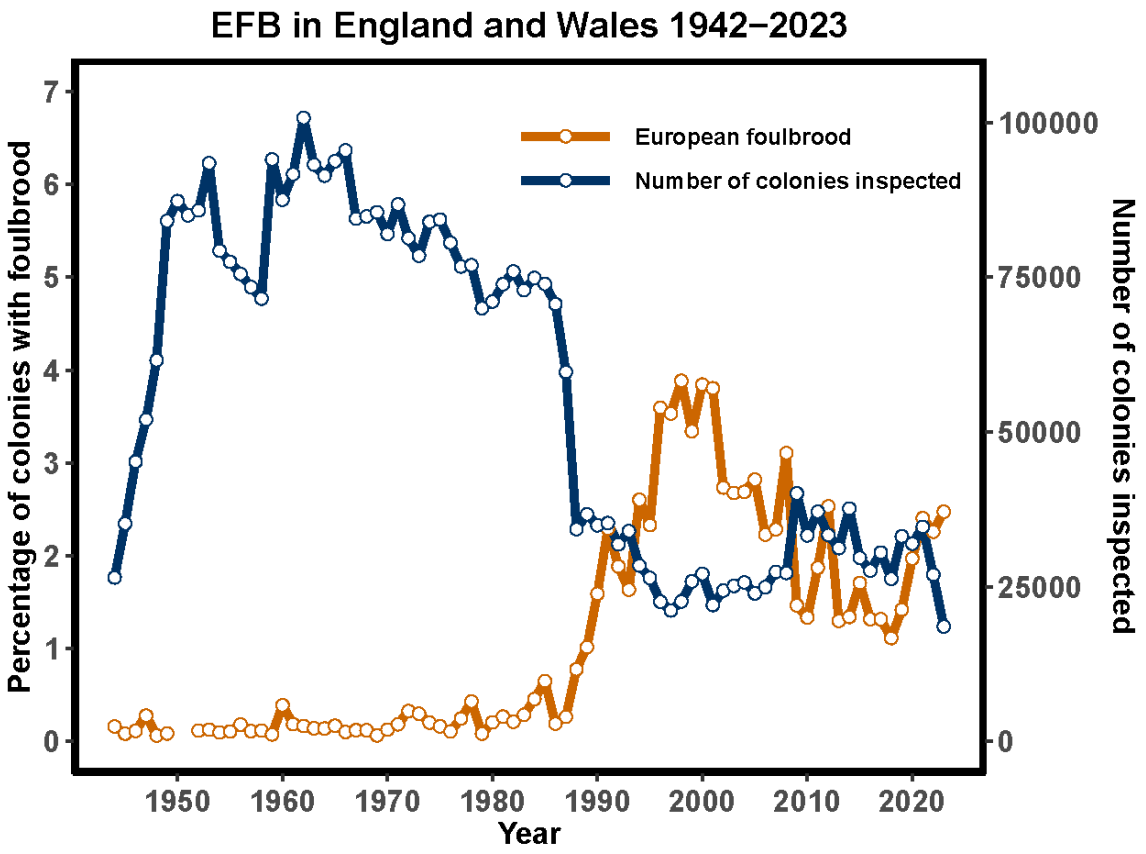


Figure 6.1: Incidences of EFB and AFB between 1942 and 2020. Image courtesy of Giles Budge using data from Annual MAFF report (until 1993) and BeeBase for recent disease and inspections data

The molecular pipeline I developed used whole genome sequencing, which has been used on cultured *M. plutonius* previously, but the new method was unique, because it utilised larval samples that were already collected in the field, for the purposes of running confirmatory LFDs specific to *M. plutonius* (Okumura et al., 2011). Obtaining sequencing data of samples from infected sites from across England and Wales in 2020 allowed the relatedness of the bacteria causing the outbreaks to be studied at a far greater resolution than existing methods. The current routine method for identifying outbreaks and studying spread is the MLST scheme, the MLST scheme uses PCR and Sanger sequencing of fragments of four genes to categorise samples into sequence types. This is a useful tool, but a majority of EFB cases in the UK are ST3, ST5 and ST23 (Budge et al., 2014). The

developed pipeline was successful in identifying localised outbreak clusters within the sequence types, highlighting the potential whole genome sequencing has for studying outbreaks, beyond the resolution provided by the MLST scheme. This information can be paired with information from the bee inspectors in the relevant areas to try and investigate transmission routes, such as in the case study of ST3 in Somerset (Figure 3.19), where one beekeeper had 2 apiary sites with the same cluster type, so transmission had occured between their two apiaries, possibly by equipment or clothing. Also, two beekeepers in close proximity were identified with the same cluster type of *M. plutonius*. A bee inspector may know of movement between the two beekeepers or if they use the same supplier of bees, and this information can be used to manage the spread by intervention. If this information is not known by the bee inspector the genetics will inform them on questions to ask to identify high risk behaviour, to form a greater picture of the transmission events at play. This allows bee inspectors to intervene if possible transmission routes are identified.

One of the key findings in this study was the identification of an imported strain, ST39. It showed to be genetically different from the strains that are established in the UK (Figure 3.8). From the survey responses in the free text box, data not shown, it was suggested that local honey imports were a reason behind EFB presence in Cambridgeshire. As ST39 is the only sequence type present in Cambridgeshire, the UK are major importers of honey from abroad, and this strain is likely to have come from abroad, local honey imports are a likely source, as *M. plutonius* can survive in honey (Mckee et al., 2003; García, 2018). There was also a route of spread identified, data not shown, that linked the spread between Norfolk and Cambridgeshire, which were the only two counties with ST39, again suggesting this information is likely correct (Anon). These two studies showed that genetics can be informative in epidemiology but alone it does not give a complete picture. The knowledge of import routes, transmission routes and the beliefs of the beekeepers were essential to better understanding the epidemiology around these disease cases. This highlighted the need for both genetic and epidemiological tools to manage and control EFB. This could inform future control policy that uses whole genome sequencing along with intensive beekeeper survey and modelling to control and manage EFB outbreaks. If genetic data suggests transmission events in a particular area, the survey and modelling could show shortfalls in that area, for example sharing equipment, and then information could be given to those areas on the risks and how to share equipment safely. If every

beekeeper with disease was sampled risks could be highlighted, however in this study the method generated final data for a quarter of the samples initially sequenced.

Not all samples from LFD buffer bottles passed the quality control necessary to pass the final stages of analysis due to low sequencing depths of *M. plutonius* obtained, resulting in missing data points. Obtaining a high enough sequencing depth was one of the challenges that came from using whole genome sequencing, from naturally infected larva collected by different bee inspectors. In theory if all outbreaks were sampled, my molecular methods and modelling would be able to highlight every single disease transmission event across England and Wales and highlight disease hotspots, as well as training needs. The training needs identified using the SEM and survey data were interesting and can be used to inform the sector. More clarity linking good biosecurity practices with EFB is required, as disease seems to be influencing practice, but not specifically influenced by EFB. More clarity on the disease risk of swarm collecting in the information provided and encouraging low risk practices is needed. Interestingly these methods identified a knowledge gap, as there was only one experiment performed in 1937 assessing risk of disease in swarm housing methods (Russel et al., 1937), so more studies need to be performed in this area, in order to inform the specific swarm housing practices to beekeepers. This shows the value of surveying and modelling methods in managing a disease where spread is a highly complex interaction between the epidemiology of EFB, population density of bees and the management to what the bees are subject. A county difference in perception of risk was observed, this shows that surveying across outbreak sites identified using the genetic methods would prove valuable in meeting the training needs not just nationally but also tailoring these needs more locally. These methods could also be used to get a bigger picture internationally.

If the methods I have developed were deployed at an international scale, they could highlight international movements of statutory notifiable disease in trade items like honey and make every international disease movement traceable, like the one shown in the potentially imported strain, ST39. An international effort allows us to know where ST39 came from, and then determine whether it was present in imported honey. Rolling out these methods routinely both locally and internationally would transform the way outbreaks are managed, both locally and internationally, allowing transmission events to be investigated and then intervened. Not only does this offer a huge potential to identify such routes, the molecular method developed can be used to go beyond this and delve deeper into other interactions involved in the infected larva itself. Once we can account

for spread, we can better understand the biological drivers for disease development in honey bees and begin to improve the resilience of honey bees.

Secondary invaders and gut microbes were identified during the process of obtaining genomic data for *M. plutonius*. Sequencing directly from *M. plutonius* cultures would only provide information about the *M. plutonius* genome, so the methods I developed add value by identifying other bacterial species present, which might be useful for understanding the epidemiology of EFB. In addition, my method provides sufficient genetic data of sufficient quantity and quality to allow bacterial genome assembly. These genomic data from the community of bacteria present, can highlight potential gene transfer events by looking at matching virulence genes and ARGs between the species and *M. plutonius*. It should be highlighted that my method is biased towards Gram-positive bacteria and would need modification to detect Gram-negative species. Including Gram-negative bacteria would be valuable as some of the secondary invaders identified fall into this category, so this would give an even clearer picture of bacterial species present for understanding the epidemiology. This pipeline highlights the potential of looking at secondary invaders and in the future this method could be further developed to include Gram-negative bacteria. Not only can this pipeline be used to identify virulence genes and ARGs in gut microbes and secondary invaders it can also be used to identify these genes in *M. plutonius*.

Identifying virulence genes in *M. plutonius* is important to understand how *M. plutonius* causes infection. Not only can this pipeline identify potential virulence genes, it can identify them on an individual case basis, potentially providing valuable insight into difference in outbreak clusters, as some strains are more virulent than others. This would be a useful tool to compare strains internationally to see if virulence genes differ. With antimicrobial resistance being a major concern, identifying ARGs in *M. plutonius* could be a valuable tool considering the pathogen could reach the food chain through honey. My novel sequencing pipeline offers a single molecular method that not only identifies localised outbreaks to help manage spread of EFB, but it can also delve deeper into the genome and identify potentially interesting genes and can also look at the effect *M. plutonius* has on other bacteria present in the larva. The single process was also designed to be cost-effective, using the cheapest and most accessible sequencing system, Oxford Nanopore technologies, as well as home-made extraction and host depletion methods (Cuber et al., 2023). It was also designed to not be time consuming, with the whole process taking less than a week. Both time and cost were important to fit within budgets

and allow scale up to facilitate national and international monitoring.

The next steps of this study would be to refine the sequencing pipeline to try and generate more sequencing depth, so that if deployed as routine it could generate as much data as possible. It would also be really interesting to deploy this method to samples from infected apiaries from other countries to see how the genetics differ in comparison to the UK strains. Generating more sequencing data from other years could be performed to provide more resolution over time, as well as just geography like in this study. A key area highlighted in this study is the need to assess the risk of different swarm housing methods, so experiments could be carried out to generate an idea of the best practice to avoid EFB. Collaborating with sectors involved in providing information to beekeepers could help to add more details or generate more literature to fill the information gaps and meet the training needs highlighted in this study.

To conclude this study has identified local outbreaks with more clarity than existing methods, using a new, cost-effective whole genome sequencing pipeline, utilising Oxford Nanopore sequencing. This method could be deployed as routine to change the way EFB is managed and controlled, by highlighting transmission routes at local, regional and national levels. My methods also showed the potential to investigate newly imported strains to pin point the route of transmission. Although this study provides potential tools to improve the control and management of EFB, it is clear that more work is needed to tackle EFB as a global threat to the health of the honey bee.

Appendix A

Chapter 2 appendix

A.1 QQplots for the generalised linear models

The normality test for the generalised linear model (Table 2.4) showed a normal distribution (Figure A.1).

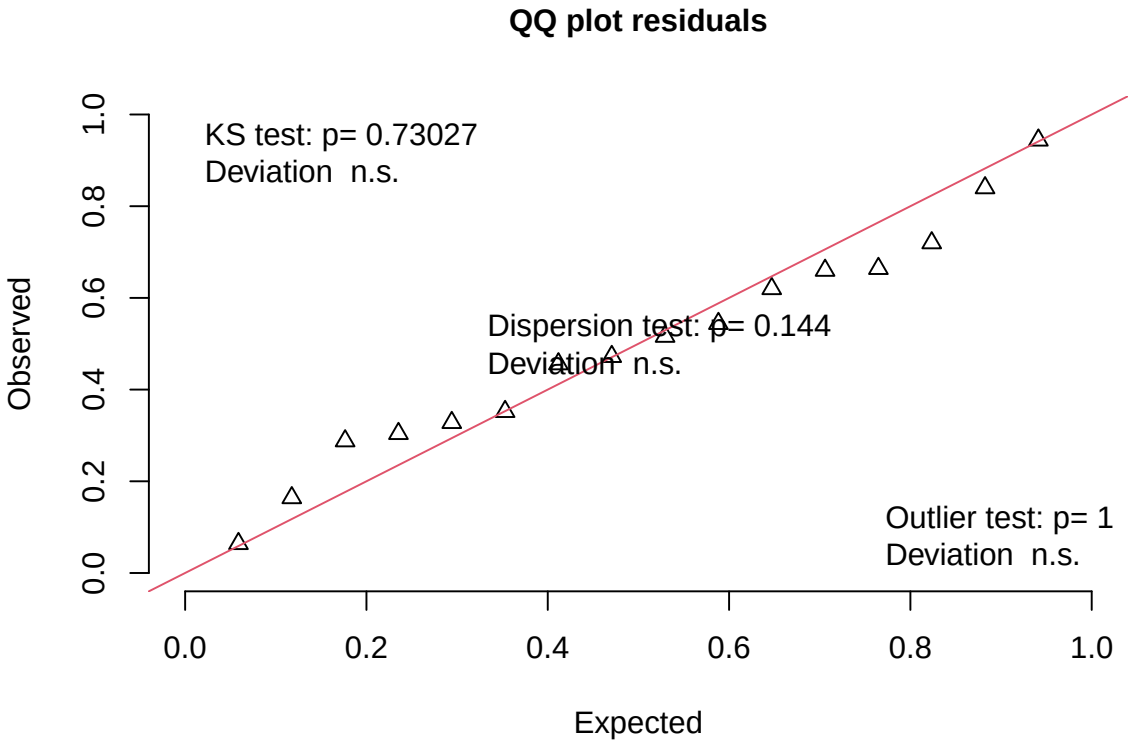


Figure A.1: QQPlot for the generalised linear model assessing the relationship of the ratio of the concentration of patogen target to host DNA in larvae containing a spiked culture of bacteria, with samples treated with different host depletion treatments (DNase, PMA and Thermo Scientific™ Pierce™ Universal Nuclease for Cell Lysis (Lysis)) and storage method (fridge or freezer). The family used for the model was Gaussian.

The normality test for the generalised linear model (Table 2.5) showed a normal distribution (Figure A.2).

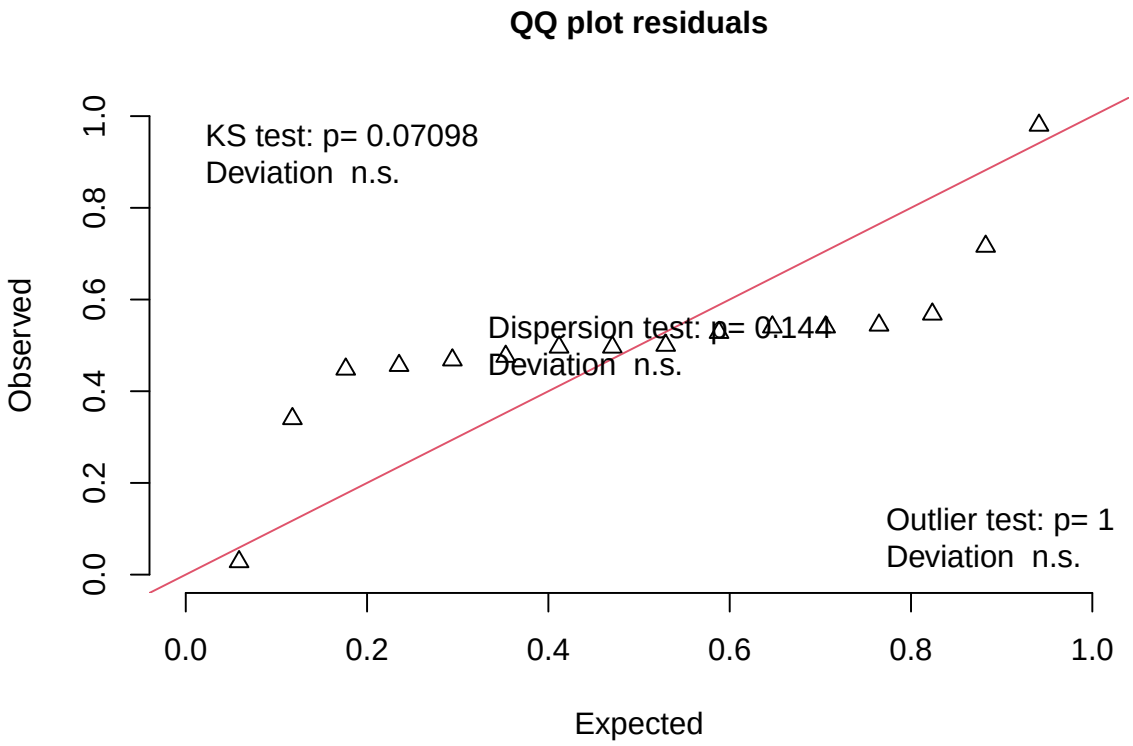


Figure A.2: QQPlot for the generalised linear model assessing the relationship of the ratio of the concentration of patogen target to host DNA in larvae containing a spiked culture of bacteria, with samples treated with different host depletion treatments (DNase, PMA and Thermo Scientific™ Pierce™ Universal Nuclease for Cell Lysis (Lysis)) and storage method (fridge or freezer). The family used for the model was Gaussian.

The normality test for the generalised linear model (Table 2.6) showed a normal distribution (Figure A.3).

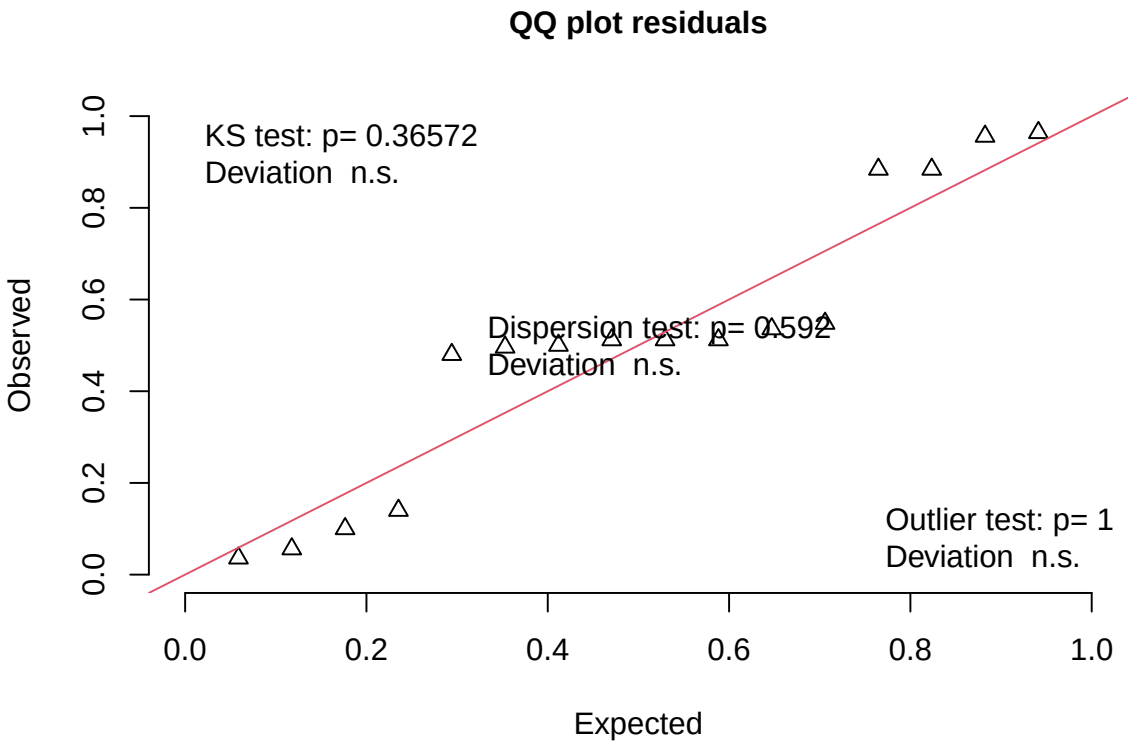


Figure A.3: QQplot for the generalised linear model assessing the relationship of the ratio of the concentration of patogen target to host DNA in larvae containing a spiked culture of bacteria, with samples treated with a combination of different lysis treatments (blood buffer, ATL buffer and water) and different host depletion treatments (DNase and PMA). The family used for the model was Gamma.

The normality test for the generalised linear model (Table 2.7) showed a normal distribution (Figure A.4).

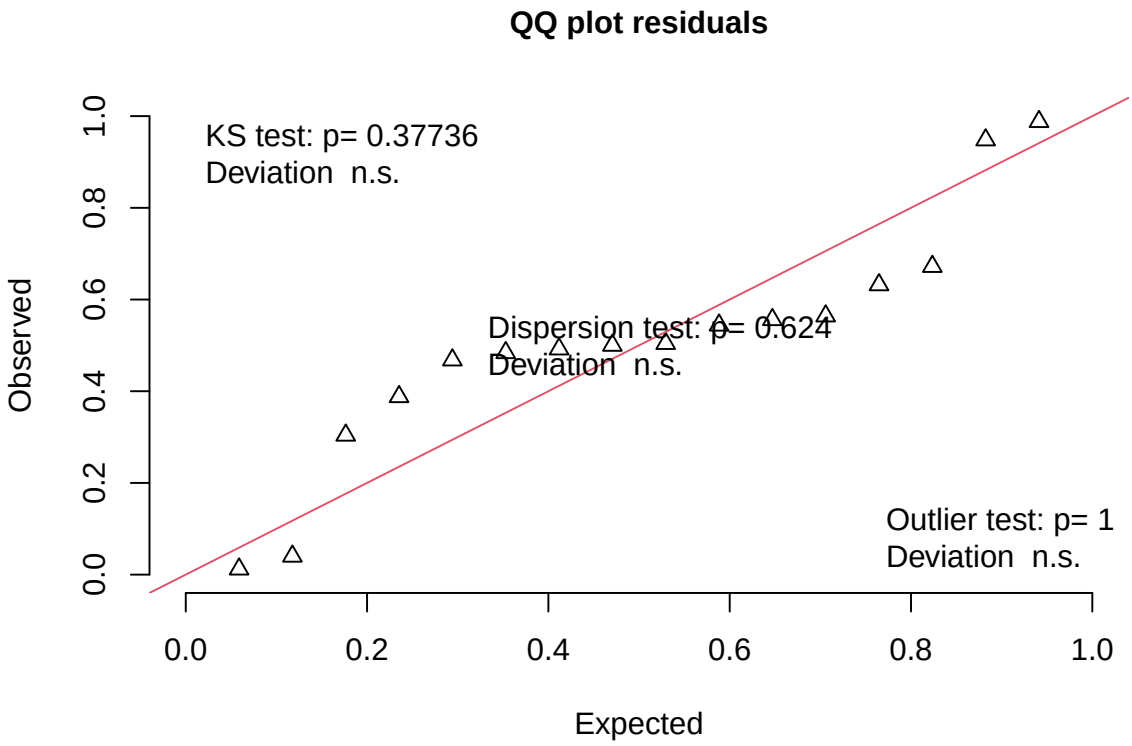


Figure A.4: QQplot for the generalised linear model assessing the relationship of the ratio on the concentration of *Melissococcus plutonius* in larvae containing a spiked culture of bacteria, with samples treated with a combination of different lysis treatments (blood buffer, ATL buffer and water) and different host depletion treatments (DNase and PMA). The family used for the model was Gamma.

The normality test for the generalised linear model (Table 2.8) showed a normal distribution (Figure A.5).

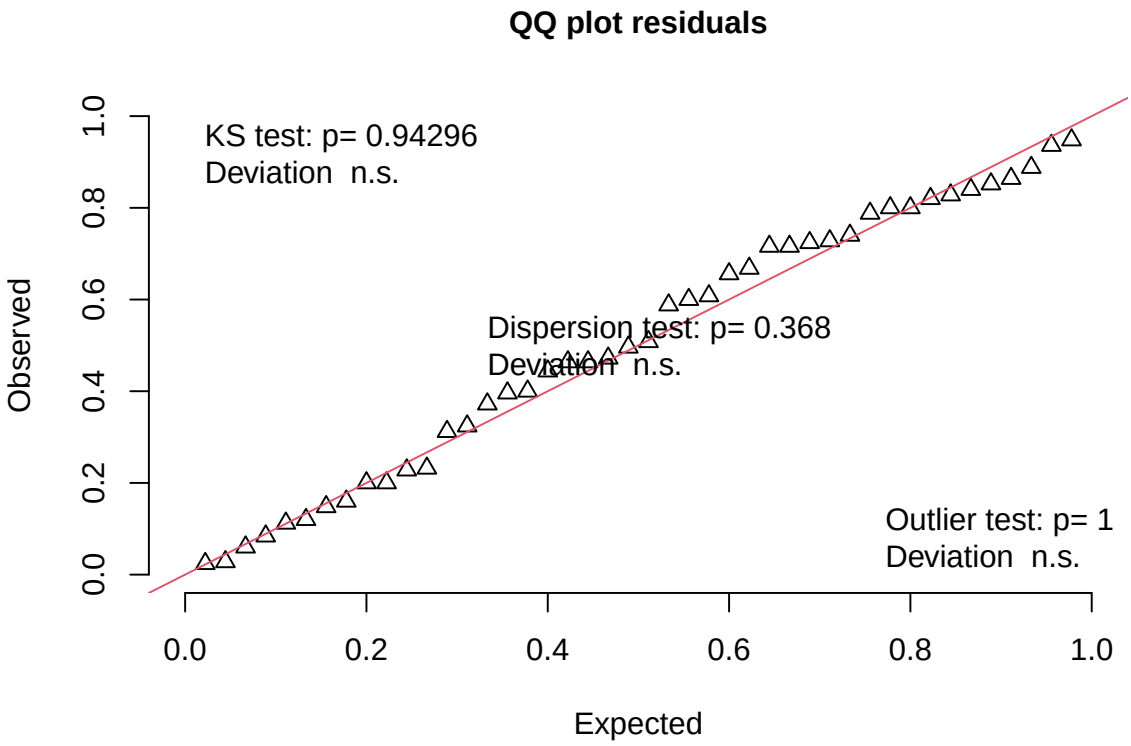


Figure A.5: QQplot for generalised linear model assessing the relationship of the ratio of the concentration of patogen target to host DNA in larvae naturally infected with *Melissococcus plutonius*, with samples treated with a combination of different lysis treatments (blood buffer and water) and different host depletion treatments (DNase and PMA).The family used for the model was Gaussian with a log link.

The normality test for the generalised linear model (Table 2.9) showed a normal distribution (Figure A.6).

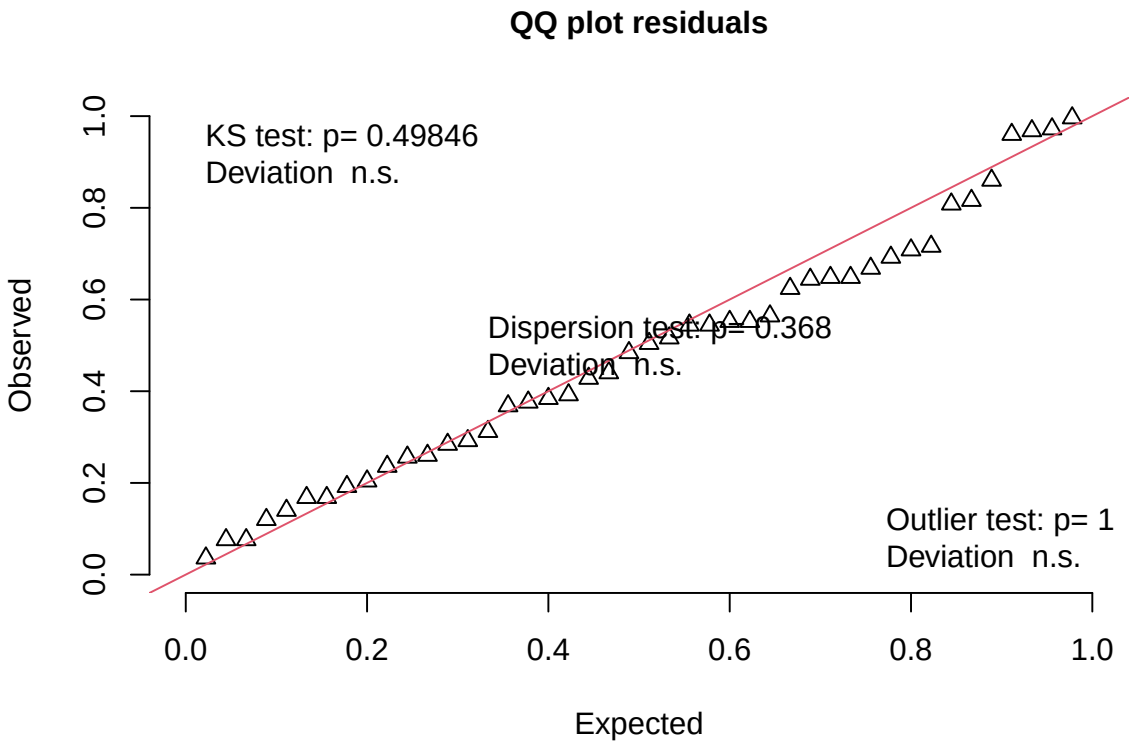


Figure A.6: QQplot for the generalised linear model assessing the relationship of the ratio on the concentration of *Melissococcus plutonius* in larvae naturally infected with *Melissococcus plutonius*, with samples treated with a combination of different lysis treatments (blood buffer and water) and different host depletion treatments (DNase and PMA). The family used for the model was Gaussian.

The normality test for the generalised linear model (Table 2.10) showed a normal distribution (Figure A.7).

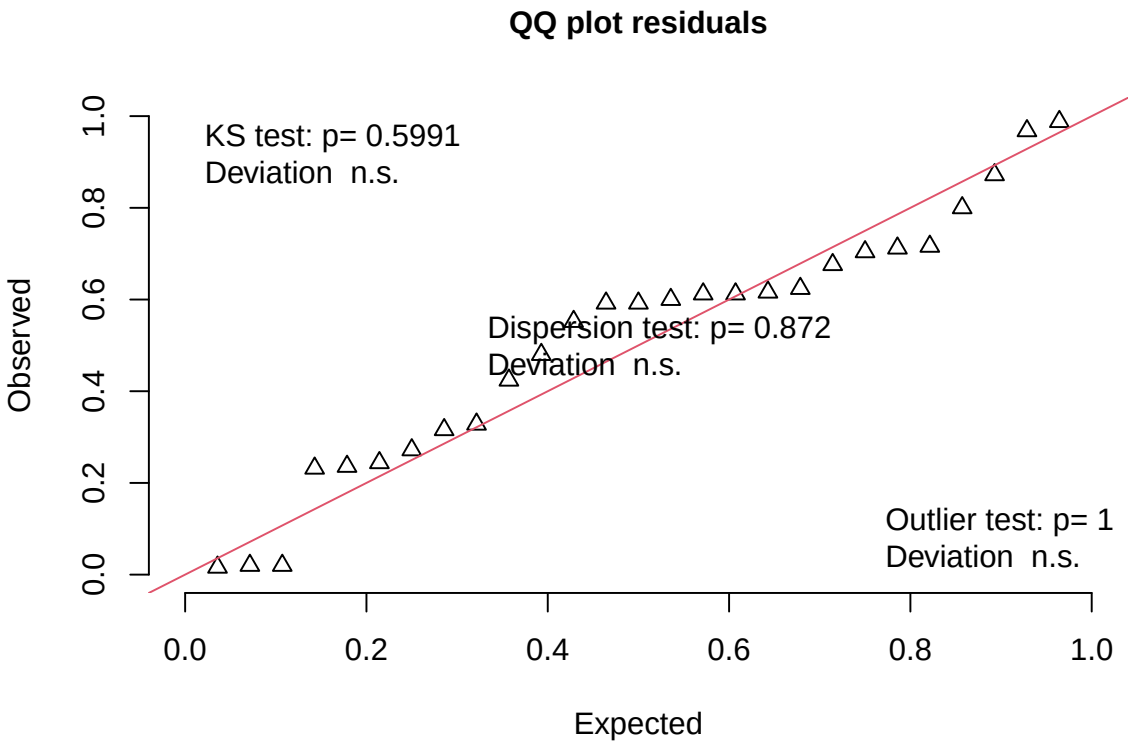


Figure A.7: The QQplot for the generalised linear model assessing the relationship of the ratio of the concentration of patogen target to host DNA in larvae naturally infected with Melissococcus plutonius, with samples treated with different host depletion treatments (PMA and NEBNext® enrichment). The family used for the model was Gaussian with a log link. (*) represents a significant difference.

The normality test for the generalised linear model (Table 2.11) showed a normal distribution (Figure A.8).

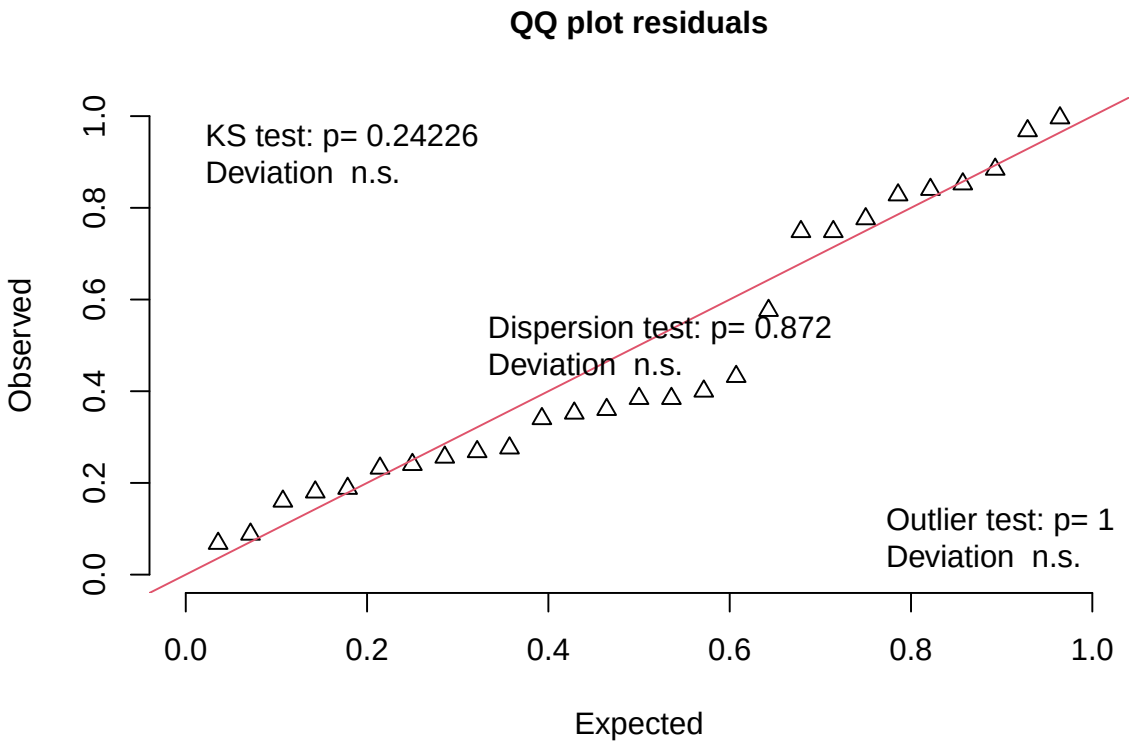


Figure A.8: The QQplot for the generalised linear model assessing the relationship of the concentration of *Melissococcus plutonius* DNA in larvae naturally infected with *Melissococcus plutonius*, with samples treated with host depletion treatments (PMA and NEBNext® enrichment). The family used for the model was Gaussian.

The normality test for the generalised linear model (Table 2.12) showed a normal distribution (Figure A.9).

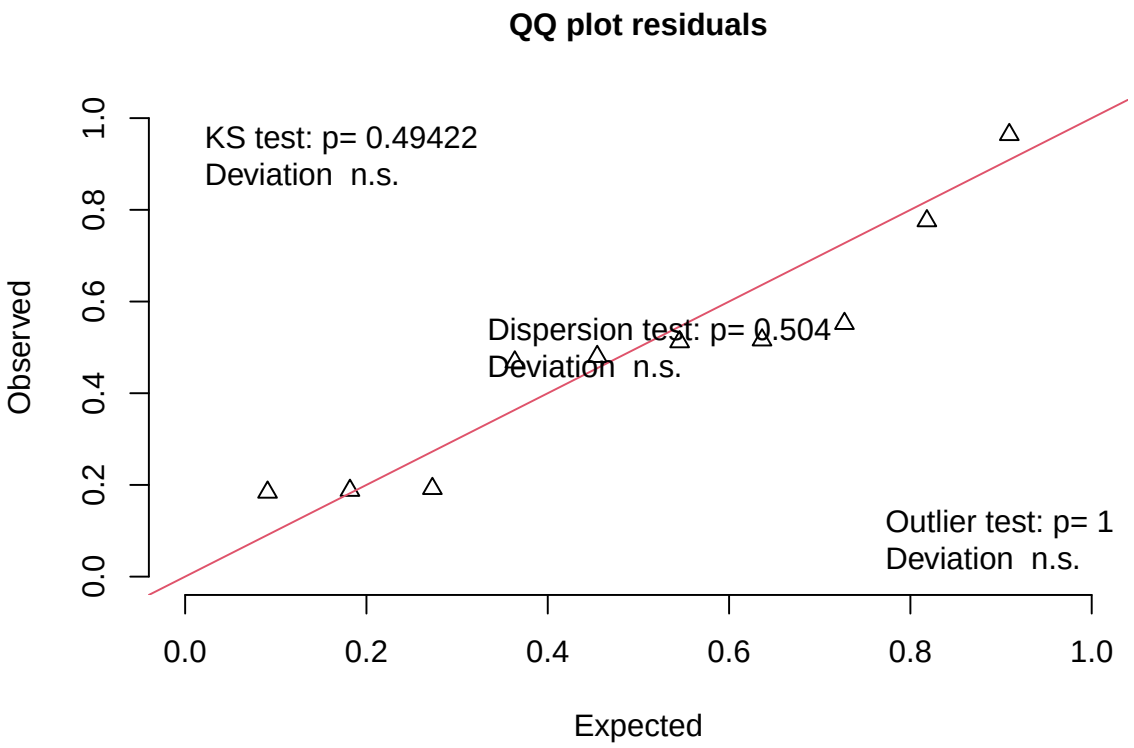


Figure A.9: The QQplot for the generalised linear model assessing the relationship of the ratio of the concentration of patogen target to host DNA in larvae naturally infected with *Melissococcus plutonius*, with samples treated with different percentages of PMA host depletion treatments compared to the reccomended volume (1%, 10% and 100%). The family used for the model was Gaussian.

The normality test for the generalised linear model (Table 2.13) showed a normal distribution (Figure A.10).

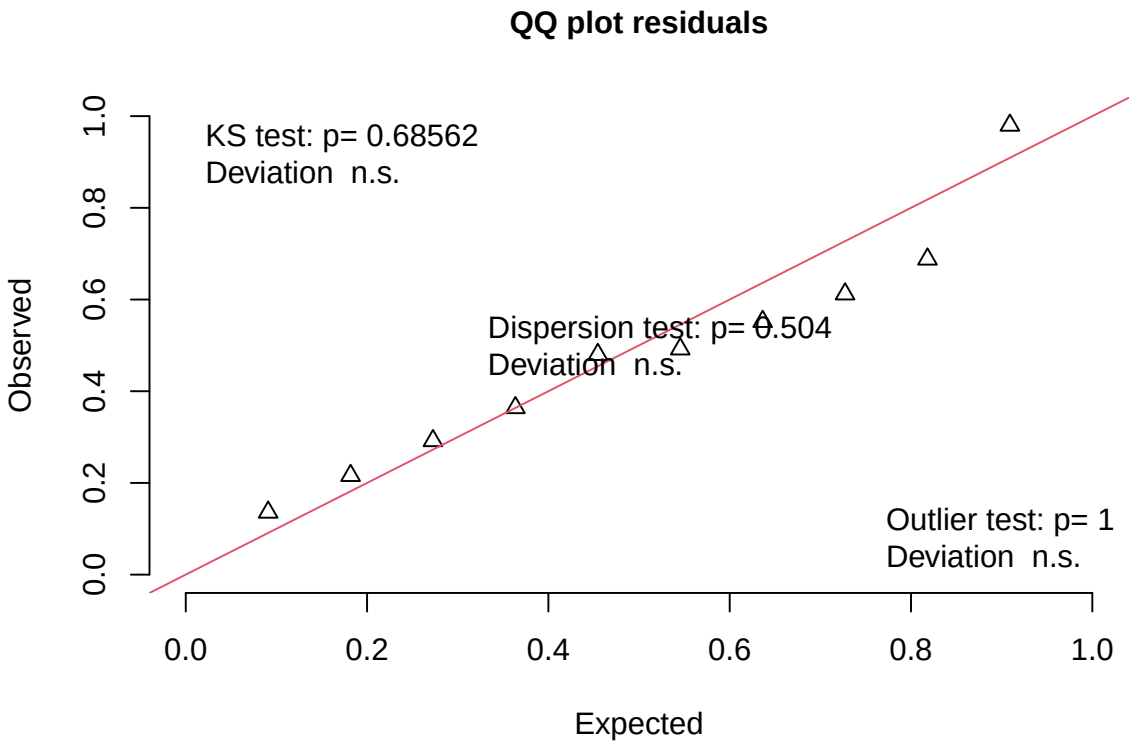


Figure A.10: The QQplot for the generalised linear model assessing the relationship of the ratio the concentration of *Melissococcus plutonius* DNA in larvae naturally infected with *Melissococcus plutonius*, with samples treated with different percentages of PMA host depletion treatments compared to the reccomended volume (1%, 10% and 100%). The family used for the model was Gaussian.

The normality test for the generalised linear model (Table 2.15) showed a normal distribution (Figure A.11).

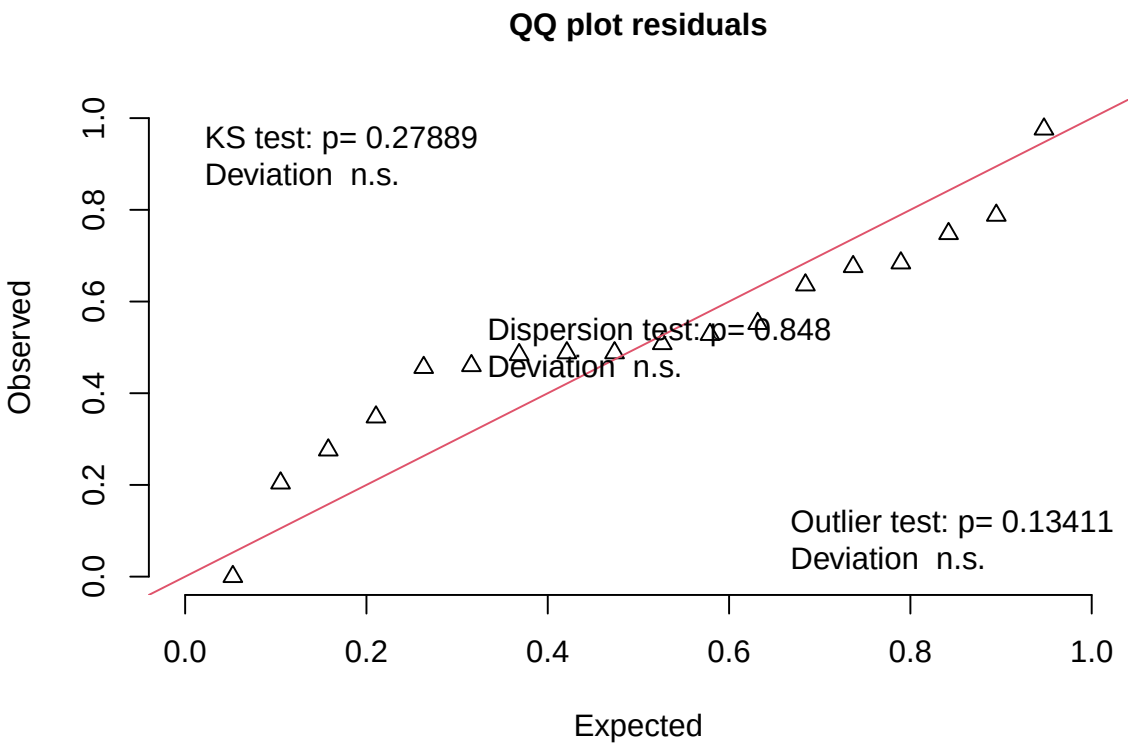


Figure A.11: The QQplot for the generalised linear model assessing the relationship of the concentration of *Melissococcus plutonius* DNA in larvae naturally infected with *Melissococcus plutonius*, with different DNA extraction methods being used (caterpillar, home-made column and Qiagen DNeasy). The family used for the model was Gaussian.

The normality test for the generalised linear model (Table 2.16) showed a normal distribution (Figure A.12).

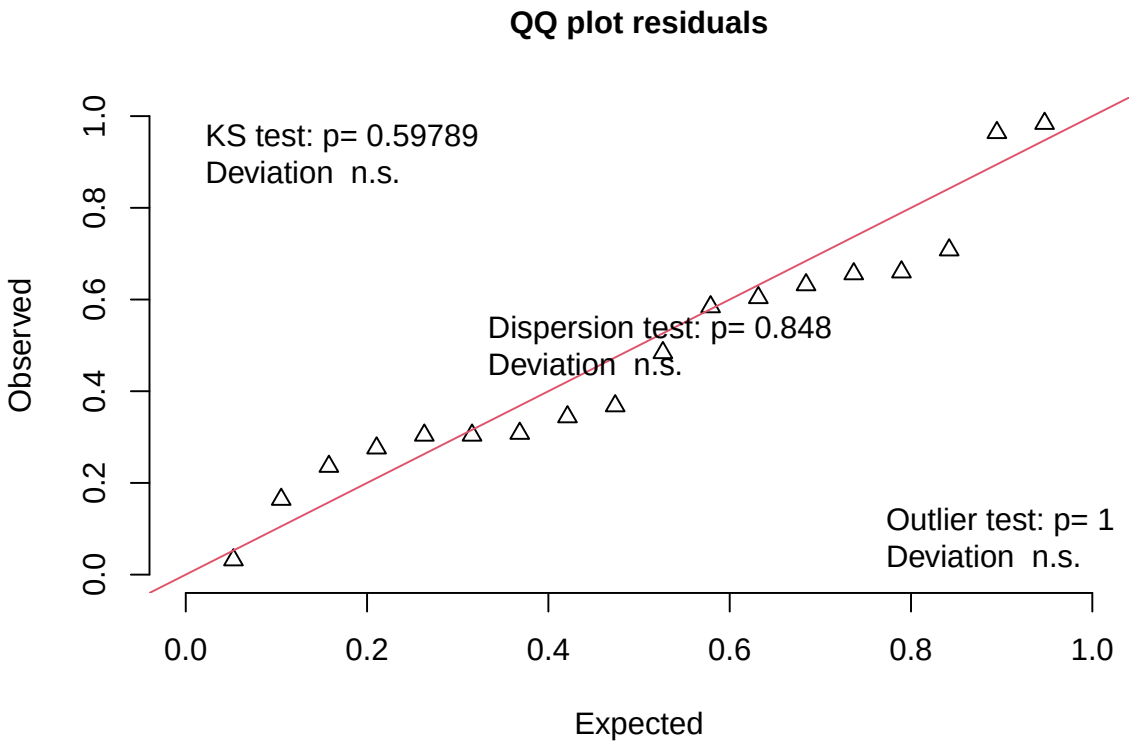


Figure A.12: The QQplot for the generalised linear model assessing the relationship of the ratio of the concentration of patogen target to host DNA in larvae naturally infected with *Melissococcus plutonius*, with different DNA extraction methods being used (caterpillar, home-made column and Qiagen DNeasy). The family used for the model was Gaussian.

The normality test for the generalised linear model (Table 2.17) showed a normal distribution (Figure A.13).

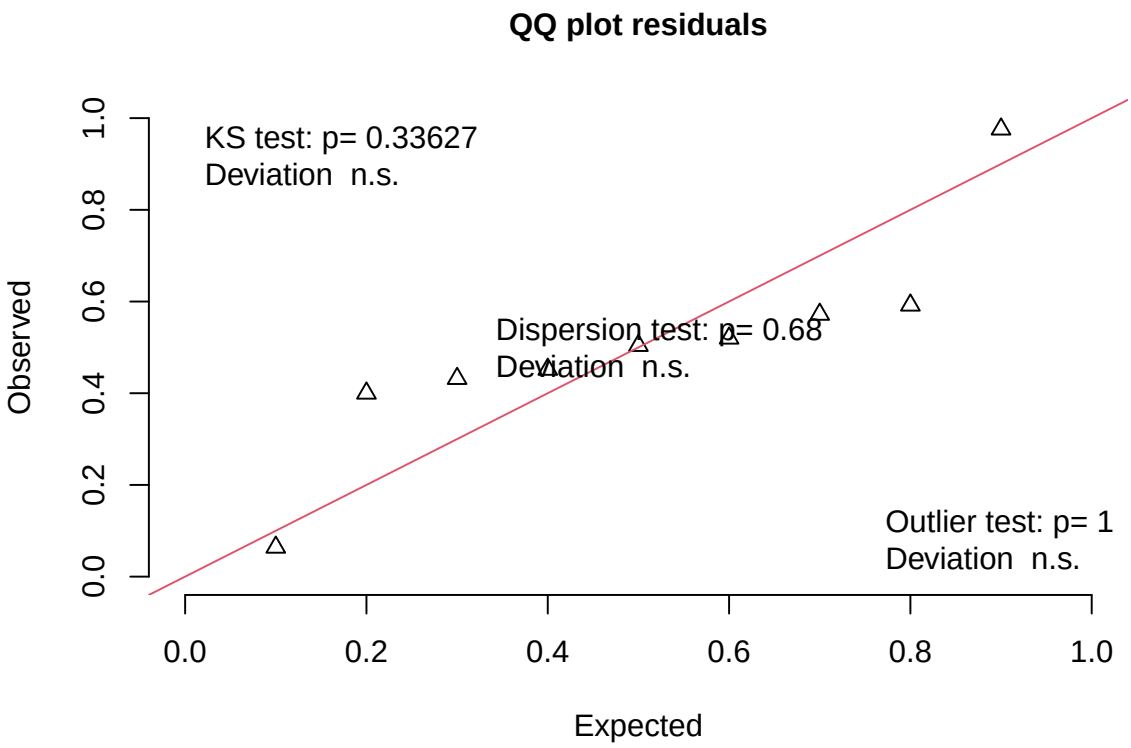


Figure A.13: The QQplot for the generalised linear model assessing the relationship of the concentration of *Melissococcus plutonius* DNA in larvae naturally infected with *Melissococcus plutonius*, with different DNA extraction methods being used (home-made bead, home-made column and Qiagen DNeasy). The family used for the model was Gaussian.

The normality test for the generalised linear model (Table 2.18) showed a normal distribution (Figure A.14).

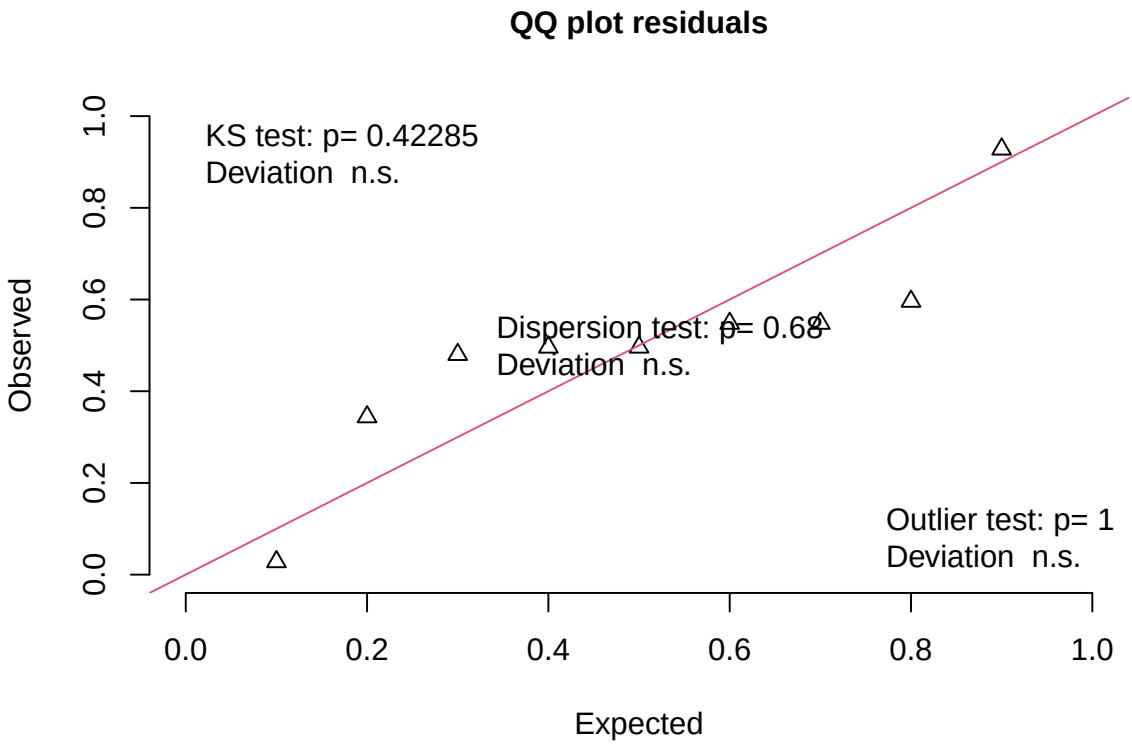


Figure A.14: QQplot for the generalised linear model assessing the relationship of the ratio of the concentration of patogen target to host DNA in larvae naturally infected with *Melissococcus plutonius*, with different DNA extraction methods being used (home-made bead, home-made column and Qiagen DNeasy). The family used for the model was Gaussian.

The normality test for the generalised linear model (Table 2.19) showed a normal distribution (Figure A.15).

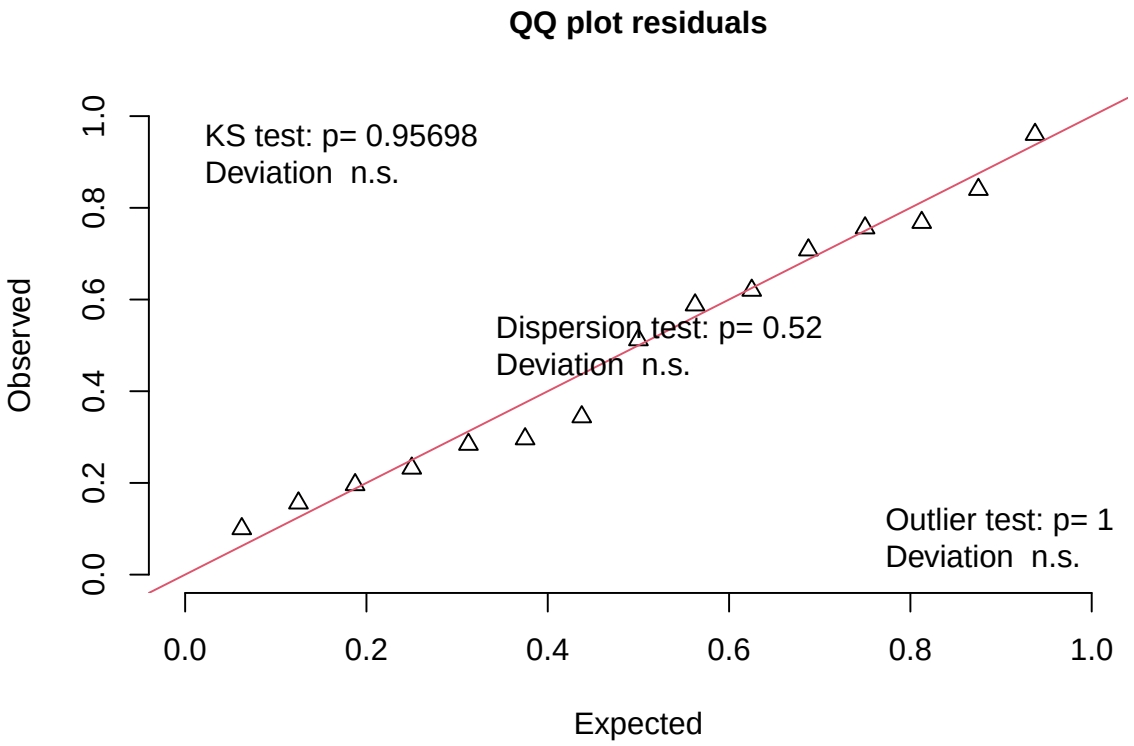


Figure A.15: QQplot for the generalised linear model assessing the relationship of the concentration of *Melissococcus plutonius* DNA in larvae naturally infected with *Melissococcus plutonius*, with different elution buffers (AE,TE,caterpillar and water), with the home-made column extraction method, as well as the standard Qiagen DNeasy kit with AE buffer. The family used for the model was Gaussian.

The normality test for the generalised linear model (Table 2.20) showed a normal distribution (Figure A.16).

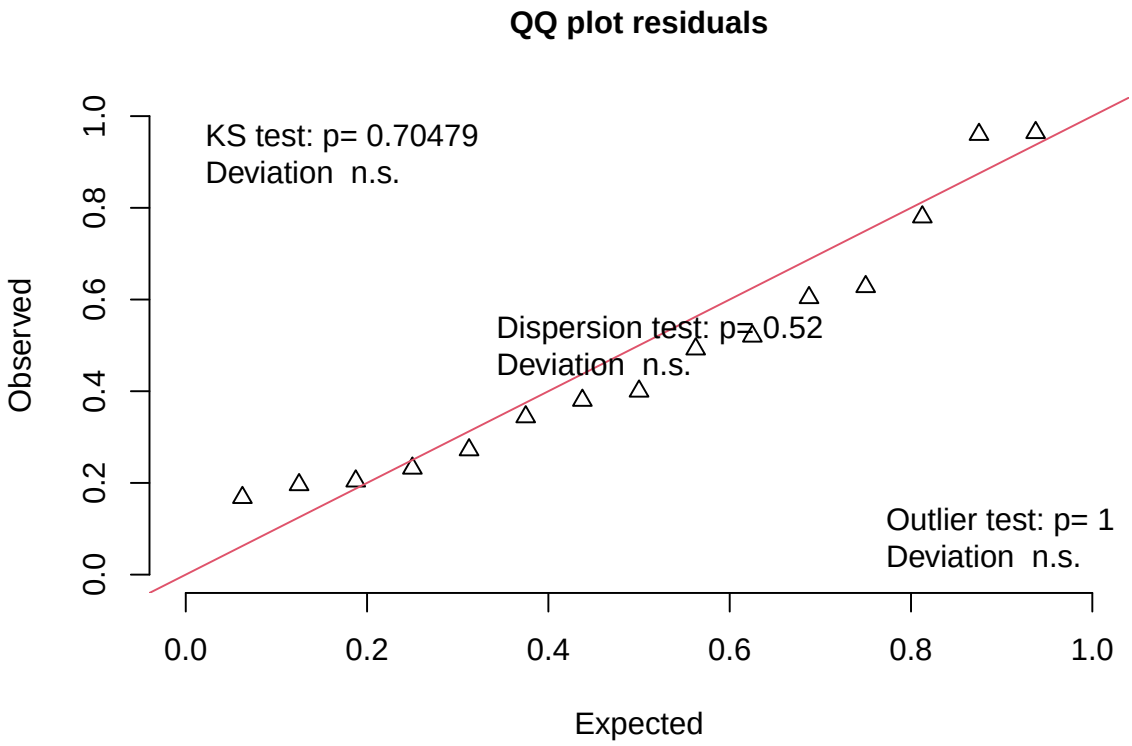


Figure A.16: The QQplot for the generalised linear model assessing the relationship of the ratio of the concentration of patogen target to host DNA in larvae naturally infected with *Melissococcus plutonius*, with different elution buffers (AE,TE,caterpillar and water), using the home-made column extraction method, compared to the standard Qiagen DNeasy kit with AE buffer for eluting DNA. The family used for the model was Gaussian.

The normality test for the generalised linear model (Table 2.21) showed a normal distribution (Figure A.17).

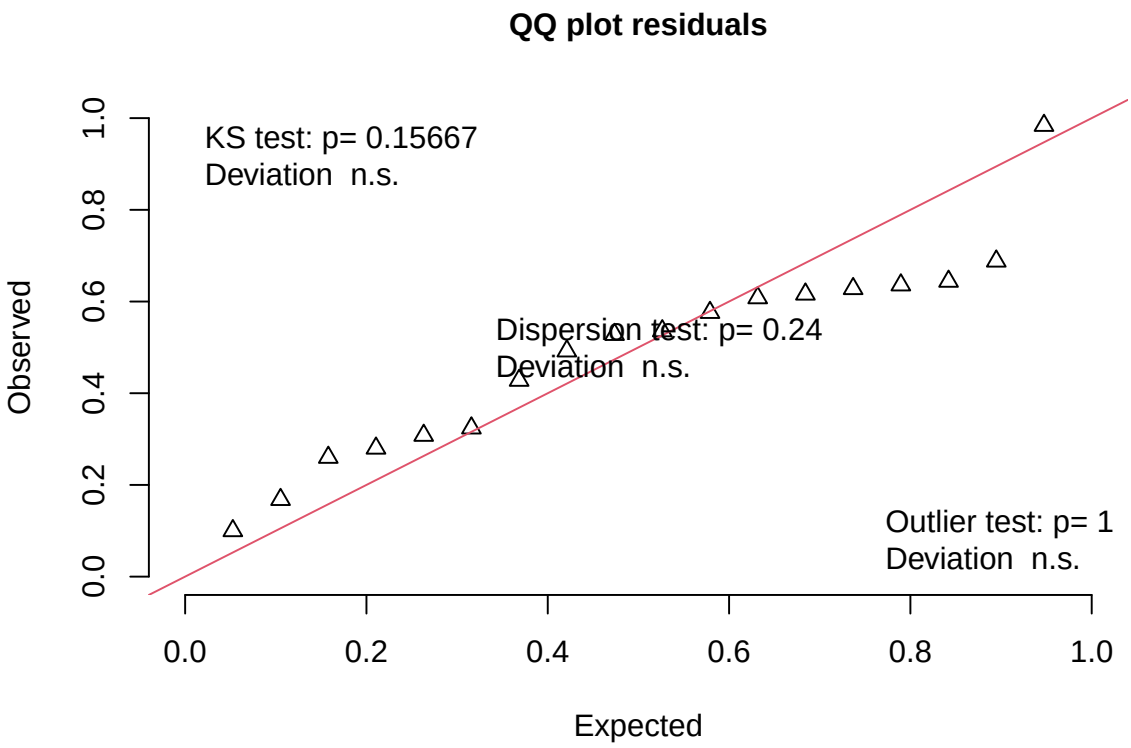


Figure A.17: The QQplot for the generalised linear model assessing the relationship of the ratio of the concentration of patogen target to host DNA in larvae naturally infected with *Melissococcus plutonius*, with samples treated different host depletion treatments (water lysis only, DNase with water lysis, 2 ul PMA, 10 μl PMA and 20 μl PMA) and sample input volume (200 μl, 1 ml and 2 ml). The family used for the model was Gaussian.

The normality test for the generalised linear model (Table 2.22) showed a normal distribution (Table A.18).

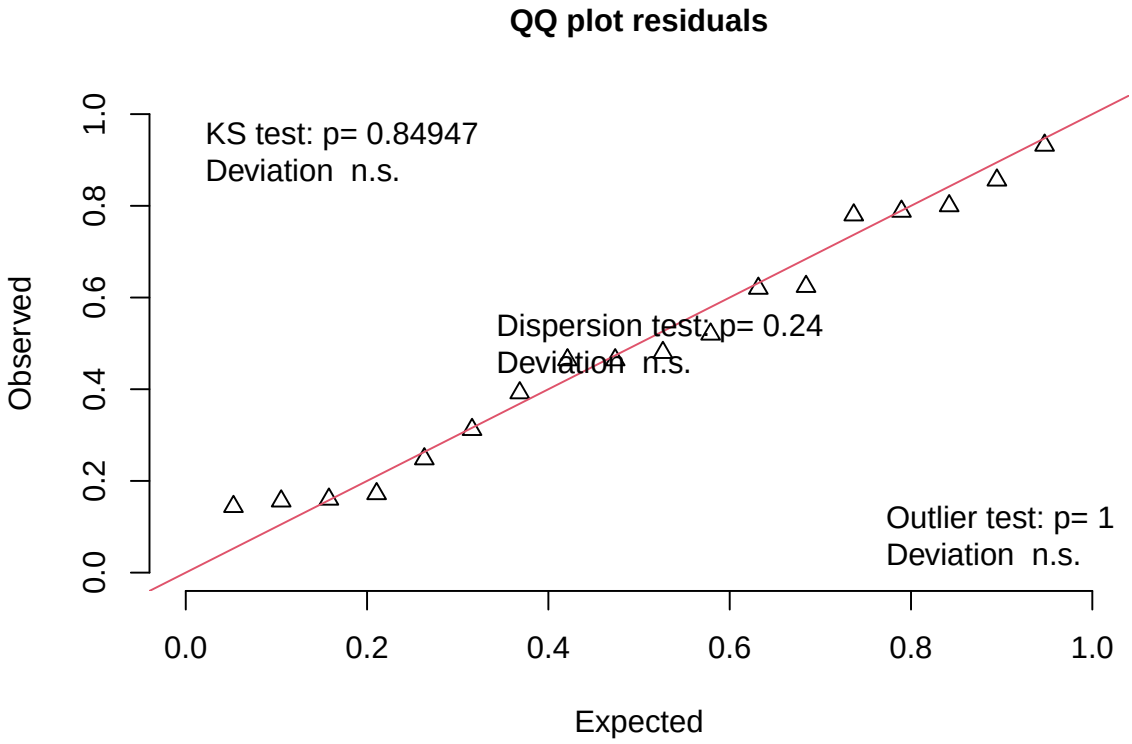


Figure A.18: The QQplot for the generalised linear model assessing the relationship of the concentration of *Melissococcus plutonius* DNA in larvae naturally infected with *Melissococcus plutonius*, with samples treated different host depletion treatments (water lysis only, DNase with water lysis, 2 μl PMA, 10 μl PMA and 20 μl PMA) and sample input volume (200 μl, 1 ml and 2 ml). The family used for the model was Gaussian.

Appendix B

Chapter 3 appendix

B.1 pMP19 phylogenetic tree generated from the SNPs

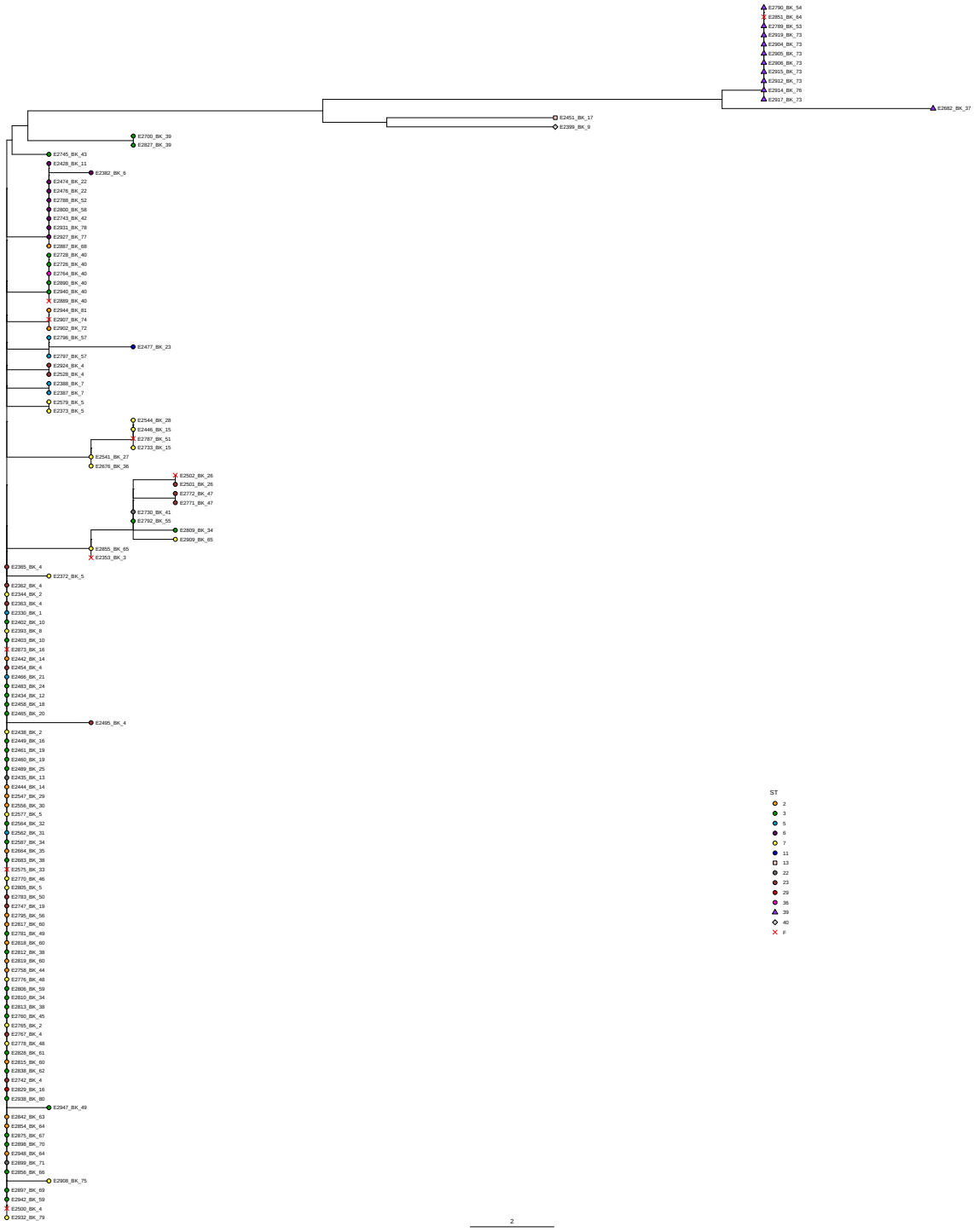


Figure B.1: A neighbourhood joining tree created using the matrix of SNPs chosen from pMP19 for further analysis. The E number is the unique identifier for the samples, BK represents beekeeper number. The scale bar represents the phylogenetic distance.

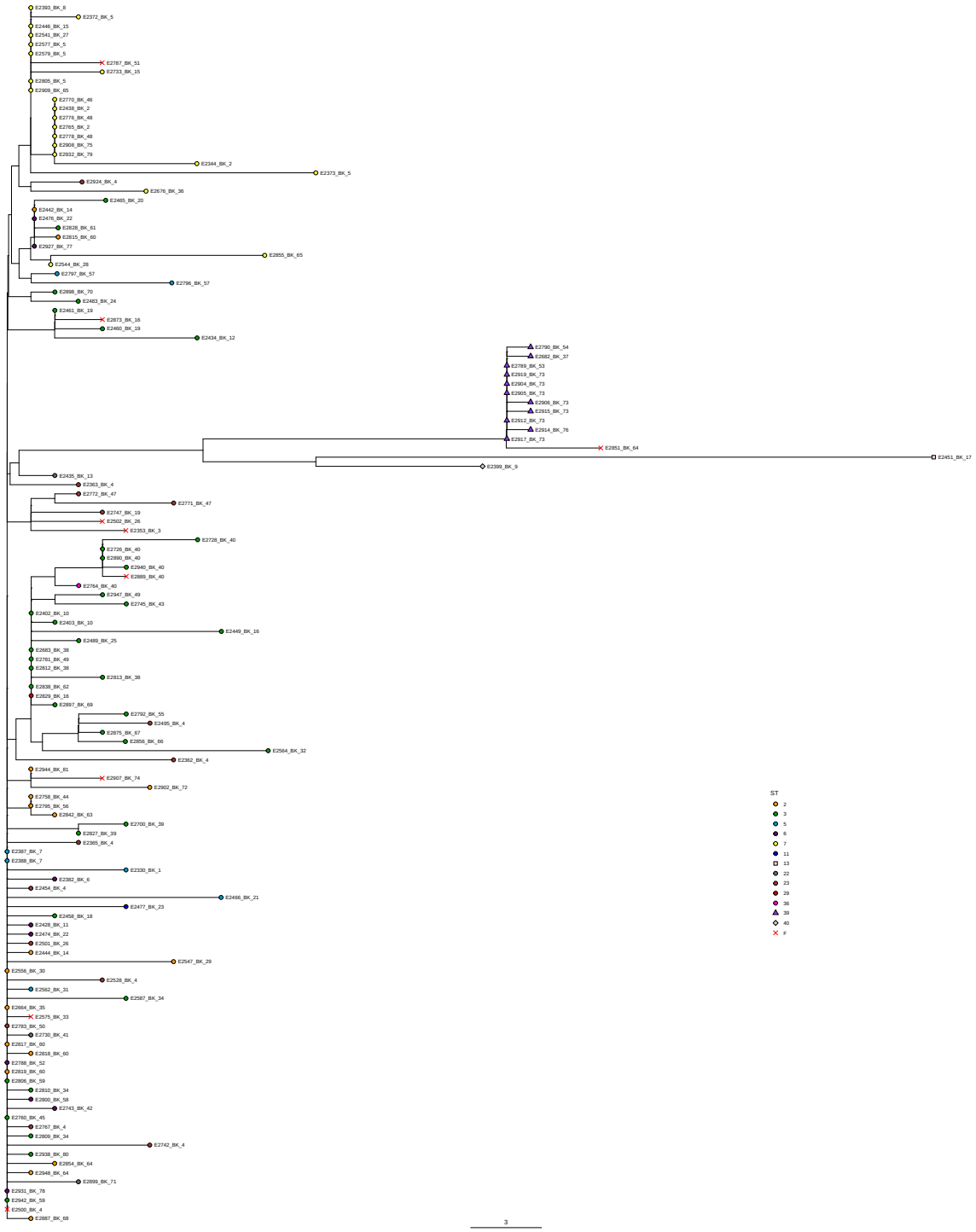


Figure B.2: A neighbourhood joining tree created using the matrix of SNPs chosen from pMP1 for further analysis. The E number is the unique identifier for the samples, BK represents beekeeper number. The scale bar represents the phylogenetic distance.

Appendix C

Chapter 5 appendix

C.1 Survey Reports

C.1.1 Somerset

1. How many production colonies do you own? A production colony is defined as a viable colony which could be used for either honey production or pollination services.

Mean colony number = 5.8976378

Number of colonies	Number of Responses
0	4
1	33
2	53
3	37
4	28
5	24
6	12
7	11

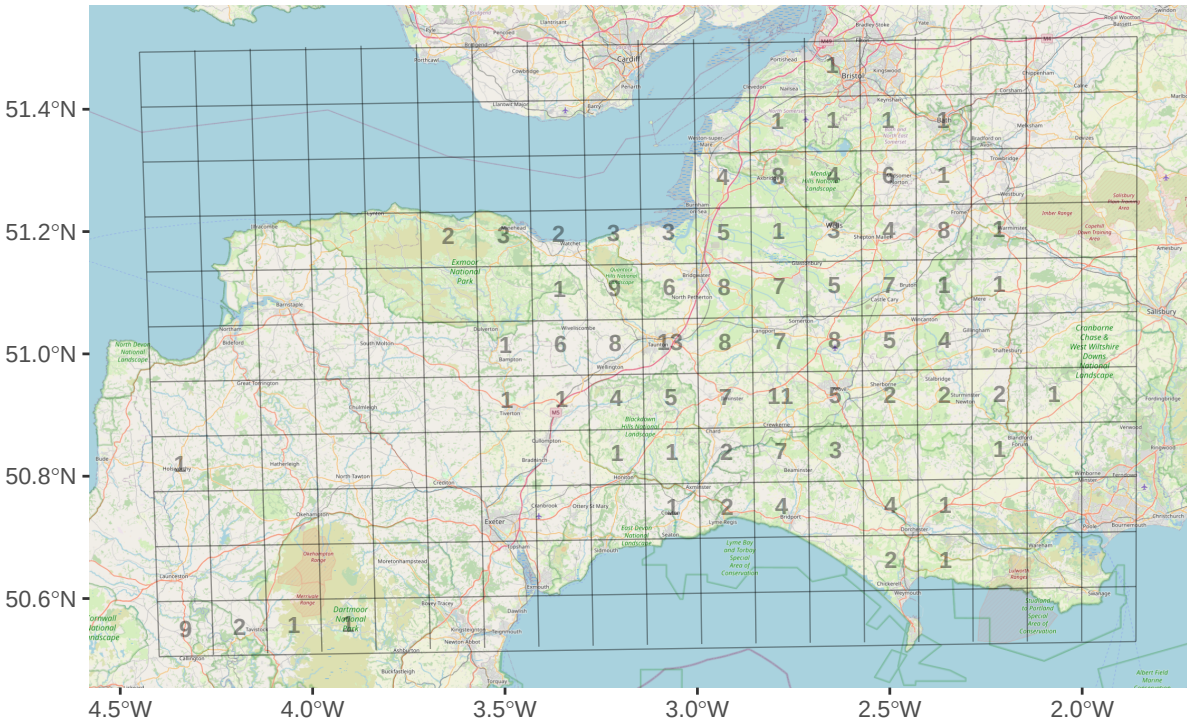
Number of colonies	Number of Responses
8	10
9	4
10	9
11	1
12	5
13	2
14	3
15	3
16	1
17	1
18	1
20	6
22	1
23	1
25	1
41	1
96	1
99	1

2. How many apiaries do you currently operate?

Mean apiary number= 1.7283465

Number of Apiaries	Number of Responses
0	2
1	165
2	55
3	11
4	10
5	3
6	3
7	1
8	1
14	2
15	1

3. Please take a moment to view the map below and indicate the number of the square that contains most of your apiaries? (Note you may need to scroll around to see all the squares)



4. Do you share any apiary sites with other beekeepers?

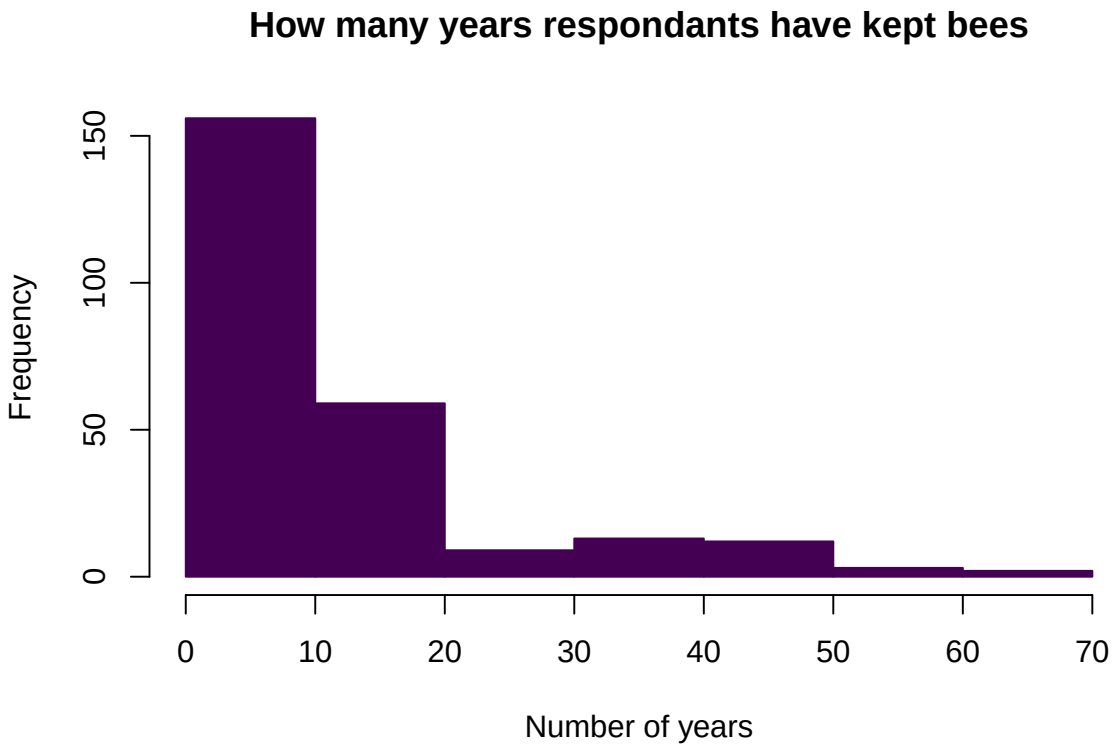
Shared sites	Number of Responses
No	229
Yes	25

5. Do you keep or tend bees at a beekeeping association apiary?

Bees kept at association apiary	Number of Responses
No	229
Yes	25

6. How many years have you kept bees?

Mean = 12.3346457



7. Are you currently a member of a local beekeeping association?

Member of an association	Number of Responses
No	3
Yes	251

7.a. If yes, please indicate which association(s) you belong to:

Association	Number of Responses
1	1
Avon and Somerset	1
B&SMNBKA	1
Bbka	1

Association	Number of Responses
Burnham & Highbridge	1
Burnham on Sea	1
Dorchester & Weymouth Beekeepers Association	6
Dorchester and Weymouth Beekeepers Association Somerset Beekeepers Association	1
FBKA	1
Frome and Bath	1
NFDBKA	1
North Dorset BKA	4
North Somerset, part of Avon And Wed more & Cheddar	1
QBKA	4
SESBKA	1
SESBKA Castle Cary	1
Sbka Mbka	1
Sherborne BKA	1
Sherborne and Yeovil	2
Somerset East Devon	1
Somerset BKA	40
Somerset BKA Somerton Division	1
Somerset BKA Burnham Division	7
Somerset BKA Central	6

Association	Number of Responses
Somerset BKA Exmoor Division	8
Somerset BKA Frome Division	12
Somerset BKA Mendip Division	4
Somerset BKA Quantock Division	15
Somerset BKA Somerton Division	32
Somerset BKA South-East Division	8
Somerset BKA South-West Division	5
Somerset BKA Taunton Division	30
Somerset BKA Wedmore & Cheddar Divison	4
Somerset BKA Wedmore and Cheddar Division	1
Somerset BKA Yeovil Division	16
Somerset Beekeepers Association	1
Somerset Beekeepers Association, British Beekeepers Association, Central Association of Beekeepers, International Bee Research Association.	1
Somerset, Exmoor Division	1
Somerton division of SBKA	1
South Somerset beekeepers association	2
South West Somerset BKA	1
TDBKA	1
WDBKA SBKA (associate)	1

Association	Number of Responses
Wedmore and Cheddar Division of Somerset Beekeepers Association	1
West Dorset Beekeepers	16
east somerset	1
somerset BKA BIBBA	1
	5

8. Do you have any beekeeping qualifications?

Beekeeping qualifications	Number of Responses
No	158
Yes	96

8.a. If yes, please indicate which association(s) you belong to:

Beekeeping Qualifications	Number of Responses
BBKA Basic	60
BBKA Basic,BBKA Bee health certificate	1
BBKA Basic,BBKA Bee health certificate,BBKA Module 1 - Honey Bee Management,BBKA Module 3 - Honey Bee Pests, Diseases and Poisoning	1
BBKA Basic,BBKA Generla certificate in beekeeping husbandry	1

Beekeeping Qualifications	Number of Responses
BBKA Basic,BBKA Generla certificate in beekeeping husbandry,BBKA Module 1 - Honey Bee Management,BBKA Module 2 - Honey Bee Products and Forage,BBKA Module 3 - Honey Bee Pests, Diseases and Poisoning,BBKA Module 5 - Honey Bee Biology,BBKA Module 6 - Honey Bee Behaviour,BBKA Module 7 - Selection & Breeding of Honey Bees	1
BBKA Basic,BBKA Generla certificate in beekeeping husbandry,BBKA Module 1 - Honey Bee Management,BBKA Module 2 - Honey Bee Products and Forage,BBKA Module 3 - Honey Bee Pests, Diseases and Poisoning,BBKA Module 5 - Honey Bee Biology,BBKA Module 6 - Honey Bee Behaviour,BBKA Module 7 - Selection & Breeding of Honey Bees,BBKA Module 8 - Honey Bee Management, Health and History,Other	1
BBKA Basic,BBKA Generla certificate in beekeeping husbandry,BBKA Module 1 - Honey Bee Management,BBKA Module 2 - Honey Bee Products and Forage,BBKA Module 3 - Honey Bee Pests, Diseases and Poisoning,BBKA Module 6 - Honey Bee Behaviour	1
BBKA Basic,BBKA Generla certificate in beekeeping husbandry,BBKA microscopy,BBKA Module 1 - Honey Bee Management	1
BBKA Basic,BBKA Generla certificate in beekeeping husbandry,BBKA microscopy,BBKA Module 1 - Honey Bee Management,BBKA Module 2 - Honey Bee Products and Forage,BBKA Module 3 - Honey Bee Pests, Diseases and Poisoning,BBKA Module 5 - Honey Bee Biology,BBKA Module 6 - Honey Bee Behaviour,BBKA Module 7 - Selection & Breeding of Honey Bees,BBKA Module 8 - Honey Bee Management, Health and History	2

Beekeeping Qualifications	Number of Responses
BBKA Basic,BBKA Generla certificate in beekeeping husbandry,BBKA microscopy,BBKA Module 1 - Honey Bee Management,BBKA Module 2 - Honey Bee Products and Forage,BBKA Module 3 - Honey Bee Pests, Diseases and Poisoning,BBKA Module 5 - Honey Bee Biology,BBKA Module 6 - Honey Bee Behaviour,BBKA Module 7 - Selection & Breeding of Honey Bees,BBKA Module 8 - Honey Bee Management, Health and History,Other	2
BBKA Basic,BBKA Module 1 - Honey Bee Management	2
BBKA Basic,BBKA Module 1 - Honey Bee Management,BBKA Module 2 - Honey Bee Products and Forage,BBKA Module 3 - Honey Bee Pests, Diseases and Poisoning	1
BBKA Basic,BBKA Module 1 - Honey Bee Management,BBKA Module 2 - Honey Bee Products and Forage,BBKA Module 3 - Honey Bee Pests, Diseases and Poisoning,BBKA Module 5 - Honey Bee Biology	1
BBKA Basic,BBKA Module 1 - Honey Bee Management,BBKA Module 2 - Honey Bee Products and Forage,BBKA Module 3 - Honey Bee Pests, Diseases and Poisoning,BBKA Module 5 - Honey Bee Biology,BBKA Module 6 - Honey Bee Behaviour,BBKA Module 7 - Selection & Breeding of Honey Bees,Other	1
BBKA Basic,BBKA Module 1 - Honey Bee Management,BBKA Module 2 - Honey Bee Products and Forage,BBKA Module 3 - Honey Bee Pests, Diseases and Poisoning,BBKA Module 6 - Honey Bee Behaviour	1
BBKA Basic,BBKA Module 1 - Honey Bee Management,BBKA Module 2 - Honey Bee Products and Forage,BBKA Module 3 - Honey Bee Pests, Diseases and Poisoning,BBKA Module 6 - Honey Bee Behaviour,BBKA Module 7 - Selection & Breeding of Honey Bees	1
BBKA Basic,BBKA Module 1 - Honey Bee Management,BBKA Module 2 - Honey Bee Products and Forage,BBKA Module 6 - Honey Bee Behaviour	1

Beekeeping Qualifications	Number of Responses
BBKA Basic,BBKA Module 1 - Honey Bee Management,BBKA Module 3 - Honey Bee Pests, Diseases and Poisoning,BBKA Module 5 - Honey Bee Biology	2
BBKA Basic,BBKA Module 1 - Honey Bee Management,BBKA Module 5 - Honey Bee Biology	1
BBKA Basic,BBKA Module 5 - Honey Bee Biology	3
BBKA Basic,BBKA microscopy,BBKA Module 1 - Honey Bee Management,BBKA Module 2 - Honey Bee Products and Forage,BBKA Module 3 - Honey Bee Pests, Diseases and Poisoning	1
BBKA Basic,BBKA microscopy,BBKA Module 1 - Honey Bee Management,BBKA Module 2 - Honey Bee Products and Forage,BBKA Module 3 - Honey Bee Pests, Diseases and Poisoning,BBKA Module 5 - Honey Bee Biology,BBKA Module 6 - Honey Bee Behaviour,BBKA Module 7 - Selection & Breeding of Honey Bees,BBKA Module 8 - Honey Bee Management, Health and History	1
BBKA Basic,BBKA microscopy,BBKA Module 3 - Honey Bee Pests, Diseases and Poisoning	1
BBKA Generla certificate in beekeeping husbandry,BBKA Module 1 - Honey Bee Management,BBKA Module 2 - Honey Bee Products and Forage,BBKA Module 3 - Honey Bee Pests, Diseases and Poisoning,BBKA Module 5 - Honey Bee Biology,BBKA Module 6 - Honey Bee Behaviour,BBKA Module 7 - Selection & Breeding of Honey Bees,BBKA Module 8 - Honey Bee Management, Health and History,Other	1

Beekeeping Qualifications	Number of Responses
BBKA Generla certificate in beekeeping husbandry,BBKA microscopy,BBKA Module 1 - Honey Bee Management,BBKA Module 2 - Honey Bee Products and Forage,BBKA Module 3 - Honey Bee Pests, Diseases and Poisoning,BBKA Module 5 - Honey Bee Biology,BBKA Module 6 - Honey Bee Behaviour,BBKA Module 7 - Selection & Breeding of Honey Bees,BBKA Module 8 - Honey Bee Management, Health and History,Other	1
BBKA Module 1 - Honey Bee Management	1
BBKA Module 1 - Honey Bee Management,BBKA Module 3 - Honey Bee Pests, Diseases and Poisoning,BBKA Module 5 - Honey Bee Biology	1
BBKA Module 1 - Honey Bee Management,BBKA Module 5 - Honey Bee Biology	1
BBKA Module 5 - Honey Bee Biology	1
Other	2
	158

9. Are you a member of the Bee Farmers Association?

Member of beefarmers association	Number of Responses
No	251
Yes	3

9.a. If yes, are you registered on the NBU Disease Accreditation Scheme for Honey bees (DASH)?

Registered on DASH	Number of Responses
No	2

10. Have you used any NBU literature for training or education?

Used any NBU literature	Number of responses
Asian Hornet ID Sheet	20
Asian Hornet ID Sheet,Common Pests, Diseases and Disorders of the Adult Honey Bee	5
Asian Hornet ID Sheet,Foulbrood Disease of Honey Bees	8
Asian Hornet ID Sheet,Foulbrood Disease of Honey Bees,Common Pests, Diseases and Disorders of the Adult Honey Bee	9
Asian Hornet ID Sheet,Foulbrood Disease of Honey Bees,Starting Right with Bees,Common Pests, Diseases and Disorders of the Adult Honey Bee	1
Asian Hornet ID Sheet,Foulbrood Disease of Honey Bees,Tropilaelaps: Parasitic Mites of Honey Bees,Common Pests, Diseases and Disorders of the Adult Honey Bee	1
Asian Hornet ID Sheet,Managing Varroa Small Hive Beetle	8
Asian Hornet ID Sheet,Managing Varroa Small Hive Beetle,Common Pests, Diseases and Disorders of the Adult Honey Bee	5
Asian Hornet ID Sheet,Managing Varroa Small Hive Beetle,Foulbrood Disease of Honey Bees	27

Used any NBU literature	Number of responses
Asian Hornet ID Sheet,Managing Varroa Small Hive Beetle,Foulbrood Disease of Honey Bees,Common Pests, Diseases and Disorders of the Adult Honey Bee	14
Asian Hornet ID Sheet,Managing Varroa Small Hive Beetle,Foulbrood Disease of Honey Bees,Starting Right with Bees	5
Asian Hornet ID Sheet,Managing Varroa Small Hive Beetle,Foulbrood Disease of Honey Bees,Starting Right with Bees,Common Pests, Diseases and Disorders of the Adult Honey Bee	3
Asian Hornet ID Sheet,Managing Varroa Small Hive Beetle,Foulbrood Disease of Honey Bees,Statutory Procedures Advisory Leaflet,Common Pests, Diseases and Disorders of the Adult Honey Bee	2
Asian Hornet ID Sheet,Managing Varroa Small Hive Beetle,Foulbrood Disease of Honey Bees,Tropilaelaps: Parasitic Mites of Honey Bees	8
Asian Hornet ID Sheet,Managing Varroa Small Hive Beetle,Foulbrood Disease of Honey Bees,Tropilaelaps: Parasitic Mites of Honey Bees,Common Pests, Diseases and Disorders of the Adult Honey Bee	23
Asian Hornet ID Sheet,Managing Varroa Small Hive Beetle,Foulbrood Disease of Honey Bees,Tropilaelaps: Parasitic Mites of Honey Bees,Starting Right with Bees	4

Used any NBU literature	Number of responses
Asian Hornet ID Sheet,Managing Varroa Small Hive Beetle,Foulbrood Disease of Honey Bees,Tropilaelaps: Parasitic Mites of Honey Bees,Starting Right with Bees,Common Pests, Diseases and Disorders of the Adult Honey Bee	8
Asian Hornet ID Sheet,Managing Varroa Small Hive Beetle,Foulbrood Disease of Honey Bees,Tropilaelaps: Parasitic Mites of Honey Bees,Starting Right with Bees,Statutory Procedures Advisory Leaflet,Common Pests, Diseases and Disorders of the Adult Honey Bee	7
Asian Hornet ID Sheet,Managing Varroa Small Hive Beetle,Foulbrood Disease of Honey Bees,Tropilaelaps: Parasitic Mites of Honey Bees,Starting Right with Bees,Statutory Procedures Advisory Leaflet,Contingency Planning Procedures Leaflet,Common Pests, Diseases and Disorders of the Adult Honey Bee	4
Asian Hornet ID Sheet,Managing Varroa Small Hive Beetle,Foulbrood Disease of Honey Bees,Tropilaelaps: Parasitic Mites of Honey Bees,Starting Right with Bees,Statutory Procedures Advisory Leaflet,Contingency Planning Procedures Leaflet,Common Pests, Diseases and Disorders of the Adult Honey Bee,Miller Queen Rearing Workbook	2
Asian Hornet ID Sheet,Managing Varroa Small Hive Beetle,Foulbrood Disease of Honey Bees,Tropilaelaps: Parasitic Mites of Honey Bees,Starting Right with Bees,Statutory Procedures Advisory Leaflet,Contingency Planning Procedures Leaflet,Common Pests, Diseases and Disorders of the Adult Honey Bee,Miller Queen Rearing Workbook,Other	1

Used any NBU literature	Number of responses
Asian Hornet ID Sheet,Managing Varroa Small Hive Beetle,Foulbrood Disease of Honey Bees,Tropilaelaps: Parasitic Mites of Honey Bees,Starting Right with Bees,Statutory Procedures Advisory Leaflet,Contingency Planning Procedures Leaflet,Common Pests, Diseases and Disorders of the Adult Honey Bee,Other	1
Asian Hornet ID Sheet,Managing Varroa Small Hive Beetle,Foulbrood Disease of Honey Bees,Tropilaelaps: Parasitic Mites of Honey Bees,Statutory Procedures Advisory Leaflet,Common Pests, Diseases and Disorders of the Adult Honey Bee	9
Asian Hornet ID Sheet,Managing Varroa Small Hive Beetle,Foulbrood Disease of Honey Bees,Tropilaelaps: Parasitic Mites of Honey Bees,Statutory Procedures Advisory Leaflet,Contingency Planning Procedures Leaflet,Common Pests, Diseases and Disorders of the Adult Honey Bee	1
Asian Hornet ID Sheet,Managing Varroa Small Hive Beetle,Other	1
Asian Hornet ID Sheet,Managing Varroa Small Hive Beetle,Starting Right with Bees	1
Asian Hornet ID Sheet,Managing Varroa Small Hive Beetle,Starting Right with Bees,Common Pests, Diseases and Disorders of the Adult Honey Bee	2
Asian Hornet ID Sheet,Managing Varroa Small Hive Beetle,Starting Right with Bees,Statutory Procedures Advisory Leaflet	1

Used any NBU literature	Number of responses
Asian Hornet ID Sheet,Managing Varroa Small Hive Beetle,Tropilaelaps: Parasitic Mites of Honey Bees	1
Asian Hornet ID Sheet,Managing Varroa Small Hive Beetle,Tropilaelaps: Parasitic Mites of Honey Bees,Common Pests, Diseases and Disorders of the Adult Honey Bee	1
Asian Hornet ID Sheet,Managing Varroa Small Hive Beetle,Tropilaelaps: Parasitic Mites of Honey Bees,Starting Right with Bees,Common Pests, Diseases and Disorders of the Adult Honey Bee	1
Asian Hornet ID Sheet,Starting Right with Bees	1
Asian Hornet ID Sheet,Statutory Procedures Advisory Leaflet	1
Asian Hornet ID Sheet,Statutory Procedures Advisory Leaflet,Contingency Planning Procedures Leaflet	1
Common Pests, Diseases and Disorders of the Adult Honey Bee	2
Foulbrood Disease of Honey Bees	1
Foulbrood Disease of Honey Bees,Common Pests, Diseases and Disorders of the Adult Honey Bee	2
Foulbrood Disease of Honey Bees,Statutory Procedures Advisory Leaflet,Contingency Planning Procedures Leaflet	1
Managing Varroa Small Hive Beetle	1
Managing Varroa Small Hive Beetle,Common Pests, Diseases and Disorders of the Adult Honey Bee	2

Used any NBU literature	Number of responses
Managing Varroa Small Hive Beetle,Foulbrood Disease of Honey Bees	2
Managing Varroa Small Hive Beetle,Foulbrood Disease of Honey Bees,Tropilaelaps: Parasitic Mites of Honey Bees	3
Managing Varroa Small Hive Beetle,Foulbrood Disease of Honey Bees,Tropilaelaps: Parasitic Mites of Honey Bees,Common Pests, Diseases and Disorders of the Adult Honey Bee	1
Managing Varroa Small Hive Beetle,Starting Right with Bees	2
Managing Varroa Small Hive Beetle,Starting Right with Bees,Common Pests, Diseases and Disorders of the Adult Honey Bee	2
Managing Varroa Small Hive Beetle,Statutory Procedures Advisory Leaflet	1
Managing Varroa Small Hive Beetle,Statutory Procedures Advisory Leaflet,Common Pests, Diseases and Disorders of the Adult Honey Bee	1
Managing Varroa Small Hive Beetle,Tropilaelaps: Parasitic Mites of Honey Bees,Starting Right with Bees,Common Pests, Diseases and Disorders of the Adult Honey Bee	1
None	22
Other	17
Starting Right with Bees	4

Used any NBU literature	Number of responses
Starting Right with Bees,Statutory Procedures Advisory Leaflet,Common Pests, Diseases and Disorders of the Adult Honey Bee	1
Statutory Procedures Advisory Leaflet,Common Pests, Diseases and Disorders of the Adult Honey Bee	1

10.a. If you selected Other, please specify:

Other literature used	Number of responses
Apiary Hygiene and Quarantine Bee Safaris The Essence of Beekeeping Hive Cleaning and Sterilisation Handling and Examining a Colony of Bees Apiary and Hive Hygiene	1
Association bee keeping course	1
BBKA literature	1
Cant remember	1
I have not used any of the NBU training material as Im not involved in training.	1
MAFF (DEFRA) productions.	1
Mbka beekeeping course	1
NBU Factsheets	1
None	6
None of above	1
None of these	1

Other literature used	Number of responses
Quantock Beekeeping course @ Brymore School	1
Somerton bee keeping course and books	1
none	2
	234

11. Do you collect swarms?

Collect swarms	Number of responses
No	117
Yes	137

11.a. If yes, in kilometres, what is the maximum distance you would travel from your apiary to collect a swarm?

Maximum distance for swarm collection	Number of responses
0	1
1	13
2	4
3	10
4	4
5	32
6	3
7	1

Maximum distance for swarm collection	Number of responses
8	7
10	29
12	2
15	10
16	2
18	1
20	12
25	3
30	2
50	1
	117

11.b. If yes, how do you house swarms?

How are the swarms housed	Number of responses
Deliberately starve for 3 days	2
Deliberately starve for 3 days,Other	1
In a box containing only foundation	36
In a box containing only foundation,Deliberately starve for 3 days	18
In a box containing only foundation,In a box with drawn comb	4

How are the swarms housed	Number of responses
In a box containing only foundation,In a box with drawn comb,Deliberately starve for 3 days	1
In a box containing only foundation,In a box with drawn comb,Initially in a quarantine apiary on their own	2
In a box containing only foundation,In a box with drawn comb,Initially in a quarantine apiary on their own,Deliberately starve for 3 days	2
In a box containing only foundation,Initially in a quarantine apiary on their own	7
In a box containing only foundation,Initially in a quarantine apiary on their own,Deliberately starve for 3 days	13
In a box with drawn comb	24
In a box with drawn comb,Deliberately starve for 3 days	4
In a box with drawn comb,Initially in a quarantine apiary on their own	5
In a box with drawn comb,Initially in a quarantine apiary on their own,Deliberately starve for 3 days	4
In a box with honey and pollen	3
Initially in a quarantine apiary on their own	7
Other	4
	117

11.b.i. If you selected Other, please specify:

How are the swarms housed	Number of responses
Cardboard box	1
In Home madePlastic collecting box	1
In a nucleus box. Empty with undreamed comb & feed with sugar syrup	1
In a top bar hive & let the bees get on with it.	1
In brood box or nuc box with undrawn and possibly some drawn foundation and 1 frame with open brood.	1
	249

12. Approximately how many swarms have left your operation in the last 2 years?

Mean = 1.6574803

How many swarms have left in the last 2 years	Number of responses
0	80
1	59
2	54
3	26
4	16
5	10
6	6
8	2
10	1

13. Do you move your bees to serve pollination contracts or to collect honey?

Do you move your bees	Number of responses
No	249
Yes	5

13.a. If yes ,please indicate which answer(s) apply

Why	Number of responses
Oilseed rape honey	3
Oilseed rape honey,Other	2
	249

13.a.i. If you selected Other, please specify:

Why other	Number of responses
Beans	1
Field beans	1
	252

14. Did you purchase honey bee colonies in the last two years?

Purchased honeybee colonies in the last 2 years	Number of responses
No	203
Yes	51

14.a. If yes, were they local in origin?

Local in origin	Number of responses
No	10
Yes	41
	203

15. Did you purchase honey bee queens in the last 2 years?

Purchased queens in last 2 years	Number of responses
No	199
Yes	55

15.a. If yes, where did your queens original from?

Where from	Number of responses
Imported	4
Locally sourced	19
Nationally sourced	31
Nationally sourced,Imported	1
	199

15.a.i. If you selected Other, please specify:

No responses

16. Do you used any shared honey extraction equipment (e.g. honey spinner or uncapping tray or settling tanks)?

Use shared honey extraction equipment	Number of responses
No	182
Yes	72

17. Are you registered as a beekeeper with the National Bee Unit?

Registered beekeeper	Number of responses
No	15
Unsure	24
Yes	215

18. Have your bees ever been inspected by a bee inspector from the National Bee Unit?

Inspection from the NBU	Number of responses
No	119
Yes	135

19. Are you registered with the National Bee Unit to receive foulbrood alerts?

Registered for foulbrood updates	Number of responses
No	24
Unsure	76
Yes	154

19.a. If yes, have you had notification of EFB in the last 2 years?

Update in the last 2 years	Number of responses
No	76
Unsure	12
Yes	66
	100

20. Do you wear gloves when handling bees?

Wear gloves when handling bees	Number of responses
No	6
Yes	248

20.a. If yes, do you use disposable gloves?

Disoposable gloves	Number of responses
No	100
Yes	148
	6

21. Do you soak your hive tool in disinfectant in between apiary visits?

Soak hive tools	Number of responses
No	77
Yes	177

22. Which foulbrood disease are you most worried about your bees catching?

Which foulbrood most worried about	Number of responses
American foulbrood (AFB)	10
Both	112
European foulbrood (EFB)	98
Neither	10
Unsure	24

23. Do you consider that European foulbrood (EFB) is a problem for beekeeping in your area?

Consider EFB a problem in area	Number of responses
No	90
Yes	164

23.a. If yes, what do you believe are the main reasons that European foulbrood (EFB) is a problem in your area?

Main reason	Number of responses
Beekeepers that are not registered on BeeBase	2
Beekeepers that are not registered on BeeBase,Beekeepers that cannot recognise the disease	3
Beekeepers that are not registered on BeeBase,Beekeepers that cannot recognise the disease,Lack of pollen,Poor weather for honey bees	1

Main reason	Number of responses
Beekeepers that are not registered on BeeBase,Beekeepers that cannot recognise the disease,Other	1
Beekeepers that are not registered on BeeBase,Beekeepers that cannot recognise the disease,Poor beekeeping practices at diseased sites	10
Beekeepers that are not registered on BeeBase,Beekeepers that cannot recognise the disease,Poor beekeeping practices at diseased sites,Feral nests	1
Beekeepers that are not registered on BeeBase,Beekeepers that cannot recognise the disease,Poor beekeeping practices at diseased sites,Honey packing plants	1
Beekeepers that are not registered on BeeBase,Beekeepers that cannot recognise the disease,Poor beekeeping practices at diseased sites,Infected swarms	13
Beekeepers that are not registered on BeeBase,Beekeepers that cannot recognise the disease,Poor beekeeping practices at diseased sites,Other	2
Beekeepers that are not registered on BeeBase,Beekeepers that cannot recognise the disease,Poor beekeeping practices at diseased sites,Unsure	1
Beekeepers that are not registered on BeeBase,Beekeepers that cannot recognise the disease,Poor weather for honey bees	1
Beekeepers that are not registered on BeeBase,Beekeepers that cannot recognise the disease,Unsure	1

Main reason	Number of responses
Beekeepers that are not registered on BeeBase,Poor beekeeping practices at diseased sites,Feral nests	1
Beekeepers that are not registered on BeeBase,Poor weather for honey bees,Feral nests	1
Beekeepers that cannot recognise the disease	7
Beekeepers that cannot recognise the disease,Feral nests,Infected swarms	1
Beekeepers that cannot recognise the disease,Infected swarms	3
Beekeepers that cannot recognise the disease,Infected swarms,Unsure	1
Beekeepers that cannot recognise the disease,Poor beekeeping practices at diseased sites	10
Beekeepers that cannot recognise the disease,Poor beekeeping practices at diseased sites,Feral nests	2
Beekeepers that cannot recognise the disease,Poor beekeeping practices at diseased sites,Feral nests,Infected swarms	2
Beekeepers that cannot recognise the disease,Poor beekeeping practices at diseased sites,Infected swarms	4
Beekeepers that cannot recognise the disease,Poor beekeeping practices at diseased sites,Infected swarms,Unsure	1
Beekeepers that cannot recognise the disease,Poor beekeeping practices at diseased sites,Other	1

Main reason	Number of responses
Beekeepers that cannot recognise the disease,Poor beekeeping practices at diseased sites,Poor weather for honey bees	1
Beekeepers that cannot recognise the disease,Poor beekeeping practices at diseased sites,Unsure	2
Beekeepers that cannot recognise the disease,Poor weather for honey bees,Feral nests	1
Beekeepers that cannot recognise the disease,Poor weather for honey bees,Infected swarms	1
Beekeepers that cannot recognise the disease,Poor weather for honey bees,Unsure	1
Beekeepers that cannot recognise the disease,Unsure	2
Failed treatments	1
Failed treatments,Beekeepers that are not registered on BeeBase,Beekeepers that cannot recognise the disease,Feral nests,Infected swarms,Unsure	1
Failed treatments,Beekeepers that are not registered on BeeBase,Beekeepers that cannot recognise the disease,Poor beekeeping practices at diseased sites	1
Failed treatments,Beekeepers that are not registered on BeeBase,Beekeepers that cannot recognise the disease,Poor beekeeping practices at diseased sites,Feral nests,Infected swarms	1
Failed treatments,Beekeepers that are not registered on BeeBase,Beekeepers that cannot recognise the disease,Poor beekeeping practices at diseased sites,Infected swarms	2

Main reason	Number of responses
Failed treatments,Beekeepers that are not registered on BeeBase,Beekeepers that cannot recognise the disease,Poor beekeeping practices at diseased sites,Unsure,Other	1
Failed treatments,Beekeepers that cannot recognise the disease,Feral nests	1
Failed treatments,Beekeepers that cannot recognise the disease,Poor beekeeping practices at diseased sites	2
Failed treatments,Beekeepers that cannot recognise the disease,Poor beekeeping practices at diseased sites,Poor weather for honey bees	1
Failed treatments,Poor beekeeping practices at diseased sites	1
Failed treatments,Unsure,Other	1
Feral nests	2
Feral nests,Infected swarms	1
Infected swarms	1
Other	2
Poor beekeeping practices at diseased sites	6
Poor beekeeping practices at diseased sites,Feral nests,Infected swarms	2
Poor beekeeping practices at diseased sites,Feral nests,Unsure	1
Poor beekeeping practices at diseased sites,Infected swarms	2
Poor beekeeping practices at diseased sites,Lack of pollen	1

Main reason	Number of responses
Poor beekeeping practices at diseased sites,Other	1
Poor beekeeping practices at diseased sites,Poor weather for honey bees,Feral nests,Infected swarms,Unsure	1
Poor beekeeping practices at diseased sites,Poor weather for honey bees,Infected swarms	1
Poor beekeeping practices at diseased sites,Poor weather for honey bees,Infected swarms,Other	1
Poor beekeeping practices at diseased sites,Poor weather for honey bees,Other	1
Poor weather for honey bees,Unsure	1
Unsure	48
	90

23.a.i. If you selected Other, please specify:

Data not shown

24. Are confident that you could recognise European foulbrood (EFB) in your colonies?

Confident to recognise EFB	Number of responses
No	24
Unsure	79
Yes	151

25. Have you ever seen European foulbrood in a honey bee colony?

Ever seen EFB in a colony	Number of responses
No	195
Unsure	3
Yes	56

26. Have your bees ever been diagnosed with European foulbrood (EFB)?

Ever had EFB diagnosed in your bees	Number of responses
No	229
Yes	25

26.a. If yes, please indicate when your bees were last diagnosed with European foulbrood?

When were bees diagnosed	Number of responses
Before 2019	15
Between 2019 and 2020	5
Within the last two years	5
	229

27. Do you know of any feral colonies within 2 km of your apiary site(s)?

Know of any feral colonies within 2km of apiary site	Number of responses
No	205
Yes	49

28. Is there any other information you wish to share with us about European foulbrood (EFB) in your region?

Data not shown

C.1.2 Cambridgeshire

1. How many production colonies do you own? A production colony is defined as a viable colony which could be used for either honey production or pollination services.

Mean colony number = 6.3669725

Number of colonies	Number of Responses
1	16
2	18
3	13
4	13
5	9
6	3
7	8
8	7
9	3
10	4
11	2
12	3
13	1
14	1
16	1
20	1
22	1

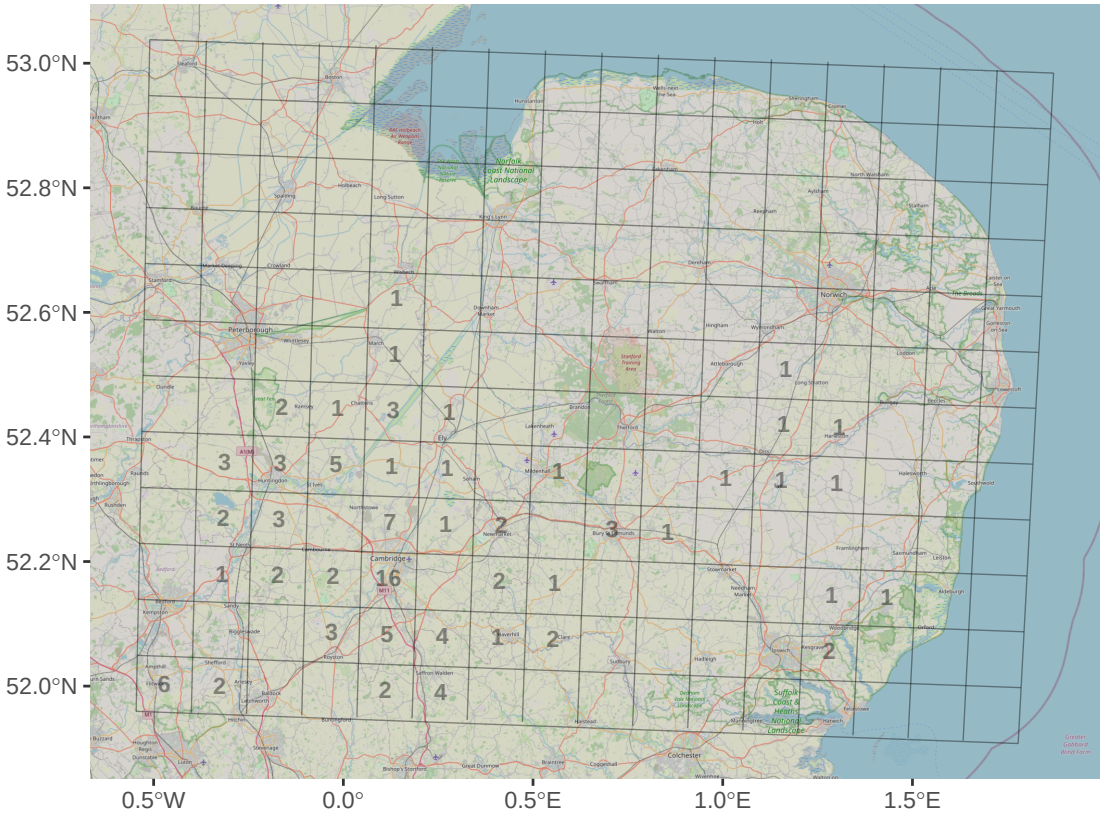
Number of colonies	Number of Responses
24	1
25	1
26	1
40	1
51	1

2. How many apiaries do you currently operate?

Mean apiary number= 1.6422018

Number of Apiaries	Number of Responses
0	1
1	67
2	27
3	8
4	3
5	2
12	1

3. Please take a moment to view the map below and indicate the number of the square that contains most of your apiaries? (Note you may need to scroll around to see all the squares)



4. Do you share any apiary sites with other beekeepers?

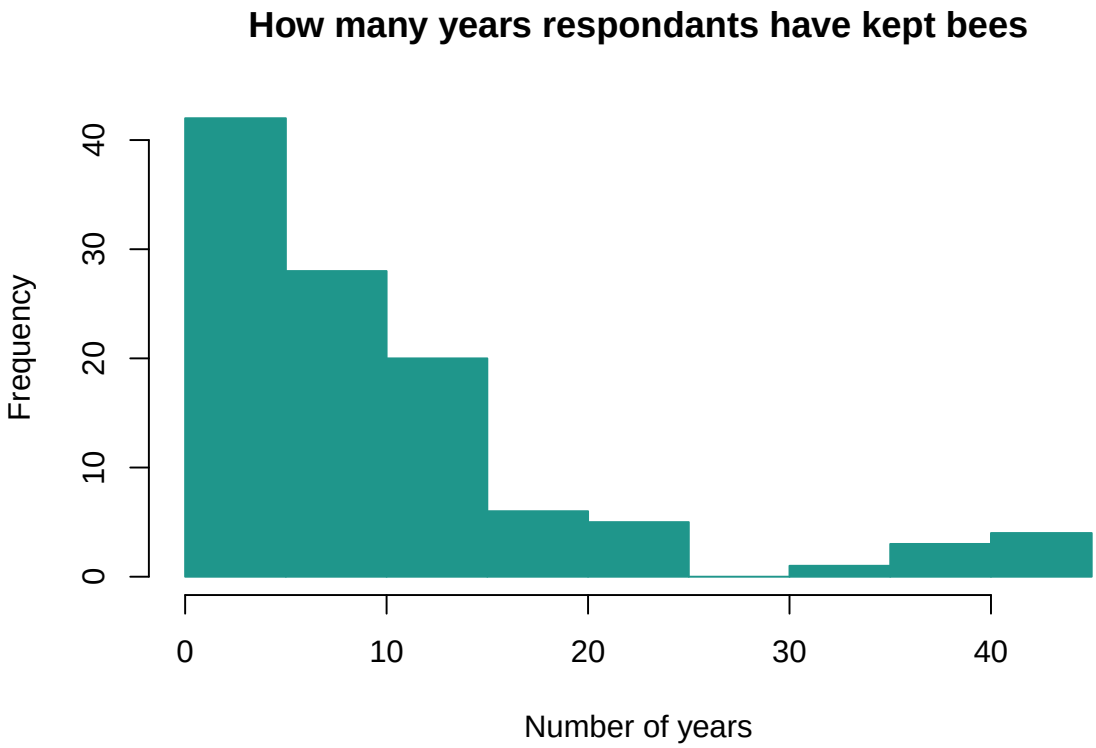
Shared sites	Number of Responses
No	97
Yes	12

5. Do you keep or tend bees at a beekeeping association apiary?

Bees kept at association apiary	Number of Responses
No	101
Yes	8

6. How many years have you kept bees?

Mean = 10.7798165



7. Are you currently a member of a local beekeeping association?

Member of an association	Number of Responses
Yes	109

7.a. If yes, please indicate which association(s) you belong to:

Association	Number of Responses
1. Dicklebees Bee-keeping support group; 2. Waveney Bee-keepers Group	1
Bedfordshire Beekeepers Association	1
CBKA	60
CBKA HBKA	1

Association	Number of Responses
CBKA & BedsBKA	1
Cambridgeshire beekeeping association Somerset beekeeping association	1
Dicklebees support group	1
Essex, Cambridge	1
HBKA	2
Harlow , Saffron Walden, CBKA, Beds BKA, Somerset BKA	1
Huntingdonshire BKA	11
Huntingdonshire BKA Cambridgeshire BKA	1
IESBKA (Suffolk Bee Keepers associtation subgroup)	1
Iesbka	1
Ipswich & East Suffolk BKA	1
Ipswich and East Suffolk	1
Ipswich and east Suffolk	1
LBKA NBKA Dickleburgh Bees	1
NBKA	1
Norfolk BKA & IESBKA	1
Norwich & District BKA	1
Peterborough & District BKA, Cambridge BKA, Somerset BKA	1
SWBKA	1
Saffron Walden	3

Association	Number of Responses
Saffron Walden Beekeeping Assoc.	1
Saffron Walden Division of Essex BK Association	1
Saffron Walden bka	1
Saffron Waldon BKA	1
Stowmarket BKA Cambridge BKA Dicklebees	1
Waveney Bee Group Dicklebees	1
Waveney Beekeepers	1
West Suffolk BKA	5
West Suffolk Beekeeping association Cambridge beekeeping association	1

8. Do you have any beekeeping qualifications?

Beekeeping qualifications	Number of Responses
No	55
Yes	54

8.a. If yes, please indicate which association(s) you belong to:

Beekeeping Qualifications	Number of Responses
BBKA Basic	33
BBKA Basic,BBKA Bee health certificate,BBKA Module 1 - Honey Bee Management,BBKA Module 3 - Honey Bee Pests, Diseases and Poisoning	1

Beekeeping Qualifications	Number of Responses
BBKA Basic,BBKA Generla certificate in beekeeping husbandry	1
BBKA Basic,BBKA Generla certificate in beekeeping husbandry,BBKA Module 1 - Honey Bee Management,BBKA Module 2 - Honey Bee Products and Forage,BBKA Module 3 - Honey Bee Pests, Diseases and Poisoning	1
BBKA Basic,BBKA Generla certificate in beekeeping husbandry,BBKA microscopy,BBKA Module 1 - Honey Bee Management,BBKA Module 2 - Honey Bee Products and Forage,BBKA Module 3 - Honey Bee Pests, Diseases and Poisoning,BBKA Module 5 - Honey Bee Biology,BBKA Module 6 - Honey Bee Behaviour,BBKA Module 7 - Selection & Breeding of Honey Bees,BBKA Module 8 - Honey Bee Management, Health and History,Other	1
BBKA Basic,BBKA Module 1 - Honey Bee Management	1
BBKA Basic,BBKA Module 1 - Honey Bee Management,BBKA Module 2 - Honey Bee Products and Forage	2
BBKA Basic,BBKA Module 1 - Honey Bee Management,BBKA Module 2 - Honey Bee Products and Forage,BBKA Module 3 - Honey Bee Pests, Diseases and Poisoning	2
BBKA Basic,BBKA Module 1 - Honey Bee Management,BBKA Module 2 - Honey Bee Products and Forage,BBKA Module 3 - Honey Bee Pests, Diseases and Poisoning,BBKA Module 5 - Honey Bee Biology,BBKA Module 6 - Honey Bee Behaviour	1
BBKA Basic,BBKA Module 1 - Honey Bee Management,BBKA Module 2 - Honey Bee Products and Forage,BBKA Module 3 - Honey Bee Pests, Diseases and Poisoning,BBKA Module 5 - Honey Bee Biology,BBKA Module 6 - Honey Bee Behaviour,BBKA Module 7 - Selection & Breeding of Honey Bees,BBKA Module 8 - Honey Bee Management, Health and History	1

Beekeeping Qualifications	Number of Responses
BBKA Basic,BBKA Module 1 - Honey Bee Management,BBKA Module 2 - Honey Bee Products and Forage,BBKA Module 3 - Honey Bee Pests, Diseases and Poisoning,BBKA Module 6 - Honey Bee Behaviour	1
BBKA Basic,BBKA Module 1 - Honey Bee Management,BBKA Module 3 - Honey Bee Pests, Diseases and Poisoning	1
BBKA Basic,BBKA Module 3 - Honey Bee Pests, Diseases and Poisoning	1
BBKA Generla certificate in beekeeping husbandry	1
BBKA Generla certificate in beekeeping husbandry,BBKA Module 1 - Honey Bee Management,BBKA Module 2 - Honey Bee Products and Forage,BBKA Module 3 - Honey Bee Pests, Diseases and Poisoning,BBKA Module 5 - Honey Bee Biology,BBKA Module 6 - Honey Bee Behaviour,BBKA Module 7 - Selection & Breeding of Honey Bees,BBKA Module 8 - Honey Bee Management, Health and History	1
BBKA Module 1 - Honey Bee Management	4
Other	1
	55

9. Are you a member of the Bee Farmers Association?

Member of beefarmers association	Number of Responses
No	108
Yes	1

9.a. If yes, are you registered on the NBU Disease Accreditation Scheme for Honey bees (DASH)?

Registered on DASH	Number of Responses
No	1

10. Have you used any NBU literature for training or education?

Used any NBU literature	Number of responses
Asian Hornet ID Sheet	4
Asian Hornet ID Sheet,Foulbrood Disease of Honey Bees	1
Asian Hornet ID Sheet,Foulbrood Disease of Honey Bees,Statutory Procedures Advisory Leaflet	1
Asian Hornet ID Sheet,Managing Varroa Small Hive Beetle	5
Asian Hornet ID Sheet,Managing Varroa Small Hive Beetle,Common Pests, Diseases and Disorders of the Adult Honey Bee	4
Asian Hornet ID Sheet,Managing Varroa Small Hive Beetle,Foulbrood Disease of Honey Bees	8
Asian Hornet ID Sheet,Managing Varroa Small Hive Beetle,Foulbrood Disease of Honey Bees,Common Pests, Diseases and Disorders of the Adult Honey Bee	2
Asian Hornet ID Sheet,Managing Varroa Small Hive Beetle,Foulbrood Disease of Honey Bees,Common Pests, Diseases and Disorders of the Adult Honey Bee,Miller Queen Rearing Workbook	1
Asian Hornet ID Sheet,Managing Varroa Small Hive Beetle,Foulbrood Disease of Honey Bees,Starting Right with Bees,Statutory Procedures Advisory Leaflet,Common Pests, Diseases and Disorders of the Adult Honey Bee	3

Used any NBU literature	Number of responses
Asian Hornet ID Sheet,Managing Varroa Small Hive Beetle,Foulbrood Disease of Honey Bees,Statutory Procedures Advisory Leaflet	1
Asian Hornet ID Sheet,Managing Varroa Small Hive Beetle,Foulbrood Disease of Honey Bees,Statutory Procedures Advisory Leaflet,Common Pests, Diseases and Disorders of the Adult Honey Bee	1
Asian Hornet ID Sheet,Managing Varroa Small Hive Beetle,Foulbrood Disease of Honey Bees,Tropilaelaps: Parasitic Mites of Honey Bees	3
Asian Hornet ID Sheet,Managing Varroa Small Hive Beetle,Foulbrood Disease of Honey Bees,Tropilaelaps: Parasitic Mites of Honey Bees,Common Pests, Diseases and Disorders of the Adult Honey Bee	7
Asian Hornet ID Sheet,Managing Varroa Small Hive Beetle,Foulbrood Disease of Honey Bees,Tropilaelaps: Parasitic Mites of Honey Bees,Common Pests, Diseases and Disorders of the Adult Honey Bee,Miller Queen Rearing Workbook	1
Asian Hornet ID Sheet,Managing Varroa Small Hive Beetle,Foulbrood Disease of Honey Bees,Tropilaelaps: Parasitic Mites of Honey Bees,Contingency Planning Procedures Leaflet,Common Pests, Diseases and Disorders of the Adult Honey Bee	1
Asian Hornet ID Sheet,Managing Varroa Small Hive Beetle,Foulbrood Disease of Honey Bees,Tropilaelaps: Parasitic Mites of Honey Bees,Starting Right with Bees,Common Pests, Diseases and Disorders of the Adult Honey Bee	1
Asian Hornet ID Sheet,Managing Varroa Small Hive Beetle,Foulbrood Disease of Honey Bees,Tropilaelaps: Parasitic Mites of Honey Bees,Starting Right with Bees,Statutory Procedures Advisory Leaflet,Common Pests, Diseases and Disorders of the Adult Honey Bee	4

Used any NBU literature	Number of responses
Asian Hornet ID Sheet,Managing Varroa Small Hive Beetle,Foulbrood Disease of Honey Bees,Tropilaelaps: Parasitic Mites of Honey Bees,Starting Right with Bees,Statutory Procedures Advisory Leaflet,Common Pests, Diseases and Disorders of the Adult Honey Bee,Other	1
Asian Hornet ID Sheet,Managing Varroa Small Hive Beetle,Foulbrood Disease of Honey Bees,Tropilaelaps: Parasitic Mites of Honey Bees,Statutory Procedures Advisory Leaflet	1
Asian Hornet ID Sheet,Managing Varroa Small Hive Beetle,Foulbrood Disease of Honey Bees,Tropilaelaps: Parasitic Mites of Honey Bees,Statutory Procedures Advisory Leaflet,Common Pests, Diseases and Disorders of the Adult Honey Bee	4
Asian Hornet ID Sheet,Managing Varroa Small Hive Beetle,Foulbrood Disease of Honey Bees,Tropilaelaps: Parasitic Mites of Honey Bees,Statutory Procedures Advisory Leaflet,Contingency Planning Procedures Leaflet,Common Pests, Diseases and Disorders of the Adult Honey Bee	1
Asian Hornet ID Sheet,Managing Varroa Small Hive Beetle,Starting Right with Bees,Common Pests, Diseases and Disorders of the Adult Honey Bee	2
Asian Hornet ID Sheet,Managing Varroa Small Hive Beetle,Tropilaelaps: Parasitic Mites of Honey Bees,Common Pests, Diseases and Disorders of the Adult Honey Bee	1
Asian Hornet ID Sheet,Starting Right with Bees,Common Pests, Diseases and Disorders of the Adult Honey Bee	1
Common Pests, Diseases and Disorders of the Adult Honey Bee	2
Foulbrood Disease of Honey Bees	1
Foulbrood Disease of Honey Bees,Common Pests, Diseases and Disorders of the Adult Honey Bee	3

Used any NBU literature	Number of responses
Managing Varroa Small Hive Beetle	4
Managing Varroa Small Hive Beetle,Common Pests, Diseases and Disorders of the Adult Honey Bee	1
Managing Varroa Small Hive Beetle,Foulbrood Disease of Honey Bees	3
Managing Varroa Small Hive Beetle,Foulbrood Disease of Honey Bees,Common Pests, Diseases and Disorders of the Adult Honey Bee	6
Managing Varroa Small Hive Beetle,Foulbrood Disease of Honey Bees,Starting Right with Bees,Common Pests, Diseases and Disorders of the Adult Honey Bee	1
Managing Varroa Small Hive Beetle,Foulbrood Disease of Honey Bees,Tropilaelaps: Parasitic Mites of Honey Bees	1
Managing Varroa Small Hive Beetle,Foulbrood Disease of Honey Bees,Tropilaelaps: Parasitic Mites of Honey Bees,Common Pests, Diseases and Disorders of the Adult Honey Bee	2
Managing Varroa Small Hive Beetle,Starting Right with Bees,Common Pests, Diseases and Disorders of the Adult Honey Bee	1
None	20
Other	5

10.a. If you selected Other, please specify:

Other literature used	Number of responses
Apiary Hygiene Hive Cleaning Sterlizing	1
CBKAs zoom seminar	1

Other literature used	Number of responses
I received training from the HBKA but I dont know what sources they used	1
None	3
	103

11. Do you collect swarms?

Collect swarms	Number of responses
No	41
Yes	68

11.a. If yes, in kilometres, what is the maximum distance you would travel from your apiary to collect a swarm?

Maximum distance for swarm collection	Number of responses
0	1
1	3
2	2
3	6
4	2
5	17
6	2
10	18
12	1

Maximum distance for swarm collection	Number of responses
15	6
16	1
18	1
20	4
23	1
25	1
30	2
	41

11.b. If yes, how do you house swarms?

How are the swarms housed	Number of responses
Deliberately starve for 3 days	1
In a box containing only foundation	15
In a box containing only foundation,Deliberately starve for 3 days	9
In a box containing only foundation,In a box with drawn comb	1
In a box containing only foundation,In a box with drawn comb,Deliberately starve for 3 days	2
In a box containing only foundation,In a box with drawn comb,Initially in a quarantine apiary on their own	1

How are the swarms housed	Number of responses
In a box containing only foundation,In a box with drawn comb,Initially in a quarantine apiary on their own,Other	1
In a box containing only foundation,Initially in a quarantine apiary on their own	4
In a box containing only foundation,Initially in a quarantine apiary on their own,Deliberately starve for 3 days	5
In a box containing only foundation,Initially in a quarantine apiary on their own,Deliberately starve for 3 days,Other	1
In a box with drawn comb	14
In a box with drawn comb,Deliberately starve for 3 days	1
In a box with drawn comb,In a box with honey and pollen,Initially in a quarantine apiary on their own	1
In a box with drawn comb,Initially in a quarantine apiary on their own	3
In a box with drawn comb,Initially in a quarantine apiary on their own,Deliberately starve for 3 days	1
In a box with honey and pollen	1
Initially in a quarantine apiary on their own	4
Other	3
	41

11.b.i. If you selected Other, please specify:

How are the swarms housed	Number of responses
Collect to pass on to other beekeepers to are on swarm wanted list.	1
Moved into a hive	1
Oxalid dribble in first week	1
foundationless frames	1
they usually go to new bee keeper who has a hive ready	1
	104

12. Approximately how many swarms have left your operation in the last 2 years?

Mean = 2.2018349

How many swarms have left in the last 2 years	Number of responses
0	32
1	20
2	23
3	11
4	6
5	9
6	3
8	3
15	2

13. Do you move your bees to serve pollination contracts or to collect honey?

Do you move your bees	Number of responses
No	103
Yes	6

13.a. If yes ,please indicate which answer(s) apply

Why	Number of responses
Borage honey	1
Heather honey,Oilseed rape honey,Other	1
Other	1
Pollination services locally	1
Pollination services locally,Borage honey,Oilseed rape honey,Other	1
Pollination services nationally,Borage honey	1
	103

13.a.i. If you selected Other, please specify:

Why other	Number of responses
Buckwheat	1
Honey	1
Sweet chestnut trees	1
	106

14. Did you purchase honey bee colonies in the last two years?

Purchased honeybee colonies in the last 2 years	Number of responses
No	93
Yes	16

14.a. If yes, were they local in origin?

Local in origin	Number of responses
No	3
Yes	13
	93

15. Did you purchase honey bee queens in the last 2 years?

Purchased queens in last 2 years	Number of responses
No	88
Yes	21

15.a. If yes, where did your queens original from?

Where from	Number of responses
Imported	2
Locally sourced	4
Locally sourced,Nationally sourced	1
Nationally sourced	14
	88

15.a.i. If you selected Other, please specify:

No responses.

16. Do you used any shared honey extraction equipment (e.g. honey spinner or uncapping tray or settling tanks)?

Use shared honey extraction equipment	Number of responses
No	80
Yes	29

17. Are you registered as a beekeeper with the National Bee Unit?

Registered beekeeper	Number of responses
No	2
Unsure	4
Yes	103

18. Have your bees ever been inspected by a bee inspector from the National Bee Unit?

Inspection from the NBU	Number of responses
No	29
Yes	80

19. Are you registered with the National Bee Unit to receive foulbrood alerts?

Registered for foulbrood updates	Number of responses
No	3
Unsure	24
Yes	82

19.a. If yes, have you had notification of EFB in the last 2 years?

Update in the last 2 years	Number of responses
No	35
Unsure	7
Yes	40
	27

20. Do you wear gloves when handling bees?

Wear gloves when handling bees	Number of responses
No	4
Yes	105

20.a. If yes, do you use disposable gloves?

Disoposable gloves	Number of responses
No	43
Yes	62
	4

21. Do you soak your hive tool in disinfectant in between apiary visits?

Soak hive tools	Number of responses
No	39
Yes	70

22. Which foulbrood disease are you most worried about your bees catching?

Which foulbrood most worried about	Number of responses
American foulbrood (AFB)	6
Both	52
European foulbrood (EFB)	42
Neither	4
Unsure	5

23. Do you consider that European foulbrood (EFB) is a problem for beekeeping in your area?

Consider EFB a problem in area	Number of responses
No	35
Yes	74

23.a. If yes, what do you believe are the main reasons that European foulbrood (EFB) is a problem in your area?

Main reason	Number of responses
Beekeepers that are not registered on BeeBase,Beekeepers that cannot recognise the disease,Feral nests,Infected swarms	2
Beekeepers that are not registered on BeeBase,Beekeepers that cannot recognise the disease,Poor beekeeping practices at diseased sites	9

Main reason	Number of responses
Beekeepers that are not registered on BeeBase,Beekeepers that cannot recognise the disease,Poor beekeeping practices at diseased sites,Feral nests	1
Beekeepers that are not registered on BeeBase,Beekeepers that cannot recognise the disease,Poor beekeeping practices at diseased sites,Feral nests,Infected swarms	1
Beekeepers that are not registered on BeeBase,Beekeepers that cannot recognise the disease,Poor beekeeping practices at diseased sites,Feral nests,Infected swarms,Other	1
Beekeepers that are not registered on BeeBase,Beekeepers that cannot recognise the disease,Poor beekeeping practices at diseased sites,Honey packing plants	1
Beekeepers that are not registered on BeeBase,Beekeepers that cannot recognise the disease,Poor beekeeping practices at diseased sites,Infected swarms	1
Beekeepers that are not registered on BeeBase,Beekeepers that cannot recognise the disease,Poor beekeeping practices at diseased sites,Infected swarms,Honey packing plants	1
Beekeepers that are not registered on BeeBase,Beekeepers that cannot recognise the disease,Poor beekeeping practices at diseased sites,Infected swarms,Other	1
Beekeepers that are not registered on BeeBase,Infected swarms	1
Beekeepers that are not registered on BeeBase,Infected swarms,Unsure	1

Main reason	Number of responses
Beekeepers that are not registered on BeeBase,Poor beekeeping practices at diseased sites,Feral nests,Infected swarms,Honey packing plants	1
Beekeepers that cannot recognise the disease	2
Beekeepers that cannot recognise the disease,Feral nests,Honey packing plants	1
Beekeepers that cannot recognise the disease,Lack of pollen,Poor weather for honey bees,Unsure	1
Beekeepers that cannot recognise the disease,Other	1
Beekeepers that cannot recognise the disease,Poor beekeeping practices at diseased sites	2
Beekeepers that cannot recognise the disease,Poor beekeeping practices at diseased sites,Feral nests	1
Beekeepers that cannot recognise the disease,Poor beekeeping practices at diseased sites,Feral nests,Honey packing plants,Unsure	1
Beekeepers that cannot recognise the disease,Poor beekeeping practices at diseased sites,Infected swarms,Honey packing plants	2
Beekeepers that cannot recognise the disease,Poor beekeeping practices at diseased sites,Infected swarms,Other	1
Beekeepers that cannot recognise the disease,Poor beekeeping practices at diseased sites,Poor weather for honey bees	1
Failed treatments,Beekeepers that are not registered on BeeBase,Beekeepers that cannot recognise the disease	1

Main reason	Number of responses
Failed treatments,Beekeepers that cannot recognise the disease,Infected swarms	1
Failed treatments,Beekeepers that cannot recognise the disease,Unsure	1
Failed treatments,Feral nests,Other	1
Honey packing plants	3
Infected swarms	2
Other	1
Poor beekeeping practices at diseased sites	2
Poor beekeeping practices at diseased sites,Honey packing plants	1
Poor beekeeping practices at diseased sites,Infected swarms	1
Poor beekeeping practices at diseased sites,Infected swarms,Honey packing plants	1
Unsure	25
	35

23.a.i. If you selected Other, please specify:

Data not shown.

24. Are confident that you could recognise European foulbrood (EFB) in your colonies?

Confident to recognise EFB	Number of responses
No	12
Unsure	28
Yes	69

25. Have you ever seen European foulbrood in a honey bee colony?

Ever seen EFB in a colony	Number of responses
No	85
Unsure	1
Yes	23

26. Have your bees ever been diagnosed with European foulbrood (EFB)?

Ever had EFB diagnosed in your bees	Number of responses
No	101
Yes	8

26.a. If yes, please indicate when your bees were last diagnosed with European foulbrood?

When were bees diagnosed	Number of responses
Before 2019	3
Within the last two years	5
	101

27. Do you know of any feral colonies within 2 km of your apiary site(s)?

Know of any feral colonies within 2km of apiary site	Number of responses
No	65
Yes	44

28. Is there any other information you wish to share with us about European foulbrood in your region?

Data not shown.

C.2 SEM output

C.2.1 Somerset

Table C.81: The output of the Structural equation model investigating the impact of education, experience and exposure to EFB on beekeeping preactices and perception in Somerset. CFI=0.847, RMSEA=0.074

Response	Predictor	Estimate	Standard error	z-value	p-value	Standard all
Swarm practices	Housing method	1.000000000	0.000			0.505
Swarm practices	Distance to swarm collect	25.400483299	3.544	7.166	0.000	0.718
Swarm practices	Collect swarms	2.179625797	0.320	6.819	0.000	0.877
Biosecurity practices	Wear disposable gloves	1.000000000	0.000			0.430
Biosecurity practices	Disinfect tools	1.193552589	0.368	3.244	0.001	0.550
Biosecurity practices	Share	0.458006853	0.211	2.166	0.030	0.215
Experience and education	Years	1.000000000	0.000			0.471
Experience and education	Read literature	0.041070457	0.008	5.421	0.000	0.563
Experience and education	Number of qualifications	0.151205686	0.030	5.070	0.000	0.494

Table C.81: The output of the Structural equation model investigating the impact of education, experience and exposure to EFB on beekeeping preactices and perception in Somerset. CFI=0.847, RMSEA=0.074

Response	Predictor	Estimate	Standard error	z-value	p-value	Standard all
Experience and education	Number of colonies	0.642227382	0.138	4.643	0.000	0.427
Perception	Experience and education	0.014801997	0.010	1.515	0.130	0.201
Swarm practices	Perception	-0.038273329	0.032	-1.214	0.225	-0.091
Biosecurity practices	Perception	-0.019382514	0.050	-0.389	0.698	-0.043
Swarm practices	Experience and education	0.005259215	0.005	1.110	0.267	0.170
Biosecurity practices	Experience and education	0.024667651	0.010	2.502	0.012	0.746
Perception	Alert in last 2yrs	0.352811420	0.062	5.700	0.000	0.328
Swarm practices	Alert in last 2yrs	0.112338088	0.035	3.236	0.001	0.248
Biosecurity practices	Alert in last 2yrs	0.013217303	0.048	0.274	0.784	0.027
Inspected by NBU	Experience and education	0.054164026	0.010	5.536	0.000	0.696
Swarm practices	Inspected by NBU	0.015900185	0.050	0.316	0.752	0.040
Biosecurity practices	Inspected by NBU	-0.074529890	0.085	-0.876	0.381	-0.175
Swarm practices	Seen foulbrood	0.069192583	0.036	1.921	0.055	0.145
Biosecurity practices	Seen foulbrood	-0.028803971	0.053	-0.541	0.588	-0.056
Seen foulbrood	Inspected by NBU	0.335761053	0.048	7.042	0.000	0.404
Perception	Inspected by NBU	-0.083864918	0.105	-0.801	0.423	-0.089
Perception	Seen foulbrood	0.198235206	0.072	2.769	0.006	0.174
Housing method	Housing method	0.115226819	0.011	10.295	0.000	0.745
Distance to swarm collect	Distance to swarm collect	23.947389343	3.398	7.047	0.000	0.485

Table C.81: The output of the Structural equation model investigating the impact of education, experience and exposure to EFB on beekeeping preactices and perception in Somerset. CFI=0.847, RMSEA=0.074

Response	Predictor	Estimate	Standard error	z-value	p-value	Standard all
Collect swarms	Collect swarms	0.056181195	0.020	2.819	0.005	0.231
Wear disposable gloves	Wear disposable gloves	0.198317112	0.023	8.616	0.000	0.815
Disinfect tools	Disinfect tools	0.147361674	0.024	6.095	0.000	0.697
Share	Share	0.193704878	0.018	10.734	0.000	0.954
Years	Years	144.078281767	14.461	9.963	0.000	0.778
Read literature	Read literature	0.149278584	0.017	9.043	0.000	0.683
Number of qualifications	Number of qualifications	2.909568501	0.298	9.777	0.000	0.756
Number of colonies	Number of colonies	76.248396815	7.427	10.266	0.000	0.818
Perception	Perception	0.184354701	0.017	10.857	0.000	0.831
Inspected by NBU	Inspected by NBU	0.128418876	0.020	6.323	0.000	0.516
Seen foulbrood	Seen foulbrood	0.143792237	0.013	11.269	0.000	0.837
Swarm practices	Swarm practices	0.034569044	0.009	3.843	0.000	0.878
Biosecurity practices	Biosecurity practices	0.027646877	0.014	1.910	0.056	0.615
Experience and education	Experience and education	41.104214962	12.339	3.331	0.001	1.000
Swarm practices	Biosecurity practices	-0.009603820	0.005	-1.945	0.052	-0.311
Alert in last 2yrs	Alert in last 2yrs	0.192324385	0.000			1.000

C.2.2 Cambridgeshire

Table C.82: The output of the Structural equation model investigating the impact of education, experience and exposure to EFB on beekeeping preactices and perception in Cambridgeshire. CFI=0.782, RMSEA=0.084

Response	Predictor	Estimate	Standard error	z-value	p-value	Standard all
Swarm practices	Housing method	1.000000000	0.000			0.451
Swarm practices	Distance to swarm collect	23.462029674	5.129	4.575	0.000	0.650
Swarm practices	Collect swarms	2.556037494	0.604	4.233	0.000	1.007
Biosecurity practices	Wear disposable gloves	1.000000000	0.000			0.526
Biosecurity practices	Disinfect tools	1.137024930	0.338	3.360	0.001	0.621
Biosecurity practices	Share	0.299984782	0.214	1.405	0.160	0.175
Experience and education	Years	1.000000000	0.000			0.393
Experience and education	Read literature	0.046172403	0.019	2.417	0.016	0.385
Experience and education	Number of qualifications	0.213394822	0.079	2.699	0.007	0.485
Experience and education	Number of colonies	1.029046140	0.365	2.817	0.005	0.548
Perception	Experience and education	-0.021226928	0.017	-1.258	0.209	-0.179
Swarm practices	Perception	-0.073606637	0.047	-1.583	0.113	-0.177
Biosecurity practices	Perception	-0.117230958	0.080	-1.475	0.140	-0.214
Swarm practices	Experience and education	0.025726842	0.012	2.158	0.031	0.523
Biosecurity practices	Experience and education	0.033151733	0.017	1.965	0.049	0.512
Perception	Alert in last 2yrs	0.020609574	0.085	0.242	0.809	0.021
Swarm practices	Alert in last 2yrs	-0.017966782	0.035	-0.519	0.604	-0.044
Biosecurity practices	Alert in last 2yrs	0.085782037	0.066	1.308	0.191	0.161
Inspected by NBU	Experience and education	0.036286076	0.018	2.022	0.043	0.326

Table C.82: The output of the Structural equation model investigating the impact of education, experience and exposure to EFB on beekeeping preactices and perception in Cambridgeshire. CFI=0.782, RMSEA=0.084

Response	Predictor	Estimate	Standard error	z-value	p-value	Standard all
Swarm practices	Inspected by NBU	-0.052011817	0.052	-0.998	0.318	-0.117
Biosecurity practices	Inspected by NBU	-0.276784614	0.103	-2.693	0.007	-0.475
Swarm practices	Seen foulbrood	-0.058164642	0.045	-1.305	0.192	-0.121
Biosecurity practices	Seen foulbrood	0.224309898	0.089	2.532	0.011	0.356
Seen foulbrood	Inspected by NBU	0.146551687	0.087	1.678	0.093	0.159
Perception	Inspected by NBU	0.382095652	0.106	3.605	0.000	0.359
Perception	Seen foulbrood	0.238432079	0.102	2.336	0.019	0.207
Housing method	Housing method	0.149870556	0.021	7.145	0.000	0.797
Distance to swarm collect	Distance to swarm collect	28.859200037	4.908	5.880	0.000	0.578
Collect swarms	Collect swarms	-0.003658565	0.035	-0.104	0.918	-0.015
Wear disposable gloves	Wear disposable gloves	0.172833713	0.031	5.662	0.000	0.723
Disinfect tools	Disinfect tools	0.136136885	0.031	4.387	0.000	0.614
Share	Share	0.188751490	0.026	7.251	0.000	0.969
Years	Years	86.662201841	13.046	6.643	0.000	0.846
Read literature	Read literature	0.193434255	0.029	6.677	0.000	0.852
Number of qualifications	Number of qualifications	2.336165875	0.383	6.105	0.000	0.765
Number of colonies	Number of colonies	38.948523199	7.009	5.557	0.000	0.700
Perception	Perception	0.181090651	0.025	7.124	0.000	0.818
Inspected by NBU	Inspected by NBU	0.174486443	0.026	6.643	0.000	0.894

Table C.82: The output of the Structural equation model investigating the impact of education, experience and exposure to EFB on beekeeping preactices and perception in Cambridgeshire. CFI=0.782, RMSEA=0.084

Response	Predictor	Estimate	Standard error	z-value	p-value	Standard all
Seen foulbrood	Seen foulbrood	0.162290406	0.022	7.382	0.000	0.975
Swarm practices	Swarm practices	0.025739441	0.011	2.343	0.019	0.673
Biosecurity practices	Biosecurity practices	0.031067439	0.020	1.570	0.116	0.469
Experience and education	Experience and education	15.784654513	9.336	1.691	0.091	1.000
Swarm practices	Biosecurity practices	-0.001848705	0.006	-0.301	0.763	-0.065
Alert in last 2yrs	Alert in last 2yrs	0.232303678	0.000			1.000

References

Agrebi, N.E., Steinhauer, N., Renault, V., Graaf, D.C. de & Saegerman, C. (2022) 'Beekeepers perception of risks affecting colony loss: A pilot survey', *Transboundary and Emerging Diseases*, 69pp. 579–590.

Ahsan, M.U., Liu, Q., Fang, L. & Wang, K. (2021) 'NanoCaller for accurate detection of SNPs and indels in difficult-to-map regions from long-read sequencing by haplotype-aware deep neural networks', *Genome Biology*, 22pp. 1–33.

Akerley, B.J., Rubin, E.J., Novick, V.L., Amaya, K., Judson, N., Mekalanos, J.J., Asai, K., Ashikaga, S., Aymerich, S., Bessieres, P., Boland, F., Brignell, S.C., Bron, S., Bunai, K., Chapuis, J., Christiansen, L.C., Danchin, A., Débarbouillé, M., Dervyn, E., et al. (2002) 'A genome-scale analysis for identification of genes required for growth or survival of haemophilus influenzae', *Proceedings of the National Academy of Sciences*, 99pp. 966–971.

Amar, Y., Lagkouvardos, I., Silva, R.L., Ishola, O.A., Foesel, B.U., Kublik, S., Schöler, A., Niedermeier, S., Bleuel, R., Zink, A., Neuhaus, K., Schloter, M., Biedermann, T. & Köberle, M. (2021) 'Pre-digest of unprotected DNA by benzonase improves the representation of living skin bacteria and efficiently depletes host DNA', *Microbiome*, 9pp. 1–14.

Anderson, K.E., Copeland, D.C., Erickson, R.J., Floyd, A.S., Maes, P.C. & Mott, B.M. (2023) 'A high-throughput sequencing survey characterizing european foulbrood disease and varroosis in honey bees', *Scientific Reports 2023 13:1*, 13pp. 1–16.

Ansari, M.J., Al-Ghamdi, A., Nuru, A., Ahmed, A.M., Ayaad, T.H., Al-Qarni, A., Alattal, Y. & Al-Waili, N. (2017) 'Survey and molecular detection of melissococcus plutonius, the causative agent of european foulbrood in honeybees in saudi arabia', *Saudi Journal of Biological Sciences*, 24pp. 1327–1335.

Arai, R., Tominaga, K., Wu, M., Okura, M., Ito, K., Okamura, N., Onishi, H., Osaki, M., Sugimura, Y., Yoshiyama, M. & Takamatsu, D. (2012) 'Diversity of melissococcus

plutonius from honeybee larvae in japan and experimental reproduction of european foulbrood with cultured atypical isolates', *PLoS ONE*, 7.

Ashbaugh, C.D., Alberti, S. & Wessels, M.R. (1998) 'Molecular analysis of the capsule gene region of group a streptococcus: The hasab genes are sufficient for capsule expression', *Journal of Bacteriology*, 180pp. 4955–4959.

Avery, O.T., Macleod, C.M. & McCarty, M. (1944) 'Studies on the chemical nature of the substance inducing transformation of pneumococcal types: Induction of transformation by a desoxyribonucleic acid fraction isolated from pneumococcus type iii', *Journal of Experimental Medicine*, 79pp. 137–158.

Bailey, L. (1956) 'Ætiology of european foul brood; a disease of the larval honey-bee [27]'. Nature 178 p. 1130.

Bailey, L. (1983) 'Melissococcus pluton, the cause of european foulbrood of honey bees (apis spp)', *Journal of Applied Bacteriology*, 55pp. 65–69.

Bailey, L. & Collins, M.D. (1982) 'Taxonomic studies on streptococcus pluton.', *Journal of Applied Bacteriology*, 53pp. 209–213.

Banerjee, A. & Davé, R.N. (2004) 'Validating clusters using the hopkins statistic', *IEEE International Conference on Fuzzy Systems*, 1pp. 149–153.

Barbuddhe, S.B. & Chakraborty, T. (2009) 'Listeria as an enteroinvasive gastrointestinal pathogen', *Current Topics in Microbiology and Immunology*, 337pp. 173–195.

Barlow, M. (2009) 'What antimicrobial resistance has taught us about horizontal gene transfer', *Methods in molecular biology (Clifton, N.J.)*, 532pp. 397–411.

Barraclough, T.G., Balbi, K.J. & Ellis, R.J. (2012) 'Evolving concepts of bacterial species', *Evolutionary Biology*, 39pp. 148–157.

BBKA (2024) *Syllabi for written exams | british beekeepers association*.

Beckly, P. (1977) *Keeping bees*.

BeeCraft (2024) *Bee craft apiary guides - my magazine subscriptions*.

BeeCraft (2022) 'Swarming and swarm controls 5a', *BeeCraft Apiary Guide*,

Behjati, S. & Tarpey, P.S. (2013) 'What is next generation sequencing?', *Archives of Disease in Childhood: Education and Practice Edition*, 98pp. 236–238.

Bertolotti, A.C., Forsgren, E., Schäfer, M.O., Sircoulomb, F., Gaïani, N., Ribière-Chabert, M., Paris, L., Lucas, P., Boisséson, C. de, Skarin, J. & Rivière, M. (2021) 'Development and evaluation of a core genome multilocus sequence typing scheme for paenibacillus larvae , the deadly american foulbrood pathogen of honeybees', *Environmental Microbiology*,

Bhagwat, A., Zhang, F., Collins, C.H. & Dordick, J.S. (2021) 'Influence of bacterial culture medium on peptidoglycan binding of cell wall lytic enzymes', *Journal of Biotechnology*, 330pp. 27–34.

Biová, J., Charrière, J.-D., Dostálková, S., Škrabišová, M., Petřivalský, M., Bzdil, J. & Danihlík, J. (2021) 'Melissococcus plutonius can be effectively and economically detected using hive debris and conventional PCR', *Insects*, 12p. 150.

Boom, R., Sol, C.J.A., Salimans, M.M.M., Jansen, C.L., Dillen, P.M.E.W.-V. & Noordaa, J.V.D. (1990) 'Rapid and simple method for purification of nucleic acids', *Journal of Clinical Microbiology*, 28pp. 495–503.

Boyd, D.A., Willey, B.M., Fawcett, D., Gillani, N. & Mulvey, M.R. (2008) 'Molecular characterization of enterococcus faecalis N06-0364 with low-level vancomycin resistance harboring a novel d-ala-d-ser gene cluster, vanL', *Antimicrobial Agents and Chemotherapy*, 52p. 2667.

Breeze, T.D., Bailey, A.P., Balcombe, K.G. & Potts, S.G. (2011) 'Pollination services in the UK: How important are honeybees?', *Agriculture, Ecosystems and Environment*, 142pp. 137–143.

Brosnan, J.T. & Brosnan, M.E. (2013) 'Glutamate: A truly functional amino acid', *Amino Acids*, 45pp. 413–418.

Bruggeling, C.E., Garza, D.R., Achouiti, S., Mes, W., Dutilh, B.E. & Boleij, A. (2021) 'Optimized bacterial DNA isolation method for microbiome analysis of human tissues', *MicrobiologyOpen*, 10.

Büchler, A.C., Wicki, M., Frei, R., Hinic, V., Seth-Smith, H.M.B., Egli, A. & Widmer, A.F. (2022) 'Matching clostridioides difficile strains obtained from shoe soles of healthcare workers epidemiologically linked to patients and confirmed by whole-genome sequencing', *The Journal of hospital infection*, 126pp. 10–15.

Budge, G., Brown, M., Pietravalle, S., Marris, G. & Inspectors, N. (2012) 'The random apiary survey (RAS) – final results', *Beecraft*, pp. 1–2.

Budge, G.E., Barrett, B., Jones, B., Pietravalle, S., Marris, G., Chantawannakul, P., Thwaites, R., Hall, J., Cuthbertson, A.G.S. & Brown, M.A. (2010) 'The occurrence of melissococcus plutonius in healthy colonies of apis mellifera and the efficacy of european foulbrood control measures', *Journal of Invertebrate Pathology*, 105pp. 164–170.

Budge, G.E., Shirley, M.D.F., Jones, B., Quill, E., Tomkies, V., Feil, E.J., Brown, M.A. & Haynes, E.G. (2014) 'Molecular epidemiology and population structure of the honey

bee brood pathogen melissococcus plutonius', *ISME Journal*, 8pp. 1588–1597.

Budge, G.E., Simcock, N.K., Holder, P.J., Shirley, M.D.F., Brown, M.A., Weymers, P.S.M.V., Evans, D.J. & Rushton, S.P. (2020) 'Chronic bee paralysis as a serious emerging threat to honey bees', *Nature Communications*, 11pp. 1–9.

Cangelosi, G.A. & Meschke, J.S. (2014) 'Dead or alive: Molecular assessment of microbial viability'. Applied and Environmental Microbiology 80 pp. 5884–5891.

Carreck, N. & Williams, I. (1998) 'The economic value of bees in the UK', *Bee World*, 79pp. 115–123.

Carvalho, F.P. (2006) 'Agriculture, pesticides, food security and food safety', *Environmental Science & Policy*, 9pp. 685–692.

Chacon-Cortes, D. & Griffiths, L.R. (2014) 'Methods for extracting genomic DNA from whole blood samples: Current perspectives', *Journal of Biorepository Science for Applied Medicine*, p. 1.

Chambers, J., Sparks, N., Sydney, N., Livingstone, P.G., Cookson, A.R. & Whitworth, D.E. (2020) 'Comparative genomics and pan-genomics of the myxococcaceae, including a description of five novel species: Myxococcus eversor sp. Nov., Myxococcus llanfairpwllgwyngyllgogerychwyrndrobwllllantysiliogogogochensis sp. Nov., Myxococcus vastator sp. Nov., Pyxidicoccus caerfyrddinensis sp. Nov., And pyxidicoccus trucidator sp. nov.', *Genome Biology and Evolution*, 12pp. 2289–2302.

Chaves, B.J. & Tadi, P. (2023) *Gentamicin*. StatPearls Publishing, Treasure Island (FL).

Chen, L., Zheng, D., Liu, B., Yang, J. & Jin, Q. (2016) 'VFDB 2016: Hierarchical and refined dataset for big data analysis–10 years on', *Nucleic acids research*, 44pp. D694–D697.

Christen, V., Mittner, F. & Fent, K. (2016) 'Molecular effects of neonicotinoids in honey bees (apis mellifera)', *Environmental Science and Technology*, 50pp. 4071–4081.

Clair, A.L.S., Clair, A.L.S., Zhang, G., Dolezal, A.G., O'Neal, M.E., Toth, A.L. & Toth, A.L. (2020) 'Diversified farming in a monoculture landscape: Effects on honey bee health and wild bee communities', *Environmental Entomology*, 49pp. 753–764.

Clark, K., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J. & Sayers, E.W. (2016) 'GenBank', *Nucleic Acids Research*, 44p. D67.

Cohn, S.K. (2018) 'Epidemics: Hate and compassion from the plague of athens to AIDS', *Epidemics: Hate and Compassion from the Plague of Athens to AIDS*, pp. 1–643.

Conte, Y.L. & Navajas, M. (2008) 'Climate change: Impact on honey bee populations and diseases', *OIE Revue Scientifique et Technique*, 27pp. 485–510.

Couston, R. (1990) *Principles of practical beekeeping.*

Crater, D.L., Dougherty, B.A. & Rijn, I.V.D. (1995) 'Molecular characterization of hasC from an operon required for hyaluronic acid synthesis in group a streptococci: Demonstration of UDP-glucose pyrophosphorylase activity', *Journal of Biological Chemistry*, 270pp. 28676–28680.

Cuber, P., Chooneea, D., Geeves, C., Salatino, S., Creedy, T.J., Griffin, C., Sivess, L., Barnes, I., Price, B. & Misra, R. (2023) 'Comparing the accuracy and efficiency of third generation sequencing technologies, oxford nanopore technologies, and pacific biosciences, for DNA barcode sequencing applications', *Ecological Genetics and Genomics*, 28p. 100181.

Dainat, B., Grossar, D., Ecoffey, B. & Haldemann, C. (2018) 'Triplex real-time PCR method for the qualitative detection of european and american foulbrood in honeybee', *Journal of Microbiological Methods*, 146pp. 61–63.

Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., Mccarthy, S.A., Davies, R.M. & Li, H. (2021) 'Twelve years of SAMtools and BCFtools', *GigaScience*, 10pp. 1–4.

Davies, J. (1996) 'Origins and evolution of antibiotic resistance.', *Microbiología (Madrid, Spain)*, 12pp. 9–16.

Davis, I. & Cullum-Kenyon, R. (2015) *The BBKA guide to beekeeping.*

Deans, A.S.C. (1949) *The beekeepers encyclopedia.*

Defra & Government, W. (2021) *Healthy bees plan 2030 implementation plan.*

Delahaye, C. & Nicolas, J. (2021) 'Sequencing DNA with nanopores: Troubles and biases', *PLoS ONE*, 16.

Dictionary.com (2020) *Molecular biology | definition of molecular biology.*

Didenko, V.V. (2001) 'Dna probes using fluorescence resonance energy transfer (fret): Designs and applications'. BioTechniques 31 pp. 1106–1121.

Djukic, M., Erler, S., Leimbach, A., Grossar, D., Charrière, J.D., Gauthier, L., Hartken, D., Dietrich, S., Nacke, H., Daniel, R. & Poehlein, A. (2018) 'Comparative genomics and description of putative virulence factors of melissococcus plutonius, the causative agent of european foulbrood disease in honey bees', *Genes*, 9pp. 1–20.

Doster, E., Lakin, S.M., Dean, C.J., Wolfe, C., Young, J.G., Boucher, C., Belk, K.E., Noyes, N.R. & Morley, P.S. (2020) 'MEGARes 2.0: A database for classification of antimicrobial drug, biocide and metal resistance determinants in metagenomic sequence data', *Nucleic Acids Research*, 48pp. D561–D569.

Earnshaw, V.A., Eaton, L.A., Kalichman, S.C., Brousseau, N.M., Hill, E.C. & Fox, A.B. (2020) 'COVID-19 conspiracy beliefs, health behaviors, and policy support', *Translational Behavioral Medicine*, 10pp. 850–856.

Edawardes, T. (1943) *Beekeeping for all*.

Eeles, R.A., Warren, W. & Stamps, A. (1992) 'The PCR revolution', *European Journal of Cancer*, 28pp. 289–293.

Ekblom, R. & Wolf, J.B.W. (2014) 'A field guide to whole-genome sequencing, assembly and annotation', *Evolutionary Applications*, 7pp. 1026–1042.

Ellis, J.D. & Munn, P.A. (2005) 'The worldwide health status of honey bees', *Bee World*, 86pp. 88–101.

Erban, T., Ledvinka, O., Kamler, M., Hortova, B., Nesvorna, M., Tyl, J., Titera, D., Markovic, M. & Hubert, J. (2017) 'Bacterial community associated with worker honeybees (apis mellifera) affected by european foulbrood', *PeerJ*, 2017.

Erler, S., Lewkowski, O., Poehlein, A. & Forsgren, E. (2018) 'The curious case of achromobacter eurydice, a gram-variable pleomorphic bacterium associated with european foulbrood disease in honeybees', *Microbial Ecology*, 75pp. 1–6.

Faison, W.J., Rostovtsev, A., Castro-Nallar, E., Crandall, K.A., Chumakov, K., Simonyan, V. & Mazumder, R. (2014) 'Whole genome single-nucleotide variation profile-based phylogenetic tree building methods for analysis of viral, bacterial and human genomes', *Genomics*, 104pp. 1–7.

Feldgarden, M., Brover, V., Haft, D.H., Prasad, A.B., Slotta, D.J., Tolstoy, I., Tyson, G.H., Zhao, S., Hsu, C.H., McDermott, P.F., Tadesse, D.A., Morales, C., Simmons, M., Tillman, G., Wasilenko, J., Folster, J.P. & Klimke, W. (2019) 'Validating the AMRFinder tool and resistance gene database by using antimicrobial resistance genotype-phenotype correlations in a collection of isolates', *Antimicrobial Agents and Chemotherapy*, 63.

Ferreira, F.A., Helmersen, K., Visnovska, T., Jørgensen, S.B. & Aamot, H.V. (2021) 'Rapid nanopore-based DNA sequencing protocol ofllfor use in surveillance and outbreak investigation', *Microbial Genomics*, 7p. 557.

Ferretti, J.J., McShan, W.M., Ajdic, D., Savic, D.J., Savic, G., Lyon, K., Primeaux, C., Sezate, S., Suvorov, A.N., Kenton, S., Lai, H.S., Lin, S.P., Qian, Y., Jia, H.G., Najar, F.Z., Ren, Q., Zhu, H., Song, L., White, J., et al. (2001) 'Complete genome sequence of an M1 strain of streptococcus pyogenes', *Proceedings of the National Academy of Sciences of the United States of America*, 98pp. 4658–4663.

Forsgren, E. (2010) 'European foulbrood in honey bees', *Journal of Invertebrate Pathology*, 103pp. S5–S9.

Forsgren, E., Budge, G.E., Charrière, J.D. & Hornitzky, M.A.Z. (2013) 'Standard methods for european foulbrood research'. Journal of Apicultural Research 52.

Foster, J.T., Beckstrom-Sternberg, S.M., Pearson, T., Beckstrom-Sternberg, J.S., Chain, P.S.G., Roberto, F.F., Hnath, J., Brettin, T. & Keim, P. (2009) 'Whole-genome-based phylogeny and divergence of the genus brucella', *Journal of bacteriology*, 191pp. 2864–2870.

Fox, J. (2022) R package version 0.8-1. *Polycor: Polychoric and polyserial correlations*.

Fraimow, H., Knob, C., Herrero, I.A. & Patel, R. (2005) 'Putative VanRS-like two-component regulatory system associated with the inducible glycopeptide resistance cluster of paenibacillus popilliae', *Antimicrobial agents and chemotherapy*, 49pp. 2625–2633.

Francis, R.V., Billam, H., Clarke, M., Yates, C., Tsoleridis, T., Berry, L., Mahida, N., Irving, W.L., Moore, C., Holmes, N., Ball, J.K., Loose, M. & McClure, C.P. (2022) 'The impact of real-time whole-genome sequencing in controlling healthcare-associated SARS-CoV-2 outbreaks', *The Journal of Infectious Diseases*, 225pp. 10–18.

Francisco, A.P., Bugalho, M., Ramirez, M. & Carriço, J.A. (2009) 'Global optimal eBURST analysis of multilocus typing data using a graphic matroid approach', *BMC Bioinformatics*, 10pp. 1–15.

Fries, I., Lindström, A. & Korpela, S. (2006) 'Vertical transmission of american foulbrood (paenibacillus larvae) in honey bees (apis mellifera)', *Veterinary Microbiology*, 114pp. 269–274.

Gaillot, O., Pellegrini, E., Bregenholt, S., Nair, S. & Berche, P. (2000) 'The ClpP serine protease is essential for the intracellular parasitism and virulence of listeria monocytogenes', *Molecular Microbiology*, 35pp. 1286–1294.

Ganda, E., Beck, K.L., Haiminen, N., Silverman, J.D., Kawas, B., Cronk, B.D., Anderson, R.R., Goodman, L.B. & Wiedmann, M. (2021) 'DNA extraction and host depletion methods significantly impact and potentially bias bacterial detection in a biological fluid', *mSystems*, 6.

García, N.L. (2018) 'The current situation on the international honey market', *Bee World*, 95pp. 89–94.

Giersch, T., Barchia, I. & Hornitzky, M. (2010) 'Can fatty acids and oxytetracycline protect artificially raised larvae from developing european foulbrood?', *Apidologie*,

41pp. 151–159.

Gilchrist, C.A., Turner, S.D., Riley, M.F., Petri, W.A. & Hewlett, E.L. (2015) 'Whole-genome sequencing in outbreak analysis', *Clinical Microbiology Reviews*, 28pp. 541–563.

Gong, Y., Li, Y., Liu, X., Ma, Y. & Jiang, L. (2023) 'A review of the pangenome: How it affects our understanding of genomic variation, selection and breeding in domestic animals?', *Journal of Animal Science and Biotechnology 2023 14:1*, 14pp. 1–19.

Govan, V.A., Brözel, V., Allsopp, M.H. & Davison, S. (1998) 'A PCR detection method for rapid identification of melissococcus pluton in honeybee larvae', *Applied and Environmental Microbiology*, 64pp. 1983–1985.

Government, H. (2023) 'Gross value added (GVA) of agriculture in the united kingdom (UK) from 2003 to 2019 (in million GBP)'. Statista. Statista Inc

Government, H. (2006) *The bee diseases and pests control (england) order 2006 (SI 2006/342)*.

Government, H. (1982) *The bee diseases control order 1982 (SI 1982/107)*.

Grädel, C., Miani, M.A.T., Barbani, M.T., Leib, S.L., Suter-Riniker, F. & Ramette, A. (2019) 'Rapid and cost-efficient enterovirus genotyping from clinical samples using flongle flow cells', *Genes 2019, Vol. 10, Page 659*, 10p. 659.

Gray, A., Adjlane, N., Arab, A., Ballis, A., Brusbardis, V., Charrière, J.D., Chlebo, R., Coffey, M.F., Cornelissen, B., Costa, C.A. da, Dahle, B., Danihlík, J., Dražić, M.M., Evans, G., Fedoriak, M., Forsythe, I., Gajda, A., Graaf, D.C. de, Gregorc, A., et al. (2020) 'Honey bee colony winter loss rates for 35 countries participating in the COLOSS survey for winter 2018–2019, and the effects of a new queen on the risk of colony winter loss', *Journal of Apicultural Research*, 59pp. 744–751.

Gray, A., Adjlane, N., Arab, A., Ballis, A., Brusbardis, V., Douglas, A.B., Cadahía, L., Charrière, J.D., Chlebo, R., Coffey, M.F., Cornelissen, B., Costa, C.A. da, Danneels, E., Danihlík, J., Dobrescu, C., Evans, G., Fedoriak, M., Forsythe, I., Gregorc, A., et al. (2023) 'Honey bee colony loss rates in 37 countries using the COLOSS survey for winter 2019–2020: The combined effects of operation size, migration and queen replacement', *Journal of Apicultural Research*, 62pp. 204–210.

Graystock, P., Goulson, D. & Hughes, W.O.H. (2015) 'Parasites in bloom: Flowers aid dispersal and transmission of pollinator parasites within and between bee species', *Proceedings of the Royal Society B: Biological Sciences*, 282.

Grossar, D., Haynes, E., Budge, G.E., Parejo, M., Gauthier, L., Charrière, J.D.,

Chapuisat, M. & Dietemann, V. (2023) 'Population genetic diversity and dynamics of the honey bee brood pathogen melissococcus plutonius in a region with high prevalence', *Journal of Invertebrate Pathology*, 196p. 107867.

Grossar, D., Kilchenmann, V., Forsgren, E., Charrière, J.D., Gauthier, L., Chapuisat, M. & Dietemann, V. (2020) 'Putative determinants of virulence in melissococcus plutonius, the bacterial agent causing european foulbrood in honey bees', *Virulence*, 11pp. 554–567.

Grüning, B., Dale, R., Sjödin, A., Chapman, B.A., Rowe, J., Tomkins-Tinch, C.H., Valieris, R. & Köster, J. (2018) 'Bioconda: Sustainable and comprehensive software distribution for the life sciences', *Nature Methods*, 15pp. 475–476.

Guindon, S., Lethiec, F., Duroux, P. & Gascuel, O. (2005) 'PHYML online—a web server for fast maximum likelihood-based phylogenetic inference', *Nucleic Acids Research*, 33pp. W557–W559.

Gupta, S.K., Padmanabhan, B.R., Diene, S.M., Lopez-Rojas, R., Kempf, M., Landraud, L. & Rolain, J.M. (2014) 'ARG-ANNOT, a new bioinformatic tool to discover antibiotic resistance genes in bacterial genomes', *Antimicrobial agents and chemotherapy*, 58pp. 212–220.

Hall, J.C. (1998) 'Review: glycine', *Journal of Parenteral and Enteral Nutrition*, 22pp. 393–398.

Hancock, L.E. & Gilmore, M.S. (2002) 'The capsular polysaccharide of enterococcus faecalis and its relationship to other polysaccharides in the cell wall', *Proceedings of the National Academy of Sciences of the United States of America*, 99pp. 1574–1579.

Hanekamp, J.C. & Bast, A. (2015) 'Antibiotics exposure and health risks: chloramphenicol', *Environmental Toxicology and Pharmacology*, 39pp. 213–220.

Hawkins, T.L., O'connor-Morin, T., Roy1, A. & Santillan, C. (1994) 'DNA purification and isolation using a solid-phase', *Nucleic Acids Research*, 22pp. 4543–4544.

Haynes, E., Helgason, T., Young, J.P.W., Thwaites, R. & Budge, G.E. (2013) 'A typing scheme for the honeybee pathogen melissococcus plutonius allows detection of disease transmission events and a study of the distribution of variants', *Environmental Microbiology Reports*, 5pp. 525–529.

Health, D. of & Care, S. (2021) *Understanding lateral flow antigen testing for people without symptoms - GOV.UK*.

Hendsch, Z.S. & Tidor, B. (1994) 'Do salt bridges stabilize proteins? A continuum electrostatic analysis', *Protein Science*, 3pp. 211–226.

Heravi, F.S., Zakrzewski, M., Vickery, K. & Hu, H. (2020) 'Host DNA depletion efficiency of microbiome DNA enrichment methods in infected tissue samples', *Journal of microbiological methods*, 170.

Hoek, A.H.A.M.V., Mevius, D., Guerra, B., Mullany, P., Roberts, A.P. & Aarts, H.J.M. (2011) 'Acquired antibiotic resistance genes: An overview', *Frontiers in Microbiology*, 2.

Holley, R.W., Apgar, J., Everett, G.A., Madison, J.T., Marquisee, M., Merrill, S.H., Penswick, J.R. & Zamir, A. (1965) 'Structure of a ribonucleic acid', *Science*, 147pp. 1462–1465.

Hornitzky, M.A.Z. & Smith, L. (1998) 'Procedures for the culture of melissococcus pluton from diseased brood and bulk honey samples', *Journal of Apicultural Research*, 37pp. 293–294.

Huang, Y.T., Liu, P.Y. & Shih, P.W. (2021) 'Homopolish: A method for the removal of systematic errors in nanopore sequencing by homologous polishing', *Genome Biology*, 22pp. 1–17.

Hunkapiller, T., Kaiser, R.J., Koop, B.F. & Hood, L. (1991) 'Large-scale and automated DNA sequence determination'. Science 254 pp. 59–67.

Iacobucci, G. (2020) 'Covid-19: Mass population testing is rolled out in liverpool', *BMJ (Clinical research ed.)*, 371p. m4268.

Ingle, D.J., Valcanis, M., Kuzevski, A., Tauschek, M., Inouye, M., Stinear, T., Levine, M.M., Robins-Browne, R.M. & Holt, K.E. (2016) 'In silico serotyping of e. Coli from short read data identifies limited novel o-loci but extensive diversity of o:h serotype combinations within and between pathogenic lineages', *Microbial Genomics*, 2.

Isa, A., Loke, Y.K., Smith, J.R., Papageorgiou, A. & Hunter, P.R. (2013) 'Mediational effects of self-efficacy dimensions in the relationship between knowledge of dengue and dengue preventive behaviour with respect to control of dengue outbreaks: A structural equation model of a cross-sectional survey', *PLOS Neglected Tropical Diseases*, 7p. e2401.

ITC (2023) 'Value of natural honey exported from the united kingdom (UK) from 2003 to 2022 (in 1,000 GBP) [graph]'. Statista

Jacques, A., Laurent, M., Ribière-Chabert, M., Saussac, M., Bougeard, S., Budge, G.E., Hendrikx, P. & Chauzat, M.P. (2017) 'A pan-european epidemiological study reveals honey bee colony survival depends on beekeeper education and disease control', *PLoS ONE*, 12pp. 1–17.

Jia, B., Raphenya, A.R., Alcock, B., Waglechner, N., Guo, P., Tsang, K.K., Lago, B.A., Dave, B.M., Pereira, S., Sharma, A.N., Doshi, S., Courtot, M., Lo, R., Williams, L.E., Frye, J.G., Elsayegh, T., Sardar, D., Westman, E.L., Pawlowski, A.C., et al. (2017) 'CARD 2017: Expansion and model-centric curation of the comprehensive antibiotic resistance database', *Nucleic acids research*, 45pp. D566–D573.

Jiang, Y., Jiang, Y., Wang, S., Zhang, Q. & Ding, X. (2019) 'Optimal sequencing depth design for whole genome re-sequencing in pigs', *BMC Bioinformatics*, 20.

Johnson, D.I. (2018) 'Bacterial virulence factors', *Bacterial Pathogens and Their Virulence Factors*, pp. 1–38.

Jolley, K.A., Bray, J.E. & Maiden, M.C.J. (2018) 'Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications.', *Wellcome Open Research*, 3.

Jones, J.G. & Morrison, G.A. (1962) 'The bacteriostatic actions of tetracycline and oxytetracycline', *Journal of Pharmacy and Pharmacology*, 14pp. 808–824.

Ju, J., Glazer, A.N. & Mathies, R.A. (1996) 'Energy transfer primers: A new fluorescence labeling paradigm for DNA sequencing and analysis'. Nature Medicine 2 pp. 246–249.

KA, W. (2016) *DNA sequencing costs: Data from the NHGRI genome sequencing program (GSP)*.

Kalinina, O., Lebedeva, I., Brown, J. & Silver, J. (1997) 'Nanoliter scale PCR with TaqMan detection', *Nucleic Acids Research*, 25pp. 1999–2004.

Kassambara, A. (2016) 'Factoextra: Extract and visualize the results of multivariate data analyses', *R package version*, 1.

Katznelson, H., Arnott, J.H. & Bland, S.E. (1952) 'Preliminary report on the treatment of european foulbrood of honeybees with antibiotics', *Scientific agriculture*, 32pp. 180–184.

Kelly, T.J. & Smith, H.O. (1970) 'A restriction enzyme from hemophilus influenzae. II. Base sequence of the recognition site', *Journal of Molecular Biology*, 51pp. 393–409.

Kim, H., Ku, B., Kim, J.Y., Park, Y.J. & Park, Y.B. (2016) 'Confirmatory and exploratory factor analysis for validating the phlegm pattern questionnaire for healthy subjects', *Evidence-based Complementary and Alternative Medicine : eCAM*, 2016.

Kirchhelle, C. (2018) 'Pharming animals: A global history of antibiotics in food production (1935–2017)', *Palgrave Communications 2018 4:1*, 4pp. 1–13.

Kitson, J. (2017) *Tissue extraction from whole caterpillars*.

Kline, R.B. (2005) *Principles and practice of structural equation modeling*. Guilford

publications.

Kuhn, M., Jackson, S. & Cimentada, J. (2022) *Corrr: Correlations in r*.

Laar, S. van, Braeken, J. & Laar, van (2021) 'Understanding the comparative fit index: It's all about the base!', *Practical Assessment, Research, and Evaluation Practical Assessment*, 26.

Laver, T., Harrison, J., O'Neill, P.A., Moore, K., Farbos, A., Paszkiewicz, K. & Studholme, D.J. (2015) 'Assessing the performance of the oxford nanopore technologies MinION', *Biomolecular Detection and Quantification*, 3pp. 1–8.

Lerma, L.L., Benomar, N., Valenzuela, A.S., Carmen Casado Muñoz, M. del, Gálvez, A. & Abriouel, H. (2014) 'Role of EfrAB efflux pump in biocide tolerance and antibiotic resistance of enterococcus faecalis and enterococcus faecium isolated from traditional fermented foods and the effect of EDTA as EfrAB inhibitor', *Food microbiology*, 44pp. 249–257.

Levine, D.P. (2006) 'Vancomycin: A history', *Clinical Infectious Diseases*, 42pp. S5–S12.

Lewis, K.A., Tzilivakis, J., Warner, D.J. & Green, A. (2016) 'An international database for pesticide risk assessments and management', *Human and Ecological Risk Assessment: An International Journal*, 22pp. 1050–1064.

Lewkowski, O. & Erler, S. (2019) 'Virulence of melissococcus plutonius and secondary invaders associated with european foulbrood disease of the honey bee', *MicrobiologyOpen*, 8pp. 1–9.

Li, D., Yang, M., Hu, J., Ren, L., Zhang, Y. & Li, K. (2008) 'Determination and fate of oxytetracycline and related compounds in oxytetracycline production wastewater and the receiving river', *Environmental Toxicology and Chemistry*, 27pp. 80–86.

Li, H. (2018) 'Minimap2: Pairwise alignment for nucleotide sequences', *Bioinformatics*, 34pp. 3094–3100.

Linde, J., Brangsch, H., Hölzer, M., Thomas, C., Elschner, M.C., Melzer, F. & Tomaso, H. (2023) 'Comparison of illumina and oxford nanopore technology for genome analysis of francisella tularensis, bacillus anthracis, and brucella suis', *BMC Genomics*, 24pp. 1–15.

Lorenz, M.G. & Wackernagel, W. (1994) 'Bacterial gene transfer by natural genetic transformation in the environment', *Microbiological Reviews*, 58pp. 563–602.

Lu, H., Giordano, F. & Ning, Z. (2016) 'Oxford nanopore MinION sequencing and genome assembly', *Genomics, Proteomics & Bioinformatics*, 14pp. 265–279.

Madera, S., McNeil, N.B., Serpa, P.B.H., Kamm, J., MPH, C.P., Caughell, C.R., Nichols,

A.R., Dynerman, D., Li, L.M., Sanchez-Guerrero, E.B., BS, M.S.P., MS, A.M.D., Neff, N., Reyes, H.B., Miller, S.A., Yokoe, D.S., DeRisi, J.L., Ramirez-Avila, L. & Langelier, C.R. (2023) 'Prolonged silent carriage, genomic virulence potential and transmission between staff and patients characterize a neonatal intensive care unit (NICU) outbreak of methicillin-resistant staphylococcus aureus (MRSA)', *Infection Control & Hospital Epidemiology*, 44pp. 40–46.

Maechler, M. (2018) 'Cluster: Cluster analysis basics and extensions', *R package version 2.0. 7–1*,

Maes, P.W., Rodrigues, P.A.P., Oliver, R., Mott, B.M. & Anderson, K.E. (2016) 'Diet-related gut bacterial dysbiosis correlates with impaired development, increased mortality and nosema disease in the honeybee (apis mellifera)', *Molecular Ecology*, 25pp. 5439–5450.

Maiden, M.C.J., Bygraves, J.A., Feil, E., Morelli, G., Russell, J.E., Urwin, R., Zhang, Q., Zhou, J., Zurth, K., Caugant, D.A., Feavers, I.M., Achtman, M. & Spratt, B.G. (1998) 'Multilocus sequence typing: A portable approach to the identification of clones within populations of pathogenic microorganisms', *Proceedings of the National Academy of Sciences of the United States of America*, 95pp. 3140–3145.

Malouin, F., Chamberland, S., Brochu, N. & Parr, T.R. (1991) 'Influence of growth media on escherichia coli cell composition and ceftazidime susceptibility', *Antimicrobial Agents and Chemotherapy*, 35pp. 477–483.

Manley, R.O.B. (1946) *Honey farming*.

Manning, R., Lancaster, K., Rutkay, A. & Eaton, L. (2007) 'Survey of feral honey bee (apis mellifera) colonies for nosema apis in western australia', *Australian Journal of Experimental Agriculture*, 47pp. 883–886.

Marotz, C.A., Sanders, J.G., Zuniga, C., Zaramela, L.S., Knight, R. & Zengler, K. (2018) 'Improving saliva shotgun metagenomics by chemical host DNA depletion', *Microbiome*, 6pp. 1–9.

Materon, I.C., Queenan, A.M., Koehler, T.M., Bush, K. & Palzkill, T. (2003) 'Biochemical characterization of beta-lactamases Bla1 and Bla2 from bacillus anthracis', *Antimicrobial agents and chemotherapy*, 47pp. 2040–2042.

Mckee, B.A., Djordjevic, S.P., Goodman, R.D. & Hornitzky, M.A. (2003) 'The detection of melissococcus pluton in honey bees (apis mellifera) and their products using a hemi-nested PCR', *Apidologie*, 34pp. 19–27.

McKee, B.A., Goodman, R.D. & Hornitzky, M.A. (2004) 'The transmission of european

foulbrood (melissococcus plutonius) to artificially reared honey bee larvae (apis mellifera)', *Journal of Apicultural Research*, 43pp. 93–100.

Mezher, Z., Bubnic, J., Condoleo, R., Jannoni-Sebastianini, F., Leto, A., Proscia, F. & Formato, G. (2021) 'Conducting an international, exploratory survey to collect data on honey bee disease management and control', *Applied Sciences 2021, Vol. 11, Page 7311*, 11p. 7311.

More, D. (1976) *The bee book: The history and natural history of the honeybee.*

Morland, D. (1934) 'Distribution of foul brood in england', *Brood diseases of Bees*,

Morse, R.A. (1975) *Bees and beekeeping.*

Mullis, K., Faloona, F., Scharf, S., Saiki, R. & Horn, G. (1986) *Specific enzymatic amplification of DNA in vitro: The polymerase chain reaction.*

Munita, J.M. & Arias, C.A. (2016) 'Mechanisms of antibiotic resistance', *Virulence Mechanisms of Bacterial Pathogens*, pp. 481–511.

Mutinelli, F. (2011) 'The spread of pathogens through trade in honey bees and their products (including queen bees and semen): Overview and recent developments', *Rev. sci. tech. Off. int. Epiz*, 30pp. 257–271.

Nakamura, K., Okumura, K., Harada, M., Okamoto, M., Okura, M. & Takamatsu, D. (2020) 'Different impacts of pMP19 on the virulence of melissococcus plutonius strains with different genetic backgrounds', *Environmental Microbiology*, 22pp. 2756–2770.

National Statistics, O. for (2022) *Population estimates for england and wales.*

Naug, D. (2009) 'Nutritional stress due to habitat loss may explain recent honeybee colony collapses', *Biological Conservation*, 142pp. 2369–2372.

NBU (2024) *APHA - national bee unit - BeeBase.*

NBU (2011) 'Apiary and hive hygiene'. Best Practice Factsheet

NBU (2017) *Foulbrood disease of honey bees and other common brood disorders.*

NBU (2018) *Hive cleaning and sterilisation.*

NBU (2013) *Review of policies on managing and controlling pests and diseases of honey bees. Evidence profile on european foulbrood (EFB).*

NBU (2016) *Starting right with bees.*

NBU (2009) *Statutory procedures for controlling honey bee pests and diseases*, pp. 1–18.

NBU (2012) 'Taking and hiving a swarm'. BBKA News

NCBI (2021) 'Genome list - genome - NCBI'. NCBI

Nocker, A., Cheung, C.Y. & Camper, A.K. (2006) 'Comparison of propidium monoazide with ethidium monoazide for differentiation of live vs. Dead bacteria by selective

removal of DNA from dead cells', *Journal of Microbiological Methods*, 67pp. 310–320.

Novak, R., Braun, J.S., Charpentier, E. & Tuomanen, E. (1998) 'Penicillin tolerance genes of streptococcus pneumoniae: The ABC-type manganese permease complex psa', *Molecular Microbiology*, 29pp. 1285–1296.

Oberacker, P., Stepper, P., Bond, D.M., Höhn, S., Focken, J., Meyer, V., Schelle, L., Sugrue, V.J., Jeunen, G.J., Moser, T., Hore, S.R., Meyenn, F. von, Hipp, K., Hore, T.A. & Jurkowski, T.P. (2019) 'Bio-on-magnetic-beads (BOMB): Open platform for high-throughput nucleic acid extraction and manipulation', *PLOS Biology*, 17p. e3000107.

Okinaka, R.T., Cloud, K., Hampton, O., Hoffmaster, A.R., Hill, K.K., Keim, P., Koehler, T.M., Lamke, G., Kumano, S., Mahillon, J., Manter, D., Martinez, Y., Ricke, D., Svensson, R. & Jackson, P.J. (1999) 'Sequence and organization of pXO1, the large bacillus anthracis plasmid harboring the anthrax toxin genes', *Journal of Bacteriology*, 181p. 6509.

Okumura, K., Arai, R., Okura, M., Kirikae, T., Takamatsu, D., Osaki, M. & Miyoshi-Akiyama, T. (2011) 'Complete genome sequence of melissococcus plutonius ATCC 35311', *Journal of Bacteriology*, 193pp. 4029–4030.

Okumura, K., Takamatsu, D. & Okura, M. (2018) 'Complete genome sequence of melissococcus plutonius DAT561, a strain that shows an unusual growth profile, obtained by PacBio sequencing', *Genome Announcements*, 6pp. e00431-18-e00431-18.

Okumura, K., Takamatsu, D. & Okura, M. (2019) 'Complete genome sequences of two melissococcus plutonius strains with different virulence profiles, obtained by PacBio sequencing', *Microbiology Resource Announcements*, 8pp. 38–57.

Oldroyd, B.P., Goodman, R.D., Hornitzky, M.A.Z. & Chandler, D. (1989) 'The effect on american foulbrood of standard oxytetracycline hydrochloride treatments for the control of european foulbrood of honeybees (apis mellifera)', *Australian Journal of Agricultural Research*, 40pp. 691–697.

Ormond Harry, A.A. (1982) *Mastering the art of beekeeping*.

Papić, B., Diricks, M. & Kušar, D. (2021) 'Analysis of the global population structure of paenibacillus larvae and outbreak investigation of american foulbrood using a stable wgMLST scheme', *Frontiers in Veterinary Science*, 8.

Paradis, E., Blomberg, S., Bolker, B., Brown, J., Claramunt, S., Claude, J., Cuong, H.S., Desper, R., Didier, G., Durand, B., Dutheil, J., Ewing, R., Gascuel, O., Guillerme, T., Heibl, C., Ives, A., Jones, B., Krah, F., Lawson, D., et al. (2023) R package version

5.7. *Ape: Analyses of phylogenetics and evolution*.

Pasquina-Lemonche, L., Burns, J., Turner, R.D., Kumar, S., Tank, R., Mullin, N., Wilson, J.S., Chakrabarti, B., Bullough, P.A., Foster, S.J. & Hobbs, J.K. (2020) 'The architecture of the gram-positive bacterial cell wall', *Nature 2020 582:7811*, 582pp. 294–297.

Pawlowski, A.C., Wang, W., Koteva, K., Barton, H.A., McArthur, A.G. & Wright, G.D. (2016) 'A diverse intrinsic antibiotic resistome from a cave bacterium', *Nature communications*, 7.

Payne, A., Holmes, N., Clarke, T., Munro, R., Debebe, B.J. & Loose, M. (2020) 'Readfish enables targeted nanopore sequencing of gigabase-sized genomes', *Nature Biotechnology 2020 39:4*, 39pp. 442–450.

Pearson, T., Busch, J.D., Ravel, J., Read, T.D., Rhoton, S.D., U'Ren, J.M., Simonson, T.S., Kachur, S.M., Leadem, R.R., Cardon, M.L., Ert, M.N.V., Huynh, L.Y., Fraser, C.M. & Keim, P. (2004) 'Phylogenetic discovery bias in bacillus anthracis using single-nucleotide polymorphisms from whole-genome sequencing', *Proceedings of the National Academy of Sciences of the United States of America*, 101pp. 13536–13541.

Pebesma, E. (2023) R package version 1.0-15. *Sf: Simple features for r*.

Pickens, L.B. & Tang, Y. (2010) 'Oxytetracycline biosynthesis', *Journal of Biological Chemistry*, 285pp. 27509–27515.

Pollie, R. (2023) 'Genomic sequencing costs set to head down again', *Engineering*, 23pp. 3–6.

Potts, S.G., Biesmeijer, J.C., Kremen, C., Neumann, P., Schweiger, O. & Kunin, W.E. (2010) 'Global pollinator declines: Trends, impacts and drivers', *Trends in Ecology and Evolution*, 25pp. 345–353.

Potts, S.G., Imperatriz-Fonseca, V., Ngo, H.T., Aizen, M.A., Biesmeijer, J.C., Breeze, T.D., Dicks, L.V., Garibaldi, L.A., Hill, R., Settele, J. & Vanbergen, A.J. (2016) 'Safeguarding pollinators and their values to human well-being', *Nature*, 540pp. 220–229.

Poyart, C., Quesne, G., Boumaila, C. & Trieu-Cuot, P. (2001) 'Rapid and accurate species-level identification of coagulase-negative staphylococci by using the sodA gene as a target', *Journal of Clinical Microbiology*, 39pp. 4296–4301.

Prestinaci, F., Pezzotti, P. & Pantosti, A. (2015) 'Antimicrobial resistance: A global multifaceted phenomenon', *Pathogens and Global Health*, 109pp. 309–318.

Project, I. (2020) *Inkscape*.

Pyörälä, S., Baptiste, K.E., Catry, B., Duijkeren, E. van, Greko, C., Moreno, M.A., Pomba, M.C.M.F., Rantala, M., Ružauskas, M., Sanders, P., Threlfall, E.J., Torren-Edo, J. & Törneke, K. (2014) 'Macrolides and lincosamides in cattle and pigs: Use and development of antimicrobial resistance', *The Veterinary Journal*, 200pp. 230–239.

Raymann, K. & Moran, N.A. (2018) 'The role of the gut microbiome in health and disease of adult honey bee workers', *Current Opinion in Insect Science*, 26pp. 97–104.

Reyneke, B., Ndlovu, T., Khan, S. & Khan, W. (2017) 'Comparison of EMA-, PMA- and DNase qPCR for the determination of microbial cell viability', *Applied Microbiology and Biotechnology*, 101pp. 7371–7383.

Rodríguez-Beltrán, J., DelaFuente, J., León-Sampedro, R., MacLean, R.C. & Millán, Á.S. (2021) 'Beyond horizontal gene transfer: The role of plasmids in bacterial evolution', *Nature Reviews Microbiology 2021 19:6*, 19pp. 347–359.

Roetschi, A., Berthoud, H., Kuhn, R. & ALP, A.I.A.L.-P. (2008) 'Infection rate based on quantitative real-time PCR of melissococcus plutonius, the causal agent of european foulbrood, in honeybee colonies before and after apiary sanitation*', *Apidologie*, 39pp. 362–371.

Rosseel, Y. (2012) 'Lavaan: An r package for structural equation modeling', *Journal of Statistical Software*, 48pp. 1–36.

Ruckert, A., Harris, F., Aenishaenslin, C., Aguiar, R., Boudreau-LeBlanc, A., Carmo, L.P., Labonté, R., Lambraki, I., Parmley, E.J. & Wiktorowicz, M.E. (2024) 'One health governance principles for AMR surveillance: A scoping review and conceptual framework', *Research Directions: One Health*, 2p. e4.

Rugalema, G., Muir, G., Mathieson, K., Measures, E., Oehler, F. & Stloukal, L. (2009) 'Emerging and re-emerging diseases of agricultural importance: Why local perspectives matter', *Food Security 2009 1:4*, 1pp. 441–455.

Russel, E.J., Tarr, H.L.A., Morison, G.D., Morgenthaler, O. & Hambleton, Jas.I. (1937) 'Diseases of bees', *ROTHAMSTED CONFERENCES*,

Russel, H., Carita, F., Gavin, D. & Robert, W. (1993) 'Kinetic PCR analysis: Real-time monitoring of DNA amplification reactions', *Biotechnology*, 11pp. 1026–1030.

Sanger, F., Nicklen, S. & Coulson, A.R. (1977) 'DNA sequencing with chain-terminating inhibitors.', *Proceedings of the National Academy of Sciences of the United States of America*, 74pp. 5463–5467.

Scheu, P.M., Berghof, K. & Stahl, U. (1998) 'Detection of pathogenic and spoilage

micro-organisms in food with the polymerase chain reaction', *Food Microbiology*, 15pp. 13–31.

Schork, N.J., Fallin, D. & Lanchbury, J.S. (2000) 'Single nucleotide polymorphisms and the future of genetic epidemiology', *Clinical Genetics*, 58pp. 250–264.

Sechirst, E.L. (1976) *Amateur beekeeping*.

Seemann, T. (2014) 'Prokka: Rapid prokaryotic genome annotation', *Bioinformatics*, 30pp. 2068–2069.

Sereika, M., Kirkegaard, R.H., Karst, S.M., Michaelsen, T.Y., Sørensen, E.A., Wollenberg, R.D. & Albertsen, M. (2022) 'Oxford nanopore R10.4 long-read sequencing enables the generation of near-finished bacterial genomes from pure cultures and metagenomes without short-read or reference polishing', *Nature Methods 2022 19:7*, 19pp. 823–826.

Sharma, M.K., Amudhan, S., Achar, M. & Vishwakarma, A. (2022) 'COVID-19 pandemic, risk, and blame attributions: A scoping review', *Indian Journal of Psychological Medicine*, 44pp. 227–233.

Slowikowski, K. (2023) R package version 0.9.4. *Ggrepel: Automatically position non-overlapping text labels with ggplot2*.

Smart, M., Pettis, J., Rice, N., Browning, Z. & Spivak, M. (2016) 'Linking measures of colony and individual honey bee health to survival among apiaries exposed to varying agricultural land use', *PLoS ONE*, 11.

Smith, E.A., Anderson, K.E., Corby-Harris, V., McFrederick, Q.S., Parish, A.J., Rice, D.W. & Newton, I.L.G. (2021) 'Reclassification of seven honey bee symbiont strains as bombella apis', *International Journal of Systematic and Evolutionary Microbiology*, 71p. 004950.

Smith, F.G. (1963) *Beekeeping*.

Smith, K.M., Loh, E.H., Rostal, M.K., Zambrana-Torrelio, C.M., Mendiola, L. & Daszak, P. (2013) 'Pathogens, pests, and economics: Drivers of honey bee colony declines and losses', *EcoHealth*, 10pp. 434–445.

Smith, P., Black, H., Evans, C., Hails, R., Thomson, A., Hesketh, H., Johnson, A., May, L., Pickup, R., Purse, B., Ashmore, M. & Others., & (2011) 'Regulating services [chapter 14]. In: UK national ecosystem assessment. Understanding nature's value to society. Technical report. cambridge', *Unep-Wcmc*, pp. 535–596.

Sprague, G.F. (1991) 'Genetic exchange between kingdoms', *Current Opinion in Genetms and Development*, 1pp. 530–533.

Stanisavljević, N., Malešević, M., Rašić, S., Vukotić, G., Gardijan, L., Obradović, M. &

Kojić, M. (2023) 'Short-term effects of brevibacillus laterosporus supplemented diet on worker honey bee microbiome: A pilot study', *Journal of Apicultural Research*,

Stoler, N. & Nekrutenko, A. (2021) 'Sequencing error profiles of illumina sequencing instruments', *NAR Genomics and Bioinformatics*, 3.

Stuart, F.S. (1947) *Beekeping practice*. 3rd edition.

Takamatsu, D., Arai, R., Miyoshi-Akiyama, T., Okumura, K., Okura, M., Kirikae, T., Kojima, A. & Osaki, M. (2013) 'Identification of mutations involved in the requirement of potassium for growth of typical melissococcus plutonius strains', *Applied and Environmental Microbiology*, 79pp. 3882–3886.

Takeuchi, F., Watanabe, S., Baba, T., Yuzawa, H., Ito, T., Morimoto, Y., Kuroda, M., Cui, L., Takahashi, M., Ankai, A., Baba, S.I., Fukui, S., Lee, J.C. & Hiramatsu, K. (2005) 'Whole-genome sequencing of staphylococcus haemolyticus uncovers the extreme plasticity of its genome and the evolution of human-colonizing staphylococcal species', *Journal of Bacteriology*, 187p. 7292.

Tams, K.W., Larsen, I., Hansen, J.E., Spiegelhauer, H., Strøm-Hansen, A.D., Rasmussen, S., Ingham, A.C., Kalmar, L., Kean, I.R.L., Angen, Ø., Holmes, M.A., Pedersen, K., Jelsbak, L., Folkesson, A., Larsen, A.R. & Strube, M.L. (2023) 'The effects of antibiotic use on the dynamics of the microbiome and resistome in pigs', *Animal Microbiome*, 5pp. 1–16.

Tamura, K., Stecher, G. & Kumar, S. (2021) 'MEGA11: Molecular evolutionary genetics analysis version 11', *Molecular Biology and Evolution*, 38pp. 3022–3027.

Tan, S.C. & Yiap, B.C. (2009) 'DNA, RNA, and protein extraction: The past and the present', *Journal of Biomedicine and Biotechnology*, 2009.

Thakkar, J.J. (2020) *Studies in systems, decision and control 285 structural equation modelling application for research and practice (with AMOS and r)*,

Thebeau, J.M., Liebe, D., Masood, F., Kozii, I.V., Klein, C.D., Zabrodski, M.W., Moshynskyy, I., Sobchishin, L., Wilson, G., Guarna, M.M., Gerbrandt, E.M., Simko, E. & Wood, S.C. (2022) 'Investigation of melissococcus plutonius isolates from 3 outbreaks of european foulbrood disease in commercial beekeeping operations in western canada', *The Canadian Veterinary Journal*, 63p. 935.

Thompson, H.M. & Brown, M. (1999) 'The role of the national bee unit in controlling statutory bee diseases', *Bee World*, 80pp. 132–139.

Thompson, H.M. & Brown, M.A. (2001) 'Is contact colony treatment with antibiotics an effective control for european foulbrood?', *Bee World*, 82pp. 130–138.

Thompson, H.M., Waite, R.J., Wilkins, S., Brown, M.A., Bigwood, T., Shaw, M., Ridgway, C. & Sharman, M. (2006) 'Effects of shook swarm and supplementary feeding on oxytetracycline levels in honey extracted from treated colonies', *Apidologie*, 37pp. 51–57.

Tinsley, J. (1945) *Beekeeping up-to-date*.

Tomkies, V., Flint, J., Johnson, G., Waite, R., Wilkins, S., Danks, C., Watkins, M., Cuthbertson, A.G.S., Carpana, E., Marris, G., Budge, G. & Brown, M.A. (2009) 'Development and validation of a novel field test kit for european foulbrood', *Apidologie*, 40pp. 63–72.

Tonkin-Hill, G., MacAlasdair, N., Ruis, C., Weimann, A., Horesh, G., Lees, J.A., Gladstone, R.A., Lo, S., Beaudoin, C., Floto, R.A., Frost, S.D.W., Corander, J., Bentley, S.D. & Parkhill, J. (2020) 'Producing polished prokaryotic pangenomes with the panaroo pipeline', *Genome biology*, 21.

Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A., Gocayne, J.D., Amanatides, P., Ballew, R.M., Huson, D.H., Wortman, J.R., Zhang, Q., Kodira, C.D., Zheng, X.H., Chen, L., et al. (2001) 'The sequence of the human genome', *Science*, 291.

Vivo, M.D., Lee, H.H., Huang, Y.S., Dreyer, N., Fong, C.L., Mattos, F.M.G. de, Jain, D., Wen, Y.H.V., Mwihaki, J.K., Wang, T.Y., Machida, R.J., Wang, J., Chan, B.K.K. & Tsai, I.J. (2022) 'Utilisation of oxford nanopore sequencing to generate six complete gastropod mitochondrial genomes as part of a biodiversity curriculum', *Scientific Reports 2022 12:1*, 12pp. 1–11.

Voelkerding, K.V., Dames, S.A. & Durtschi, J.D. (2009) 'Next-generation sequencing:from basic research to diagnostics'. Clinical Chemistry 55 pp. 641–658.

Waine, A.C. (1975) *Background to beekeeping*.

Waite, R.J., BROWN, M.A., THOMPSON, H.M. & BEW, M.H. (2007) 'Controlling european foulbrood with the shook swarm method and oxytetracycline in the UK', *Apidologie*, 38pp. 67–76.

Wallberg, A., Bunikis, I., Pettersson, O.V., Mosbech, M.B., Childers, A.K., Evans, J.D., Mikheyev, A.S., Robertson, H.M., Robinson, G.E. & Webster, M.T. (2019) 'A hybrid de novo genome assembly of the honeybee, apis mellifera, with chromosome-length scaffolds', *BMC genomics*, 20.

Wang, W., Xu, Y., Gao, R., Lu, R., Han, K., Wu, G. & Tan, W. (2020) 'Detection of SARS-CoV-2 in different types of clinical specimens'. JAMA - Journal of the American

Medical Association 323 pp. 1843–1844.

Watson, J.D. & Crick, F.H.C. (1953) 'Molecular structure of nucleic acids: A structure for deoxyribose nucleic acid', *Nature*, 171pp. 737–738.

Weaver, W. (1970) 'Molecular biology: Origin of the term [10]'. Science 170 pp. 581–582.

Webb, A. (1945) *Beekeeping for profit and pleasure*.

White, G.F. (1912) 'The cause of european foul brood', *Bureau of Entomology circular*, pp. 1–15.

Wick, R.R., Judd, L.M., Gorrie, C.L. & Holt, K.E. (2017) 'Completing bacterial genome assemblies with multiplex MinION sequencing', *Microbial genomics*, 3.

Wickham, H. (2016) *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York.

Wilkins, S., Brown, M.A. & Cuthbertson, A.G.S. (2007) 'The incidence of honey bee pests and diseases in england and wales', *Pest Management Science*, 63pp. 1062–1068.

Winston, Mark.L. (1987) *The biology of the honey bee*.

Wong, T.Y., Fernandes, S., Sankhon, N., Leong, P.P., Kuo, J. & Liu, J.K. (2008) 'Role of premature stop codons in bacterial evolution', *Journal of Bacteriology*, 190pp. 6718–6725.

Wu, H.J., Wang, A.H.J. & Jennings, M.P. (2008) 'Discovery of virulence factors of pathogenic bacteria', *Current Opinion in Chemical Biology*, 12pp. 93–101.

Xian, Z., Li, S., Mann, D.A., Huang, Y., Xu, F., Wu, X., Tang, S., Zhang, G., Stevenson, A., Ge, C. & Deng, X. (2022) 'Subtyping evaluation of salmonella enteritidis using single nucleotide polymorphism and core genome multilocus sequence typing with nanopore reads', *Applied and Environmental Microbiology*, 88.

Yu, G. (2020) 'Using ggtree to visualize data on tree-like structures', *Current Protocols in Bioinformatics*, 69(1), p. e96.

Zacepins, A., Kviesis, A., Komasilovs, V. & Brodschneider, R. (2021) 'When it pays to catch a swarm—evaluation of the economic importance of remote honey bee (apis mellifera) colony swarming detection', *Agriculture 2021, Vol. 11, Page 967*, 11p. 967.

Zankari, E., Hasman, H., Cosentino, S., Vestergaard, M., Rasmussen, S., Lund, O., Aarestrup, F.M. & Larsen, M.V. (2012) 'Identification of acquired antimicrobial resistance genes', *The Journal of antimicrobial chemotherapy*, 67pp. 2640–2644.

Zhou, Z., Alikhan, N.F., Sergeant, M.J., Luhmann, N., Vaz, C., Francisco, A.P., Carriço, J.A. & Achtman, M. (2018) 'GrapeTree: Visualization of core genomic relationships among 100,000 bacterial pathogens', *Genome Research*, 28pp. 1395–1404.