Epidemiology and Genomics of European Foulbrood (*Melissococcus plutonius*) of Honey Bees

Edward George Haynes

PhD

The University of York

Biology

September 2013

Abstract

European Foulbrood (EFB) is an important disease of honey bee larvae that has increased in prevalence in recent years, in both the UK and other countries. EFB is caused by the gram-positive bacterium *Melissococcus plutonius*. To date, most molecular epidemiology studies on *M. plutonius* have concentrated on developing detection methods, and using these to identify the bacteria in honey bees and honey bee hive products, though recently two genomes of *M. plutonius* have been published. In this thesis a genome sequence for the Type Strain is generated, and used to draw inferences about the accuracy of the published sequences. Genome sequence for other, field-collected isolates were generated and used to identify mobile genetic elements and to elucidate the evolutionary history of *M. plutonius*. The genome sequences were also used to design the first strain typing scheme for this pathogen, despite this pathogen being previously described as genetically homogenous.

Previously undetectable diversity of *M. plutonius* is explored at a landscape level, showing geographical structuring of populations of the bacterium both within and among countries. The drivers of the observed structure are investigated, with both anthropogenic movements by beekeepers and natural transmission by bees implicated in the maintenance of *M. plutonius* population structure. This thesis demonstrates the role of the beekeeper in spreading the bacterium through the sale of live bees and through contaminated equipment. Asymptomatic larvae are shown to be carriers of the bacterium (and to go on to develop disease) and a potential role for social wasps as a vector of the pathogen was discovered.

List of Contents

List of Figures

List of Tables

Acknowledgements

A great many people made invaluable contributions to this thesis, through their support, help and advice. My parents have always been there to help pick me up and calm me down, and it is certain that I would not be where I am today without them. I also owe my brother Henry thanks, for keeping me grounded with gentle mockery.

Much credit for the content of this thesis goes to my many supervisors, who have been great sources of advice and intellectual stimulation. Thanks to Professor Peter Young, for sharing his immense experience and expertise in bacterial genetics and genomics; Dr Richard Thwaites for giving invaluable technical help in microbiology and molecular biology; Dr Giles Budge for an unparalleled knowledge of bee diseases, and for applying intellectual rigour to my ideas; and Dr Thorunn Helgason, for imparting her great knowledge of microbial ecology, and for helping me develop as a scientist. Thanks also to Bernard Diaper, my supervisor at Bee Disease Insurance, for his support, interest, and commitment to my work.

I owe a great deal to so many of my colleagues at the University of York and the Food and Environment Research Agency. There are far too many to name, so I will restrict myself to a few brief mentions. Thanks to Jan Bew and Dr Jennifer Hodgetts, for inducting me into the world of the microbiology lab. Thanks to Nitin Kumar and Dr Toby Hodges, for all their help on the road to becoming a bioinformatician. A big thank you to everyone in the NBU team, but especially Mike Brown, Dr Gay Marris, and Ben Jones, all of whom have helped me enormously over the last four years. Thanks as well to all the NBU Bee Inspectors, particularly Julian Routh and Terry Ashley, for all the help with sampling.

Finally thanks to all my friends for being there for me through the years. To Amanda, for giving me love and support when I've been irritable and stressed. To Rose and Adam, for managing to live in the same house as me for three years. And to everyone else, who I couldn't possibly name here, for all the much needed distractions from the PhD.

Author's Declaration

The work in this thesis has contributed to the following publications:

Haynes, E., Helgason, T., Young, J.P.W., Thwaites, R., Budge, G.E. (2013) "A typing scheme for the honey bee pathogen *Melissococcus plutonius* allows detection of disease transmission events and a study of the distribution of variants." *Environmental Microbiology Reports*, **5:** 525-529

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*" *The ISME Journal*, doi:10.1038/ismej.2014.20

All work presented herein is the author's own, with the following exceptions:

Chapter 2

The original draft *M. plutonius* genome was supplied at the start of the project by Dr Matthew Holden, of the Wellcome Trust Sanger Institute. Preparation of extracted *M. plutonius* DNA for sequencing on the Roche 454 FLX pyrosequencer was performed by Ummey Hany, of Fera. Assembly of the resulting reads into contigs using Newbler was performed by Rachel Glover, of Fera. Preparation of extracted *M. plutonius* DNA for sequencing on the Ion Torrent PGM Sequencer was performed by Celina Whalley, then of the University of York.

Chapter 3

Help identifying traditional MLST loci was provided by Professor Edward J Feil, University of Bath.

The IPI sampling scheme was designed by Dr Giles Budge, Fera, to fulfil the requirements of the, "Modelling systems for managing bee disease: the epidemiology of European foulbrood," project. For these samples, field sampling was performed by the Seasonal Bee Inspectors, Fera. Sampling of larvae from combs was performed by Ben Jones, Fera. The first round of culture of *M. plutonius* from larvae, DNA extraction by Chelex® and testing by Taqman® assay were performed by Emiline Quill, Fera. Processing of all 2012 IPI isolates, and ST assignment using the BioNumerics software, was performed by Victoria Tomkies, Fera. Distribution maps for England and Wales and for Scotland were created with Dr Giles Budge, Fera. All beekeeper ownership network analysis and original creation of Figure 4.7 was performed solely by Dr Mark Shirley, Newcastle University (tone and sharpness of Figure 4.7 subsequently enhanced by author to improve clarity).

1. Introduction

This introduction provides an overview of honey bee biology, the importance of honey bees and the severity of the threats they face, illustrating the necessity of the study of honey bee diseases. The current state of our understanding of European Foulbrood is described, outlining the knowledge base from which this project started. There then follows a brief introduction to the concepts of molecular epidemiology and genomics, which are central to all the work performed in this thesis. Finally, the research undertaken in the other chapters of this thesis is outlined.

1.1. Honey Bees

1.1.1. Species and Distribution

Honey bees are eusocial insects in the genus *Apis*, which produce a vertical nest comb made of wax cells. These cells are used for storing brood and honey (Engel, 1999). Early human species have hunted honey from *Apis* bees since *Homo habilis* first emerged in Africa, around five million years ago, and the earliest evidence of the management of honey bees in hives dates from 2,400 BCE in Egypt (Crane, 1999). Between three and 24 species of honey bee have been described over the years, but most studies now identify nine or ten distinct species (Engel, 1999; Arias and Sheppard, 2005). These can be subdivided into three clear groups, or subgenera; the dwarf bees (*Micrapis*) *Apis florea* and *Apis andreniformis*; the giant bees (*Megapis*) *Apis dorsata*, *Apis binghami* and *Apis laboriosa;* and the cavity-nesting bees (*Apis*) *Apis mellifera*, *Apis cerana*, *Apis koschevnikovi*, *Apis nuluensis* and *Apis nigrocincta*. The monophyly of these three groups is supported by DNA sequence data, though the differentiation of some species within them is still unresolved (Arias and Sheppard, 2005).

Honey bees are primarily Asian species. Different species of giant honey bee are distributed from the Himalayas to the islands of Sulawesi and the Philippines. The dwarf bees occur from China to Malaysia and Indonesia, with *A. florea* extending westwards to the Arabian Peninsula. Three of the cavity-nesting species, *A. koschevnikovi*, *A. nuluensis* and *A. nigrocincta* are confined to South East Asia (Arias et al., 1996; Engel, 1999). The bees with the widest natural ranges are also the most commercially important. *A cerana*, the Eastern Hive Bee, is found from Afghanistan through India, China and Korea, and across the islands of East and Southeast Asia from Japan to Malaysia and Indonesia (Engel, 1999). The Western Honey Bee, *A. mellifera*, appears to have arisen in Africa, and has dispersed naturally though much of the continent, as well as through the Middle East and Europe (Engel, 1999; Whitfield et al., 2006).

In historical times, honey bees were absent from the Americas, Australia and New Zealand (although a 14 million year old fossilised specimen of the newly described *Apis nearctica* has been identified from North America (Engel et al., 2009)). Many species of hornet and stingless bee were reported from the Americas by early explorers and settlers, but none were recognizably the same as the European honey bee (Purchas, 1657; Pellet, 1924). The first introduction of Western honey bees (*A. mellifera*) into the New World was to Bermuda in 1617, and the transport of bees into Virginia took place from the 1620s onwards (Oertel, 1976a; Crane, 1999). Further introductions took place, largely from England (but occasionally and haphazardly from other Northern European countries, like Sweden (Oertel, 1976b)), before feral bees began to swarm and spread across the continent by themselves. Indeed, the Native Americans recognised these insects as a European introduction, the arrival of which often foreshadowed the movement of settlers into an area; "the white man's fly," (Jefferson, 1787). Similar introductions took place to Australia in the 1820s, and New Zealand in the 1830s (Crane, 1999). *A. mellifera* is now globally distributed, present on every continent except Antarctica, and is displacing *A. cerana* in temperate regions of Asia (Sakagami, 1959; Bailey, 1981). Conversely in Australia, *A. cerana* is considered a potentially harmful invasive species (Radloff et al., 2010; Carr, 2011).

1.1.2. The Keeping of Honey Bees

Two honey bee species are kept in important numbers by humans. *A. cerana* is managed in several countries across Asia. It is a slightly smaller bee than *A. mellifera* (Figure 1.1) with a smaller foraging range. Colonies are usually managed in wooden boxes, on wax combs that are traditionally fixed in position, but moveable in more modern hives (Hisashi, 2010).

A. mellifera is the bee species that is most widely kept around the world. In the UK, a colony of bees consisting of a single queen, tens of thousands of female workers and several hundreds or thousands of male drones (Winston, 1987) is usually kept in a wooden colony box or hive. Several variations on the simple box exist, but generally each box contains a number of movable wooden frames, or combs, on which the bees can draw out a layer of hexagonal wax cells in which to store brood, honey or pollen. When the queen lays an egg in a brood cell it hatches into a larva after three days, and is fed honey and hypopharyngeal gland secretions by brood-tending nurse bees for a further five days, whereupon its cell is capped over by the nurse bees and the larva pupates. After another 13 days or so the new adult bee emerges (Bailey, 1981). The newly emerged bee first engages in cell cleaning activity as it finalises its development, then progresses through nurse bee, middle aged bee (performing tasks such as nest building and colony defence) and, finally forager. At this stage the bees leave the hive to search for sources of pollen and nectar (Johnson, 2010).

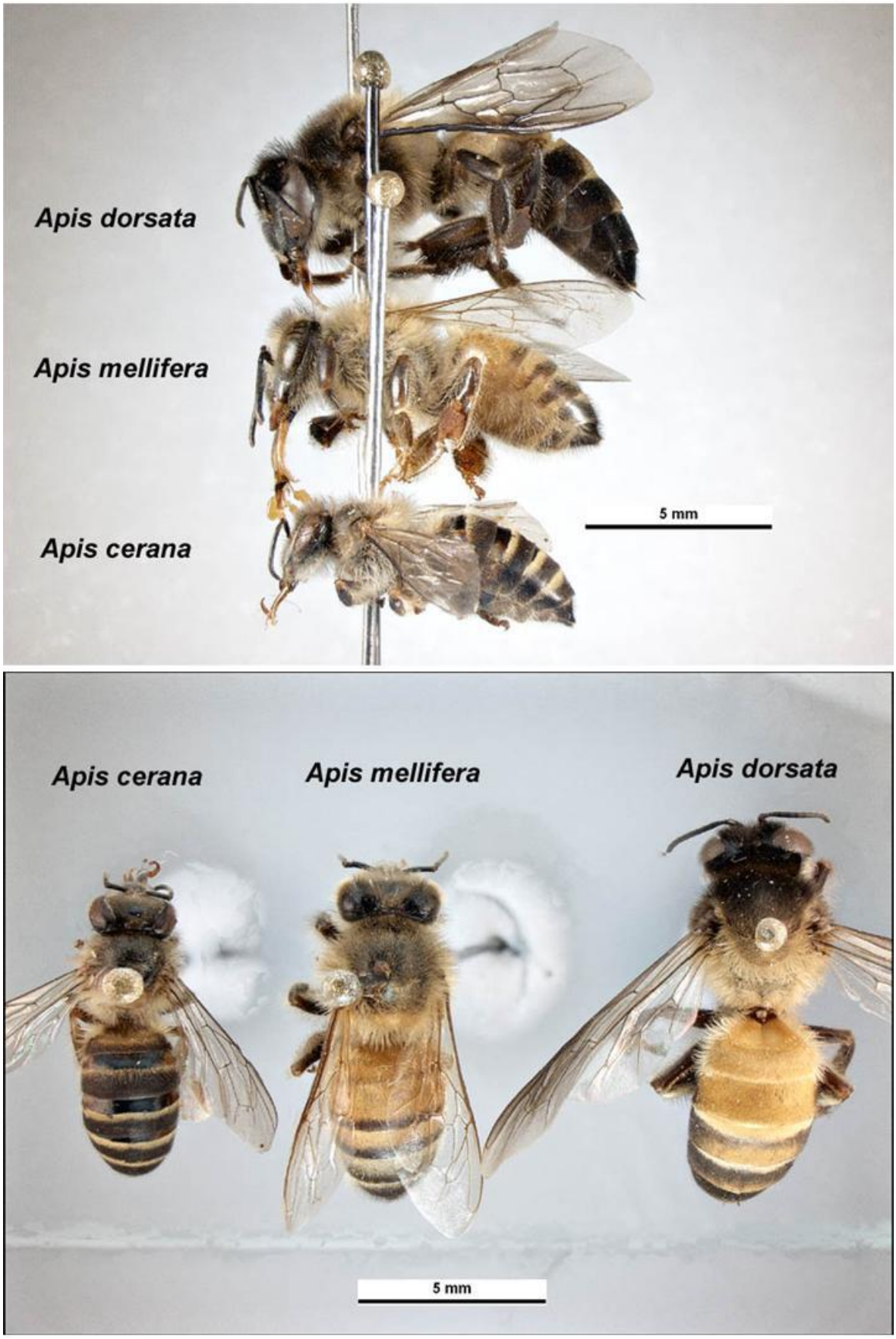


Figure 1.1. Lateral and dorsal views of workers of *Apis mellifera*, *Apis cerana* and *Apis dorsata*. *A. cerana* is noticeably smaller than *A. mellifera*. *A. dorsata* is much larger, and similar to the closely related *Apis laboriosa*. Images by Ken Walker, Museum Victoria, used under Creative Commons Attribution 3.0 Australia.

The management of the colony for honey production can be enhanced by the use of one or more super boxes. This is a smaller box that sits on top of the brood box, which contains shorter frames that the workers fill with honey. The queen is prevented from laying brood in the super frames by the use of a queen excluder – a grill that is big enough to allow the workers through, yet small enough to prevent the larger queen from passing (Figure 1.2). Many aspects of beekeeping are focussed on reducing the difficulty and hazard of management. To that end, personal protective equipment (a bee suit and veil) are almost uniformly worn when dealing with bees, and smoke is often pumped into a colony (using a "smoker") to reduce aggression (probably through a combination of sensory inhibition, and inducing a gorging response in preparation of relocation as a swarm away from the fire (Visscher et al., 1995)). Bee management is also aided by grouping several colonies together on one site, known as an apiary (or, in the US especially, a bee yard). Some beekeepers specialise in the movement of colonies between sites, sometimes across large distances, to provide pollination services or take advantage of nectar flow (Corbet et al., 1992).

1.1.3. The Importance of Bees

Honey bees are extremely important economically and ecologically through their role as pollinators; the yields of 96% of animal-pollinated crops can be increased by the visits of honey bees (Potts, Biesmeijer, et al., 2010). Honey bees are the most economically valuable pollinators of monoculture crops worldwide (Klein et al., 2007), and contribute billions of dollars annually to the economy of the US alone (Southwick and Southwick, 1992), including hundreds of millions of dollars in direct bee products such as honey and wax (Gallai et al., 2009). Other more esoteric benefits, such as nitrate fertilization of plants with bee faeces (Mishra et al., 2013), are just beginning to be investigated.

Apis bees are of course not the only animal pollinators of flowering plants. Species known to pollinate crops for human consumption include bumble bees, solitary bees, stingless bees, wasps, flies, beetles, thrips, birds and, possibly, bats (Klein et al., 2007). There has been a suggestion that, through exploitative competition, introduced honey bees can have a detrimental impact on native pollinators and their relationships with the plants they pollinate (Aizen and Feinsinger, 1994; Kato et al., 1999; do Carmo et al., 2004). However, the widespread nature of such effects is not proven (Goulson, 2003; Paini, 2004), and synergistic interactions between *Apis* and non-*Apis* pollinators have been demonstrated (Greenleaf and Kremen, 2006).

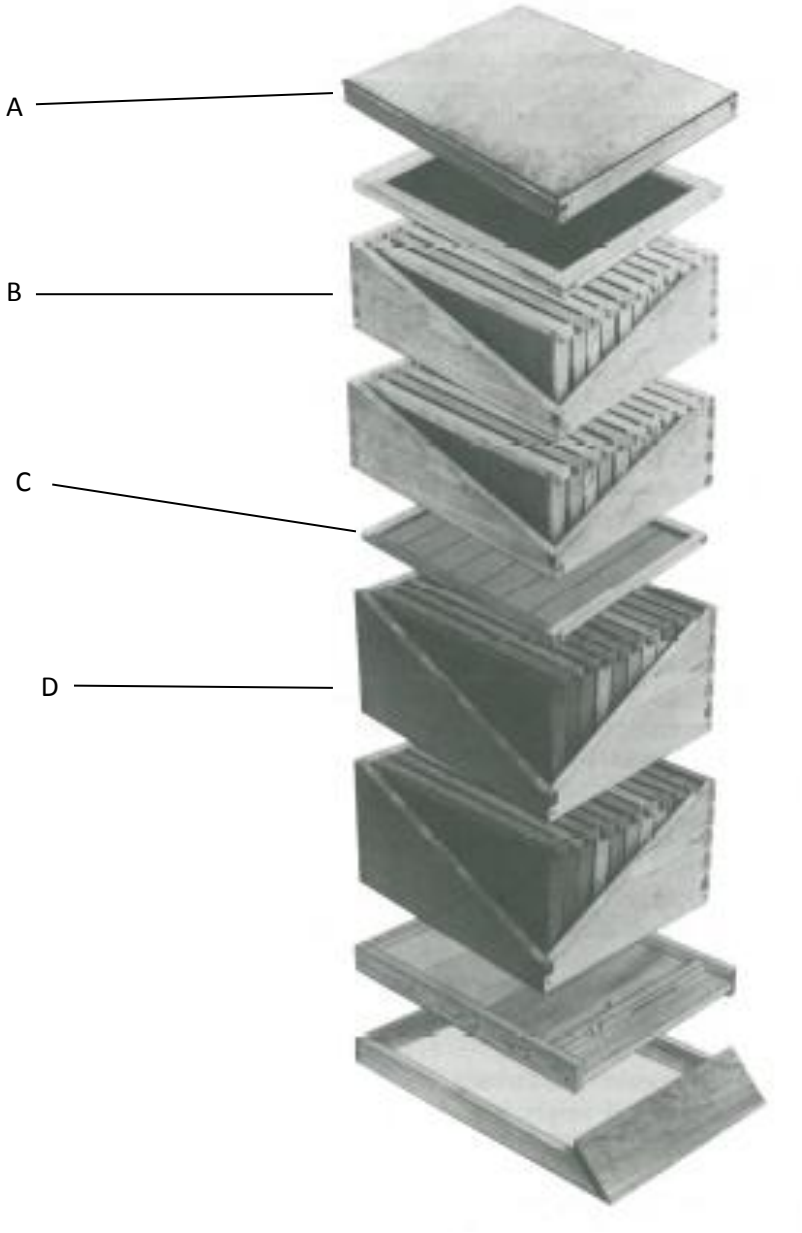


Figure 1.2. Cut away view of a basic modern hive. Some parts are highlighted, including; A, colony roof; B, super boxes, containing super frames; C, queen excluder; D, brood boxes, where brood combs are kept. Photograph adapted from “The Hive and the Honey Bee” (2000) Dadant & Sons, Hamilton, Illinois (use permitted with acknowledgment).

1.1.4. Threats to the Honey Bee

The decline of pollinator populations in North America and Europe, especially those of managed honey bees, has been repeatedly noted in recent years (Biesmeijer et al., 2006; De La Rua et al., 2009; Potts, Biesmeijer, et al., 2010; Potts, Roberts, et al., 2010). There is no single cause of these declines, but instead many interacting factors play a role. Pesticides are often cited as a cause of bee declines, and the European Commission has recently agreed to restrict the use of three systemic pesticides within the EU (The European Commission, 2013). While most countries have a robust legal framework to reduce bee exposure to directly-applied pesticides (VanEngelsdorp and Meixner, 2010) the effects of systemic pesticides (pesticides which are applied as a seed dressing, and then migrate through the plant) are harder to monitor. This is because the pesticides are present in pollen and nectar in sub-lethal doses, which do not lead to mass bee deaths in fields but which can still have important deleterious effects on the bees (Desneux et al., 2007), including serious impairment of their ability to navigate their surroundings (Henry et al., 2012). This has led to the conclusion that pesticides may play an important role in Colony Collapse Disorder (VanEngelsdorp et al., 2008; Henry et al., 2012).

Colony Collapse Disorder (CCD) is a relatively recent phenomenon, characterised by the rapid disappearance of virtually all adult bees from a colony. CCD has contributed to the decline of honey bee populations in the United States (VanEngelsdorp et al., 2008), and has recently been observed in Europe for the first time (Dainat et al., 2012). As well as a hypothesised relationship with pesticides, CCD has been correlated with high colony density (VanEngelsdorp et al., 2008), pathogen load (VanEngelsdorp et al., 2009) and the presence of specific viral pathogens (Cox-foster et al., 2007). Honey bees are made more susceptible to virus infection, through both immune suppression and vectored transmission, by the presence of the ectoparasitic mite *Varroa destructor* (Sumpter and Martin, 2004; Shen et al., 2005). Consequently the impact on honey bee survival of *V. destructor*, now ubiquitous in Europe and North America, has been extreme (Currie et al., 2010; Genersch et al., 2010; Guzmán-Novoa et al., 2010).

1.1.5. Honey Bee Pathogens

As well as pathogenic drivers of large-scale colony losses, honey bees suffer from a vast array of other parasites and pathogens. Many different viruses infect honey bees throughout all developmental stages, though often field infections are latent, causing no disease symptoms (Chen et al., 2006). Important honey bee viruses in the UK include chronic bee paralysis virus (which causes paralysis), sacbrood virus (which deforms and then kills the bee larva), black queen cell virus (which discolours the queen cell, and kills the developing queen pupa) (Bailey, 1981),

and deformed wing virus (which deforms the honey bee's wings, and is associated with colony collapse) (Ribiere et al., 2005; de Miranda and Genersch, 2010).

Several eukaryotic organisms can cause serious disease in honey bees. Fungal pathogens include *Ascosphaera apis*, the cause of the larval disease Chalkbrood (Bailey, 1981), and *Nosema apis* and *Nosema ceranae*, microsporidians that considerably shorten the lifespan of adult bees (Bailey, 1981; Forsgren and Fries, 2010; Fries, 2010). The most important metazoan parasite is the aforementioned *V. destructor*, though the tracheal mite *Acarapis woodi* can also have an impact on colony health, reducing a colony's chance of surviving the winter (Bailey, 1981).

There are only two important bacterial diseases of honey bees, both of which infect the larval stage of the insect. American foulbrood (AFB) is caused by the spore-forming *Paenibacillus larvae* (Genersch et al., 2006). AFB is a devastating disease for a honey bee colony, with clinical infections ultimately proving fatal without treatment (Hansen and Brødsgaard, 1999). In a larva infected with AFB, brood death usually occurs after the larval cell has been capped over (a step necessary for pupation) with these cappings often appearing darkened and perforated. As adult bees remove infected larvae from their cells, brood often end up in a scattered pattern across the comb. Dead larvae will decay into a viscous, brown mass, which dries to a hard scale that is difficult to remove from the cell (Hansen and Brødsgaard, 1999). AFB is found on every continent where bees are kept (Bailey, 1981), however infection is relatively rare; in Great Britain, in 2012, only 40 apiaries tested positive for AFB (NBU, 2012a). Because of the serious consequences of AFB infection, all colonies that do test positive in Britain are destroyed by burning (Waite, Brown, et al., 2003a).

The other bacterial disease of honey bees is European Foulbrood.

1.2. European Foulbrood

European foulbrood (EFB) has different characteristics to AFB. Symptomatic larvae first become displaced in their cells, then lose their definition, discolour, die and decompose. This sometimes forms a scale, but these are much easier to remove from the cell than those produced by AFB (Bailey, 1981). EFB preferentially kills younger, four to five day old larvae, and has a seasonality (a higher prevalence in early summer) that is lacking in AFB. This summer peak occurs when bee colony growth is at its most rapid, and fewer resources are available for individual larvae. EFB has therefore been referred to as a stress disease. As such, symptoms can spontaneously disappear when conditions improve, and infection is not necessarily fatal for a colony (Bailey, 1981).

EFB is also found on every continent where bees are managed. Prevalence has increased in recent years in some countries (such as the UK and Switzerland (Wilkins et al., 2007; Roetschi et al., 2008)), and others that were thought to be free of disease (e.g. Scotland and Norway) have suffered recent outbreaks. EFB is much more common in Britain than AFB, with over 200 infected apiaries in England, Scotland and Wales in 2012 (NBU, 2012a).

The economic cost of European Foulbrood infection is considerable, both to the individual beekeeper (through lost honey production and pollination contracts, and replacing infected bees and equipment) and to the government, through the costs of inspection and treatment. In the UK the cost to a beekeeper of a single EFB-infected colony ranges from £400 to £800. The cost to the government from inspections in England and Wales (which look for both EFB and AFB) appears more difficult to determine, but figures of up to £200 for each inspection requiring treatment and follow up, and around 700 treatments required per year (Defra, 2013) amount to approximately £140,000 per year. This doesn't include surveillance of hives which turn out to be EFB-negative, nor does it include the cost of antibiotics (circa £2,400 in 2011) (Defra, 2013). Aside from AFB, the other honey bee diseases endemic to the UK are not statutorily reportable, and therefore do not legally require inspection or treatment. Hence, while they may reduce colony productivity, their associated costs to government and the beekeeper will be smaller than EFB and AFB.

EFB is caused by the gram-positive, non-spore forming bacterium *Melissococcus plutonius* (Bailey, 1983). This bacterium was first identified as the cause of the disease in the early twentieth century (White, 1912), when it was named *Bacillus pluton*. However, this observation was controversial at the time, with some authors being unable to distinguish *B. pluton* from a lactic acid bacterium referred to at the time as *Streptococcus apis* (Tarr, 1935). It wasn't until the 1930s that *B. pluton* was shown by cultivation and experimental infection to fulfil Koch's postulates as the cause of EFB (Tarr, 1937). In the 1950s the species' name was changed to *Streptococcus pluton*, based on morphology and an inability to form spores (Bailey, 1956). Further work was done to characterise nutritional requirements and serology of the bacterium (Bailey, 1957; Bailey and Gibbs, 1962), before the species was given its own monospecific genus as *Melissococcus pluton* (Bailey and Collins, 1982), and its name later amended to *Melissococcus plutonius* (Truper and de' Clari, 1998).

There is still a lack of clarity about of the exact mechanism of pathogenicity of *M. plutonius*. It is generally thought that the bacteria compete with the host for nutrients in the larval midgut (Bailey, 1981), however it has been suggested that bacterial invasion of the peritrophic membrane (McKee et al., 2004) or secretion of toxins such as tyramine (Kanbar et al., 2004) may play a role.

M. plutonius can persist in larvae without showing clinical symptoms (McKee et al., 2003; Forsgren et al., 2005; Budge et al., 2010), and there is some evidence that secondary bacteria can have effects on the symptomology of EFB. Some bacteria are thought to exacerbate disease (Bailey, 1957), and it has proven easier to reproduce disease symptoms with directly inoculated disease material than with pure cultures of *M. plutonius* (McKee et al., 2004). Other bacteria, especially lactic acid bacteria such as *Lactobacillus kunkeei*, have been shown to suppress symptoms (Endo and Salminen, 2013; Vásquez et al., 2012). The mechanisms of inhibition are currently unknown, but related lactic acid bacteria in other systems can produce potent antimicrobial metabolites (Axelsson et al., 1989). At any rate, *M. plutonius* infections are often accompanied by other bacteria such as *Achromobacter eurydice*, *Paenibacillus alvei* and *Enterococcus faecalis* (Bailey, 1981).

In the UK, the standard treatments for EFB are the application of the antibiotic oxytetracycline (OTC) or shook swarm (the transfer of all adult bees in a colony to new combs in a clean brood box, and the scorching or destruction of the previous equipment). Experimentally a combination of shook swarm and OTC application resulted in less disease recurrence than OTC alone (Waite, Brown, et al., 2003b), and colonies that were subjected to shook swarm treatment without application of OTC had measurably less *M. plutonius* present in the spring following infection than OTC-treated colonies (Budge et al., 2010). Because a 16 to 18 week waiting period after treatment is required for OTC in honey to return to saleable levels (Thompson et al., 2006), shook swarm on its own is frequently used. Despite the fact that OTC has been widely used to treat EFB, there is no evidence that *M. plutonius* has acquired antibiotic resistance (Hornitzky and Smith, 1999; Waite, Jackson, et al., 2003). This is perhaps surprising, given the fact that OTC resistance has been observed in the other bacterial brood pathogen, *P. larvae* (Evans, 2003).

M. plutonius is most closely related to the Enterococci (Cai and Collins, 1994). Historically within *M. plutonius* little genetic variation has been seen, with only very minor differences in RFLP (Restriction Fragment Length Polymorphism) patterns (Djordjevic et al., 1999), culture characteristics and serology (Allen and Ball, 1993) being reported among isolates from diverse sources. This despite the pathogen being found in three different bee species, *A. mellifera*, *A. cerana* (Bailey, 1974) and *A. laboriosa* (Allen et al., 1990). Very recently, a subtype of *M. plutonius* was isolated in Japan. This was designated “atypical” *M. plutonius*, as its less stringent culture characteristics, higher virulence *in vitro* and more diverse PFGE (Pulsed Field Gel Electrophoresis) patterns distinguished it from the type strain and all previously identified “typical” *M. plutonius* (Arai et al., 2012; Takamatsu et al., 2013).

Much recent work on *M. plutonius* has focussed on developing molecular tools for its study. Many techniques for *M. plutonius* detection now exist, including a conventional PCR assay (Govan et al., 1998) and a hemi-nested PCR assay. The hemi-nested assay uses two reactions, the second reaction performed on the products of the first reaction using a third, internal primer, with the aim of increasing specificity (Djordjevic et al., 1998). An antibody-based lateral flow device (LFD) has been developed (Tomkies et al., 2009), which is now used as standard in the field in the UK to confirm infection as EFB. A limited study has shown some, possibly weaker reaction between the LFD and samples of atypical *M. plutonius* (Takamatsu et al., 2012). A gold nanoparticle assay has been developed that allows rapid, sensitive and visual confirmation of the presence of *M. plutonius* (Saleh et al., 2012). The most sensitive detection assays use real-time PCR, which also allows a quantification of the amount of target present (Roetschi et al., 2008; Budge et al., 2010). The number of whole genome sequences published for this bacterium has increased from zero to two over the last two years. The first sequence published was from isolate ATCC 35311 (Okumura et al., 2011). This is the type strain of *M. plutonius*, and the same as LMG 20360 (studied in this thesis), but from a different culture collection. The second genome sequence published was from isolate DAT561 (Okumura et al., 2012). ATCC 35311 is a member of the typical *M. plutonius* group, and DAT561 is an example of atypical *M. plutonius*.

1.3. Molecular Epidemiology

Molecular epidemiology has been defined many times, but the general feature of the discipline is the use of molecular techniques to study the distribution and determinants of disease occurrence in populations. There are two broad areas of epidemiology to which molecular techniques have been extensively applied; pathogen identification and strain typing (Foxman and Riley, 2001). These objectives are similar, strain typing being a more discriminative form of pathogen identification, but they have a difference in emphasis. Identification is used to confirm the presence of pathogen, or aid disease diagnosis, which can be especially important in a notifiable disease such as EFB. Strain typing is used to identify links between cases or outbreaks of a specific disease, to elucidate transmission routes or mechanisms.

1.3.1. Pathogen Detection

Traditionally pathogen identification involved culture characteristics and substrate utilization patterns, as until recently molecular techniques were considered too cumbersome for widespread use (Sullivan and Coleman, 1998). Over the last decade however, molecular techniques have

become the norm for pathogen identification. Antibody-based techniques are commonly used to detect pathogens, for example in Enzyme-Linked Immunosorbent Assays (ELISA) (Crowther, 1995). The commonly used sandwich-ELISA protocol involves immobilisation of an antibody (specific to an antigen from the target bacterium) in a well. Antigens are added and adsorb to the antibodies. Further antibodies are added, this time labelled with an enzyme, and these adsorb to the immobilised antigens. Finally the enyzme's substrate is added and a colour change observed in the vicinity of the now immobilized antigens (Lazcka et al., 2007). The aforementioned LFDs work on similar principles. In a sandwich lateral flow immunoassay, a suspension of the substrate to be analysed (for example, bacterial cells covered in target antigens) moves by capillary action along a membrane strip. On moving along the strip, the suspension picks up target-specific antibodies that are labelled with nanoparticles (often made out of latex). A control line of immobilised antispecies immunoglobulin binds to the antibodies on any free nanoparticles, and the concentration of particles gives a visual signal (in the form of a dark line) that the technology is working. If the substrate contains the bacterium of interest, a test line of antigen-specific antibodies binds to the target antigen, which in turn is bound to the antibody-nanoparticle complex. This forms another visual signal, indicating that the test is positive (Posthuma-Trumpie et al., 2009).

Many detection techniques are based on nucleotide sequences. Some involve sequencing genes which have species-specific nucleotide sequence. Often the bacterial 16S rRNA gene is sequenced, though issues such as lack of knowledge of interspecies and even intragenomic sequence divergence rates can make this problematic (Janda and Abbott, 2007). PCR-based techniques look for the presence of an amplification product using primers specific to a sequence only found in the pathogen of interest (Govan et al., 1998). A more modern variant of the PCR-based approach is quantitative PCR (qPCR), which can measure the relative abundance of product generated in the reaction, and therefore make inferences about infection levels (Roetschi et al., 2008; Budge et al., 2010). Detection technologies continue to develop increased rapidity and ease of use. For example, LAMP (loop-mediated isothermal amplification) assays amplify target DNA without the need for a thermal cycling stage (Mori and Notomi, 2009). As such they are rapid and can be used in the field, and are consequently being developed for a range of different pathogens (Hara-Kudo et al., 2005; Hodgetts et al., 2011; Boonham et al., 2013).

1.3.2. Strain Typing

Strain typing of bacteria is important for understanding the spread and origins of infectious disease epidemics. If clusters of disease are caused by the same pathogen variant, a link can be inferred, and similarly a novel variant being detected in a region can suggest a new import of

disease (e.g. Gopal et al., 2006). Many techniques exist for distinguishing between different variants of a pathogen, and new techniques are being developed all the time. Therefore what follows is only a limited account of some of the more historically important typing systems, with the emphasis on bacterial pathogens.

Early work focussed on serotyping bacteria, looking for an agglutination reaction between an antibody and specific antigens present on the bacterial cell surface. This technique has been used for many decades (Lancefield, 1933) and several important bacterial pathogens are still divided into serotypes (e.g. *Salmonella enterica*), but isolates are now identified to serotype level based on nucleotide differences rather than agglutination reactions. Protein-based information traditionally used for strain typing included differences in the electrophoretic migration rate of extracted enzymes (Goullet, 1977), while modern mass spectrometry techniques allow easier, rapid differentiation of subspecies (Barbuddhe et al., 2008).

Of many techniques that study nucleotide variation some, such as RFLP, involve enzymatic digestion of bacterial DNA and visualisation of the resulting fragments by gel electrophoresis. Because the restriction enzymes used to cut DNA do so at sequence-specific sites, different sized DNA fragments are produced for different bacterial variants. Since larger DNA fragments migrate more slowly through a gel when an electric current is applied, different bacterial variants will produce different banding patterns when these gels are visualised (Olsen et al., 1994). A variant of this technique, PFGE, runs digested DNA through a gel in an electric current that periodically switches direction (Schwartz and Cantor, 1984). This allows longer DNA fragments to be distinguished from one another and gives great discrimination, and has become a standard technique for investigating outbreaks of foodborne disease (http://www.cdc.gov/pulsenet/).

A suite of strain typing methods exploit PCR, the amplification aspect of which allows the analysis of tiny amounts of DNA, or the DNA from non-cultivable organisms. The targets of these PCRs can be strain-specific (Blum-Oehler et al., 2003), so any amplification product indicates presence of the strain, or PCRs can have random targets (as in RAPD – Randomly Amplified Polymorphic DNA) such that relationships between isolates are assigned based on matching patterns of amplified fragment length (Genersch et al., 2006). A technique that uses both PCR and DNA sequencing is Multi Locus Sequence Typing (MLST) (Maiden et al., 1998), where fragments of housekeeping genes are amplified with PCR and then sequenced. This is a widely-used typing method for bacterial (Bygraves et al., 1999; Suerbaum et al., 2001; Lowder et al., 2009), and sometimes fungal (Litvintseva et al., 2006), pathogens of humans and other animals, and is easy to use, portable (information is easily exchanged between laboratories), discriminative and allows an assessment of recombination (Urwin and Maiden, 2003). In recent years sequencing technology has advanced

so far that multi-genome-scale investigations are being undertaken in real time during outbreaks. In these cases transmission between individuals, or shared origin of infection, is inferred by inspection of the number of nucleotide substitutions between pathogen isolates across their entire genomes (Harris et al., 2010; Grad et al., 2012; Biek et al., 2012; Allard et al., 2013). MLST still has the advantages of ease of use, affordability and no requirement for specialist bioinformatics support. However, the expansion of cheaper high throughput sequencing technologies and automated pipelines for analysis suggests that, in the future, whole-genome sequencing of pathogens during outbreaks will become routine.

1.4. Bacterial Genomics

Some of the first free living organisms to have their genomes sequenced were pathogenic bacteria (Fleischmann et al., 1995; Fraser et al., 1995). These organisms presented the advantage of having relatively short genomes (1.8 megabases in the case of *Haemophilus influenza*, 580 kilobases for *Mycoplasma genitalium*), which allows sequencing using traditional Sanger-sequencing methods (Sanger et al., 1977). Now, the generation of genome-scale sequence has been made commonplace by the advancement of so called Next Generation Sequencing (NGS) techniques.

1.4.1. Next Generation Sequencing

Several different technologies exist for large-scale nucleotide sequencing. Each individual sequencing reaction on one of these machines produces a single string of DNA letters, known as a read. The length and number of different reads produced in each run of a machine depends on the technology used. Once reads are produced, there are several ways of analysing them. One option is to assemble reads into longer contiguous regions of sequence, known as contigs. Assembly programs look for overlapping sequences between the end of one read and the start of another. The length of contig these programs will be able to produce depends on a number of factors, including the length of the individual reads (longer reads are easier to assemble, as they are more likely to have overlapping regions) and the depth of coverage (the number of reads produced, proportional to the length of the genome being sequenced, with greater sequencing depth giving longer contigs). The *de novo* assembly approach is necessary if you are working with non-model organisms, or organisms which have not had their genomes sequenced previously. It can also help identify novel mobile genetic elements.

Another approach is to map sequence reads to an already sequenced genome. This is plausible for species for which a representative has already had a genome sequence published or produced. In

this mapping or “resequencing” approach individual reads are aligned to the corresponding region of the reference genome sequence, with a tolerance for differences caused by either sequencing error or genuine, biological polymorphisms between samples. The advantages of resequencing are that it is faster than *de novo* assembly and doesn’t require such high coverage (so more samples can be sequenced in the same run). It also works well with shorter read technologies. However mapping reads to a reference sequence will not show the presence of mobile genetic elements that are not present in the reference genome.

The different technologies each operate with their own chemistries, and therefore produce reads of different length, quality and quantity. The first NGS machine was the 454 pyrosequencer, now manufactured by Roche (Glenn, 2011). DNA samples for pyrosequencing are first fragmented (to make fragments which are the appropriate length for sequencing) and then ligated to adaptors. The fragments are then attached to DNA capture beads, at concentrations such that each bead carries a unique, single-stranded fragment. Beads are then emulsified in a water-oil mixture to trap individual beads within a micro drop of water. These then act as amplification microreactors for emulsion PCR, whereby each bead becomes covered in millions of clonal copies of the original fragment. For the sequencing reaction beads are loaded on to a picotitre plate, which is covered with wells sized to accommodate a single bead. For sequencing, individual nucleotides are washed sequentially across the wells. The incorporation of a base releases inorganic phosphate, which is combined with APS (3’-Phosphoadenosine-5’-phosphosulphate) by the enzyme sulfurylase to form ATP. This ATP then reacts with luciferin (catalysed by luciferase) to release light. It is this light, detected by a high resolution camera pointed at the plate, that records the incorporation of a particular base, which is how the sequence of the fragments on each bead is worked out (Roche, 2013). 454 produces relatively long read lengths (500-800 bases), but has a propensity for errors when long sequences of the same base are present. This is because the 454 detects the incorporation of multiple bases as an increased light emission. When large numbers of the base are present, it becomes difficult to distinguish the precise amount of light emitted, and therefore the number of bases. The Ion Torrent ™ machine (Life Technologies™) has very similar initial steps to the 454. However, instead of detecting light, it detects the H^+ ion flux (as a change in pH) caused by the incorporation of a nucleotide (Life Technologies, 2013). Because it doesn’t detect light, it doesn’t require expensive optical equipment, and has a cheaper initial outlay than many other machines (Glenn, 2011). However it also suffers from the same problems of multiple base calling as the 454.

The Illumina® MiSeq® is a technology that has recently been embraced by the Food and Drug Administration (FDA) for the routine genome sequencing of foodborne pathogens (Illumina, 2012). This chemistry of the MiSeq® is different from the 454, though it is another sequencing-by-

synthesis method. DNA is fragmented, and adapter oligonucleotides are ligated to both ends of the fragment. Fragments are then size selected and purified. The next stage takes place on the Flow Cell, where the sequencing itself will occur. The Flow Cell is covered in a lawn of oligonucleotides, which bind to the oligonucleotide adaptors on the DNA fragments. Both ends of the fragment are the covalently bound to the Flow Cell surface, and each fragment is amplified (in a process known as Bridge PCR) to create hundreds of millions of unique clusters of DNA fragments. The reverse strand is then cleaved and washed off, and primers are attached. The actual sequencing then takes place with labelled, reversibly-terminated nucleotides. To sequence, all four nucleotides are washed over the Flow Cell and compete to bind to the fragments, but because of a blocking group attached to each nucleotide only one nucleotide binds to each fragment during each washing cycle. The Flow Cell is washed clean of free nucleotides and the clusters are then excited by a laser, with the colour emitted by the labelled base indicating which nucleotide has been incorporated. The label and blocking group are removed (the reversible-termination stage) and the cycle repeats (Illumina, 2013a). The MiSeq® is capable of generating many more reads per run than either the 454 or the Ion Torrent machines (and in the case of the related HiSeq®, many orders of magnitude more data) but the main disadvantage is the time it takes to run: around a day in the case of the MiSeq®, up to eight days for the HiSeq® (Glenn, 2011), though this is compensated for by a shorter sample preparation time.

Another, interesting technology is the PacBio RSII. This is exciting because it sequences a single molecule of DNA, rather than a cluster of identical fragments as in the 454, Ion Torrent and MiSeq®, and it also produces the longest reads of any current NGS technology. This makes it an useful proposition for generating genome sequences *de novo* (Glenn, 2011). The so-called SMRT cell contains thousands of tiny pits, the bottom 20-30 nanometres of which are illuminated from underneath the cell. A DNA template-polymerase complex is immobilised at the bottom of the cell, and nucleotides (each labelled with a differently coloured fluorophore) are introduced to the cell. When a nucleotide is held in the polymerase the fluorophore is excited by the light, and the colour detected. The fluorophore is then cleaved off by the process of nucleotide incorporation, and the next nucleotide can move into the polymerase (Pacific Biosciences, 2013a). The main advantage, as mentioned, is the read length (average read length over 8kb, up to a maximum of over 30kb (Pacific Biosciences, 2013b)), but the error rate is very high; up to 16% compared to the 0.1% of the MiSeq® (Glenn, 2011).

In addition to all these technologies for producing single reads, many of them can be adapted to produce paired-end reads. Producing paired-end reads involves sequencing both ends of a DNA fragment on a next generation sequencer. The fragment length is known, and assembly algorithms can use this information to precisely align reads, and overcome difficult to assemble

regions such as repeats (Illumina 2013b). A specific type of paired-end read sequencing, mate pair sequencing, can sequence the ends of very long inserts of DNA. In mate pair sequencing DNA is fragmented and size selected, up to several kilobases long. The ends of the fragment are repaired and labelled, and then the fragment is circularized. Circularized DNA is then fragmented and the labelled DNA purified and paired-end sequenced. This technique is very useful for genome assembly (Illumina 2013c).

What becomes clear is that several technologies exist for sequencing the whole genomes of microorganisms, each with inherent advantages and disadvantages. Before any project begins, a worthwhile exercise would be to consider the appropriate sequencing technology for your needs. For example for *de novo* genome assembly one might choose the 454 or PacBio machines, to take advantage of their longer read length. To resequence one or multiple genomes a shorter read technology is sufficient, which could either have a cheaper initial outlay (e.g. Ion Torrent™) or produce larger numbers of sequences, to sequence more isolates or bigger genomes (e.g. MiSeq® or HiSeq®).

1.4.2. Horizontal Gene Transfer

Comparative genomics is well suited to studying the phenomenon of horizontal gene transfer. Horizontal gene transfer is the transmission of genetic material between organisms in the environment, rather than vertical transmission from parent to daughter cell. This phenomenon allows bacterial strains to acquire important genes from their other lineages, which can enable the exploitation of novel niches. For example, antibiotic resistance genes are frequently acquired via horizontal gene transfer (Butaye et al., 2003; Nandi et al., 2004). In the pre-genomics era horizontal gene transfer could be identified by incongruence between phylogenies of different genes within an individual genome. Many statistical tests exist to test incongruence, to examine if incongruence is due to incorrect trees (with the genes actually sharing the same evolutionary history) or whether to the trees are correct, and genes therefore have different evolutionary histories. Character congruence analyses are particularly useful for this, as they take both tree topology and underlying support for the topology into account (Planet, 2006).The principle of testing incongruence is easily applied in MLST studies, where multiple housekeeping genes are sequenced (Maiden et al., 1998; Enright and Spratt, 1999). In the genomic era, the number of multi-gene comparisons that can be made is rapidly increasing. These can often best be displayed in a phylogenetic network rather than in traditional phylogenetic trees. These networks display taxa as nodes and evolutionary relationships as edges, and are useful for showing relationships where genetic material is transferred between taxa (Huson and Bryant, 2006). Horizontal gene transfer can also be inferred from genomes by examining them for anomalous nucleotide

statistics. Anomalous gene regions can be characterised by, for example, GC content of the sequence, codon usage within genes and amino acid usage within proteins that are different from the surrounding genome (Karlin, 2001). These statistics indicate that the region comes from a source other than the host genome.

NGS also allows the identification of mobile genetic elements (such as bacteriophages, plasmids and pathogenicity islands) within a genome. Such islands can be recognised by nucleotide statistics different from the host genome, by the presence of open reading frames with sequence similarity to bacterial integrases, and insertion at or near the sites of short repeats and/or tRNA genes (Ochman et al., 2000). By comparing these elements to previously identified sequences the donor species can be inferred. Whole genome sequencing like this has shown transfer of genetic material between bacterial species (Allard et al., 2013), between bacteria and archaea (Nelson et al., 1999) and between bacteria and eukaryotes (Dunning Hotopp et al., 2007).

1.4.3. Inferring Gene Function

The gold standard for elucidating the function of a particular gene is through a phenotypic assay. This may involve the deletion or inactivation of genes, followed by phenotypic assays to check for substrate utilization etc. Some such assays have even been performed on a genome scale (Winzeler et al., 1999). However, now that so much genomic data is being generated it is not feasible to perform these experiments in every study. Bioinformatic (DNA sequence analysis) approaches must be used instead. The most straightforward of these involve comparison of sequence to a database of sequences of known function, for example using the BLAST® (Basic Local Alignment Search Tool) algorithm (Altschul et al., 1990) against a database such as GenBank® (Benson et al., 2012) or the UniProt databases for proteins (Bairoch et al., 2005). The results of such comparisons must be treated with a degree of caution – for example, gene duplication events can lead to genes with similar sequences diverging to quite different function (Hurles, 2004). The orthologous nature of genes (where genes in two species originate by vertical descent from a common ancestor) can be established by methods such as reciprocal best BLAST hits, if genes are each other's best BLAST hits in two different genomes (Moreno-Hagelsieb and Latimer, 2008). Other tools, such as InterProScan, look for the presence of certain functional groups such as signal peptides or transmembrane domains within a gene, allowing properties of the encoded protein to be inferred (Zdobnov and Apweiler, 2001).

Automated gene function assignment has the drawback of relying on matching patterns of sequence to those that have already been identified. By definition, novel genes unrelated to known genes cannot have their functions assigned in this way. Aside from gene knockouts and

phenotypic assays, methods other than sequence matching do exist for inferring the function of unknown genes. These include inferences based on the genomic context of the gene, such as whether genes are fused to others of known function, co-occur in operons with them, or co-occur with them among genomes (Huynen et al., 2000).

At the genome scale many pipelines exist to automate and speed up the process of gene function prediction, including xBASE (Chaudhuri et al., 2008), RAST (Aziz et al., 2008), KAAS (Moriya et al., 2007) and MaGe (Vallenet et al., 2006). Comparative genomics can also help unravel the genetic basis for differences in phenotype. For example, typical *M. plutonius* requires media supplemented with potassium in order to grow in culture, whereas atypical *M. plutonius* does not. Comparison of genomes of each variant suggested that the loss of functional *napA* and *ctaP* genes in typical *M. plutonius* might underlie this requirement, an observation that was confirmed with complementation analysis (Takamatsu et al., 2013).

1.5. This Thesis

This thesis will study the biology and epidemiology of *M. plutonius* at the broadest possible range of scales, from the sub-cellular to the global. Several studies have used molecular techniques to identify *M. plutonius*, usually for disease confirmation, or to study the prevalence of the bacterium in different life stages of the honey bee. However, only limited work has been done attempting to distinguish variants within *M. plutonius*, with most investigations concluding that the bacterium is highly homogenous. Amongst other things, the study presented here appears to be the first to apply robust and easily repeatable molecular strain typing techniques to such a broad samples of isolates.

In the second chapter of this thesis the genome of *M. plutonius* is analysed. Whole genome sequences are generated and compared among different isolates of *M. plutonius*, and compared with the published genomes. In the third chapter the genome sequence is used to design the first typing scheme for *M. plutonius*. We examine the global diversity of the pathogen, and use the scheme in epidemiological studies to confirm the transmission of the bacterium within and between countries. The fourth chapter comprises an in-depth study of the diversity and spread of *M. plutonius* around England and Wales, and a broader study of the bacterium's diversity across several European countries. Finally the fifth chapter looks at the spread of *M. plutonius* at an apiary level, examining different reservoirs and transmission routes for the pathogen. Novel insights about the epidemiology of European Foulbrood are described at all these scales, and it is hoped that these can be used to design interventions or management changes to reduce the burden of this disease.

2. The Genome Sequence of *Melissococcus plutonius*, and the Comparative Analysis of Multiple Genomes

2.1. Introduction

2.1.1. The Genome

The genetic information an organism requires to perform all the functions necessary for survival and reproduction is stored on its genome (Yadav, 2007), which in the vast majority of organisms is encoded in the biomolecule DNA (Deoxyribonucleic Acid) (Watson and Crick, 1953). In the domain Bacteria, genome sequences can be partitioned between different elements of DNA that replicate independently of one another, known as replicons. The largest single replicon in a bacterial cell is the chromosome, and this carries the so called core, or essential genes (Krawiec and Riley, 1990). Other genetic material can be stored on plasmids, smaller circular DNA molecules with different mechanisms of replication (del Solar et al., 1998). These carry the accessory genome, the genes that may encode nonessential products that offer a selective advantage in certain situations, such as antibiotic resistance genes. Plasmids can exist at higher copy number per cell than the single chromosome, and multiple different plasmids can be maintained within the same bacterium (e.g. Young et al., 2006). Plasmids can often move between different chromosomal lineages of bacteria by horizontal gene transfer (Frost et al., 2005). A more recently identified replicon, the chromid, can carry core genes and has similar nucleotide composition to the chromosome but has the replication and partitioning systems of plasmids. These were previously known as second chromosomes, or megaplasmids (Harrison et al., 2010).

The exchange of plasmids between bacteria (known as conjugation) is not the only mechanism of horizontal gene transfer. DNA can be taken up directly from the environment (transformation), or moved between cells by a bacteriophage (transduction) (Thomas and Nielsen, 2005). Lysogenic bacteriophages are those that are able to either incorporate themselves into the host genome (becoming a prophage), or to form independently replicating replicons within the cytosol. These phages can then excise themselves from the host genome, usually after exposure to a stressor such as UV light, and lyse the host cell (Lwoff, 1953). Occasionally, the genes necessary for excision become deactivated, and the phage remains immobilised within the host genome.

2.1.2. Genomics

The study of genomes (genomics) has become possible, and indeed widespread, due to the rapid recent increase in sequencing technologies; so called next generation sequencing (NGS) (Mardis, 2008). The first complete bacterial genome sequence, *Haemophilus influenza*, (Fleischmann et al., 1995), was undertaken using conventional Sanger sequencing, a technology which can produce a maximum of 0.06 Mb (Megabases) of data per run. The first commercially available next-generation sequencing technology was 454 pyrosequencing, with the Roche 454 FLX+ now capable of producing 900 Mb of data in a 20 hour run. The Illumina HiSeq 2000, one of the highest throughput sequencers currently available, can produce up to 600,000 Mb of sequence over the course of a ten day run (Glenn, 2011). This phenomenal increase has revolutionised the field of genomics, especially for microbial species. Around nine hundred bacterial and archaeal genomes were submitted to Genbank between 2003 and 2009 (Benson et al., 2004, 2010), while over three and a half thousand species have had genome sequences submitted since then, with many of these species having multiple strains represented (http://www.ncbi.nlm.nih.gov/genome/browse/). These advances are leading to the routine use of genomics in microbiology, as seen in the 100k Foodborne Pathogen Genome Project, which aims to sequence the genomes of 100,000 strains of important foodborne pathogens such as *Salmonella* and *Campylobacter* (http://100kgenome.vetmed.ucdavis.edu/).

Sequencing the entire genomes of related bacteria allows horizontal gene transfer to be observed. This can be inferred from observations of phylogenetic networks (Huson and Bryant, 2006), and by comparing fully assembled genomes to identify replicons present in some strains but absent in others. This has shown, for example, the transfer of plasmids from an *Escherichia coli* 0157 strain to a *Salmonella enterica* enterica strain (Allard et al., 2013). These technologies have even been used during the course of individual outbreaks, for example showing that a large outbreak of virulent *E. coli* O104:H4 in Germany in 2011 was caused by a novel variant containing genes from both enteroaggregative *E. coli* and enterohemorrhagic *E. coli* (Mellmann et al., 2011). Genome sequencing can also be used to directly investigate the epidemiology of disease outbreaks, either by informing the design of more traditional typing schemes (Haynes et al., 2013), or by direct analysis of SNPs from whole genome sequences (Grad et al., 2012; Köser et al., 2012; Allard et al., 2013). This allows discrimination between previously indistinguishable isolates, and enables the investigation of epidemiological processes.

Before this project was initiated, very little was known about the genomics of *M. plutonius*. Melting point analysis had shown the genome's GC content to be 29-30% (Bailey and Collins, 1982), and some ribosomal RNA gene sequence had been produced (Cai and Collins, 1994;

Djordjevic et al., 1998; Behr et al., 2000), as well as a single sequence of a superoxide dismutase gene (Roetschi et al., 2008). An RFLP (Restriction Fragment Length Polymorphism) study had shown isolates of *M. plutonius* to be genetically very similar (Djordjevic et al., 1999), but much of the earlier of the work on *M. plutonius* had concentrated on traditional biochemical assays (Bailey and Gibbs, 1962; Bailey, 1984; Allen and Ball, 1993).There was much scope, therefore, to advance our knowledge of this pathogen's genome.

In this project a whole genome sequence for the Type Strain of *M. plutonius* was generated, and compared to two genome sequences released during the course of the project (Okumura et al., 2011, 2012). From this, conclusions could be drawn about the accuracy of the assembly of the published genomes. We also compared genome sequences from multiple isolates taken from the field; firstly to attempt to use whole genome sequences to elucidate the evolutionary history of a traditionally homogenous bacterium, and secondly to identify and characterise any mobile genetic elements present.

2.2. Methods

2.2.1. Isolates Sequenced

Three separate runs of next generation sequencing were performed. The first was on the Type Strain (LMG 20360). This comes from the same original isolation as ATCC 35311, the first published *M. plutonius* genome, but is from a different culture collection. The second sequencing run was on 11 field-sampled isolates of *M. plutonius* isolated in 2010 (isolates 7512, 7526, 7533, 7534, 7540, 7595, 7605, 7606, 7610, 7611 and 7613). The third sequencing run was on a single isolate of *M. plutonius*, also isolated in 2010 (isolate 7596). This isolate is part of the atypical *M. plutonius* group (see Chapter Three), as is the source of the second published *M. plutonius* genome, isolate DAT561. All isolates came from southern and eastern England (Table 2.1).

2.2.2. Sequencing

Isolates were streaked and subcultured anaerobically on M110 media (Forsgren et al., 2013), and confirmed as *M. plutonius* using an *M. plutonius*-specific Taqman® assay (Budge et al., 2010). Whole genomic DNA was extracted using the QIAGEN DNeasy ® Blood & Tissue kit, including an additional enzymatic lysis and proteinase K incubation step, recommended by the kit for gram-positive bacteria. For the first run, DNA was sequenced on 2*1/8 of a Titanium plate of a Roche 454 FLX pyrosequencer, and reads were assembled into contigs using Roche's Newbler software. For the second sequencing run DNA was sequenced on a Roche 454 FLX pyrosequencer, using 11 MID tags (out of 12 used in the run, with one isolate sequenced not being from this project) split across a whole Titanium plate. Reads were again assembled into contigs using Roche's Newbler

Isolate	Geographical Origin
LMG 20360	England (no finer detail available)
7512	Greater London
7526	Greater London
7533	Greater London
7534	Norfolk
7540	Norfolk
7595	Greater London
7596	Oxfordshire
7605	Essex
7606	Essex
7610	Greater London
7611	Suffolk
7613	Suffolk

Table 2.1. Geographical origins of the whole-genome sequenced isolates.

software. For the third sequencing run, isolate 7596 was sequenced on half a chip of an Ion Torrent ™ PGM sequencer (Life Technologies ™).

2.2.3. Generating the Type Strain Sequence

2.2.3.1. Alignment, Visualisation and Rearrangement of Sequences

Contigs generated in the first sequencing run (for the Type Strain, LMG 20360) were aligned to a draft LMG 20360 scaffold (provided by the Sanger Institute at the start of the project) using the megablast algorithm on the NCBI BLAST server (http://blast.ncbi.nlm.nih.gov/Blast.cgi), and visualised in the Artemis sequence visualisation and annotation tool (Rutherford et al., 2000). As further genome sequences became available, for the Type Strain (ATCC 35311, from another culture collection (Okumura et al., 2011)) and a field-isolated strain (DAT561 (Okumura et al., 2012)), our genome sequence for LMG 20360 was aligned to them using WebACT (Abbott et al., 2005) and visualised in the Artemis Comparison Tool (ACT) (Carver et al., 2005). Arrangement of these genomes was also examined using CGview (Grant and Stothard, 2008). After each of these alignments, contigs from our draft genome were rearranged, and gaps closed, based on their positions relative to the new sequence information. This rearrangement, and incorporation of new sequence, was performed by manually editing the genome FASTA file.

2.2.3.2. Gap-closing PCR

Conventional PCRs were performed to try and close 18 gaps identified in our *M. plutonius* genome sequence. Primers were designed for flanking regions (Rozen and Skaletsky, 1999), and are given in Table 2.2. Briefly, 2 µl of DNA sample was added to 12.5 µl of Fermentas 2 x PCR Master Mix (Massachusetts, USA) with 10 pmols of each primer and made up to a final volume of 25 µl. Each reaction was run on an Applied Biosystems 2720 Thermal Cycler (New York, USA) at 94°C for 2 minutes followed by 30 cycles of 94°C for 30 seconds, 59°C for 30 seconds, and 72°C for 30 seconds with a final elongation step at 72°C for 7 minutes. PCR products were mixed 4:5 with 10x Orange G loading dye, and run on a 2% agarose gel at 120V, with Bioline Hyperladder IV (100b). Products that were successfully amplified were purified using QIAquick PCR Purification kits, unless multiple bands were observed. In those instances, bands that were of the predicted size (based on the published ATCC 35311 genome) were gel extracted using the MO-BIO UltraClean® Gelspin® DNA Extraction Kit. All PCR products were eluted in 30µl of the relevant extraction buffer to increase final DNA concentration. Amplicons were then sequenced with the forward primer (and, if the fragments were large, with the reverse primers as well) on an ABI 3730xl 96-capillary DNA Analyzer (MWG Eurofins).

Primer	Primer Sequence
Gap1 L	TGT TGG TCG TGC TTT AGA AGA
Gap1 R	TTC AAC CTT TCA CCT CCC TA
Gap2 L	GAT GCC CCT TCT ATT CGT GA
Gap2 R	AAG AAG CGT TCG ACT TGC AT
Gap3 L	ATG GTT GGA GCG GAC TGA TA
Gap3 R	AAT AGA TGG GCC ATT TCT GC
Gap4 L	AAC AGG CTG CCG ATT TTA TG
Gap4 R	AAT AGC AGT TCG GCC AAT TC
Gap5 L	ACT GGG AAA ATC CAT CCA AA
Gap5 R	TAA GTG CAC GAC CAA TTC CA
Gap6 L	GCG TCG GTG CTA GTG ATA TG
Gap6 R	TTA TTG TCG GCA TTC CCA TT
Gap7 L	CGC TGG CAT TTA CAA AAT GAT
Gap7 R	TTG CAA TTG GCA CAG AGA AC
Gap8 L	TTG ATC TCA CGA CCA AAT CG
Gap8 R	AGC GAT GAG CCT AGT GAA CC
Gap9 L	AGT CAG GGT GGA ACT GCC TA
Gap9 R	CAG CAG ATG GTG AAA TGA ACC
Gap10 L	CGT CCA GCT ACT CGA CCA AT
Gap10 R	CAA ACG CTG CAG AAT AAG CA
Gap11 L	TTG GTT GCC TTC TCA TTT CC
Gap11 R	GCT TCA GGC GCT AGT GGT AG
Gap12 L	ATT TAC AGT CCG TCG CGT TT
Gap12 R	TTC CCA TGT CGA ACA CAG AA
Gap13 L	CGC TCG GTT GTC AAC TCT TT
Gap13 R	AAT AAG GGA GGA AAT GGC CTA
Gap14 L	CAA AAG TTG TTT TCT TAT CTT AAT CG
Gap14 R	GGC TAT CAA CAA GAA GGC AAA
Gap15 L	CTC GTA ACG CGT AGG TCA CA
Gap15 R	TCA AGG CGT TAC AAG AAG CA
Gap16 L	TGC AAA TGT CAA AAT CAA AGA TG
Gap16 R	CGA ATT TTT GGA ATG GAC AAA
Gap17 L	CCC CTA AAA AGA TCG GGA AG
Gap17 R	AGT GTG AAG CCC ACC TCA AG
Gap18 L	TTT TTC AAC CTC GCT TAC CG
Gap18 R	GAG CAT TCT TAC CGT CAT AGG C

Table 2.2 Primer sequences (5’ to 3’ direction) of chromosome gap PCRs.

2.2.3.3. Long Range PCR

Rather than implement another NGS run on the same isolate, using mate pair sequencing to investigate large-scale chromosomal re-arrangements, another PCR-based strategy was decided upon. Long range PCR was performed with the Roche Expand Long Template PCR System, using conditions set out for System 3 (for difficult amplicons). Primers for long range PCR are given in (Table 2.3). Briefly, 24 µl of DNA sample (measured at 20.79 ng μl^{-1} with a NanoDrop ND-1000 (Thermo Scientific, Delaware, USA)) was added to 5 µl of 10x PCR Buffer 3 with $MgCl_2$ (Roche, Mannheim, Germany) with 15 pmols of each primer, 25 pmols of each dNTP, 0.75 µl of Expand Long Template Enzyme Mix (Roche, Mannheim, Germany) and made up to a final volume of 50 µl. Each reaction was run on an Applied Biosystems 2720 Thermal Cycler (New York, USA) at 92°C for 2 minutes followed by 10 cycles of 92°C for 10 seconds, 45°C for 30 seconds, and 68°C for 8 minutes, with a further 25 cycles of 92°C for 10 seconds, 45°C for 30 seconds, and 68°C for 8 minutes (+ 20 seconds for each successive cycle),with a final elongation step at 68°C for 7 minutes. To visualise, 5µl of PCR product were mixed with 1µl of Fermentas 6x Loading Dye, and run on a 2% agarose gel at 100 volts, in 1x TBE buffer, with Bioline Hyperladder I (1kb). Products were run for approximately two hours, and then visualised with a UV transilluminator.

When it was confirmed that the PCR had produced amplicons of the expected size (circa 6kb), 30µl of PCR product was run again on a 1.5% gel, and extracted using the QIAquick Gel Extraction Kit (including a kit-recommended optional cleaning stage of an additional spin of QG buffer). Products were eluted in 30µl EB, to increase DNA concentration. Amplicons were then sequenced with the forward and reverse primers, on an ABI 3730xl 96-capillary DNA Analyzer (Eurofins MWG).

Primer Name	Primer Sequence
1F	CCG CCA ACC GGT GGA TTA G
1R	CAG CTG AGC TAC ACC GCG AAG
2F	TTT TTC GAT GAT GGG TCA AAC A
2R	ACT GCC ACA TTT CCC TAC AAA GAG T
3F	GCG TAA TAA ATC AAG GAA CGG TTT TTC
3R	TTC GGC CTC TCG GAC AAC TC
4F	GGG CCA TTT CTG CTC GCC TA
4R	GCA TTC TTA CCG TCA TAG GCT TTC A

Table 2.3. Primer sequences (5’ to 3’ direction) for long range PCR.

2.2.3.4. Annotation of Genome Sequence

The rearranged genome sequences were automatically annotated using the RAST (Rapid Annotation using Subsystem Technology) server (Aziz et al., 2008).

2.2.4. Phylogenetic Network

2.2.4.1. Mapping Reads to the Reference Genome

Reads from the second and third runs of sequencing (isolates 7512, 7526, 7533, 7534, 7540, 7595, 7605, 7606, 7610, 7611, 7613 and 7596) were mapped to our final rearrangement of the LMG 20360 chromosome, and the ATCC 35311 plasmid sequences using segemehl version 0.1.6 (Hoffmann et al., 2009). Samtools version 0.1.19 (Li et al., 2009) was then used to convert each SAM file produced by segemehl into a BAM file, and to then sort and index that BAM file. The BAM files produced could then be examined using Tablet 1.13.05.17 (Milne et al., 2010).

The mpileup command of samtools was then used to create a VCF (variant call format) file from the indexed BAM file. This VCF file is a list of all the positions that varied between the reference genome and the reads. VCFtools version 0.1.10 (Danecek et al., 2011) was then used to remove all indels from the VCF file and leave only SNPs. This step was performed as the primary errors in both 454 and Ion Torrent sequencing are indels (Glenn, 2011). The resulting, modified VCF file was then compressed and indexed with the tabix package of samtools, and could then be viewed in Artemis over the reference genome and accompanying annotation.

This process was repeated for isolate 7596, to map its reads to the atypical DAT561 chromosome and plasmid sequences.

2.2.4.2. Chromosome SNP Phylogeny

VCF files from the mappings of all isolates were visualised over the LMG 20360 chromosome sequence, with the annotation file produced by RAST. SNPs were filtered to remove any with a quality score <50, and to remove any that contained multiple alleles. This gave the most conservative measure of divergence between isolates. Sequences from some isolates were too contaminated with non-*M. plutonius* sequence and did not contain enough *M. plutonius* sequence for further analysis (see section 2.3.6). For each of the other isolates, and the reference LMG 20360, sequences for every coding sequence (CDS) in the annotation were extracted and concatenated in Artemis. A neighbour-net tree was produced using SplitsTree4 (Huson and Bryant, 2006).

2.2.5. Identification of Putative Mobile Genetic Elements

The chromosome sequences of LMG 20360 and DAT561 were compared to each other using BLAST alignments, which were visualised in CGView using the CGView Server (Grant and Stothard, 2008). Regions with no homology were identified, and the genes present extracted from the annotation file produced on the RAST server.

As well as being used for mapping reads to reference sequences, contigs from the 454 sequence were generated *de novo* using Roche's Newbler software. Contigs from individual isolates were first assigned a species identity by BLASTing all the 16S rRNA gene sequences present against the NCBI nucleotide database. Contigs for each isolate which contained only *M. plutonius* sequence (i.e. no 16S rRNA genes from contaminant bacteria were present in the contig set for that isolate) were aligned to a concatenated sequence of the LMG 20360 chromosome and the ATCC 35311 plasmid, and rearranged to match its order, using the progressiveMauve algorithm (Darling et al., 2010). The Mauve visualisation was then examined to look for evidence of insertions of genetic material in either the reference LMG 20360 chromosome and ATCC35311 plasmid, or within contigs of the 454 sequence.

2.3. Results

2.3.1. First Rearrangement of the *M. plutonius* Genome

The draft genome supplied at the start of the project consisted of 2.01 Mb of sequence, in 62 contigs. Contigs ranged from 1032b to 194kb. A total of 1984 CDS were identified in the genome. From the 454 sequencing, 2.03 Mb of assembled sequence was produced, in 68 contigs. Contigs from the 454-produced sequence covered virtually all of the draft genome. All large contig segments had 99-100% identity to the draft. Many putative gaps in the draft sequence were internal to 454 contigs, and were closed (either deleted, or filled with new sequence). When the 454 sequences were included, the *M. plutonius* genome increased to 2.03Mb, and the number of contigs was reduced to 18. 2009 CDS were identified when the new sequence was included.

2.3.2. Second Rearrangement of the *M. plutonius* Genome

The ACT comparison between the published ATCC 35311 chromosome and plasmid, and the first rearrangement of our LMG 20360 sequence, showed clearly that part of our LMG 20360 genome sequence matched the ATCC 35311 plasmid. In the rest of the genome, which matched the ATCC 35311 chromosome, there were several rearranged and/or inverted sections of sequence (Figure 2.1). Using this alignment, plasmid sequence was removed from the LMG 20360 chromosome

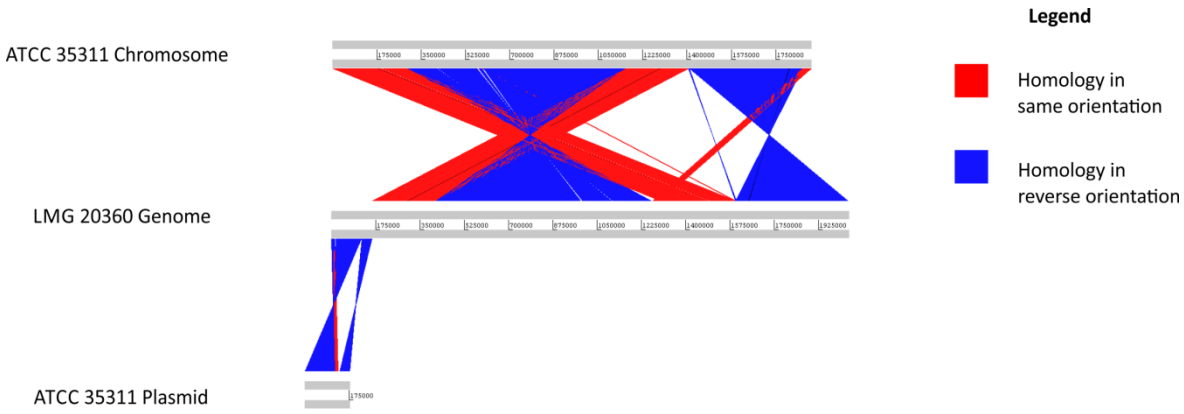


Figure 2.1. ACT visualisation, showing BLAST hits between our First Rearrangement of the *M. plutonius* LMG 20360 genome and the published chromosome and plasmid sequences for ATCC 35311. This image shows how the LMG 20360 genome sequence is divided into regions that match the chromosome and the plasmid. It also shows how, within these different replicons, the DNA sequence is rearranged and flipped in several places with respect to the published sequence.

sequence, which was then rearranged to match the published sequence. This new alignment was therefore then all in the same order and orientation as the ATCC 35311 chromosome. Some short regions have homology to multiple positions in the chromosome, and these correspond to the ribosomal RNA (*rrn*) operons (Figure 2.2).

Of the 18 chromosome gap crossing PCRs, 14 generated products of the expected size, which when sequenced covered both gap-flanking regions and filled the gaps with additional sequence. The reactions which did not work were those for gaps five, nine, eleven and thirteen.

2.3.3. Third rearrangement of the *M. plutonius* Genome

A second genome sequence for a different strain of *M. plutonius* (DAT561) was also published during the course of this project (Okumura et al., 2012). This was compared to the previous sequence with ACT, and with CGView (Stothard and Wishart, 2005), which allows circular genomes to be visualised with genomic statistics (such as GC content and GC skew) overlaid.

When visualised in ACT and CGview, it became clear that the published genomes for ATCC 35311 and DAT561 were not syntenous. The DAT561 chromosome seemed much more likely to be correctly arranged, based on its pattern of GC skew. The DAT561 chromosome had a roughly half-and-half split of positive GC skew (in the first half of the chromosome) and negative GC skew (in the second half), whereas the ATCC 35311 genome was much less evenly split (Figure 2.3). In many bacteria, the chromosome displays a 50/50 divide of GC skew around the terminus (Mrázek and Karlin, 1998). Our sequence for LMG 20360 was therefore rearranged to match the DAT561 sequence. Plasmid sequences from ATCC 35311 were also compared to DAT561 using webACT, and visualised in ACT.

The ACT comparison between the second rearrangement of the *M. plutonius* 20360 chromosome sequence (that is, the same orientation as the published *M. plutonius* ATCC 35311 (NC_015516)) and the published chromosome sequence for atypical *M. plutonius* DAT561 (NC_016938) shows the two chromosomes to be syntenous at the start and end, but the majority of the middle portion of the sequences are inverted relative to each other (Figure 2.4). The *M. plutonius* chromosome sequence was rearranged to match the orientation of the DAT561 chromosome. The comparison between the ATCC 35311 and DAT561 plasmids shows them to be assembled in largely the same orientation, such that no further rearrangement was required (Figure 2.5). As the ATCC 35311 plasmid sequence appears to be correctly arranged, it was used in all subsequent analyses.

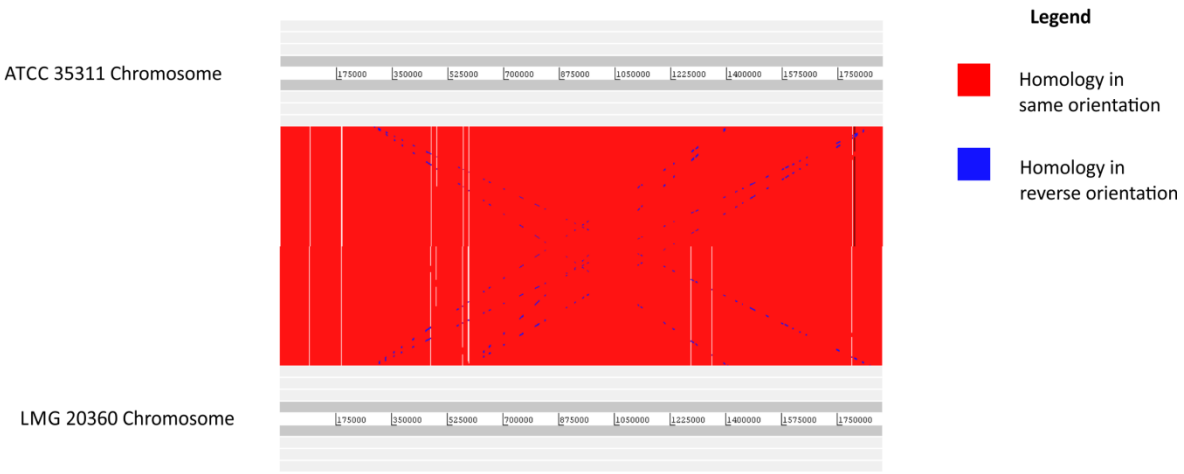


Figure 2.2. ACT visualisation, showing BLAST hits between our Second Rearrangement of the *M. plutonius* LMG 20360 genome and the published chromosome sequence for ATCC 35311. The new sequence corresponds to the published sequence, shown in red. Some smaller blue reverse orientation matches can be seen crossing around the genome, corresponding to the ribosomal RNA operons.

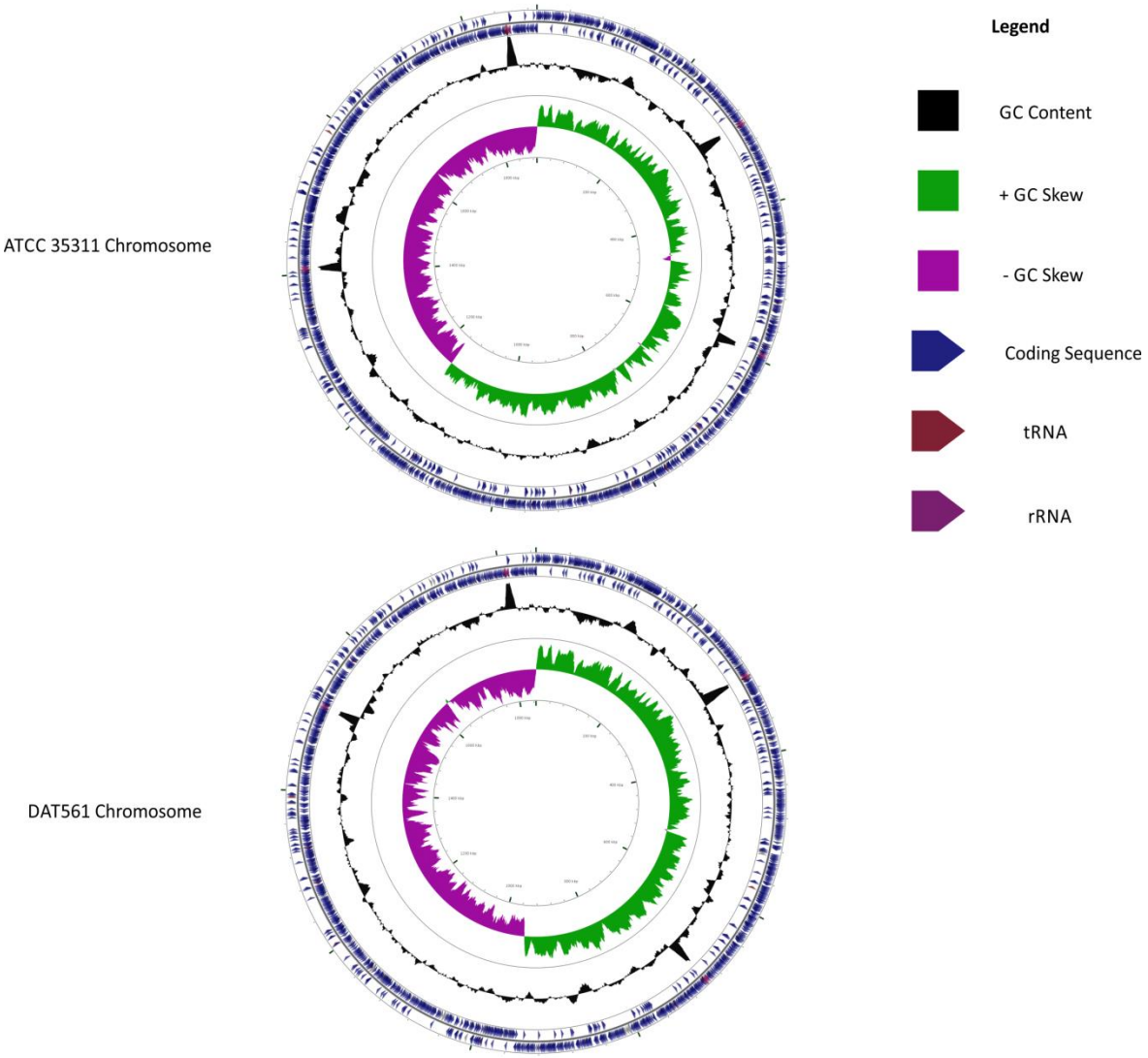


Figure 2.3. Circle plot, showing the published chromosome sequences for *M. plutonius* ATCC 35311 (top) and DAT561 (bottom). The differences in GC skew, and coding sequence strand location, can clearly be seen between the two chromosomes. The four large peaks of GC content on each chromosome correspond to the ribosomal RNA operons.

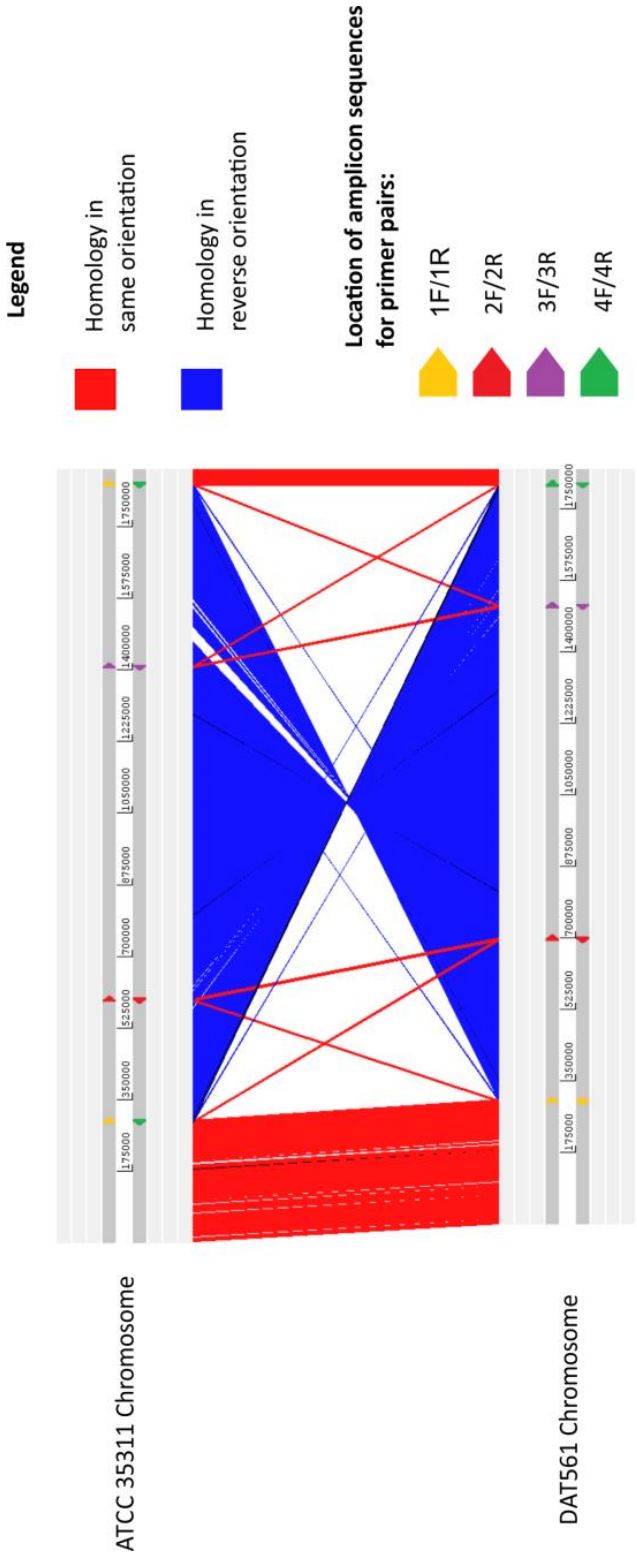


Figure 2.4. ACT visualisation, showing BLAST hits between our Second Rearrangement of the *M. plutonius* LMG 20360 chromosome and the published DAT561 chromosome sequence. Also shown are the results of the long-range PCRs. The coloured arrows on the genome show the location of the longest BLAST hits for each of the sequences from the long range PCRs. It is clear that the arrangement of DAT561 agrees with the PCR products, whereas PCR products 1F/1R and 4F/4R contradict the arrangement of ATCC 35311.

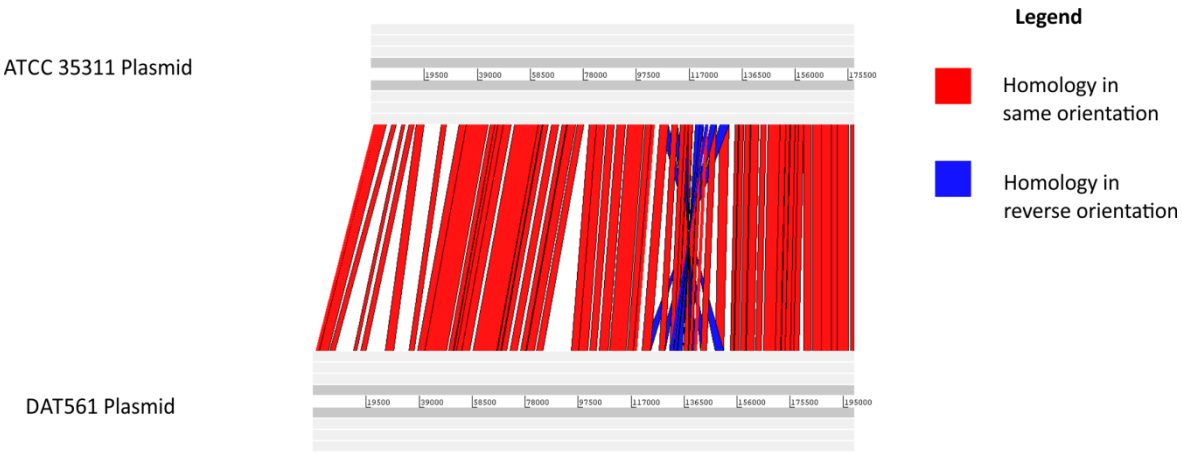


Figure 2.5. ACT comparison between the *M. plutonius* ATCC 35311 and DAT561 plasmid sequences. While there is a region of blue (from circa 100kb to 130kb) indicating an inversion, there are also matches of the same orientation spanning this region.

2.3.4. Long Range PCR

Examination of contig sequences in ACT revealed that the discrepancies between the DAT561 and ATCC 35311 chromosome sequences were due to uncertain arrangement of sequence on either side of the ribosomal RNA (*rrn*) operons. These are four 5kb-long operons of identical sequence that cannot be bridged by short read technologies. For this reason, long range PCR was performed between primers designed to the non-identical flanking regions (Rozen and Skaletsky, 1999). The reorientation of the chromosome sequence is such that, if the DAT561 orientation is correct, each amplification should yield a circa 6kb long product, whereas if the ATCC 35311 orientation is correct the reactions between 1F and 1R, and 4F and 4R, should not yield products.

Amplicons generated by long range PCRs spanning the *rrn* operons yielded products of the expected size (circa 6kb) when performed in the orientation predicted by the DAT561 chromosome sequence (Figure 2.6). Four 6kb amplicons were gel extracted, as well as one smaller, presumed non-specific amplicon from the reaction between primers 2F and 2R. These products were sequenced with the F and R primers, and aligned by BLAST against ATCC 35311 and our rearranged sequence (which was now oriented to match DAT561).

This confirmed that the PCRs had amplified the sequences targeted in the primer design, as the sequences up to the 3' end matched the *rrn* operons. There were short regions from the 5' end of each F and R sequence that also allowed the flanking regions to be confirmed. The position of these long PCR products on both the ATCC 35311 and DAT561 genomes confirm that these products could only be produced if the LMG 20360 chromosome was syntenous with the DAT561 chromosome, and not the published Type Strain genome (ATCC 35311) (Figure 2.4). The shorter amplicon from the reaction between 2F and 2R hit to the plasmid sequence.

2.3.5. Read Mapping and SNP Phylogeny

Isolates 7512, 7526 and 7533 were highly contaminated (likely due to the clumping behaviour of *M. plutonius* (Tarr, 1938) making it difficult to obtain pure cultures), and did not have enough *M. plutonius* sequence present to be analysed in the phylogenetic network. The network clearly shows strong differentiation between isolate 7596 and the rest of the isolates. Within the rest of the isolates, there is evidence of two groupings, with the Type Strain LMG 20360 and isolates 7534 and 7540 in one group, and isolates 7595, 7605, 7606, 7610, 7611 and 7613 in the second group. Variation within the first group is higher than in the second group, within which the isolates cluster much more closely, with the exception of 7605 (Figure 2.7). The strong differentiation between isolate 7596 and the rest is not an artefact of using a different sequencing

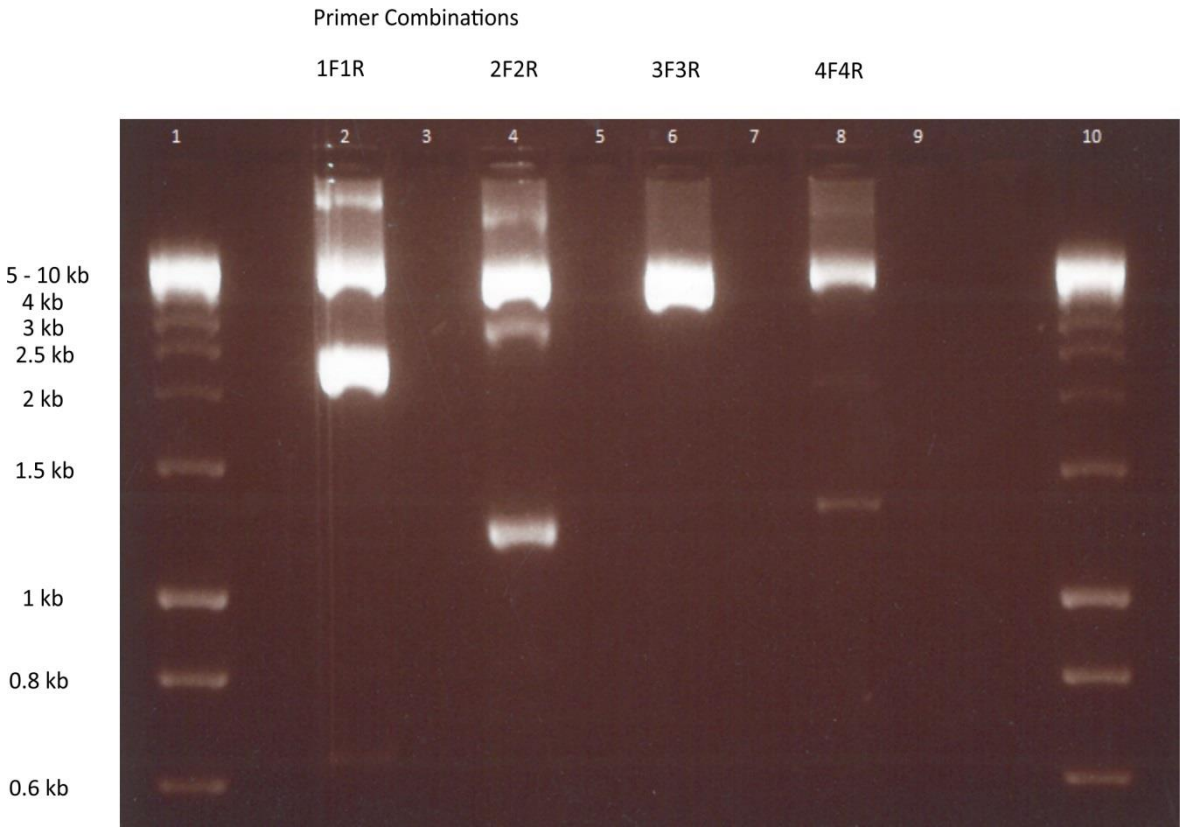


Figure 2.6. Electrophoretogram of long range PCR products. Lanes 1 and 10 contain Bioline Hyperladder 1kb. Each lane with visible products is followed by a lane containing a negative control (no DNA). Strong amplification is observed around 6kb (though the size of the products is difficult to judge for certain, as the ladder did not separate completely). Some non-specific amplification was observed, but either much smaller or much larger than expected. The secondary fragment from primers 2F and 2R that was also extracted and sequenced is seen on the gel, around 3kb.

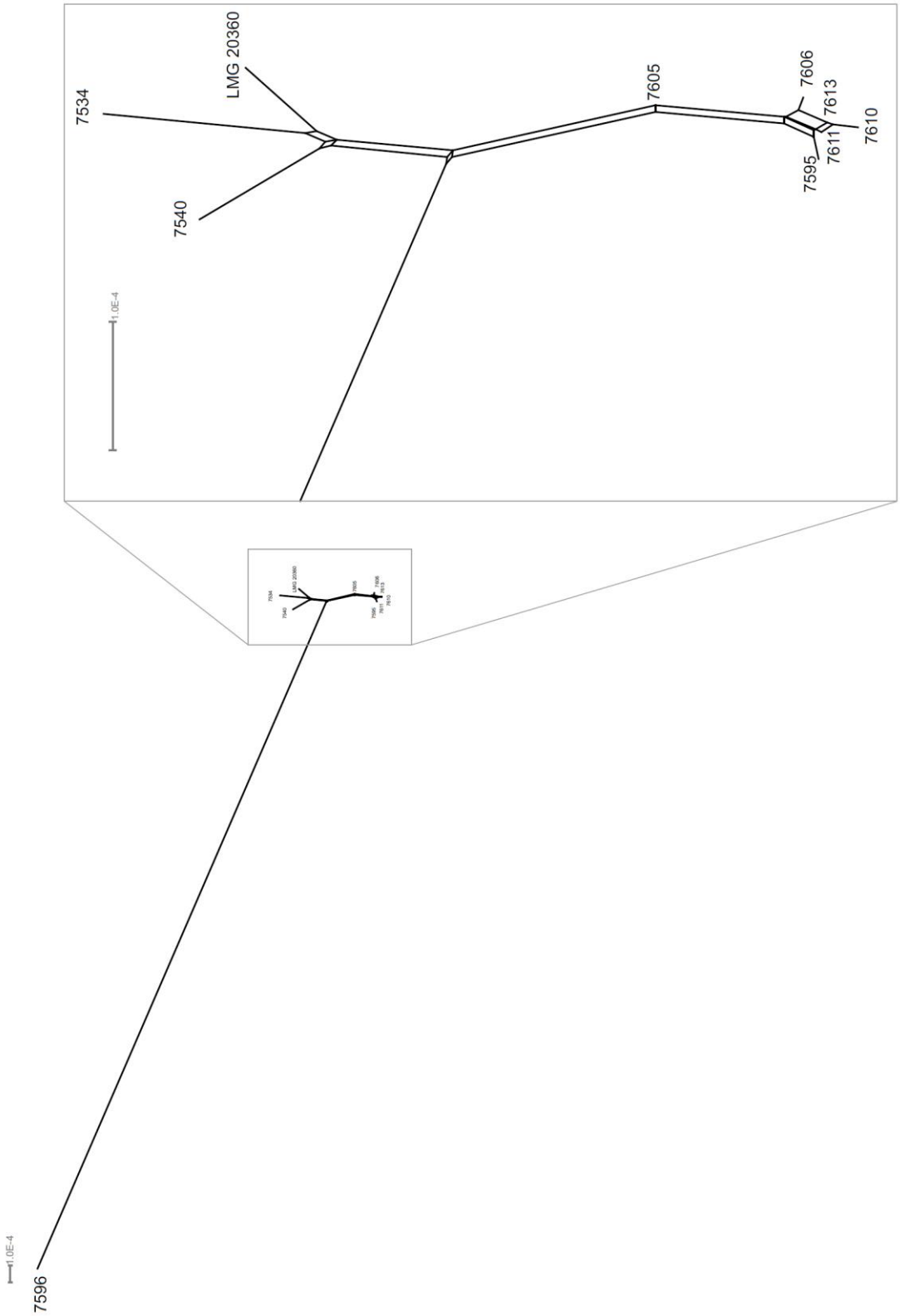


Figure 2.7. Neighbour-net tree showing evolutionary relationships between ten strains of *M. plutonius*. The tree clearly illustrates the differentiation between atypical (7596) and typical (the rest) *M. plutonius*. The portion of the network depicting the typical *M. plutonius* clade has been enlarged, enabling some structure among isolates of typical *M. plutonius* to be seen.

technology. This is illustrated by comparing the number of SNPs identified in the VCF files from the mapping of 7596 to LMG 20360 (8593 SNPs), and 7596 and DAT561 (162 SNPs).

All isolates except 7534 had instances of two different nucleotides being found at the same position in different reads mapping to the same chromosomal location. This was investigated by visualising the BAM file from 7596 against LMG 20360 in Tablet. Individual regions containing multiple alleles were visualised to examine the read population (Figure 2.8), and the depth of read coverage was graphed across the chromosome (Figure 2.9). The regions with multiple alleles clearly show two populations of reads; one group shares the same sequence as the reference allele and the second contains the alternate alleles, for multiple variant nucleotides. The coverage graph shows depth of coverage across the genome. There are four major peaks, corresponding to the *rrn* operons, and one very short peak nearer the start of the sequence.

2.3.6. Identification of Mobile Genetic Elements

In the chromosome sequence of LMG 20360, with BLAST hits from DAT561 overlaid, there are three regions in LMG 20360 from circa 575-585kb, 590-600kb, and 640-680kb that have no corresponding hits in the other strain. The third region corresponds to a region of GC content that differs from the surrounding genome. Magnified images of these sections clearly show the presence of genes associated with mobile genetic elements (Figure 2.10). The products encoded by the genes present in these regions are shown in Appendix 1.

Of the ten isolates from around England that were sequenced on the 454 FLX, only six solely contained 16S rRNA sequences with a closest match to *M. plutonius*. Isolate 7512 only contained a 16S rRNA gene sequence with a closest match to *Lactobacillus kunkeei*, isolate 7540 contained sequences from *M. plutonius*, *Paenibacillus alvei* and *Enterococcus spp.*, and isolates 7533 and 7605 contained sequences from *M. plutonius*, *L. kunkeei* and an uncultured *Lactobacillus*. Of the seven remaining *M. plutonius* isolates, 7526 and 7595 did not produce enough contig sequence to be properly analysed. A progressiveMauve alignment of contigs from the remaining five isolates shows that they have genome sequences of almost the same length as LMG 20360/ATCC 35311, and that there are no obvious regions of inversion within their contigs. The mobile genetic elements identified earlier appear to have moved within the genomes of some of the isolates (Figure 2.11).

A region was identified in the ATCC 35311 plasmid which appeared to have no homology in the other isolates, and a region in the plasmids of the other isolates was identified that appeared to have no homology in the ATCC 35311 plasmid. To determine the genes these regions were

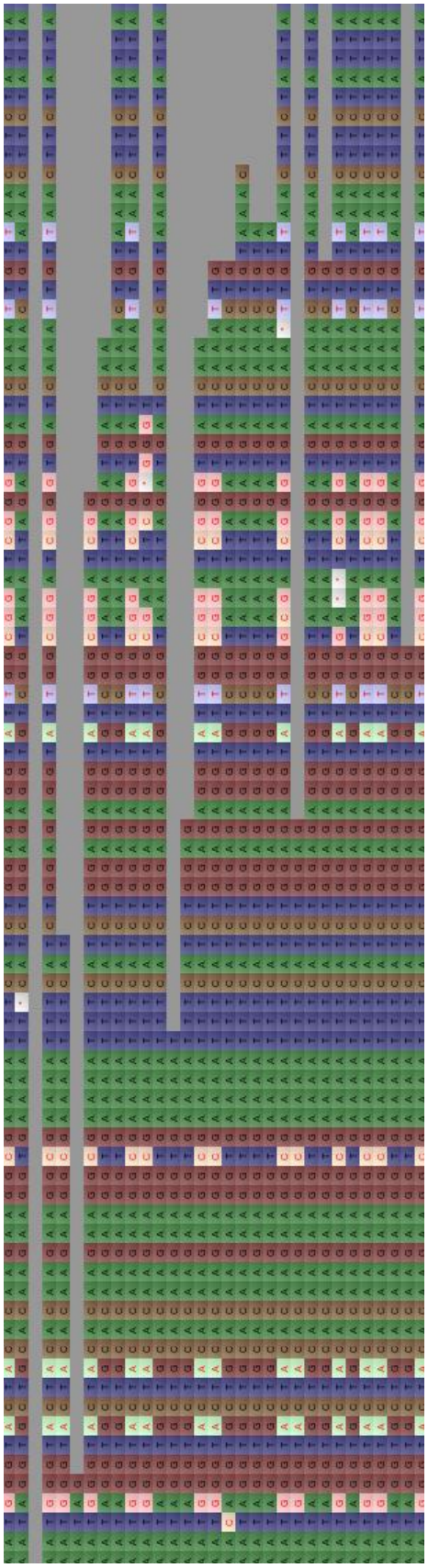


Figure 2.8. Magnified image of a section of reads from Isolate 7596 mapped to the LMG 20360 chromosome. Position is circa base 84840. Variant nucleotide positions are highlighted. It is clear that there are two populations of reads, some containing the reference alleles (dark) and some containing the alternate alleles (highlighted).

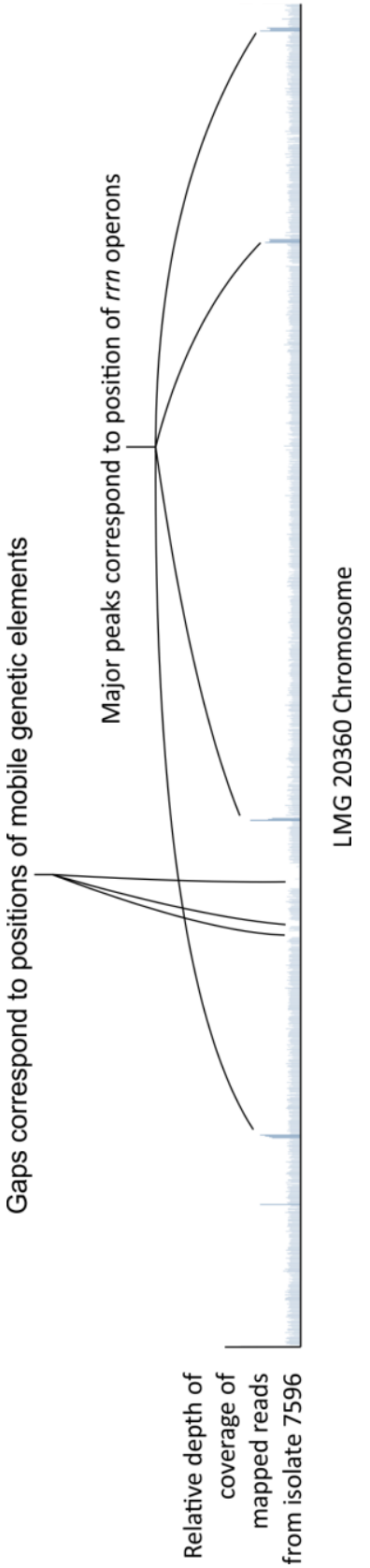


Figure 2.9. Coverage graph of reads from Isolate 7596, mapped to the LMG 20360 chromosome. The four major peaks of coverage correspond to the positions of the ribosomal RNA operons. The three gaps about a third of the way through the sequence correspond to three mobile elements present in LMG 20360 but absent in 7596, described in the following section (Figure 2.10).

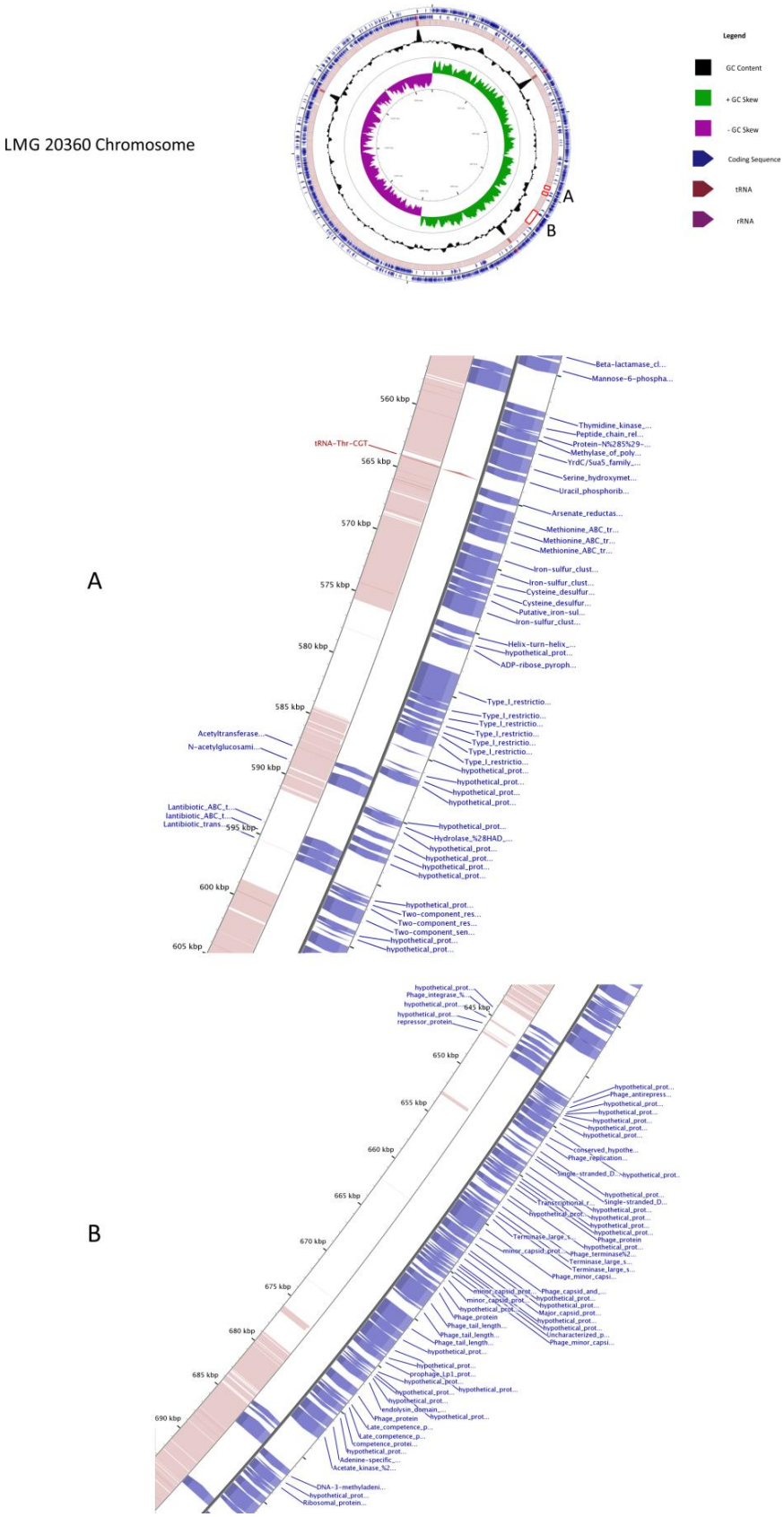


Figure 2.10. Circle plot showing the Third Rearrangement of the LMG 20360 chromosome (outermost dark grey line), with BLAST hits against the DAT561 chromosome (inner, light red circle). The mobile genetic elements identified are highlighted and enlarged (A & B), and the products coded by their coding sequences are annotated.

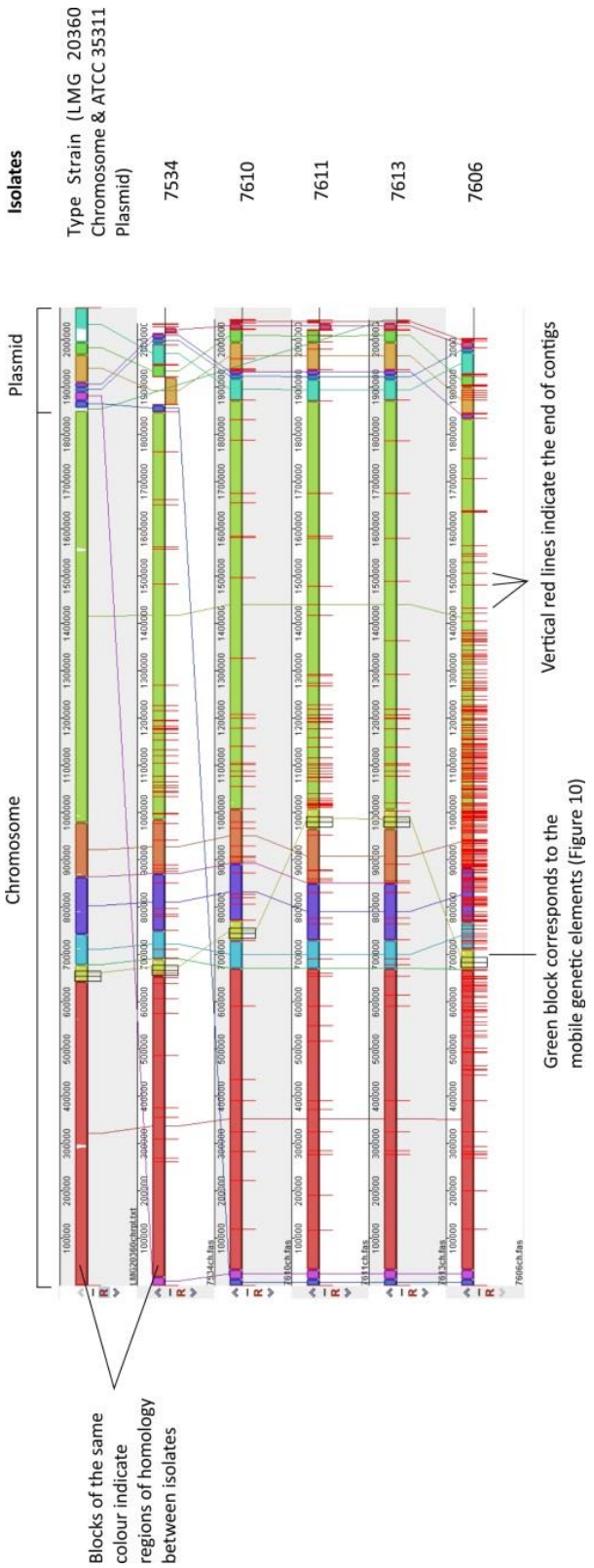


Figure 2.11. progressiveMauve alignment of the genomes of LMG 20360 and five recently isolated *M. plutonius*. Most of the isolates seem to have a genome sequence of almost the same length as LMG 20360/ATCC 35311, which can be arranged in the same orientation. The smaller blocks of homology, beyond 1.85Mb, correspond to the plasmid sequence. Here there appears to be evidence of insertion of novel elements (in the turquoise block in the top sequence, and the short, pink block in the plasmid sequence for each of the five lower isolates).

carrying, the annotation of their CDS were examined (Appendix 1). From examining the CDS it is clear that these regions contain genes encoding the same products. BLAST comparisons revealed the genes to have the same gene sequences in ATCC 35311 as the other isolates, and it was concluded that these regions correspond the same section of plasmid. The lack of mauve alignment between ATCC 35311 and the other isolates may be an artefact of the short length of the contigs.

2.4. Discussion

2.4.1. Genome Orientation

The comparison between the GC skew profiles of the chromosomes of *M. plutonius* ATCC 35311 and DAT561 revealed clear differences between the two arrangements. DAT561 is in agreement with the half way split around the terminus that would be predicted, whereas the ATCC 35311 is much closer to 60% positive GC skew, 40% negative. These fifty-fifty splits are seen in many species of bacteria, and some viruses (McLean et al., 1998; Mrázek and Karlin, 1998), and are probably due to asymmetry of DNA replication and repair (Sueoka, 1962; Muto and Osawa, 1987) and the accumulation of mutations on the non-transcribed strand during transcription (McLean et al., 1998).

Examination of the two chromosomes shows that differences in the direction of large tracts of sequence that lie between ribosomal RNA (*rrn*) operons are the cause of the observed discrepancy. This should not come as a surprise; the *rrn* operons are more than 5kb of identical sequence. These are impossible to cross with a single read from the FLX titanium machines used to sequence these genomes, a system which typically produces read lengths of around 500bp (Glenn, 2011). As such, whilst the flanking regions of *rrn* operons can be identified, it is impossible to know which flanking regions are present either side of an individual operon without further investigation.

The long range PCRs undertaken in this study were performed with primers designed specifically to individual *rrn* flanking regions. Successful amplification was achieved in the orientation described by the DAT561 chromosome sequence, and the ends of the flanking regions were sequenced to confirm the locations of the products. We have therefore demonstrated that the Type Strain of *M. plutonius* is in fact arranged in the same orientation as DAT561, rather than the published ATCC 35311. In this study strain LMG 20360 was used rather than ATCC 35311, but both are clones from the original NCDO 2443 (Bailey and Collins, 1982; Forsgren et al., 2013). Genome orientation may seem an academic point, but it is relevant to an understanding of the biology of

M. plutonius. Important bacterial genes, such as rRNA and tRNA genes, are found almost solely on the leading DNA strand. This is because of the effect on the transcript of collisions between RNA polymerase and DNA polymerase when the same gene is being transcribed and replicated in different directions (Rocha and Danchin, 2003). The copy number of a gene also depends on its position on the chromosome (Bremer and Dennis, 1996). In fact, artificial inversions of bacterial chromosomes around ribosomal operons have led to significant reductions in bacterial growth (Srivatsan et al., 2010). Misunderstanding the orientation of a bacterial chromosome may therefore lead to genes being thought to be closer to or further from the origin of replication than in reality, and more or less important to the bacterium. Furthermore, genuine inversions of the bacterial chromosome have been seen to occur symmetrically around the origin of replication (Eisen et al., 2000). The fact that these inversions occur symmetrically around the origin underscores the importance of genes retaining their relative proximity to the origin (apparently regardless of which side of the origin they are on). Such inversions also illustrate the importance of having a correctly arranged reference chromosome sequence so that genuine inversions can be detected. At a smaller scale, misassembly resulting in inversions or rearrangements of part of a chromosome could prevent the identification of gene cassettes or operons.

2.4.2. Genomic Comparisons

Genomic comparisons between different *M. plutonius* variants has confirmed the differentiation between two recently identified subtypes of the bacterium, typical and atypical *M. plutonius* (Okumura et al., 2012; Arai et al., 2012; Takamatsu et al., 2013). The long branch-length on the phylogenetic network between 7596 and the typical *M. plutonius* shows that 7596 is much less closely related to the typical *M. plutonius* isolates than any of them are to each other. The tiny number of SNPs (162) identified across the whole chromosome between 7596 and DAT561 show these isolates have a recent common ancestor, certainly compared to the 8593 SNPs between isolate 7596 and LMG 20360. Since DAT561 was isolated from infected larvae in Japan (Arai et al., 2012), and 7596 from infected larvae in the UK, we must hypothesise that infectious *M. plutonius* is being transported globally and released into the environment, likely through the trade in honey bees or honey bee products (Mutinelli, 2011).

Comparisons of the whole chromosome sequence of LMG 20360 and DAT561, and of LMG 20360 with the contig sequences from the recently-isolated typical *M. plutonius*, show regions that are absent in DAT561, and appear to have been acquired in other isolates by horizontal gene transfer. This supposition is supported by the proximity of the regions to a tRNA site (shown in Figure 2.10). tRNAs frequently act as integration sites for genetic elements (Williams, 2002). The largest region in LMG 20360 that has no hits in DAT561 is a prophage, demonstrated by the presence of

annotated bacteriophage genes in the region. One of the smaller regions, by the annotation of the genes it carries, is likely to be a Type I Restriction Modification (RM) System. These are systems that contain restriction enzymes and modification methylases, and whose primary role is to protect the bacteria from invasion by foreign DNA, including bacteriophages. However there is strong evidence that RM systems can themselves act as transposons, and facilitate horizontal gene transfer and genome rearrangement (Kobayashi, 2001). In *M. plutonius* it appears that the Type I RM system has become deactivated. Type I systems consist of three proteins; R (restriction), M (modification) and S (specificity) subunits (Wilson, 1991). In *M. plutonius* the R subunit is divided by stop codons into two CDS, the M subunit into three, and the R subunit into two.

The third region contains genes that encode lantibiotic transporters, two-component response regulators and a nisin biosynthesis sensor (*nisK*). Nisin and other lantibiotics are complex molecules with antimicrobial properties, that are used by bacteria to suppress competitors (Cheigh and Pyun, 2005; Willey and van der Donk, 2007). These genes would potentially be beneficial to *M. plutonius* in competing with protective *Bifidobacterium* and *Lactobacillus spp.* in the honey bee gut (Vásquez et al., 2012; Wu et al., 2013). However, none of the other genes in the *nisA* promoter region are seen in *M. plutonius*, and these are necessary for the production of nisin (de Ruyter et al., 1996). Neither are there genes present with homology to those encoding other, related linear lantibiotics such as subtilin, epidermin, pep5 or lacticin (de Vos et al., 1995). It could therefore be possible that the sensor nisK detects nisin produced by lactic acid bacteria co-occurring with *M. plutonius*, which via the two-component response regulator leads to the secretion of an as yet unidentified nisin-degrading enzyme. Such defence mechanisms against nisin have been suggested in other bacterial species (Brul and Coote, 1999). Since nisin appears to act at the cell wall by both forming pores, leading to the rapid efflux of small metabolites, and interfering with peptidoglycan biosynthesis (Héchard and Sahl, 2002) rather than being taken up by the cell, it is currently unclear what role the identified lantibiotic transporters play in *M. plutonius*.

We have shown that *M. plutonius* contains mobile genetic elements, including a plasmid, prophage and Type I RM system. However, on none of these elements do we see any evidence of antibiotic resistance genes. These genes are frequently found on mobile genetic elements in gram-positive bacteria (Butaye et al., 2003; Nandi et al., 2004), a potentially significant fact given the use of the antibiotic oxytetracycline (OTC) in the treatment of EFB (Waite, Brown, et al., 2003b; Wilkins et al., 2007). The absence of antibiotic resistance genes in the *M. plutonius* genomes we have studied agrees with the continued susceptibility of this pathogen to OTC (Hornitzky and Smith, 1999; Waite, Jackson, et al., 2003). Nonetheless, our observations of mobile

genetic elements, and the presence of OTC-resistance in other honey bee brood bacteria (Evans, 2003), suggests the potential for transfer of the relevant genes into *M. plutonius* in the future.

Finally, we observed that several isolates contained more than one nucleotide in the same position in different reads mapping to the same chromosomal location, at several genes around the genome. When the alignment (BAM) files were examined in Tablet, many of these instances appear to be biologically genuine. This is because the proportion of reads containing each different nucleotide was approximately equal (therefore unlikely to be sequencing error), the reads in question seem to be confined to certain genes (not spread randomly through the genome), and in several cases multiple positions with more than one nucleotide were found within individual reads. There are a number of possible explanations for this observation – not all the isolations were perfect, and non-*M. plutonius* sequence is present in some of the sets of reads. If the non-*M. plutonius* sequence had high enough sequence homology to the reference genome, it could have been mapped to the reference and caused the secondary allele. Alternatively, there could be multiple variants of *M. plutonius* infecting the same larva, and polymorphisms between the two strains could have been causing the two populations of reads. Finally, these could be due to gene-duplication events in the sequenced isolates that did not occur in the reference genome, and the duplicated genes were being mapped to their closest relative in the reference genome.

From the graph of coverage depth it is clear that the only regions that have greatly higher coverage than the rest of the genome are the *rrn* operons. This is probably due to contamination with non-*M. plutonius* sequence, as rRNA genes are highly conserved and *rrn* sequence from contaminant bacteria would be the sequences most likely to be mapped to the reference genome. The fact that only these regions show much higher coverage implies that it is unlikely that contamination with non-*M. plutonius* has led to this phenomenon, as we would also expect genes present in the contaminant to show higher depth of coverage, perhaps approaching that of the *rrn* genes (as reads from the duplicated gene in the sequenced isolate would only have the paralog in the reference genome to map to). The fact that these variant reads are seen in an approximately 1:1 ratio implies that they are paralogs within the same genome, as each genome is likely to contain one of each copy of the gene. If the variants were the result of contamination with another strain of *M. plutonius* it is unlikely that the strains would be present in such a ratio. Additionally, natural variation across the genome makes it difficult to use read depth to make predictions about the presence of paralogs. Finally, the instances of multiple variant nucleotides within individual genes shows more polymorphism than would be expected between different strains of *M. plutonius*, as later work (see Chapter Three) has shown polymorphism between isolates at individual genes to be low.

In this study of the genomes of *M. plutonius* several important aspects of its biology have been elucidated. Firstly, a correctly-arranged genome sequence of the Type Strain was produced. Mobile Genetic Elements containing genes that are likely to have an impact on *M. plutonius*' ability to cause disease were identified. The distribution of these elements was found to differ among isolates, being present in the examined typical *M. plutonius* and absent in the atypical *M. plutonius*. The differences between these groups, as well as differences within the typical *M. plutonius*, were further highlighted by the phylogenetic network. An ability to distinguish genetically between different variants of the pathogen will prove invaluable in investigating its epidemiology.

3. A typing scheme for the honey bee pathogen *Melissococcus plutonius* allows detection of disease transmission events and a study of the distribution of variants

3.1. Introduction

European Foulbrood of honey bees (EFB) is an important disease of honey bee larvae, which is found on every continent where bees are managed. Infected larvae become displaced in their cells and develop a waxy sheen, before decomposing and discolouring (Bailey, 1981). Whilst not always the case, the disease can prove fatal to a colony (Bailey, 1981). EFB is caused by the Gram-positive bacterium *Melissococcus plutonius* (White, 1912; Bailey, 1957) which despite being globally distributed has been found to exhibit extremely low levels of genetic diversity (Allen and Ball, 1993; Djordjevic et al., 1999). Indeed, isolates from the UK and Australia have proven indistinguishable by RFLP (Restriction Fragment Length Polymorphism). In spite of apparently low genetic diversity in the pathogen population, some differences have been observed in biochemical and physical characteristics (Allen and Ball, 1993). More recently a subtype of *M. plutonius*, showing genetic and metabolic differences from previously reported *M. plutonius*, was reported in Japan (Arai et al., 2012), suggesting *M. plutonius* may contain more genetic variation than previously thought.

Disease prevalence varies annually and regionally (Wilkins et al., 2007; Roetschi et al., 2008; Budge et al., 2010). *M. plutonius* is able to persist in a honey bee without causing symptoms (Forsgren et al., 2005; Roetschi et al., 2008; Budge et al., 2010), which may be exacerbated by the presence of secondary bacteria (Bailey, 1981). Regional differences in prevalence are often found: for example in the UK, in 2012, 1.9% of inspected apiaries in Wales were EFB-positive, compared with 6.6% in England (NBU, 2012a). Recent evidence suggests that levels of EFB infection are increasing in Switzerland and the UK (Wilkins et al., 2007; Roetschi et al., 2008; NBU, 2012b) and countries thought to be disease-free have suffered recent outbreaks (e.g. Norway and Scotland). Despite this growing risk, the epidemiology of EFB is not fully understood.

Multi locus sequence typing (MLST) has proved a successful and powerful way of distinguishing pathogen variants and inferring understanding of disease aetiology (Killgore et al., 2008). A traditional MLST scheme uses sequence variation in six to ten housekeeping genes, with each gene representing a different locus and any variation representing different alleles (Maiden et al., 1998). Such a scheme uses housekeeping genes because they are conserved enough to be present

in all isolates, but should show enough variability to enable different variants to be identified (Urwin and Maiden, 2003). Genetically distinct isolates can then be grouped into variants or Sequence Types (ST) based on different allelic profiles across all loci. MLST schemes have been developed which do not exactly correspond to the original concept of an MLST, as they contain non-housekeeping gene loci (Ahmed et al., 2006), and the scheme developed here makes use of genomic comparisons to identify appropriate loci when dealing with a genetically homogeneous bacterium. The lack of a typing scheme for *M. plutonius* has severely limited epidemiological investigations into this important disease.

We present the first usable typing scheme for *M. plutonius*, which is able to distinguish multiple STs of the bacterium and to identify epidemiologically relevant transmission routes. We also characterise the geographical distribution of STs.

3.2. Methods

3.2.1. Bacterial growth and DNA extraction

Isolates of *M. plutonius* from the Fera (Food and Environment Research Agency) culture collection were plated out on to M110 agar and grown under anaerobic conditions at 35°C ± 2°C. To isolate bacteria from individual infected larvae, larvae were mixed with 0.5 ml PBS and vortexed, with a 10µl loop used to streak the resulting suspension on M110 plates and incubated as before. After growth, individual colonies were spread on to fresh M110 plates, and left to grow for 7-14 days (Forsgren et al., 2013).

DNA was extracted from *M. plutonius* colonies using the Promega Wizard ® Magnetic DNA Purification System for Food. In some cases, diseased larvae did not yield viable bacteria, so a suspension of larva in PBS underwent DNA extraction by the same method.

3.2.2. Identification of loci

Loci were tested for the same genes as the MLST schemes for *Enterococcus faecalis* (Ruiz-Garbajosa et al., 2006) and *Enterococcus faecium* (Homan et al., 2002), and for genes that had predicted membrane-spanning domains. Subsequently loci were identified based on Mauve (Darling et al., 2010) alignments between contigs of *M. plutonius* genome sequence, from the type strain LMG 20360 and 9 isolates collected in England in 2010 (contigs having previously been produced using newbler (Roche) with reads from a Roche 454 pyrosequencer (see Chapter 2)). The Export SNPs function of Mauve was used to locate polymorphic sites. These were verified by manual BLASTn alignments between contigs from all isolates. Loci were chosen that showed three

or more different sequences between 10 isolates over a 300bp region. Primers were designed to the *M. plutonius* type strain sequence, using primer3 (Rozen and Skaletsky, 1999), and were synthesised by MWG Eurofins.

3.2.3. Testing of locus variability

PCR primers were first used to amplify fragments from DNA from a test set of *M. plutonius* isolates, including culture-collection isolates and those isolated from the field in the UK. Briefly, 2 µl of DNA sample was added to 12.5 µl of Fermentas 2 x PCR Master Mix (Massachusetts, USA) with 10 pmols of each primer and made up to a final volume of 25 µl. Each reaction was run on an Applied Biosystems 2720 Thermal Cycler (New York, USA) at 94°C for 2 minutes followed by 35 cycles of 94°C for 30 seconds, 50°C for 30 seconds, and 72°C for 1 minute with a final elongation step at 72°C for 1 minute. PCR products were visualised on a 2% agarose gel using Fermentas ® 6X DNA Loading Dye, and Bioline HyperLadder ™ 100bp (London, UK). PCR products were purified using the Qiagen® Qiaquick PCR purification kit, and sequenced in one direction on an ABI 3730xl 96-capillary DNA Analyzer (MWG Eurofins).

Loci based on previous MLST schemes and the presence of membrane-spanning domains did not show sufficient variation to merit inclusion in a scheme. Loci identified from 454 data, *galK*, *argE*, *gbpB*, and *purR*, did reveal higher levels of variation, and were able to distinguish between multiple STs. These loci were then used as the final version of the MLST scheme, and the primers (Table 3.1) were used to amplify PCR products for all further isolates.

3.2.4. International Isolates

After initial validation on a test set of culture collection and UK-collected isolates, a range of globally sourced samples were tested with the MLST primers. Samples were tested from Australia, Brazil, Denmark, France, India, Italy, the Netherlands, Poland, the Republic of Ireland, Tanzania, Thailand, and the United States of America, and sequences for the MLST loci were downloaded from the Genbank genome sequence of a sample of atypical *M. plutonius* from Japan (NC_016938.1).

Primer	Primer Sequence
galK L	TTT CCA GCA GCA ATT ACA A
galK R	GGG TAG GGA TTT TTG AAG AG
argE L	GGT GGG ACA TTT AGA CGT AG
argE R	AAA TTA AGA CCC AAC CCT TC
gbpB L	AGC AGC TAA ACA GAA TGA GC
gbpB R	GCC AAC GTC TAA CAG ATA CC
purR L	ACC ACC AAG TGC CAG TAT TA
purR R	CGA TTT TGT TCT GAT AAC CTG

Table 3.1. PCR primer sequences (5’ to 3’ direction) for the *M. plutonius* MLST scheme. Expected product sizes were; 579 nt argE; 565 nt galK; 386 to 632 nt gbpB; 507 nt purR.

3.2.5. Case Studies of Suspected Honey bee Movements

To demonstrate the effectiveness of the MLST scheme for making epidemiological inferences, several anonymised cases of EFB linked to known or suspected honey bee movements or disease recurrence were identified by discussion with National Bee Unit field staff (G. Budge, *Pers. comm.*). *M. plutonius* cultures were grown from infected larvae taken from these outbreaks, and typed. All case studies occurred in 2010.

3.2.6. Data Analysis

Potential patterns of evolutionary descent between STs were calculated using the goeBURST algorithm in the PHYLOViZ program (Francisco et al., 2012) a modification of the earlier eBURST algorithm (Feil et al., 2004). Measures of differentiation between STs present in the UK and the rest of the world were calculated in R (R Core Team, 2013) using the vegan package (Okansen et al., 2013).

3.3. Results and Discussion

3.3.1. Locus selection using traditional methods

Loci used in MLST schemes for the closely related *Enterococcus faecalis* (Ruiz-Garbajosa et al., 2006) and *Enterococcus faecium* (Homan et al., 2002) were tested initially. In addition, some genes encoding proteins with predicted transmembrane domains were tested, because proteins with extracellular regions often show high diversity driven by interactions between pathogen and host (Gupta and Maiden, 2001). We tested 15 such loci on a test set of 21 isolates. These isolates included the Type Strain (LMG 20360) and 18 other epidemiologically unrelated isolates from a broad geographic area across the UK, as well as single isolates from Thailand and Australia.

Two loci in the test set showed no variation on sequencing. Nine of the loci had two alleles each, but the second allele was only present in isolate 7596, later shown to belong to the atypical group. The remaining four loci each had three alleles. When all 15 loci were used, they split the test set of isolates into six STs.

3.3.2. Locus selection using genomic comparison

Subsequently, four loci were identified by genomic comparisons among the Type Strain and nine isolates sampled from the UK in 2010 (Table 3.2). These loci split the same set of 21 isolates into ten STs. The 15 loci tested earlier provided no additional information, so the decision was made to continue with the scheme using only the four loci identified using genomic comparison (Figure 3.1).

These more informative loci are not all internal fragments of housekeeping genes, as frequently used in traditional MLST schemes. Two loci, *argE* and *gbpB*, are found within the coding regions of genes (encoding acetylornithine deacetylase and a putative secreted antigen respectively). The *gbpB* product is related to proteins which appear to be essential virulence factors for other cocci (Teng et al., 2003; Stipp et al., 2008). *gbpB* in *M. plutonius* includes a VNTR (Variable Number Tandem Repeat), which appears to be in an unstructured region of the protein. *argE* may be important (Harris and Singer, 1998), though perhaps not essential (Kobayashi et al., 2003) for arginine metabolism in *M. plutonius*. The remaining two loci, *galK* and *purR*, include intergenic regions, and therefore cannot be considered traditional MLST loci. The *galK* locus spans a region between two galactokinase fragments separated by a region of stop codons, and the *purR* locus begins upstream of the 5' end of the *purR* (purine operon repressor) coding sequence. Whilst it is therefore possible that *argE* and *gbpB* might have been located by a continued process of trial and error of testing housekeeping genes, the *galK* and *purR* loci could only have been located by comparing genomes for highly polymorphic regions. This scheme is therefore considered a modified MLST, as it uses fewer loci than is conventional and not all loci are housekeeping genes. However for brevity the term MLST is used throughout this thesis.

3.3.3. Typing UK and International Isolates

The four locus typing scheme was tested on a further 42 isolates, sourced from the UK and internationally. The STs identified so far reveal a total of 26 alleles over four loci (Table 3.3). These sequence data have been submitted to the EMBL database under the accession numbers HF569117 to HF569142. A list of isolates typed is shown in Table 3.2.

Fera ID	Alternative ID	Country of Origin	Donated By	ST
6404*	LMG 20360	England and Wales		1
7087*		England and Wales		3
7102*		England and Wales		2
7148*		Thailand	1	1
7149		Thailand	1	1
7154*	NCFB 2442	Australia		4
7363		France		4
7365		France		4
7366		France		4
7369		France		4
7483*		England and Wales		7
7509*		England and Wales		3
7511*		England and Wales		7
7512†		England and Wales		6
7515*		England and Wales		7
7516*		England and Wales		3
7517*		England and Wales		3
7521*		Scotland		2
7523*		England and Wales		3
7524*		England and Wales		3
7526*†		England and Wales		6
7531*		England and Wales		7
7533†		England and Wales		6
7534*†		England and Wales		8
7540*†		England and Wales		9
7595†		England and Wales		6
7596*		England and Wales		10
7599*		England and Wales		3
7604		England and Wales		8
7605†		England and Wales		11
7606		England and Wales		5
7609*		England and Wales		5
7611†		England and Wales		11
7612		England and Wales		5
7613†		England and Wales		5
7641		England and Wales		9
8061		England and Wales		20
8214		USA	2	3
8217		USA	2	10
8220		USA	2	3
8222		USA	2	3
8224		USA	2	3
8396		Denmark	3	13
8397		Poland (sampled in UK)		14
8401		Republic of Ireland	4	3
8469	NCFB 2440	Brazil		16
8470	NCFB 2439	India		15
8472	NCFB 2441	Tanzania		17
8473		Scotland	5	18
8475		Scotland	5	1
8498		The Netherlands	6	3
8500		France		3
8513		The Netherlands	6	19

Table 3.2

8516		Poland	7	13
8517		Poland	7	13
	EFB C+70818	Italy	8	1
	EFB 1185/1	Italy	8	3
	EFB 1185/2	Italy	8	3
	US21	USA	2	3
	US18	USA	2	12
	P4	Poland	7	13
	DAT 561	Japan	‡	10

Table 3.2. Isolates used in this study. Samples without a Fera ID were not cultured, and PCR was performed on a whole larval DNA extract, or on a culture extract supplied by the donor. Isolates highlighted with a * are those that form the initial 21 isolates on which all loci were screened. Isolates highlighted with a † are those that were used for genomic comparisons.

Donors;

1) Dr Panuwan Chantawannakul, Department of Biology, Faculty of Science, Chiang Mai University, Chiang Mai, Thailand

2) Prof Jeff Pettis and Mr I Barton Smith, Bee Research Laboratory, USDA, Beltsville, MD, USA

3) Dr Eva Forsgren, Department of Ecology, Swedish University of Agricultural Sciences, Uppsala, Sweden

4) Mr Eoghan Mac Giolla Coda, Galtee Bee Breeding Group, Republic of Ireland

5) Dr Fiona Highet and Ms Mairi Carnegie, Science and Advice for Scottish Agriculture (SASA), Edinburgh, UK

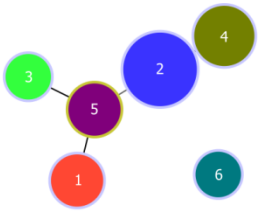
6) Dr Sjef van der Steen and Dr Bram Cornelissen, Wageningen University, Netherlands

7) Dr Krystyna Pohorecka and Mr Andrzej Bober, National Veterinary Research Institute, Pulawy, Poland

8) Dr Anna Granato, Istituto Zooprofilattico Sperimentale delle Venezie, viale dell'Università, Legnaro, Italy

‡ Allele sequences from strain DAT 561 are taken from NC_016938.1. Note; at positions 36-41 in all *purR* alleles except that from NC_016938.1, 6 thymine bases are present. In NC_016938.1 there is a seventh thymine inserted, but this is believed to be an artefact of sequence assembly, as 454 technology can have difficulty resolving long tandem sequences of thymines. This assumption is supported by the fact that the type strain sequence published by the same group (NC_015516) also has 7 thymine bases present, whereas our Sanger sequencing of the type strain allele shows 6 thymine bases. We have therefore not treated this extra thymine in DAT 561 as indicating a novel allele.

A



B

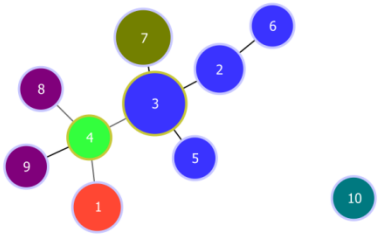


Figure 3.1. goeBURST (Francisco et al., 2009) trees of STs. Each circle represents a different ST, with lines linking closest relatives. Black lines indicate a single allelic change between STs, light grey lines indicate differences at two loci. Circles ringed with a yellow outline indicate putative founder genotypes. 1A shows putative pattern of descent based upon the 15 traditional MLST loci. 1B shows the putative pattern of descent based upon the four loci identified by genomic comparisons. In both 1A and 1B the numbers in the circles show the ST as defined by the four locus scheme, and the colours show the STs as defined by the 15 traditional loci – that is if a circle has the same colours in 1A and 1B, they contain the same isolates.

ST	galK	argE	gbpB	purR	Number of isolates
1	1	1	1	1	5
2	3	2	2	2	2
3	3	2	2	4	17
4	1	1	2	4	5
5	3	2	4	4	4
6	2	2	2	2	4
7	3	2	5	4	4
8	1	1	7	3	2
9	1	1	6	4	2
10	4	4	3	4	3
11	3	3	4	4	2
12	4	4	3	5	1
13	1	1	1	4	4
14	5	1	8	4	1
15	1	5	1	4	1
16	6	4	9	4	1
17	3	1	1	4	1
18	1	1	2	1	1
19	4	4	10	5	1
20	1	1	8	4	1

Table 3.3. *M. plutonius* MLST typing scheme. Sequence Types (ST) identified in this study are shown, and the four columns, *galK* (galactokinase), *argE* (acetylornithine deacetylase), *gbpB* (secreted antigen) and *purR* (purine operon repressor), show the alleles present at each of these four loci in each ST. Within a column each number represents a unique allele (that is, a unique DNA sequence) at that locus. Each ST has a unique allelic profile across all loci. The number of isolates found so far in each ST is stated in the final column.

In total, 12 STs were found in the UK (England and Wales, and Scotland), and 11 in the rest of the world (Figure 3.2). The Sørensen similarity index was used to compare differentiation in ST distribution between the UK and the rest of the world, as it requires only presence/absence data. The international isolates were not sampled in a systematic way and therefore measures of ST abundance are not meaningfully comparable with the UK. The Sørensen Index between the UK and the rest of the world is 0.2609, which, on a 0-1 scale of similarity, indicates a low level of similarity between *M. plutonius* populations. The measure β_{sim} (Lennon et al., 2001) takes into account the percentage of unshared variants found in each area. β_{sim} values increase with a decline in the percentage of species (or in this case, bacterial variants) shared between two focal areas, and achieves their maximum when the percentages gained and lost when moving between the two areas are similar. When the percentage of shared species is low, and the percentage of variants found exclusively in one focal area is also low, β_{sim} is particularly sensitive to small changes in the percentage of unique variants at an individual site. β_{sim} therefore performs well at reflecting gain and loss of species (Koleff et al., 2003). The β_{sim} index for these samples is 0.7273, a high value of dissimilarity, and demonstrates that the differentiation among communities is due to each one having a similar proportion of unshared variants; one is not just a subset of the other.

Isolates identical or similar by MLST to the Japanese atypical genome have been found in Brazil, the UK, the USA and the Netherlands, suggesting that they are in fact widely distributed. Interestingly, the Brazilian isolate (NCDO 2440) (Bailey, 1984) described in the 1990s (Allen and Ball, 1993) showed similar, but not identical, culture requirements to the Japanese atypical isolates (Arai et al., 2012). Our MLST profile of this Brazilian isolate, ST16, places it in the atypical group, as it is similar but not identical to that of the Japanese atypical isolate DAT 561 (ST10).

3.3.4. Local epidemiology

We have identified and sequence typed isolates from four outbreaks for which we have detailed epidemiological data, in order to investigate the utility of this MLST scheme for confirming links between infections.

In Case 1, a beekeeper with disease was known to have sold honey bees to a beekeeper 54 km away, and diseased material was sampled from both the seller (6th September 2010) and buyer (29th June 2010). A 10 km grid square around the buyer had been free of EFB for the previous ten years. Bacteria from both outbreaks were found to be ST9, an ST which was not found anywhere else in the UK.

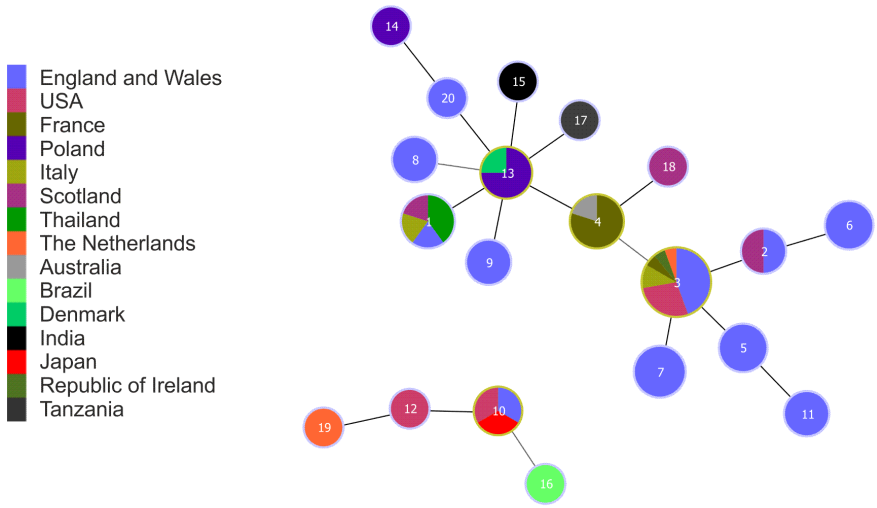


Figure 3.2. goeBURST Tree of STs. Each circle represents a different ST, with lines linking closest relatives. Black lines indicate a single allelic change between STs, light grey lines indicate differences at two loci. Circles ringed with a yellow outline indicate putative founder genotypes. Colours within circles show the proportion of isolates of a particular type that were found in the countries indicated.

In Case 2, a different beekeeper with disease sold bees to another beekeeper 84 km away. Bacteria from both seller (26th May 2010) and buyer (28th May 2010) outbreaks were ST11. Again, this ST has been found nowhere else in the UK. The cases of ST9 and ST11 indicate that, when a rare ST is identified, such information can be used as evidence to support hypothesised routes of transmission.

In Case 3, diseased material was sampled from a shared apiary at three time points (27th April 2010, 24th June 2010 and 21st June 2011). The bacteria recovered from all three outbreaks were found to be ST6, a type only found at one other location in England, 3 km away from Case 3. The case of ST6 is more complicated than the examples of the sale of bees. ST6 was seen to persist in the apiary monitored in this study, despite treatment success at the colony level. There are two possible explanations for this. Latent *M. plutonius* could be present in untreated asymptomatic colonies (Budge et al., 2010). Such colonies would then have become symptomatic after the initial visit to control the disease. However, the close proximity of the other ST6 apiary (3 km) indicates that this could be an example of local, natural transmission events between the two sites. Here the typing scheme has allowed us to generate data in support of hypotheses for infection of colonies that would have been unsubstantiated without the genetic data.

Case 4 was comprised of a single apiary where all bees and equipment were imported into the UK from Poland, six weeks prior to disease being observed in 40% of colonies. EFB had not been found within 10 km of the apiary location in the previous 18 years. The bacteria recovered from this outbreak were ST14, which was not isolated from any other UK samples. It is possible that this instance depicts the import of novel STs into the UK. Three samples of *M. plutonius* were obtained from Poland to try to confirm this, but all belonged to ST13. It is therefore clear that more extensive sampling is required in the UK, and also Poland, before a firmer inference can be made. Indeed, more extensive sampling in the UK would improve our understanding of all of these case studies. If the STs involved were found to be absent in the buyers' regions then the conclusions would be strengthened, and if the STs were more widespread they would be weakened. Despite these caveats, these case studies suggest that human movement of bees is a route of dissemination of *M. plutonius* at large spatial scales. To that end, better hygiene testing (incorporating molecular testing of asymptomatically infected larvae, rather than relying on observation of symptoms) may reduce the risk of *M. plutonius* spread. The MLST scheme presented here has therefore shown that *M. plutonius* is not genetically homogenous across its geographical range, and is able to discriminate multiple types within groups of typical and atypical *M. plutonius*. Furthermore, the scheme's utility for making epidemiological inferences has been demonstrated.

4. Investigating the diversity of *Melissococcus plutonius* at a landscape level

4.1. Introduction

Understanding the geographical distribution of a disease is important in unravelling its epidemiology. Mapping a disease allows an immediate visualization of the extent of the problem, and can be used to plan control strategies on different levels. Disease maps can also provide a baseline for future assessment of interventions (Hay et al., 2013). The task of mapping diseases, and the analysis of spatial patterns in disease distribution, has become more practical due to the increasing use of geographic information systems (GIS) (Clarke et al., 1996; Cromley, 2003), automated tools for storing, mapping and analysing data. These systems, combined with improved data gathering and storage abilities, provide a standard against which to compare real-time outbreak alerts (Hay et al., 2013).

A well-studied example of the use of mapping to study a livestock disease is that of Bovine Tuberculosis, a serious, notifiable disease of cattle in the UK and Ireland with important wildlife reservoirs (including the European Badger, *Meles meles*), caused by the bacterium *Mycobacterium bovis*. Mapping of disease prevalence in badgers has shown geographical clustering of outbreaks, and allowed the impact of control methods to be assessed (Jenkins et al., 2007). Adding genetic data to these datasets can show clustering of pathogen genotypes in the landscape (Smith et al., 2003), similarities between local genotypes in different host species (Woodroffe et al., 2009; Goodchild et al., 2012; Balseiro et al., 2013), and introduction of non-local types to an area (Gopal et al., 2006). Even genomic data can be incorporated, allowing inferences to be drawn about disease spread at an incredibly fine scale, between multiple farms within 5km of each other (Biek et al., 2012).

Distribution of variants within and among regions can also be studied. Taxa-area relationships have been shown for bacteria (i.e. more bacterial taxonomic units are found in larger areas) (Horner-Devine et al., 2004), and patterns of between-site (β) diversity have been shown to change over different spatial scales. Salt-marsh Nitrosomonadale bacterial communities have been shown to be increasingly different from one another over distances of centimetres to hundreds of kilometres, though this pattern does not hold true at the continental (thousands to tens of thousands of kilometres) scale (Martiny et al., 2011). One driver of α (within-site) and β diversity seems to be heterogeneity of environment, especially the animal species present for animal-associated pathogenic (Dunn et al., 2010) and non-pathogenic bacteria (Dunn et al., 2013).

Another important driver is dispersal limitation, with organisms that have low dispersal rates being unable to override regional metacommunity processes that lead to population or community differentiation (Telford et al., 2006; Martiny et al., 2011).

Presented here is an investigation into whether *M. plutonius* populations showed geographical structuring, or whether all types were distributed across the study area of Western Europe. Investigation took place at two spatial scales; an intensive study of diversity in one region, England and Wales, and a comparison of diversity across four regions (England and Wales, Scotland, Switzerland and the Netherlands). The outcome revealed population structuring of *M. plutonius* both within and among countries, and allowed inferences about disease dispersal and control to be made.

4.2. Materials and Methods

4.2.1. Sampling

4.2.1.1. Sample Locations

M. plutonius isolates came from four different regions; England and Wales, the Netherlands, Scotland, and Switzerland. Isolates from England and Wales were obtained as part of a larger project, for which I developed and demonstrated the *M. plutonius* typing scheme (Chapter 3), and then typed around half of the isolates. For samples from other countries, I developed collaborations with international honey bee researchers to gain access to infected larval samples (Scotland and the Netherlands) or *M. plutonius* DNA extracts (Switzerland) (Table 4.1). I then processed these at Fera.

Large-scale sampling of diseased brood combs from across England and Wales took place in 2011 and 2012 as part of the, "Modelling systems for managing bee disease: the epidemiology of European foulbrood," project of the Insect Pollinators Initiative (IPI), led by Dr Giles Budge, at the Food and Environment Research Agency (Fera). Wales (low EFB prevalence) and Norfolk/Suffolk (high EFB prevalence) were sampled intensively, with a comb of brood taken from every infected colony. The rest of the country was sampled less intensively, and a comb of brood from every fourth infected apiary was sampled.

Country	Sample Reference	Culture Collection Reference	Year Collected	ST
The Netherlands	NL2205		2007	3
The Netherlands	NL2206	8512	2007	3
The Netherlands	NLJH1		2008	3
The Netherlands	NL2281	8510	2008	3
The Netherlands	NL2286	8513	2008	19
The Netherlands	NL2365-6	8515	2009	3
The Netherlands	NL2365-P		2009	3
The Netherlands	NL2491	8509	2010	3
The Netherlands	NL2492		2010	3
The Netherlands	NL2493		2010	3
The Netherlands	NL2566		2011	3
The Netherlands	NL2567		2011	19
The Netherlands	NL2579	8498	2011	3
The Netherlands	NL2642		2012	3
The Netherlands	NL2655	8507	2012	3
The Netherlands	NL2657-3	8501	2012	3
Scotland	FS09/114	7521	2009	2
Scotland	Scottish Honey	8566	2011	1
Scotland	FS11/15		2011	1
Scotland	FS11/28		2011	1
Scotland	FS11/32		2011	1
Scotland	FS11/56		2011	18
Scotland	FS11/67		2011	1
Scotland	FS11/68		2011	1
Scotland	FS11/71		2011	1
Scotland	FS11/84		2011	1
Scotland	FS12/06	8473	2012	18
Scotland	FS12/13	8570	2012	1
Scotland	FS12/26	8571	2012	1
Scotland	FS12/45	8573	2012	1
Scotland	FS12/52	8475	2012	1
Scotland	FS12/47		2012	1
Scotland	FS12/58	8574	2012	1
Scotland	FS12/71	8575	2012	1
Scotland	FS12/145		2012	1
Scotland	FS12/161		2012	1
Scotland	FS12/169		2012	18
Switzerland	37.6			3
Switzerland	38.3			3
Switzerland	39.3			27
Switzerland	41.4			3
Switzerland	42.4			7
Switzerland	44.4			7
Switzerland	45.4			3
Switzerland	46.4			25
Switzerland	49.4			3
Switzerland	50.3			3
Switzerland	51.3			3
Switzerland	53.4			3
Switzerland	55.4			7
Switzerland	56.4			3
Switzerland	57.4			26

Table 4.1

Switzerland	59			27
Switzerland	77			28
Switzerland	62			7
Switzerland	63			13
Switzerland	65			13
Switzerland	66			29
Switzerland	69			13
Switzerland	70			3
Switzerland	71			20
Switzerland	72			7
Switzerland	73			30
Switzerland	74			13
Switzerland	75			30
Switzerland	76			30
Switzerland	49.3			3
Switzerland	90			13

Table 4.1. Samples typed from the Netherlands, Scotland and Switzerland. Year of collection is shown for every isolate for which it is known. Samples from the Netherlands and Scotland without a culture collection reference were not isolated as cultures, and were typed from larval DNA extracts. All samples from Switzerland were supplied as DNA extracts from cultured bacteria. EFB-positive larvae from the Netherlands were supplied by Dr Sjef van der Steen and Dr Bram Cornelissen of Wageningen University, the Netherlands. EFB-positive larvae from Scotland were supplied by Dr Fiona Highet, Mairi Carnegie and Steve Sunderland of Science & Advice for Scottish Agriculture (SASA). DNA extracts from Swiss *M. plutonius* samples were supplied by Dr Laurent Gauthier of Agroscope, Switzerland.

4.2.1.2. Sample Processing for England and Wales

From sampled combs, three diseased larvae were picked at random from each comb, and transferred to individual sterile 1.5ml Eppendorf tubes using a fresh sterile matchstick for each larva. Larval homogenates were then plated out, and *M. plutonius* cultured anaerobically on M110 media (Forsgren et al., 2013). Colonies were identified as *M. plutonius* using an *M. plutonius*-specific Taqman® assay (Budge et al., 2010), DNA was extracted using Chelex ®(Bio-Rad), and a sample of bacteria stored at -80°C using the Protect Microorganism Preservation System (Technical Service Consultants Ltd).

The DNA extract from each comb that had the lowest CT value (that is, the highest amount of target) from the *M. plutonius* Taqman® assay was typed using the *M. plutonius* MLST scheme (Haynes et al., 2013). PCR was performed using Fermentas ® PCR Master Mix (2X). Products were assessed on a 2% agarose gel, using Fermentas ® 6X DNA Loading Dye, and Bioline HyperLadder ™ 100bp. PCR products were purified using the QIAGEN QIAquick PCR Purification Kit, and Sanger sequenced in one direction at MWG Eurofins using an ABI 3730xl 96-capillary DNA Analyzer. Any MLST loci that failed in either the PCR or the sequencing stage were first re-PCRed. If this failed, the bacteria were regrown from the Protect system, and DNA extracted using the Promega Wizard ® Magnetic DNA Purification System for Food. Any samples that failed again using this method were abandoned.

4.2.1.3. Sample Processing for the Netherlands, Scotland, and Switzerland

Samples of EFB-positive larvae from the Netherlands were vortexed in PBS, streaked and incubated anaerobically on M110 agar (Forsgren et al., 2013). Individual colonies were picked and spread on M110, then incubated. Resulting colonies were then directly tested with the *M. plutonius* Taqman® assay (Budge et al., 2010). Bacteria that tested positive for *M. plutonius* DNA underwent DNA extraction using the Promega Wizard ® Magnetic DNA Purification System for Food, and bacteria from each sample were stored at -80°C using the Protect Microorganism Preservation System. For samples from which culture could not be obtained, a DNA extract of larval homogenate was performed using the Promega Wizard ® Magnetic DNA Purification System for Food.

Larval samples from Scotland were treated as above, though some later samples had DNA extracted directly from larvae, without an initial culturing attempt. One isolate from Scotland was grown directly from a honey sample. In this instance a 10µl loop of honey was streaked on M110 agar and incubated anaerobically (Forsgren et al., 2013). Individual colonies were then spread on

to M110 agar and incubated before being identified as *M. plutonius* with the *M. plutonius* Taqman® assay (Budge et al., 2010). This honey sample was one of four honey samples purchased directly from a commercial beekeeper suspected of having *M. plutonius* in their operation. Two samples were of, "Blossom Honey," (neither containing cultivable *M. plutonius*), and two were of, "Heather Honey," (one containing cultivable *M. plutonius*). Samples were received in June 2012, so collection of honey from bee colonies must have taken place in the previous summer or autumn. *M. plutonius* was cultivated in November 2012, so the bacterium must have remained viable in the honey for approximately 18 months.

DNA extracts from Swiss *M. plutonius* were diluted 1:24 and then typed (Haynes et al., 2013). Some sequencing reactions failed, so original DNA was diluted to 1:99, and PCR performed again. All reactions were then successful. Samples were described as covering a broad geographical area of Switzerland.

4.2.2. Sequence Analysis

DNA sequences were aligned in clustalX2 to reference allele sequences to calculate allele designations and hence STs. Evolutionary relationships between all STs found so far, from Haynes et al., (2013) and this chapter, were calculated in the Phyloviz program (Francisco et al., 2012). Isolates from the IPI project were also assigned to Sequence Types using BioNumerics software (www.applied-maths.com/bionumerics), and all allele designation agreed with those identified manually using clustalX2.

4.2.3. Geographical Distribution and Diversity of STs

Distribution of STs in England and Wales, Scotland and the Netherlands was mapped using ArcMap 10.1. Evidence for geographical clustering of STs in England and Wales was assessed using Nearest-Neighbour Contingency Tables, with the dixon2002 command in the ecespa package (De la Cruz, 2008) in R (R Core Team, 2013). To assess clustering, all STs present in only one location were removed from the analysis. Additionally, all instances of the same ST being found in the same apiary on multiple occasions (either at different times or in different colonies) were removed, as this may have constituted a type of pseudoreplication.

The diversity of STs in England and Wales (all isolates, both from the IPI and Haynes et al., 2013), the Netherlands, Scotland and Switzerland was compared using rarefaction curves in the vegan package (Okansen et al., 2013) of R (R Core Team, 2013). Rarefaction analyses look at the number of species (or in this case STs of *M. plutonius*) found in an area as a function of the sample size. The ST sample pools for each country are randomly resampled for various sample sizes, to

calculate the number of expected STs per unit sampling effort. This enables a comparison of countries which have been sampled at different intensities.

4.3. Results

4.3.1. Typing of Isolates

In total, 30 STs have so far been identified (Table 4.2). Of the 206 isolates from the IPI sampling in England and Wales, 15 different STs were found. Four novel STs were found in the IPI samples that do not appear in Haynes et al., (2013). Three STs were detected in the Scottish samples and two from the Dutch samples, all of which were previously reported in Haynes et al., (2013). From the Swiss DNA extracts, 11 different STs were identified of which six were novel.

Four new alleles were identified at locus *galK* and five at *gbpB*. No new alleles were discovered at loci *argE* or *purR* (Appendix 2). The goeBURST output showing evolutionary relationships between the STs identified in the IPI samples (this study) and those found in Chapter 3 (Haynes et al., 2013) clearly illustrates the effect of the intensity of the IPI sampling, finding six STs in England and Wales for the first time (Figure 4.1). It also clearly shows the split between typical and atypical *M. plutonius* (Arai et al., 2012), which is confirmed when all STs identified to date are included (Figure 4.2).

Clear differences in the distribution of STs in each of our focal countries can also be seen. Both Scotland and the Netherlands have one endemic ST each (ST 18 and ST 19 respectively), and only one or two more common STs. Strikingly, Switzerland has more than twice as many STs belonging to Clonal Complex 13 (that is, single- or double-locus variants of ST 13) as Clonal Complex 3 (single- or double-locus variants of ST 3), whereas in England and Wales more STs belonging to Clonal Complex 3 are found. Atypical *M. plutonius* (Clonal Complex 12; single-, double- or triple-locus variants of ST 12) was not found in either Scotland or Switzerland.

4.3.2. Geographical Distribution of STs

The geographical distribution of STs in England and Wales is distinctly non-homogenous. The most common type, ST 3, is widespread across Southern England, the Midlands, Wales and Northern England. However, distinct local clusters of many other types are seen, which are especially evident in the focal areas of Wales and Norfolk/Suffolk. Particularly clear clusters of disease types are ST 2 in the West Country, ST 23 in the Welsh borders, and STs 5 and 13 in Norfolk/Suffolk (Figure 4.3).

ST	*galK*	*argE*	*gbpB*	*purR*
1	1	1	1	1
2	3	2	2	2
3	3	2	2	4
4	1	1	2	4
5	3	2	4	4
6	2	2	2	2
7	3	2	5	4
8	1	1	7	3
9	1	1	6	4
10	4	4	3	4
11	3	3	4	4
12	4	4	3	5
13	1	1	1	4
14	5	1	8	4
15	1	5	1	4
16	6	4	9	4
17	3	1	1	4
18	1	1	2	1
19	4	4	10	5
20	1	1	8	4
21	4	4	11	5
22	7	2	2	2
23	3	2	12	4
24	3	2	13	4
25	3	2	14	4
26	9	1	15	4
27	8	1	1	4
28	10	1	11	4
29	8	1	2	4
30	10	1	1	4

Table 4.2. Allele designations of all STs found to date. Sequence Types (ST) identified in this study are shown, and the four columns, *galK* (galactokinase), *argE* (acetylornithine deacetylase), *gbpB* (secreted antigen) and *purR* (purine operon repressor), show the alleles present at each of these four loci in each ST. Within a column each number represents a unique allele (that is, a unique DNA sequence) at that locus. Each ST has a unique allelic profile across all loci.

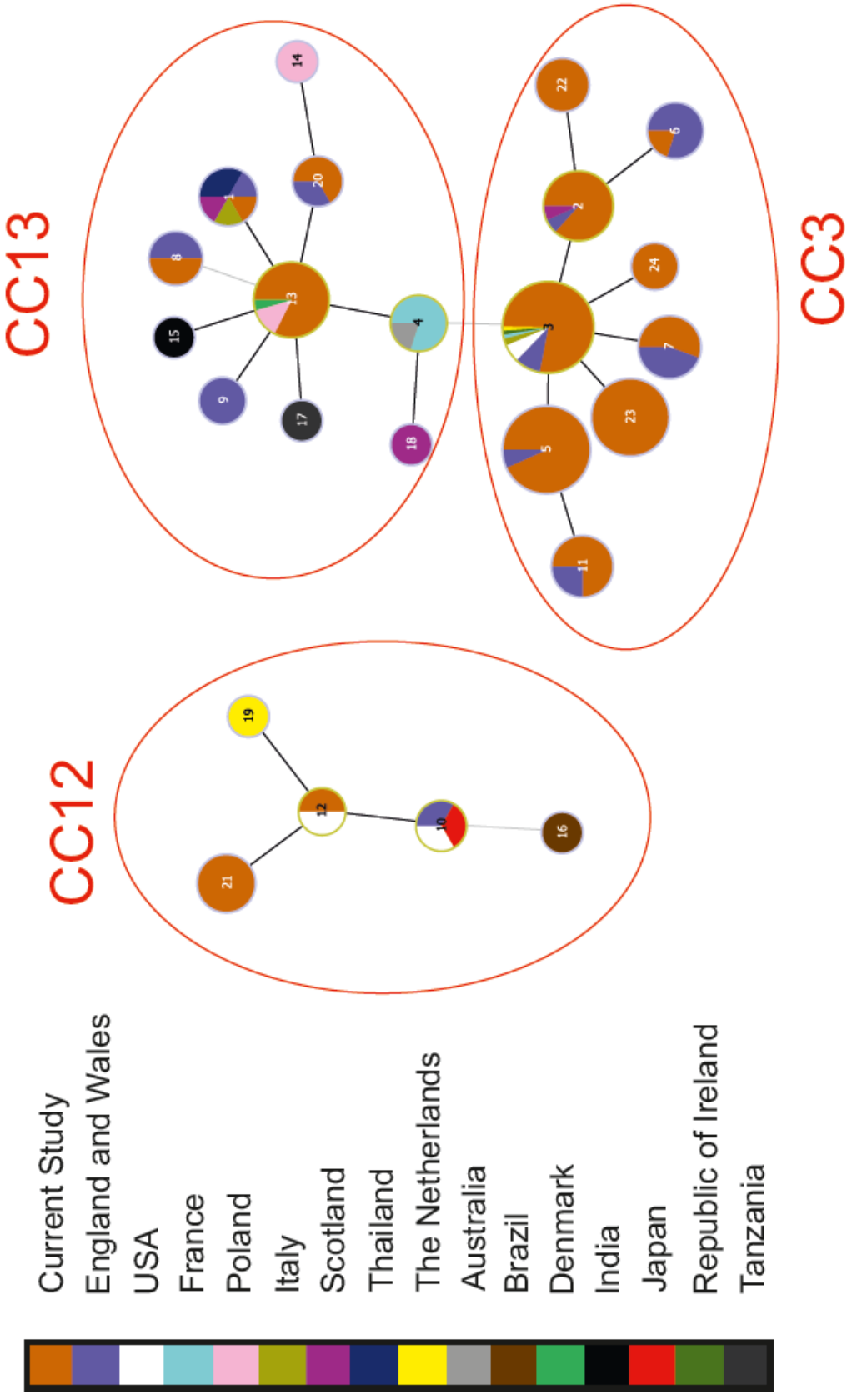


Figure 4.1. goeBURST figure showing evolutionary relationships of STs identified in Haynes et al., 2013 and the IPI samples (this study). Also labelled are the clonal complexes (CCs) to which different STs belong.

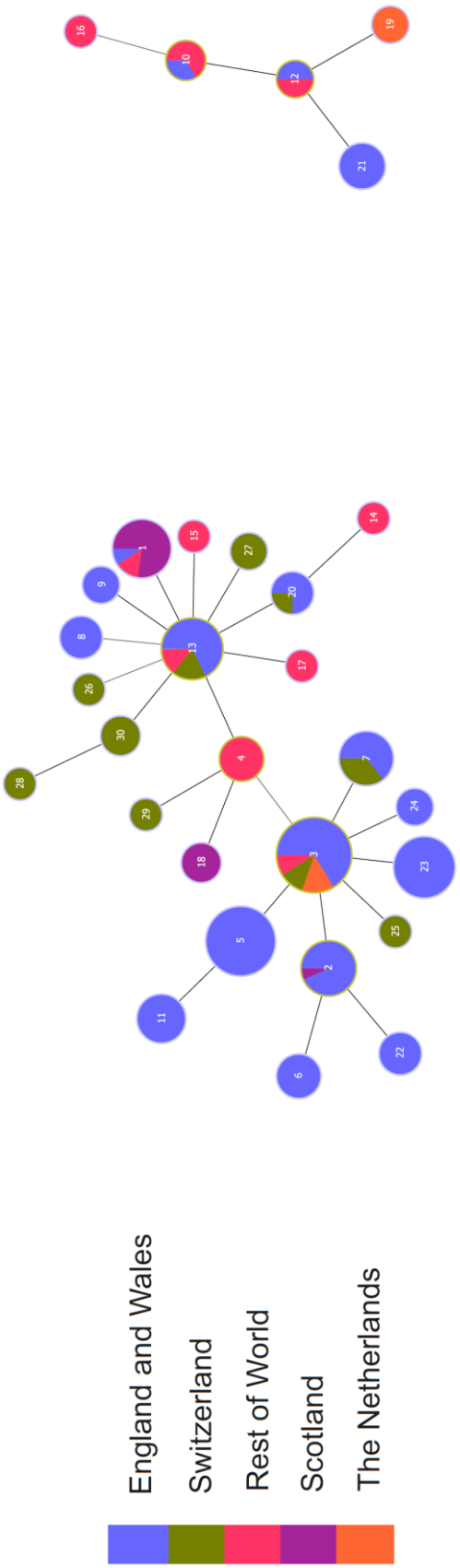


Figure 4.2. goeBURST figure showing evolutionary relationships of all STs identified to date. Clear differences in the distribution of STs in each of our focal countries can be seen.

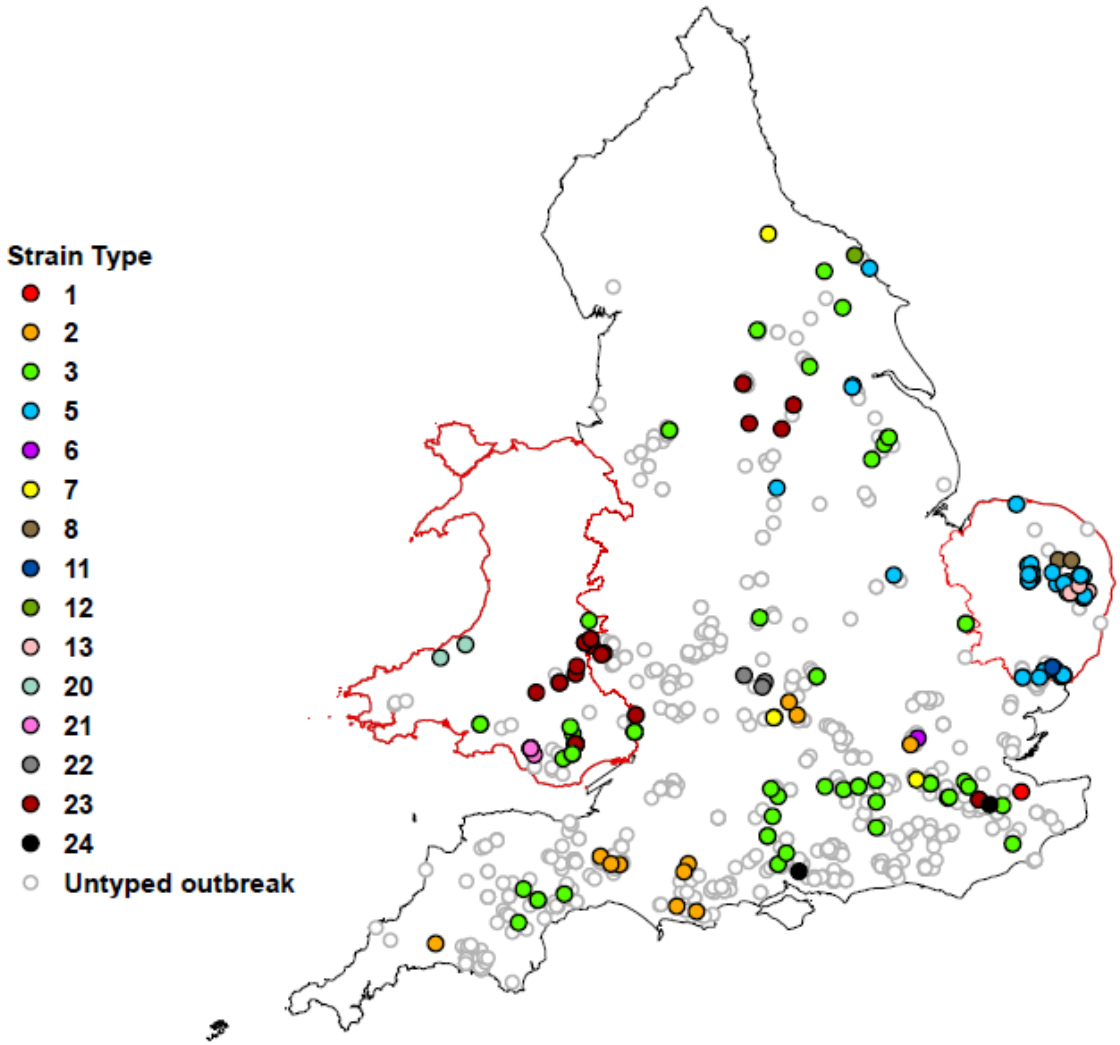


Figure 4.3. Distribution map of STs of *M. plutonius* found in England and Wales, from 2011 and 2012. Colours show the STs found at each location, and hollow grey circles represent untyped outbreaks. Regions that were sampled at high intensity (Wales and Norfolk/Suffolk) are outlined in red.

The distribution of STs in the Netherlands is very different. The majority of samples across the country, and across the years of sampling, are ST 3. Only two isolates of ST 19 were identified, and these are not geographically close to one another (Figure 4.4). In Scotland in 2011-12 all but three isolates from across the sample area, and across multiple beekeepers, were ST 1. The only other ST found in these years was ST 18 (Figure 4.5). Location data for Swiss samples were not provided.

4.3.3. Clustering of STs

The spatial segregation of STs was investigated using nearest-neighbour contingency table analysis (Dixon, 1994, 2002), and was performed on the IPI samples (Table 4.3). Segregation of an ST means it is more likely than expected to have a nearest neighbour that is of the same ST. STs 2, 3, 5, 7, 8, 13, 20, 21, 22, and 23 all show significant segregation. However, care should be taken when interpreting these results for STs with very low numbers of observed nearest neighbours of the same ST. For example, STs 7, 8, 20 and 21 only have two such points each. Positive or negative association of STs means that an ST is more or less likely than expected to have a nearest neighbour that is of a different ST. Significant negative associations are seen between ST 3 and ST 5, and ST 5 to ST 3 (i.e. ST 3 is less likely than expected to be near an ST 5, and *vice versa*), from ST 3 to ST 13 and ST 13 to ST 3, from ST 2 to ST 5 and from ST 23 to ST 5. These negative associations probably reflect the restricted geographical range, but extreme local abundance of ST 13 and ST 5 (totally and largely restricted to Norfolk respectively). This also underlies the sole positive association, of ST 13 to ST 5.

4.3.4. Diversity of STs

Rarefaction curves for each of the four regions show that, for a given sample size, many more STs would be expected to be found in England and Wales and Switzerland than in the Netherlands and Scotland. For example, with a sample size of 16, one would expect between two and three different STs in the Netherlands and Scotland, but around seven different STs in England and Wales and Switzerland. Furthermore, rarefaction curves for England and Wales, Scotland and the Netherlands are starting to plateau, indicating that few additional STs would be expected in these countries even if more sampling were to take place. The curve for Switzerland has yet to plateau, indicating that more STs are likely to be found there (Gotelli and Colwell, 2001)(Figure 4.6).

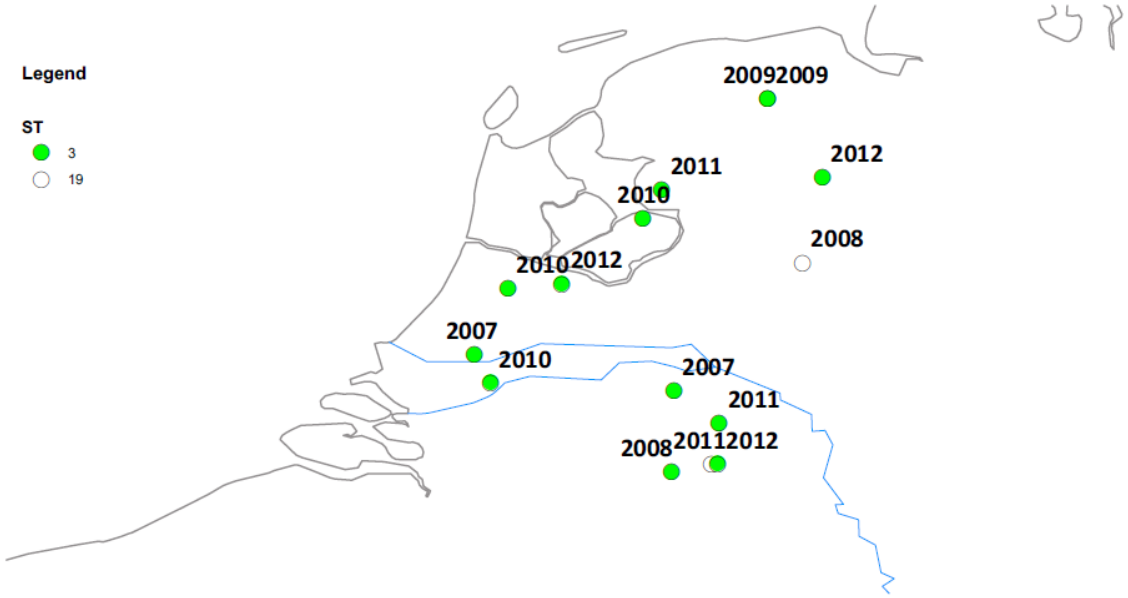


Figure 4.4. Distribution map of STs of *M. plutonius* found in the Netherlands. Points are labelled with their year of sampling. One isolate, NLJH1, did not have location data.

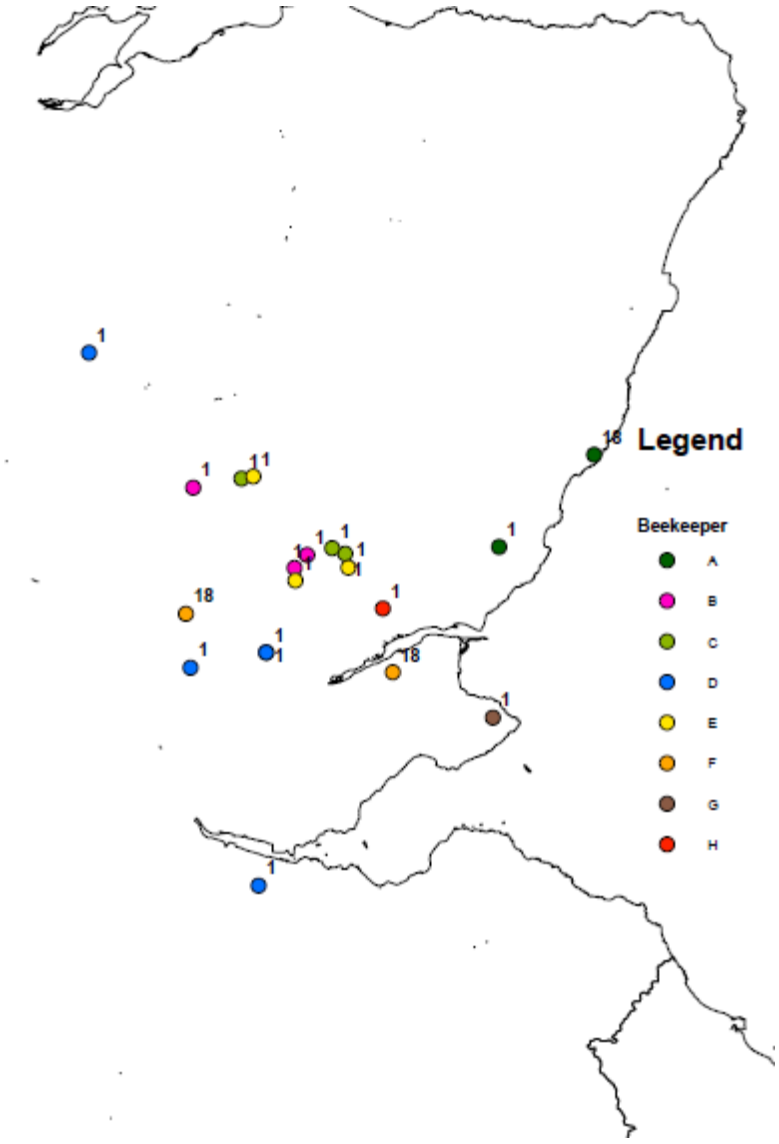


Figure 4.5. Distribution map of STs of *M. plutonius* found in Scotland in 2011 and 2012. Points are labelled with their ST, and coloured according to beekeeper ownership.

From	To	Observed Count	Expected Count	Z	p-val.as	p-val.rnd
2	2	9	0.97	6.83	0	0.0099
2	3	1	3.88	-1.78	0.0745	0.0792
2	5	0	2.91	-1.97	0.0483	0.0149
2	7	1	0.44	0.87	0.3845	0.5594
2	8	0	0.18	-0.43	0.6675	0.5099
2	11	1	0.35	1.12	0.2619	0.203
2	13	0	1.06	-1.09	0.275	0.2129
2	20	0	0.18	-0.43	0.6675	0.5198
2	21	0	0.18	-0.43	0.6675	0.5099
2	22	0	0.35	-0.61	0.5406	0.5099
2	23	0	1.32	-1.23	0.217	0.1832
2	24	0	0.18	-0.43	0.6675	0.5099
3	2	1	3.88	-1.61	0.1065	0.0594
3	3	31	13.91	4.83	0	0.0198
3	5	1	10.68	-3.5	0.0005	0.0099
3	7	1	1.62	-0.52	0.5997	0.7079
3	8	0	0.65	-0.86	0.3888	0.5545
3	11	0	1.29	-1.23	0.2203	0.2277
3	13	0	3.88	-2.17	0.0297	0.0248
3	20	0	0.65	-0.86	0.3888	0.3861
3	21	1	0.65	0.47	0.6383	0.5743
3	22	1	1.29	-0.28	0.7806	0.6584
3	23	7	4.85	1.09	0.2777	0.2178
3	24	1	0.65	0.47	0.6383	0.5545
5	2	1	2.91	-1.22	0.2228	0.1782
5	3	2	10.68	-3.26	0.0011	0.0099
5	5	23	7.76	5.29	0	0.0099
5	7	1	1.21	-0.21	0.8368	0.703
5	8	0	0.49	-0.73	0.4629	0.5297
5	11	1	0.97	0.03	0.9747	0.5941
5	13	3	2.91	0.06	0.9551	0.7426
5	20	0	0.49	-0.73	0.4629	0.5693
5	21	0	0.49	-0.73	0.4629	0.5297
5	22	1	0.97	0.03	0.9747	0.6386
5	23	1	3.64	-1.52	0.1283	0.104
5	24	0	0.49	-0.73	0.4629	0.5495
7	2	0	0.44	-0.7	0.4847	0.3119
7	3	2	1.62	0.37	0.7144	0.6535
7	5	1	1.21	-0.22	0.8235	0.6683
7	7	2	0.15	3.87	0.0001	0.0198
7	8	0	0.07	-0.27	0.7836	0.505
7	11	0	0.15	-0.39	0.6956	0.5099
7	13	0	0.44	-0.7	0.4847	0.5396
7	20	0	0.07	-0.27	0.7836	0.505
7	21	0	0.07	-0.27	0.7836	0.505
7	22	0	0.15	-0.39	0.6956	0.505
7	23	0	0.55	-0.79	0.4291	0.5446
7	24	0	0.07	-0.27	0.7836	0.505
8	2	0	0.18	-0.44	0.6596	0.5099
8	3	0	0.65	-0.98	0.3279	0.5396
8	5	0	0.49	-0.8	0.4231	0.3168
8	7	0	0.07	-0.28	0.7821	0.505

Table 4.3

8	8	2	0.01	12.89	0	0.0099
8	11	0	0.06	-0.25	0.8053	0.505
8	13	0	0.18	-0.44	0.6596	0.505
8	20	0	0.03	-0.17	0.8626	0.505
8	21	0	0.03	-0.17	0.8626	0.505
8	22	0	0.06	-0.25	0.8053	0.505
8	23	0	0.22	-0.5	0.6182	0.505
8	24	0	0.03	-0.17	0.8626	0.505
11	2	2	0.35	2.91	0.0036	0.0545
11	3	0	1.29	-1.38	0.1663	0.198
11	5	2	0.97	1.2	0.229	0.1733
11	7	0	0.15	-0.39	0.6949	0.505
11	8	0	0.06	-0.25	0.8062	0.505
11	11	0	0.09	-0.24	0.8129	0.5198
11	13	0	0.35	-0.62	0.5324	0.5099
11	20	0	0.06	-0.25	0.8062	0.505
11	21	0	0.06	-0.25	0.8062	0.5099
11	22	0	0.12	-0.35	0.7267	0.505
11	23	0	0.44	-0.71	0.4799	0.5297
11	24	0	0.06	-0.25	0.8062	0.505
13	2	0	1.06	-1.09	0.275	0.203
13	3	0	3.88	-2.4	0.0163	0.0099
13	5	7	2.91	2.77	0.0056	0.0099
13	7	0	0.44	-0.69	0.4923	0.5297
13	8	0	0.18	-0.43	0.6675	0.5149
13	11	0	0.35	-0.61	0.5406	0.5248
13	13	5	0.97	3.43	0.0006	0.0248
13	20	0	0.18	-0.43	0.6675	0.5198
13	21	0	0.18	-0.43	0.6675	0.5099
13	22	0	0.35	-0.61	0.5406	0.5297
13	23	0	1.32	-1.23	0.217	0.2277
13	24	0	0.18	-0.43	0.6675	0.5099
20	2	0	0.18	-0.44	0.6596	0.5149
20	3	0	0.65	-0.98	0.3279	0.2723
20	5	0	0.49	-0.8	0.4231	0.5198
20	7	0	0.07	-0.28	0.7821	0.505
20	8	0	0.03	-0.17	0.8626	0.505
20	11	0	0.06	-0.25	0.8053	0.505
20	13	0	0.18	-0.44	0.6596	0.505
20	20	2	0.01	12.89	0	0.0099
20	21	0	0.03	-0.17	0.8626	0.505
20	22	0	0.06	-0.25	0.8053	0.505
20	23	0	0.22	-0.5	0.6182	0.505
20	24	0	0.03	-0.17	0.8626	0.505
21	2	0	0.18	-0.44	0.6596	0.505
21	3	0	0.65	-0.98	0.3279	0.3168
21	5	0	0.49	-0.8	0.4231	0.2921
21	7	0	0.07	-0.28	0.7821	0.505
21	8	0	0.03	-0.17	0.8626	0.505
21	11	0	0.06	-0.25	0.8053	0.505
21	13	0	0.18	-0.44	0.6596	0.5099
21	20	0	0.03	-0.17	0.8626	0.505
21	21	2	0.01	12.89	0	0.0149
21	22	0	0.06	-0.25	0.8053	0.5099

Table 4.3

21	23	0	0.22	-0.5	0.6182	0.505
21	24	0	0.03	-0.17	0.8626	0.505
22	2	0	0.35	-0.62	0.5324	0.5099
22	3	0	1.29	-1.38	0.1663	0.1782
22	5	1	0.97	0.03	0.9726	0.6485
22	7	0	0.15	-0.39	0.6949	0.505
22	8	0	0.06	-0.25	0.8062	0.505
22	11	0	0.12	-0.35	0.7267	0.505
22	13	0	0.35	-0.62	0.5324	0.5198
22	20	0	0.06	-0.25	0.8062	0.505
22	21	0	0.06	-0.25	0.8062	0.505
22	22	3	0.09	7.81	0	0.0099
22	23	0	0.44	-0.71	0.4799	0.5297
22	24	0	0.06	-0.25	0.8062	0.505
23	2	0	1.32	-1.22	0.2207	0.2277
23	3	4	4.85	-0.47	0.6366	0.7475
23	5	0	3.64	-2.21	0.027	0.0248
23	7	0	0.55	-0.77	0.4409	0.5594
23	8	0	0.22	-0.48	0.6296	0.5099
23	11	0	0.44	-0.69	0.4921	0.5446
23	13	0	1.32	-1.22	0.2207	0.2475
23	20	0	0.22	-0.48	0.6296	0.5198
23	21	0	0.22	-0.48	0.6296	0.5099
23	22	0	0.44	-0.69	0.4921	0.5594
23	23	10	1.54	5.81	0	0.0099
23	24	1	0.22	1.7	0.0883	0.0891
24	2	0	0.18	-0.44	0.6596	0.5099
24	3	1	0.65	0.53	0.5936	0.5297
24	5	0	0.49	-0.8	0.4231	0.5248
24	7	0	0.07	-0.28	0.7821	0.505
24	8	0	0.03	-0.17	0.8626	0.505
24	11	0	0.06	-0.25	0.8053	0.505
24	13	0	0.18	-0.44	0.6596	0.5099
24	20	0	0.03	-0.17	0.8626	0.505
24	21	0	0.03	-0.17	0.8626	0.505
24	22	0	0.06	-0.25	0.8053	0.505
24	23	1	0.22	1.76	0.0782	0.1089
24	24	0	0.01	-0.1	0.9239	0.505

Table 4.3. Output from the Nearest-Neighbour Contingency Table analysis. The From and To columns show the direction of the association being tested. Observed Count is the number of instances where the ST in “From” had the ST in “To” as a Nearest Neighbour, and Expected count is the expected number of instances, if the STs were randomly distributed around the network of sample locations. The Z score shows the strength and direction of the association; a large, positive Z means that STs are strongly positively associated (i.e. are more likely than expected to be found near each other), and a large, negative Z means that STs are strongly negatively associated. The p-values show whether the segregation/association is significant. p-val.as is the p-value of the asymmetric chi-squared test (an approximate fit). p-val.rnd is obtained through Monte Carlo simulation (for making decisions in marginal cases).

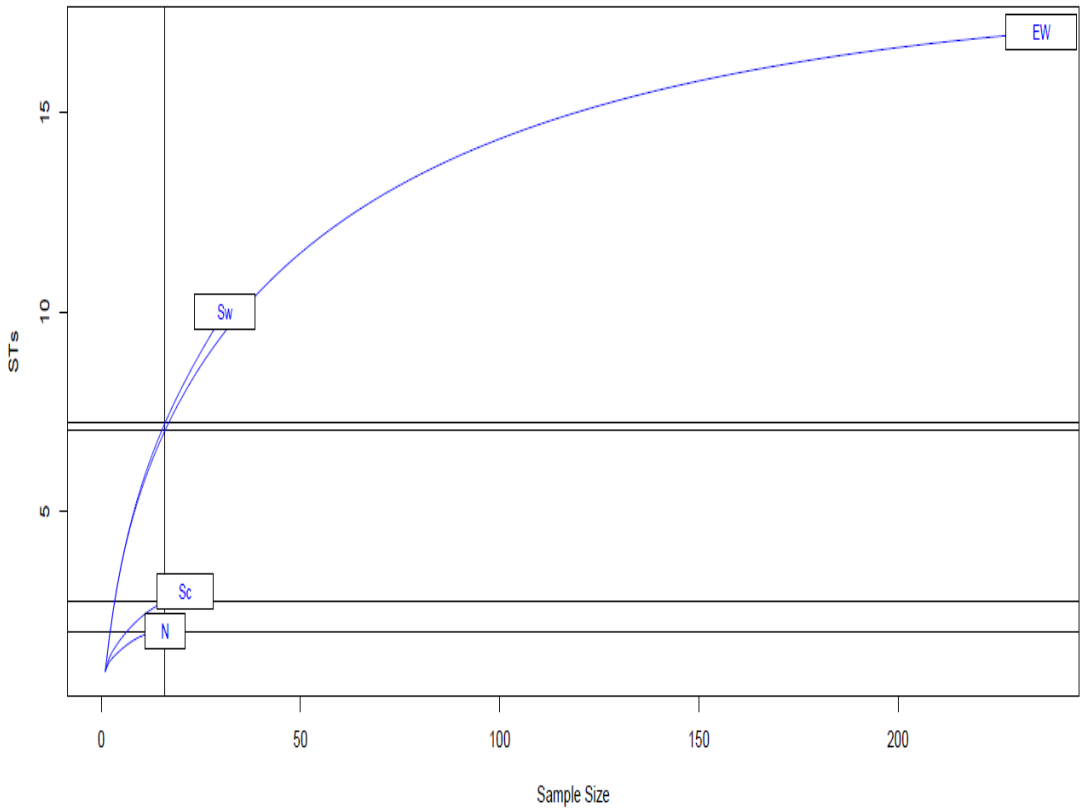


Figure 4.6. Rarefaction curves of the STs of *M. plutonius* found in England and Wales (EW), Scotland (Sc), Switzerland (Sw) and the Netherlands (N). Y-axis shows the number of STs expected in a country for a given number of samples taken (x-axis). The curves show that, for a given sample size, more STs would be expected in England and Wales and Switzerland than in the Netherlands and Scotland.

4.3.5. Beekeeper Ownership

A beekeeper ownership network with ST data was created by considering each sample as a vertex, and then joining all samples that contain a common owner with an edge. The colour of each vertex shows its ST. The network shows 97 owners with a single ST, four beekeepers with two STs, and two beekeepers with three STs (Figure 4.7).

4.4. Discussion

4.4.1. EFB in England and Wales

In comparison with the only previous, detailed study of the population genetics of UK *M. plutonius* (Haynes et al., 2013), this study has found a greater number of STs in England and Wales. Four new STs have been identified (ST 21, which grouped with atypical *M. plutonius*, and STs 22, 23 and 24, which grouped with typical *M. plutonius*), with one new allele at the *galK* locus and three at *gbpB*. In addition, two STs (atypical ST12 and typical ST13) were found in the UK for the first time, having previously been isolated from the USA, and Denmark and Poland respectively (Haynes et al., 2013). This is consistent with the greatly increased sampling effort compared with the previous study, and demonstrates that the present investigation gives a fuller and more accurate picture of *M. plutonius* diversity in England and Wales. This study is also the first to identify the Type Strain (ST 1) in the wild in England since it was first isolated by Bailey in the mid-twentieth century (Bailey, 1957). This ST has now been found in England, Scotland, Italy and Thailand.

The *M. plutonius* communities in England and Scotland differed markedly in the STs present. Scotland was dominated by ST 1 (found in England only in a single isolate from Kent) and ST 18, which was not found in England at all. This implies extremely limited mixing of populations across the border. There is also no evidence for the earlier presence of ST 14 in the UK. This was a type thought to be imported into the UK from Poland in 2012 (Haynes et al., 2013), and its absence before the import event supports this hypothesis.

Additionally, more light can be shed on a specific incidence of the spread of EFB around England. One of the outbreaks identified by Haynes et al. (2013) demonstrated the transport of ST 9 from one beekeeper to another over 54km. This outbreak occurred in one of our high-intensity sampling areas, and thus every infected colony in the region has been sampled. During the two years following the outbreak, no further ST 9 is found, either in this region or in England and Wales as a whole. Since one beekeeper involved has ceased operations, and the other is free of

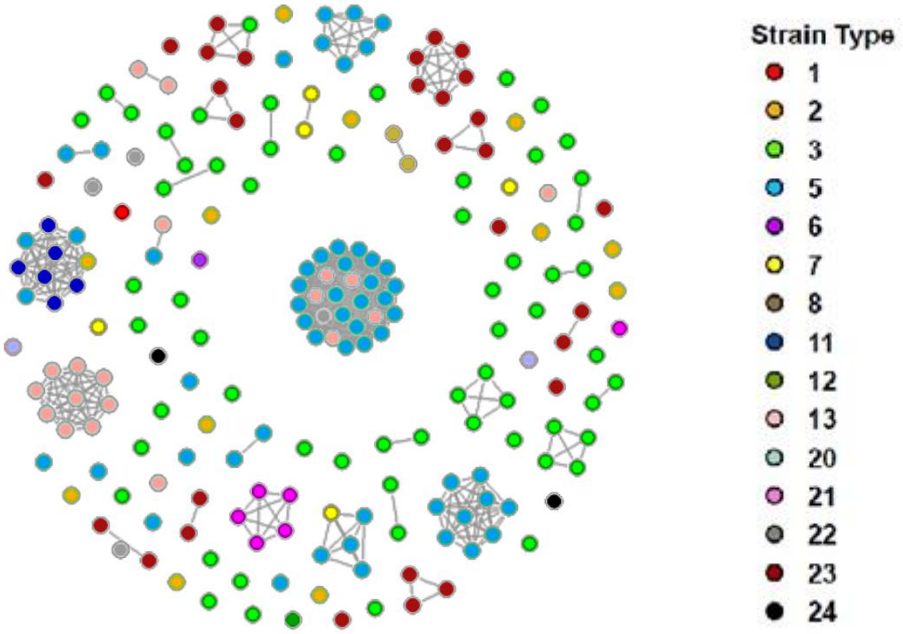


Figure 4.7. Beekeeper ownership network. Each point in the figure shows a sample of *M. plutonius* that has been typed. Colours show the ST identified. Lines link samples from bees that have a common owner.

ST9, it is suggested that this type has successfully been eradicated from the UK. The absence of this type from the rest of the country, and beekeeper records which show that the source of infection in this outbreak has no history of honey bee imports, suggest the very recent evolution of this type. This argument is strengthened by the fact that ST 9 only differs from ST 13 (a locally circulating type) by a length variation in the VNTR (Variable Number Tandem Repeat) at *gbpB*. VNTRs have been found to show variation in even highly genetically homogenous bacteria (Jackson et al., 1997).

As the beekeeper ownership network shows, several beekeepers with large networks of colonies have the same (or largely the same) type of *M. plutonius* across their operation. This demonstrates that in many instances, disease persists within a beekeeper's operation, and suggests that the beekeeper may play a role in this. Inferences can also be made about the spread of disease between beekeepers; where types are largely found within one person's hives (e.g. ST13) this might be an example of partially successful control measures, with disease spreading through one owner's apiaries before being detected, and prevented from being transmitted to neighbours. However, when a beekeeper (e.g. the central beekeeper in the ownership network) has different colonies infected with multiple types which are simultaneously infecting their neighbours, this strongly suggests the natural spread of disease between beekeepers (for example by drifting of drones between apiaries, or robbing of honey by worker bees).

Mapping ST distribution shows apparent clustering of types within the landscape of England and Wales. The clustering of many types is shown to be statistically significant by the nearest neighbour contingency table analysis, and when considered relative to the beekeeper ownership/disease relationship discussed above it can be seen that different clusters are probably maintained either by beekeepers spreading within their operations (e.g. ST13, largely confined to one beekeeper, with some spill over), or bees spreading a type around a local area (e.g. ST2, of which there is a large cluster in South West England that is spread across many beekeepers).

The maps also allow inference about the history of EFB spread to the UK. ST3 is the most widespread variant, and was also the most common ST in England and Wales. This implies that ST3 has been present in England and Wales longer than many of the other STs, and indeed is ancestral to some of them (Smith et al., 2003) (e.g. ST5, which is one SNP different from ST3 and largely confined to Norfolk, but locally highly abundant). Despite ST 3 having a wide geographical distribution, and appearing to give rise to localised variants such as ST 5, it does not seem to be the oldest variant. The Type Strain (NCDO 2443) is an ST 1, and was isolated by Bailey in England in the 1950s (Bailey, 1957; Bailey Pers. Comm.). For this to have been isolated from the wild, it is likely to have been much more abundant 60 years ago than it is now. It has now been almost

entirely displaced by ST 3, which could be either a recent introduction or could be better suited to survive under the current conditions.

At the other end of the spectrum, ST13 is also locally abundant in Norfolk but is confined to a few beekeepers. This type is not closely related to any other that is widely circulating in England and Wales, but ST13 is found on continental Europe (Denmark and Poland, (Haynes et al., 2013) and Switzerland (this study)). It therefore appears that ST13 in England shows a more recent introduction of an *M. plutonius* variant into the UK (like ST14 (Haynes et al., 2013)), but in this case control measures have failed to prevent it becoming locally established.

4.4.2. EFB across different regions

We see large differences in the diversity of *M. plutonius* found across four regions of Europe. In general, *M. plutonius* populations in England and Wales and Switzerland were comparatively diverse, and *M. plutonius* populations in Scotland and the Netherlands were comparatively depauperate. The rarefaction curves demonstrate that this is not an artefact of smaller sample size in the less diverse countries, an observation supported by the spatial and temporal distribution of samples from the Netherlands (spread across the whole country from 2007-2012) and Scotland (from a broad area of the country where almost all EFB is found (NBU, 2013), from 2009, 2011 and 2012). Figure 4.2 also shows dissimilarity between the STs found in each country. This implies limited mixing of STs between countries, and shows, for example that the Scottish *M. plutonius* population is not just a subset of the wider UK population, but is distinct from that found in England and Wales.

Environmental heterogeneity is a potential driver of differentiation of bacterial populations. With no known environmental reservoir of *M. plutonius*, the only place where the bacterium is capable of reproducing is inside the honey bee gut. *M. plutonius* can infect at least three species of honey bee, *Apis mellifera*, *Apis cerana* (Bailey, 1974) and *Apis laboriosa* (Allen et al., 1990), but the only honey bee species present in Europe is *A. mellifera*. However, within the *A. mellifera* species there is considerable genetic variability. In Europe there are two major mitochondrial lineages which spread northwards from refugia after the last glacial maximum; the M lineage to Western Europe and the C lineage to Eastern Europe (De La Rua et al., 2009). These lineages are further subdivided into different races of bees. Lineage M contains *A. m. mellifera* and *A. m. iberiensis* (itself a hybrid from the Iberian peninsula of *A. m. mellifera* and the North African *A. m. intermissa* (De La Rua et al., 2009)) and lineage C contains *A. m. ligustica*, the Italian bee, and *A. m. carnica* (Whitfield et al., 2006).

With the advent of modern beekeeping, these races have been extensively transported around Europe. Queens of *A. m. ligustica*, the Italian Bee, have been widely introduced to Northern Europe (De La Rua et al., 2009), and English honeybee populations show significant introgression of Italian genetic material (Jensen et al., 2005). In Central Europe, much of the native *A. m. mellifera* stock has been totally replaced with imported *A. m. carnica*, though in Switzerland strenuous attempts are being made to maintain some pure *A. m. mellifera* (De La Rua et al., 2009). On the other hand, a population of honeybees from North-West Scotland was shown to be relatively pure *A. m. mellifera* (Jensen et al., 2005). The Netherlands, by contrast, despite not being listed in the literature as a country with high levels of *A. m. mellifera* replacement (Meixner et al., 2010), shows some evidence of introgression of *A. m. carnica* and Buckfast (a hybrid) bees in the country (van der Zee and Pisa, 2011).

What this means is that, while it is possible that the diversity of the host population differs among these countries, the situation is predictably complex. Populations of honeybees in England and Switzerland contain large amounts of genetic material from two or more races, but those of the Netherlands and Scotland may or may not have higher proportions of *A. m. mellifera*. Whether this has any impact on the variants of *M. plutonius* that are able to persist in these countries is itself highly speculative. It is worth noting, however, that there has been some suggestion that *A. m. mellifera* is more susceptible to EFB than *A. m. ligustica* (VanEngelsdorp and Meixner, 2010). *In vivo* inoculation assays may provide a partial answer. Inoculating larvae from each of the different races found in Europe with an ST from each of the clonal complexes would show whether any race is more or less susceptible to particular lineages of *M. plutonius*.

Another driver of differentiation of bacterial populations is limited dispersal ability – if bacteria are not able to disperse easily between areas, the homogenising effects of movement will not be able to overcome the evolutionary processes that cause variation in populations (Telford et al., 2006; Martiny et al., 2011). From the geographical clustering seen within England and Wales, it may be inferred that some limit on dispersal of *M. plutonius* exists, especially if the predominant transmission mechanism is beekeepers transferring disease within their own operations. At an international level, the risk of transmission of EFB through honey bees and honey bee products has been assessed as low (Mutinelli, 2011). Conversely, there are known instances where *M. plutonius* has been transferred anthropogenically over tens, hundreds (Haynes et al., 2013) and even thousands (see Chapter 2) of kilometres, within and between countries. It has also been shown here that viable *M. plutonius* can survive in honey for 18 months, making it a potentially important reservoir or transmission route.

All four countries either have high levels of EFB infection, *M. plutonius*-presence in colonies, or have suffered recent outbreaks (Belloy et al., 2007; Wilkins et al., 2007; Sunderland et al., 2013; van der Steen, pers. comm.), and all have declining numbers of both beekeepers and bee colonies (Potts, Roberts, et al., 2010) indicating host population dynamics in all countries are similar. However, they do not all treat EFB in the same way. In Switzerland and England and Wales, an active, risk-based inspectorate regime inspects honey bee colonies for signs of disease, and inspects all apiaries within a certain radius of an infected hive (NBU, 2009; Agroscope, 2009). In Scotland, the system relies more on self-reporting of infections (The Bee Diseases and Pests Control (Scotland) Order 2007), and in the Netherlands EFB is not notifiable, and no government-backed inspectorate exits (van der Steen, pers. comm.). Within England and Wales, the targeted inspection regime may be successful in some areas in preventing transmission between beekeepers. If this is the case it seems possible that countries with a more active inspectorate have *M. plutonius* populations with reduced dispersal ability and a continual supply of treated hives, which are free of pathogen but susceptible to infection by invading types. In a low inspecting country, the invasion of new types could be inhibited by priority effects.

Inhibitory priority effects are seen when an organism moves into a new niche and modifies it or utilises available resources in such a way that subsequent invasion or competition by another type of organism is suppressed (Fukami et al., 2005). These effects are seen in microbial communities (Kennedy and Bruns, 2005; Kristin and Miranda, 2013), and some bacteria have shown strong priority effects only with respect to close phylogenetic relatives (Tan et al., 2012). If established *M. plutonius* within a colony were able to some extent to inhibit the invasion of a novel type, then it is possible that this could contribute to the patterns of diversity witnessed here. Consider the following scenarios; if introductions of *M. plutonius* into a country are assumed to be rare (Mutinelli, 2011), and if a country with low-intensity inspections has increased within-country dispersal of *M. plutonius*, then it would be predicted that most bee colonies in a low inspection intensity country (that are susceptible to EFB) rapidly become infected with a single type of *M. plutonius*. Inhibitory ecological priority effects would then reduce the likelihood of a novel type of *M. plutonius* successfully invading. In a high inspection intensity country, not only is the realised dispersal of the disease potentially reduced, but available niches (that is, susceptible but uninfected honey bee colonies) are continually being opened by the treatment of diseased hives. These niches can then be reinfected by new types of *M. plutonius*.

There are other potential explanations of the differences in diversity seen between these regions, including differences in disease virulence, or latency until symptoms appear. STs that spread faster, produce more infectious particles or have shorter incubating periods (the time between infection and the onset of symptoms) are more likely to spread rapidly to dominate a system with

no treatments. In a system with treatments, these are the STs that are most likely to show symptoms, and therefore be detected and treated. Pathogen variants that have longer latency periods or produce weaker symptoms are more likely to be missed by inspections, and will have the opportunity to persist and spread. What one would therefore expect to see would be a pattern of few, highly virulent types in low-inspection areas, and many, low-virulence types in high-inspection areas.

Alternatively, an established *M. plutonius* population in a country without disease treatments may evolve towards lower virulence. This would be due to the trade-off of high virulence causing increased transmissibility, whilst simultaneously reducing the susceptible host population. Evolution towards lower virulence may be expected over long timescales, but over the short term avirulent pathogens will be outcompeted by their more virulent (and transmissible) relatives (Lenski and May, 1994).

At any rate, preliminary observations do not support a simple relationship between virulence and control regime. Data from Swiss collaborators shows ST 3 to have high virulence, and ST 1 and ST 13 to have low virulence (Gauthier, pers comm.), an observation supported by Budge et al., 2014. ST 3 (high virulence) accounts for the vast majority of Dutch isolates. However, the Scottish isolates are almost all ST 1 (low virulence). Similarly, the most common STs in the England and Wales and Switzerland are from both Clonal Complex 3 and Clonal Complex 13.

Nonetheless, these simple high and low virulence attributions do not take into account disease latency. It is also possible that the isolates circulating in these countries have acquired mobile genetic elements that would affect their virulence, and would be undetected by the MLST. Further work, including *in vitro* assays of virulence and latency of *M. plutonius* isolated from the countries in question, is necessary to see what role if any virulence may play.

These suggestions remain highly speculative, and further work is required to assess their validity. This includes assessment of the frequency with which multiple types of *M. plutonius* are found in the same apiary/colony/larva. Assays should also be performed to test whether *M. plutonius* does exhibit priority effects, both *in vitro* in culture or larval bioassays and *in vivo* at the bee colony level. If *M. plutonius* does show priority effects, and they are strongest between closely related types (Tan et al., 2012), then that may explain why the only two types present in the totally non-inspecting Netherlands are distantly related variants of typical and atypical *M. plutonius*.

The work presented here shows for the first time that different variants of *M. plutonius* cluster geographically in the landscape, and that these clusters can often be maintained by spread by

beekeepers. Major differences are shown in the diversity of *M. plutonius* in different countries, and several, possibly connected hypotheses are offered as to how these differences in populations are maintained. However, further work is necessary to elucidate the precise mechanisms.

5. Reservoirs and Transmission Routes of *M. plutonius*

5.1. Introduction

Melissococcus plutonius is an important bacterial pathogen of honey bees, causing the serious larval disease European Foulbrood (EFB) (Forsgren, 2010). Most work on the spread of *M. plutonius* to date has focussed on the role of the hosts themselves, looking at where on bees' bodies, and at what point in their lifecycle, the bacteria can be detected. The only internal tissue in living bees in which multiplying bacteria have been found is in the lumen of the larval midgut (Bailey, 1959b; Forsgren, 2010). Some larvae are able to survive infection and pupate, and these new adults can carry *M. plutonius* on the exterior of their bodies (Bailey, 1959a) and in their faeces, wherein the bacteria can survive desiccation for many months (Bailey, 1959b) despite being non-sporeforming (Bailey and Collins, 1982). Bees that are involved in hygienic behaviour (the cleaning of cells and the removal of infected larvae) may themselves pick up and transmit pathogens around a colony (Evans and Spivak, 2010; Evans and Schwarz, 2011), and *M. plutonius* has been detected on both nurse bees (adults who remain on brood comb, tending the larvae) and forager adults (those which leave the hive to forage for food) (Roetschi et al., 2008). The bacterium is also capable of persisting asymptomatically in larvae (Forsgren et al., 2005; Budge et al., 2010). Thus, the bees themselves are likely to be an important reservoir of disease. The pathogen is also found in honey (see Chapter 4, and McKee et al., 2003), suggesting robbing of honey by adult bees as a potential inter-colony transmission route.

In comparison to the honey bees and their products, very little work has been done identifying alternative reservoirs and transmission mechanisms of the pathogen. Natural spread by non-*Apis* insect species appears non-existent. There has been a suggestion that Varroa mites can spread *M. plutonius* to individual bees (Kanbar et al., 2004), but the mite itself is spread between colonies by robbing and drifting of bees (Sammataro et al., 2000) so would be unlikely to contribute to bacterial spread more than the actions of the bees themselves. Presence of an *M. plutonius* reservoir in other social hymenopteran genera has not been demonstrated in either targeted screening (van der Steen and Blom, 2010; Graystock et al., 2013) or broader gut microbiome investigations (Reeson et al., 2003; Mrázek et al., 2008; Koch and Schmid-Hempel, 2011; Martinson et al., 2011), and the current understanding is that the bacterium only multiplies in the larval honey bee gut (Forsgren, 2010). Furthermore, little work has been done on the anthropogenic transmission of disease. Anecdotal evidence suggests that beekeeping practises play a role in spread, and beekeepers have been shown to perpetuate variants of *M. plutonius*

within their own operations (see Chapters 3 and 4). However, no study appears to have screened an apiary or beekeeper for possible *M. plutonius* transmission routes.

Many methods have been developed for the detection of *M. plutonius*, including the use of conventional PCR (Djordjevic et al., 1998; Govan et al., 1998), antibody-based lateral flow devices (Tomkies et al., 2009), gold nanoparticles (Saleh et al., 2012) and real-time PCR (Roetschi et al., 2008; Budge et al., 2010). The major advantage of real-time PCR is that it allows a quantitative analysis of the amount of target DNA present. Several chemistries exist, including dye- and probe-based systems. Both *M. plutonius* real-time PCR assays are Taqman® assays, which are probe-based, with one targeting the *M. plutonius* 16S rRNA gene (Budge et al., 2010), and one the *sodA* gene (Roetschi et al., 2008). Briefly, Taqman® involves the annealing of an oligonucleotide probe to the target DNA sequence. The probe is covalently attached to a fluorophore dye at the 5' end (e.g. FAM, or 6-carboxyfluorescein), and a quencher molecule at the 3' end (e.g. TAMRA, or tetramethylrhodamine) that prevents fluorescence. PCR primers anneal to the target DNA on either side of the probe's annealing site, and the 5' to 3' exonuclease activity of DNA polymerase (during PCR amplification) causes degradation of the probe. This releases the fluorophore from the proximal influence of the quencher (Shipley, 2006) and provides an increase in fluorescence. The amount of target DNA present can be quantified, either relative to other samples or absolutely to a known standard, because the amount of fluorescence released during each PCR cycle is proportional to the amount of product amplified in each PCR cycle.

In this study we use a sensitive real-time PCR assay (Budge et al., 2010) to identify the presence of *M. plutonius* DNA to infer pathogen reservoirs and routes of transmission within and between apiaries. Natural spread by the host and owner induced spread, due to husbandry practices, were each considered.

5.2. Methods

5.2.1. Validation of DNA extraction methods

Many of the substrates studied had not previously been validated for detecting the presence of *M. plutonius* and therefore some initial validation was required. For each extraction a pellet of *M. plutonius* cells from culture was suspended in 100 µl molecular grade water (MGW), and serially diluted four times (10 µl of cell suspension into 90 µl MGW).

5.2.1.1. Soil samples

Soil samples from within the grounds of the Food and Environment Research Agency (Fera) were autoclaved for 45 minutes at 121°C and then used as a substrate for the soil DNA extraction protocol. Each of four soil samples (4 g) was inoculated with one of the serially diluted *M. plutonius* cell suspensions and processed using a modified DNA extraction protocol (Budge et al., 2009; Woodhall et al., 2012).

Briefly, 4g of each soil sample was deposited in a 60 ml plastic Nalgene® bottle (to which the *M. plutonius* suspension was added) containing 8 ml of soil lysis buffer (120 mM sodium phosphate buffer pH 8, 2% CTAB, 1.5 M NaCl), 250 µl antifoam B emulsion (Sigma-Aldrich®) and ten steel ball bearings (10 mm). The bottle was then shaken on an automix high speed paint mixer for 2 minutes to grind and homogenise the sample, reducing intrasample variance. Next, 4 ml of lysate was aliquoted into a 5 ml tube and centrifuged for 2 minutes at 2,000g. Supernatant (1 ml) was then removed and transferred to a fresh 2 ml tube containing 250 µl of Buffer B (Promega Wizard ® Magnetic DNA Purification System for Food). This mixture was then vortexed before the addition of 750 µl precipitation solution (Promega Wizard ® Magnetic DNA Purification System for Food) and mixed by inversion. Each sample was then spun in a microcentrifuge at approximately 8,000 g for 10 minutes prior to the removal of 750µl supernatant. Supernatant was transferred to a clean 2 ml tube containing 50 µl of vortexed Magnesil beads (Promega Wizard ® Magnetic DNA Purification System for Food) and 600 µl isopropanol. The sample was vortexed and 1 ml of the mixed solution added to the first well of a KingFisher mL Magnetic Particle Processor (Thermo Scientific®). The beads were then passed through Buffer B, twice through 70% Ethanol before being eluted into 200 µl TE Buffer. Eluted samples were stored at -20°C until required.

5.2.1.2. Adult and Larval bees

The protocol for extracting DNA from honey bees was developed to detect the presence of *Nosema apis* and *Nosema ceranae*, eukaryotic parasites of bees (Budge, unpublished). Bee homogenates were produced by grinding larvae (n=20) and adult honey bees (n=30) in Long Extraction Bags (Bioreba) using a Lenze grinder. PBS 7.3 (5 ml) was added to the bag, which was then ground. The samples was filtered through the permeable membrane within the bag and transferred to a 10 ml tube. The homogenate was retained, in the event that an *M. plutonius* culture was required. In total, 510 µl of homogenate was then removed and added to a screw cap tube containing approximately 500 mg of 0.5 mm silica beads and 600 µl lysis buffer (as for soil extraction) with 10% antifoam B emulsion (Sigma Aldrich®). The serially diluted *M. plutonius* cell suspensions (90 µl) were inoculated into the larval and adult bee homogenate/lysis buffer mix.

Screw cap tubes were beaten on a Precellys 24 lysis and homogenization bead beater (Bertin technologies) at 5000 bpm, for 2 minutes, with a 45 second pause in the middle. Approximately 1 ml was then removed, and the DNA extracted using the soil extraction method as described above.

5.2.1.3. Water samples

Honey bees often visit local sources of water from which to drink and obtain water to dilute stored honey. Water samples were inoculated with 100 µl of *M. plutonius* cell suspensions before being drawn through 0.45 µm filters using a vacuum pump. The filter pad and paper were then cut in half with a sterile scalpel, and half placed in a labelled 50 ml centrifuge tube. Lysis buffer (8 ml, as above) and 250 µl antifoam B (Sigma Aldrich®) were added, and the tubes placed in a PTR-60 Rotator (Grant bio). Tubes were then mixed for 1, 2, 5, 10 or 15 minutes, with standard conditions (Table 5.1). The resulting solution (4 ml) was aliquoted out into a 5 ml tube, and then centrifuged at 2,000 g for 2 minutes, as per the soil extraction method. The extraction was then performed as per the soil extraction.

5.2.1.4. Beekeeping equipment

Wooden- (Copan Italia) or plastic-handled (Sterilin® Limited) cotton swabs were used to remove residues from apiary equipment and subsequently test for the presence of *M. plutonius*. Swabs were inoculated with 100 µl of *M. plutonius* cell suspensions (as above). The wadding was removed and placed into a 15 ml tube containing 2 ml PBS 7.3. Swab samples were placed in a PTR-60 Rotator (Grant bio) for 30 minutes and processed as previously described for water samples.

5.2.1.5. Real-time PCR

All resulting DNA preparations were subjected to *M. plutonius* testing using species-specific real-time PCR with Taqman® chemistry (Budge et al, 2010). Duplicate real-time PCR reactions were run for each sample on an ABI 7500 real-time PCR machine. Reactions were performed with either Taqman® Environmental Master Mix 2.0 (Applied Biosystems®) or AmpliTaq Gold ® (Applied Biosystems®) depending on reagent availability. For Environmental Master Mix 2.0, 5 µl of DNA sample was added to 12.5 µl of Environmental Master Mix 2.0 (Applied Biosystems ®) with 10 pmols of each primer, 5 pmols of probe and made up to a final volume of 25 µl. For AmpliTaq Gold ®, 5 µl of DNA was added to 2.5 µl Buffer A (Applied Biosystems ®) with 7.5 pmols of each primer, 5 pmols of probe, 275 pmols $MgCl_2$, 0.125 µl AmpliTaq Gold ® (Applied Biosystems ®) and made up to a final volume of 25 µl. Generic reaction conditions were used (95 °C for 10 min and

Orbital (rpm)	Reciprocal (deg.)	Vibro/pause	Time
48	30	5	0:25
02	05	1	STOP

Table 5.1. Rotation conditions of the PTR-60 Rotator, as they appear in the input screen. These conditions were the default settings for the PTR-60 Rotator at Fera.

40 cycles of 60 °C for 1 min plus 95 °C for 15 s). *M. plutonius* DNA positive controls were included in all qPCR assays, and were always detected. In a sub-sample of adult and larval bee extracts, an *Apis mellifera*-specific Taqman® assay was used to validate the extraction, under the same reaction conditions as the *M. plutonius* assay (Ward et al., 2007). *A. mellifera* DNA was detected in samples positive and negative for *M. plutonius* DNA with little observable variance (Data not shown), so the *A. mellifera* assay was not carried out in all experimental extractions.

5.2.2. Collection of field samples

Apiaries with a recurring history of EFB were selected for intensive sampling, to increase the likelihood of *M. plutonius* detection. Visits were made either by myself in the company of a National Bee Unit Appointed Bee Inspector (ABI), or samples collected by an ABI working alone in the field. Field samples were subjected to the appropriate validated DNA extraction as described above.

5.2.2.1. Case Study One

The first case study was located in England where the colonies were owned by a beekeeper (Beekeeper One) who had suffered recurring EFB for five years. Their Home Apiary site was free from bees at the time of the first visit in April 2011, as all colonies had been destroyed due to disease. At this site visit, thirty swabs of different articles of beekeeping equipment were taken. Two soil samples were taken; one from the site of former, infected colonies, and one from the site of numerous colony burnings over a number of years. Two water samples were taken; one from a water tub and one from the ABI's hive tool wash bucket, which also contained washing soda. As no colonies were present at this site only two bee samples were taken; one sample of dead bees from an outbuilding and one sample of foraging bees from an unknown source. Another insect sample of ants (which can sometimes enter bee hives as a pest) was also taken. All insect samples were extracted as per the honey bee extraction protocol.

During the first visit the apiary that was to supply new bees to this beekeeper was also sampled. Of the ten colonies that were to be sold to the case study beekeeper, ten adult bee samples and eight larval samples were obtained. A single additional water sample, from a water trough at which bees were seen to drink, was taken from the supplier apiary.

The second sampling took place in July 2011. At this point it became apparent that Beekeeper One also owned two other apiaries, and these were sampled as well. The Second Apiary was 4.8 km from the Home Apiary, the Third Apiary was 1.4 km from the Home Apiary, and the Second and Third Apiaries were 5.1 km apart. Adult bee samples (ten) and larval samples (nine) were

taken at the Home Apiary. At the Second Apiary three adult bee and two larval samples were taken. At the Third Apiary one adult bee and three larval samples were taken.

The third sampling took place in May 2012. On this occasion, 11 adult bee and larval samples were taken from the Home Apiary. At the Second Apiary one adult bee and larval sample was taken, and at the Third Apiary two adult bee and larval samples were taken.

Samples of larvae (n=20 per colony) were extracted from cells with a sterile matchstick and collectively placed in a 50 ml sterile plastic collection tube, and circa 30 nurse adults scooped off the brood comb into another, identical tube. Nurse adults were chosen, as they have been shown to have higher bacterial loads of *M. plutonius* than workers at the flight entrance (Roetschi et al., 2008). Twenty larvae was regarded by the Beekeepers and ABIs as the maximum number that could be taken without damaging the colony, and thirty adults was the optimum number that could easily be captured in one go. No larvae sampled were symptomatic for EFB. Adults were sampled from all sufficiently strong colonies, larvae from those colonies where sufficient brood was present, as judged by the ABI.

5.2.2.2. Case Study Two

Case Study Two was also located in England and the commercial beekeeper involved (Beekeeper Two) had again had a long history of recurring EFB (approximately 15 years). Due to the large number of apiaries owned, two problematic apiaries were sampled; one small (five colonies) and one large (18 colonies), 25.6 km apart.

For Beekeeper Two only one visit took place, in June 2011. Larval and Adult bees were sampled from a subset of five colonies at the Home Apiary. This was because it was the large apiary, and the bees belonging to this beekeeper were particularly aggressive, so the Bee Inspector was not confident in safely obtaining bees from all colonies at the site. Swabs were taken from 21 articles of beekeeping equipment at the Home Apiary. A sample of dead social wasps from an equipment store was also taken.

Bees were sampled from all colonies at the Second Apiary (five adult bee and four larval samples). Another insect sample, of a bumblebee species *Bombus lucorum*, was also taken from this apiary.

5.2.2.3. Case Study Three

Case Study Three was an apiary operated by a beekeeping association in England which had been diseased for two years, and in which a single colony had been treated for disease (by shook

swarm (Budge et al., 2010)) a week before the first sampling visit. The first visit to this apiary took place in June 2011. Adult bee samples (15) and larval samples (17) were taken. All colonies at the apiary were treated with shook swarm in July 2011.

A follow up sampling took place in April 2012. Adult bee samples (20) and larval samples (17) were again taken.

5.2.3. Wider environmental sampling

The non-honey bee insect samples were followed-up by further social wasp samples requested from ABIs around the country, from apiaries that are currently infected with EFB. The number of individual wasps obtained was 27, from 13 sites. These additional wasp samples were treated differently from previous insect samples, with wasps pressed into M110 agar which was then streaked, and incubated anaerobically (Forsgren et al., 2013). This was performed to distinguish *M. plutonius* on the outside of the wasps from any that may have been ingested. Any resulting cultures were tested with the *M. plutonius* Taqman assay®, and typed with the *M. plutonius* MLST scheme (Chapter 3). In addition, swab samples were acquired from another ABI, unrelated to either of these beekeepers. Swabs and a swabbing protocol were sent to the ABI, and swabs were returned for processing.

5.3. Results

5.3.1. Extraction Validation

For all assays, only samples for which both replicate real-time PCR reactions were positive were judged to be positive.

5.3.1.1. Soil

The soil DNA extraction method was successful in removing detectable *M. plutonius* DNA, with DNA detectable at all but the lowest concentration. Detection became less reliable at lower concentrations, and the C_T value did not increase with decreasing concentration of inoculated cell suspension, as would have been expected (Table 5.2). Taqman® negative controls (water) were negative for the presence of *M. plutonius* DNA.

Soil Serial Dilution	First C_T value	Second C_T value
1	33.3885	31.0780
2	32.6645	31.5012
3		36.0415
4		
-		

Table 5.2. *M. plutonius* DNA in soil validation assay. *M. plutonius* cell suspension was diluted serially from 1 (highest concentration) to 4 (lowest concentration). Assay is able to detect *M. plutonius* DNA in soil over at least three orders of magnitude.

5.3.1.2. Bees

DNA extraction was very successful for larval samples, and was partially successful for the adult bees. The larval extractions clearly showed a relative increase in *M. plutonius* DNA concentration with increasing concentration of the cell suspension (Figure 5.1, $R^2 = 0.99$), but the adult bee samples did not (Table 5.3). High concentrations of target DNA can lead to inhibition of PCR, for example if there is a massive excess of DNA then the supply of primers and probes may be exhausted before detection. At any rate, detection of *M. plutonius* in adult bees is qualitative, and should be treated with caution. Taqman® negative controls (water) were negative for the presence of *M. plutonius* DNA.

5.3.1.3. Water

The water filtration method was capable of filtering and detecting *M. plutonius* in water. Length of time rotated appeared to bear little relationship to the amount of *M. plutonius* DNA extracted (Table 5.4). Taqman® negative controls (water) were negative for the presence of *M. plutonius* DNA. For field samples of water, the longest rotator time (15 minutes) was used.

5.3.1.4. Swabs

Swab DNA extractions worked, and showed a clear relationship between *M. plutonius* cell suspension concentration and the amount of DNA detected (Table 5.5, Figure 5.2, $R^2 = 0.98$). Taqman® negative controls (water) were negative for the presence of *M. plutonius* DNA.

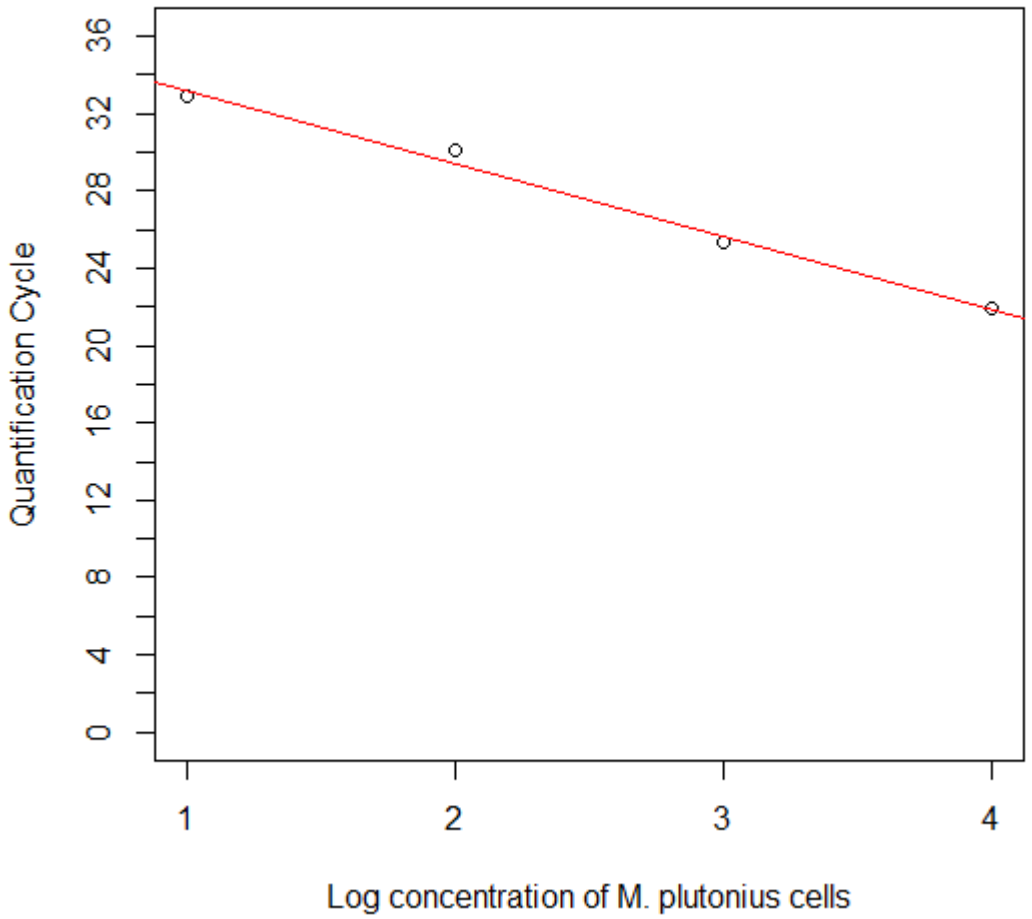


Figure 5.1. The Quantification Cycle (C_T value) at which each concentration of *M. plutonius* cells was detected, for the larval bee DNA extraction validation. Log *M. plutonius* cell concentration is used, as cell suspension was serially diluted.

Bee Serial Dilution	Larvae First C_T value	Larvae Second C_T value	Adults First C_T value	Adults Second C_T value
1	22.1705	21.6576		39.8012
2	25.1766	25.4267		39.2121
3	29.6632	30.5180	37.5688	37.5797
4	33.0254	32.7869	37.8720	36.4043
-		39.8011		

Table 5.3. *M. plutonius* DNA in honey bee extraction validation assay. *M. plutonius* cell suspension was diluted serially from 1 (highest) to 4 (lowest) concentrations. – is a negative control of bees uninoculated with *M. plutonius*.

Time in Rotator	First C_T value	Second C_T value
1 min	33.5493	31.7319
2 min	32.3374	31.8365
5 min	31.3064	31.1134
10 min	32.0812	31.7352
15 min	31.1714	30.9640

Table 5.4. *M. plutonius* DNA in water extraction validation assay.

Swab Serial Dilution	First C_T value	Second C_T value
1	25.1441	25.2025
2	29.1568	29.1567
3	32.7513	32.5831
4	35.2762	34.3619
-		

Table 5.5. *M. plutonius* DNA in swab extraction validation assay. Swab extraction protocol is able to detect *M. plutonius* DNA across at least four orders of magnitude of *M. plutonius* cell suspension concentration.

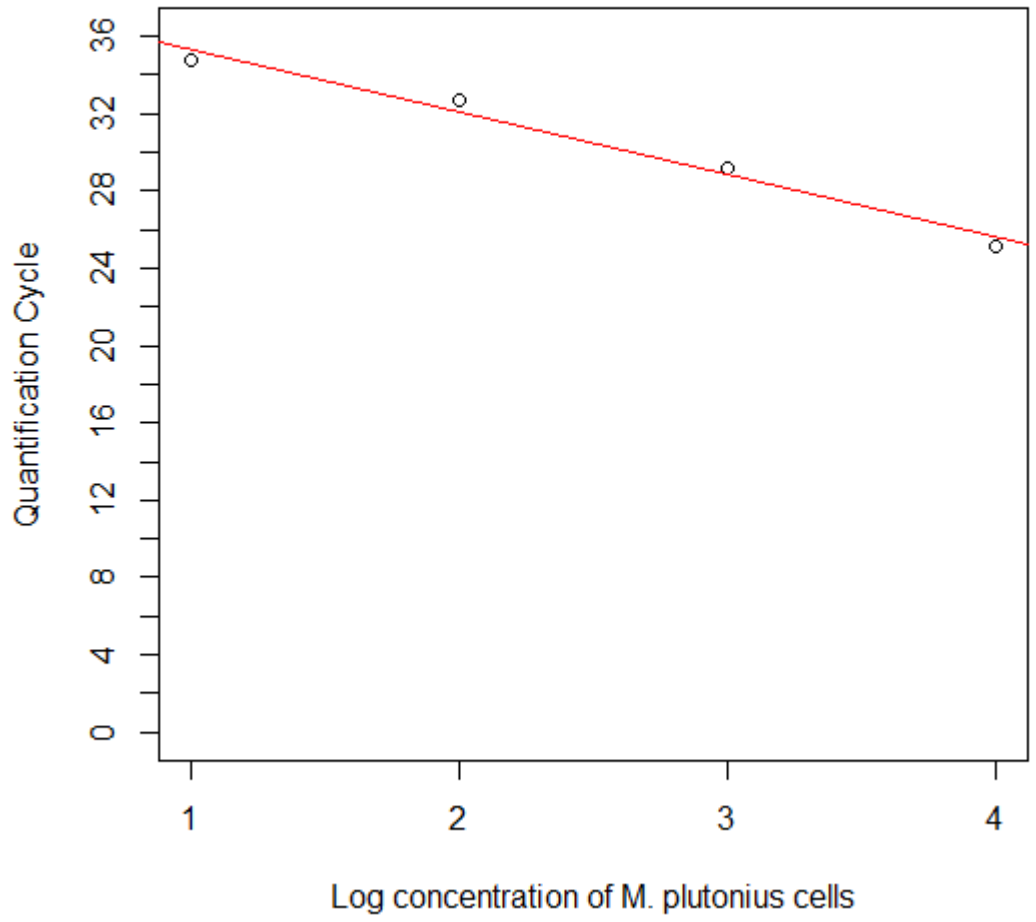


Figure 5.2. The Quantification Cycle (C_T value) at which each concentration of *M. plutonius* cells was detected, for the swab (beekeeping equipment) DNA extraction validation. Log *M. plutonius* cell concentration is used, as cell suspension was serially diluted.

5.3.2. Field Samples

5.3.2.1. Case Study One

On the first visit to the Home Apiary, all soil and water samples were negative for *M. plutonius*. Ant samples, dead bees and foraging bees all tested negative. Swabs that contained detectable levels of *M. plutonius* DNA were from a crown board which had previously been scorched, an eke (a frame used to extend or enlarge a colony box), and a smoker and car interior belonging to the ABI (Table 5.6). All bees from the Supplier Apiary were *M. plutonius*-negative, as was the water sample from that apiary.

Of the samples from the second visit, no colonies showed detectable *M. plutonius* levels.

On the third visit, three colonies from the Home Apiary showed high levels of *M. plutonius* in larvae, and *M. plutonius* was detected in adult bees (Table 5.7). No colonies at the other apiaries had detectable *M. plutonius*.

5.3.2.2. Case Study Two

Of the honey bee samples taken from Beekeeper Two's apiaries, none of the colonies from the Home Apiary showed detectable levels of *M. plutonius*. Of the colonies from the Second Apiary, three out of five (colonies 26, 42 and 104) had varying levels of *M. plutonius* in larvae, and *M. plutonius* was also detected in adult bees (Table 5.8). Additional honey bee samples were taken from outside the colonies. Dead bees from a box store at the Home Apiary were negative for *M. plutonius*, but dead bees from the beekeeper's car were positive (CTs 37.53/37.82).

Of the swabs taken at Beekeeper Two's Home Apiary, the only swab to show detectable *M. plutonius* DNA was from the Beekeeper's car boot (Table 5.9). Two other insect samples were obtained from this beekeeper's apiaries. One was a white-tailed bumblebee (*Bombus lucorum*) which was negative for *M. plutonius*. The other were dead wasps (*Vespula spp.*) taken from the box store. These were positive for *M. plutonius* (38.45/37.30).

5.3.2.3. Case Study Three

On the first visit to this apiary, *M. plutonius* was detected in seven adult bee samples and four larval samples (Table 5.10), with one larval sample showing high levels of pathogen.

On the second visit to the apiary, all samples were negative for *M. plutonius*.

Substrate Swabbed	First C_T	Second C_T
1) Caustic soda tub		
2) Brood box (scorched)	34.5269	35.5085
3) Crown board (scorched)		
4) Brood box (scorched)		38.6776
5) Crown board (scorched)		
6) Hive tool		
7) Old crown board		
8) Old crown board		
9) Wellington Boots (ABI)		
10) Clipboard (ABI)		
11) Bee suit (ABI)		
12) Eke	38.5975	39.4123
13) Smoker (ABI)	39.2055	38.2818
14) Honey extractor	38.6473	
15) Metal stand		
16) Breeze block		
17) Boots (ABI)		
18) Dripping tap	39.0669	
19) Brood box		
20) Gutter		39.8099
21) Wasp trap		
22) Brood box	39.2348	
23) Crown board		
24) Crown board		39.0693
25) Brood box	37.7485	
26) Smoker (ABI)		
27) Suit (ABI)		
28) Spacers		
29) Rear interior of car (ABI)	37.4834	36.2613
30) Unused swab (wood)		
31) Hive tool (ABI)		

Table 5.6. Swab *M. plutonius* DNA results from Beekeeper One Home Apiary. Swabs with two reactions, and lower C_T values indicate reliable presence of *M. plutonius* DNA. The unused swab was included as a negative control.

Apiary and Colony	Larvae First C_T value	Larvae Second C_T value	Adults First C_T value	Adults Second C_T value
Home 4	17.2529	17.0380	26.0074	26.1483
Home 6				
Home 8				
Home 10		39.1503		
Home 12	38.6844			
Home 17				
Home 19				
Home 20				
Home 21	35.0449	35.5769	37.9870	39.1206
Home 25	16.9157	17.0115	23.3146	23.8277
Home 26				
Second Apiary 14				
Third Apiary 9				
Third Apiary 24				

Table 5.7. *M. plutonius* in honey bee samples from colonies in apiaries belonging to Beekeeper One, from the third visit.

Apiary and Colony	Larvae First C_T value	Larvae Second C_T value	Adults First C_T value	Adults Second C_T value
Home 5				
Home 31				
Home 34				
Home 35				
Home 96	39.4361		39.4422	
Second Apiary 22				
Second Apiary 26	18.5264	18.2505	23.7801	23.7077
Second Apiary 42	22.649	22.5777	28.7482	28.8357
Second Apiary 77	-	-		
Second Apiary 104	27.3617	25.0852	31.7481	30.5087

Table 5.8. *M. plutonius* in honey bee samples from colonies in apiaries belonging to Beekeeper Two. – indicates sample which could not be taken.

Substrate Swabbed	First C_T	Second C_T
1) Box store super frame		
2) Box store door handle		
3) Box store light switch		
4) Old hive tool, brood box Store		
5) Box store door mesh		
6) Puddle water/mud, yard		
7) Super box, box store		
8) Large honey warmer, exterior	37.3047	
9) Radiation-sterilised brood combs		
10) Honey warmer #1		
11) Queen excluder, box store		
12) Box store brood box		
13) Clipboard (ABI)		
14) Honey warmer #2		
15) Box store brood box frames		
16) Bee suit (prior to inspection)		
17) Bee suit (after inspection)		
18) Beekeeper car boot	37.3072	37.7405
19) Beekeeper suit		38.2846
20) gloves (after inspection)		38.8194
21) Small honey warmer, exterior		

Table 5.9. Swab *M. plutonius* DNA results from Beekeeper Two Home Apiary. Swabs with two reactions, and lower C_T values indicate reliable presence of *M. plutonius* DNA.

Colony	Larvae First C_T value	Larvae Second C_T value	Adults First C_T value	Adults Second C_T value
1	23.3535	22.4932	33.1124	32.1181
2				
3			38.6204	
5			39.8156	
6				
7				
8	39.0713	39.7356	28.3871	28.4696
9	34.9849	34.9429	28.8245	28.5854
10	38.0193	37.859	30.4917	30.7469
12			-	-
13			-	-
14	39.0799		36.795	38.5768
15				
16				
20				
21				
Top Left			36.5914	36.7674
N2	-	-	20.7629	20.7502

Table 5.10. *M. plutonius* in colonies from the Case Study Three apiary. – indicates samples which could not be taken.

5.3.3. Wider Environmental Sampling

Of the 27 individual wasps (*Vespula spp.*) that were plated out, only one produced cultivable *M. plutonius*. This sample was from Oxfordshire, and when Sequence Typed (see Chapter 3) was an ST 22, the same as the local type (Figure 4.3). Of the swabs taken from the additional ABI samples, *M. plutonius* was found in the vehicle glove box and door handle (Table 5.11).

5.4. Discussion

The validated extraction methods were useful for monitoring *M. plutonius* in environmental samples. Both the swab (beekeeper equipment) and honey bee larval extractions allowed quantitative measurement of the amount of *M. plutonius* present, demonstrated by C_T values which showed a strong linear relationship to a log increase in the amount of *M. plutonius* inoculant ($R^2 \geq 0.98$). However, the adult bee, soil and water extractions did not show similar relationships and are therefore useful only in a qualitative sense. These extractions allow the user to determine the presence of *M plutonius*, but not to specify the relative amount. For the soil extraction there is a discrepancy between this result and the use of the same protocol for other organisms, which did allow a quantitative description (Budge et al., 2009; Woodhall et al., 2012). The lack of quantitative measurement is probably due to the clumping behaviour of *M. plutonius*, which makes it extremely difficult to produce a uniform suspension (Tarr, 1938). Another factor may be the presence of PCR inhibitors in the form of humic acid in the soil (Tsai and Olson, 1992), and melanin (which accumulates during pupation (Zufelato et al., 2000)) in the adult bees (Eckhart et al., 2000).

The observation with potentially the most impact from this study was the presence of cultivable *M. plutonius* on social wasps (*Vespula spp.*). In one case a social wasp sample from one of the apiaries was positive for *M. plutonius* DNA, and one of the wasps submitted by the ABIs had viable *M. plutonius* on its surface. Care must be taken in interpreting the second observation. The sample was taken by an ABI inspecting an EFB-positive apiary and the sampling was not observed by the author, so it could possibly have been contaminated during the sampling process. Nevertheless, two independent observations of *M. plutonius* in *Vespula spp.* wasps indicate that these insects could be a transmission route.

Because of their lifecycle, social wasps such as these are unlikely to be an alternate host species for *M. plutonius*. In temperate climates only the queens survive the winter, in diapause, and then emerge from hibernation in spring to build a new nest from scratch. The old nest is not reused, and no workers survive the winter (Greene, 1991). The likelihood of the bacteria surviving in this

Substrate Swabbed	First C_T	Second C_T
1) Honey, from van floor		
2) Storage box, from van		
3) Smoker bellows		
4) Smoker body		
5) Wellington boots		
6) Kit buckets		
7) Van floor, front		39.8449
8) Van floor, side	39.6880	
9) Glove box	39.1701	39.2250
10) Van door handle	39.4514	39.4146
11) Van floor, rear		
12) Dead bees, from van		

Table 5.11. Swab *M. plutonius* DNA results from Beekeeper Two Home Apiary. Swabs with two reactions, and lower C_T values indicate reliable presence of *M. plutonius* DNA.

single queen to reinfect the next year's colony is therefore low. This conclusion is supported by the fact that a wasp-specific variant of *M. plutonius* was not found. More likely is that the wasp workers have picked up *M. plutonius* whilst robbing honey from the infected bee colony. Robbing behaviour has been observed in both of the *Vespula* species present in the UK, *V. vulgaris* and *V. germanica* (Clapperton et al., 1989). Indeed, the carriage by robbing wasps of a bee parasite has been observed with the mite *Varroa jacobsoni*. In this case mites were found on the exterior of wasps, within the wasp nest, but not reproducing on the wasp brood (Sammataro et al., 2000). Within a season, however, it is possible that these wasps could be responsible for mechanically transmitting the bacteria between colonies. Inter-apiary transmission is also possible, with *Vespula spp.* wasps recorded making foraging trips of up to 4 km (Beggs et al., 1998). For wasps to represent a viable transmission route, *M. plutonius* must be transferred from the outside of the wasp to the honey bee gut. It seems likely that when robbing honey from a honey bee colony the wasp could transfer *M. plutonius* to either remaining honey or emptied honey cells. *M. plutonius* could then persist in the honey (see Chapter 4) and infect the bees when they next fed from the cells.

An assay to confirm these suppositions is required. Such an assay could be performed by isolating a wasp nest in a glass house or similar in late summer/early autumn (when robbing behaviour by wasps is at its highest). The nest could then be exposed for a period of time to a honey bee colony carrying high levels of *M. plutonius*. After robbing behaviour takes place, the bee colony would be removed and the area it was stored in and any forage plants present would be isolated (to prevent carryover of *M. plutonius* on surfaces exposed to infected bees). The wasp nest would then be exposed to a new honey bee colony that is free from *M. plutonius*. Once robbing of this second colony had occurred, it could then be assayed for the presence of *M. plutonius*. It would be necessary to temporally separate the exposures of the different bee colonies, as bees also engage in robbing and drifting behaviour. Additionally, as wasps and bees are of a similar size it would be difficult to physically exclude bees without preventing the wasp robbing behaviour. If *M. plutonius* were found in the second colony it would only appear possible through wasp-mediated transmission, a supposition that could be supported by sequence typing the bacteria in both of the bee colonies.

Asymptomatic infection (McKee et al., 2003; Forsgren et al., 2005; Budge et al., 2010) was confirmed in all three case studies, and for the first time field data are presented showing that asymptomatically infected colonies became diseased at a later date. At the Second Apiary of Case Study Two, three colonies were found to be asymptomatically infected. Coherent records exist for colonies 42 and 104. Colony 42 was diseased in April 2011 and treated by shook swarm (Budge et al., 2010), while the last occasion pre-sampling that colony 104 showed disease symptoms was in

May 2010. Both colonies were free from disease on the 7th of June 2011, when the samples presented here were taken. Both were heavily infected, but were not detected as symptomatic until six weeks later on the 20th July 2011. It is therefore demonstrated that asymptomatic colonies, carrying high levels of *M. plutonius* do sometimes become diseased, and that the six week follow-up protocol followed by the NBU did detect the disease. This particular apiary was under a standstill notice due to its recurring EFB infections, but in other (less well monitored) cases such infections could represent a window during which colonies could have been moved to another apiary or sold to another beekeeper. It is known that the sale of colonies can transmit *M. plutonius* (Chapter 3), and indeed all colonies at the Case Study One Home Apiary were signed off for movement by the ABI shortly after the third sampling period, despite three of them asymptomatically carrying high levels of *M. plutonius*. An inspection policy that is based on disease symptomology (Wilkins et al., 2007) will be entirely unable to detect infected but asymptomatic colonies, or to treat or prevent the movement of such colonies. Moving to a certification for sale based on molecular testing for *M. plutonius* would likely help to reduce the long distance transmission of the disease. Additionally, if asymptomatic colonies are able to transmit disease before developing symptoms, this represents a risk of infection to uninfected colonies in the apiary. If an apiary contains diseased colonies it is more likely to have asymptomatically infected colonies as well (Belloy et al., 2007; Budge et al., 2010), and any treated colonies that are returned to an apiary with asymptomatically infected colonies could be subject to intense reinfection pressure. This would render colony-based treatments of such apiaries ineffective and would therefore suggest EFB should be seen as a disease of the apiary. This is supported by Case Study Three. In this instance, the confirmation of asymptomatic infection directly led to the decision to treat the entire apiary with shook swarm. In the two years after treatment, no further disease has occurred in this apiary. If all colonies in infected apiaries were treated, instead of just those which showed symptoms, it might increase the success of treatments.

These are very preliminary observations, and the durations over which a colony can maintain a subclinical infection needs to be thoroughly investigated. This could be elucidated with sampling of much finer temporal resolution, from the start of the season, so that the entire period from uninfected, to infected, to symptomatic can be seen. Furthermore, the latent period, or time between infection and infectiousness, must be determined. This could be investigated by the repeated removal and replacement of non-infected contact colonies from the apiaries studied above, with the colonies subsequently monitored for the presence of *M. plutonius*. Needless to say, strong biosecurity measures would have to be put in place, to prevent infection from outside the apiary.

Residual *M. plutonius* was detected on numerous pieces of beekeeping equipment, especially in vehicles used to transport bees and equipment. This could represent a route for transmission of *M. plutonius* into new colonies (from hive box and inspection equipment) and apiaries (from vehicles). However, the levels of *M. plutonius* detected on these objects are low compared to those detected in the larvae. We also do not know if the *M. plutonius* cells detected were alive (and therefore potentially infectious) or dead, as the qPCR assay only detects DNA and not the presence of viable organisms. To understand the importance of these transmission routes, more intensive sampling across more apiaries should be performed, and culturing should be untaken to determine the live status of the pathogen.

This chapter has illustrated several potentially important points; potential transmission routes for *M. plutonius* have been suggested (Figure 5.3), including the previously unsuspected route of a non-honey bee insect vector. DNA extraction and Taqman ® detection methods were validated as either quantitative or qualitative descriptors of *M. plutonius* presence and/or relative abundance. However, this work acts as more of a starting point for further work. The latent period needs defining more robustly, and the role of social wasps in *M. plutonius* transmission needs to be confirmed or refuted.

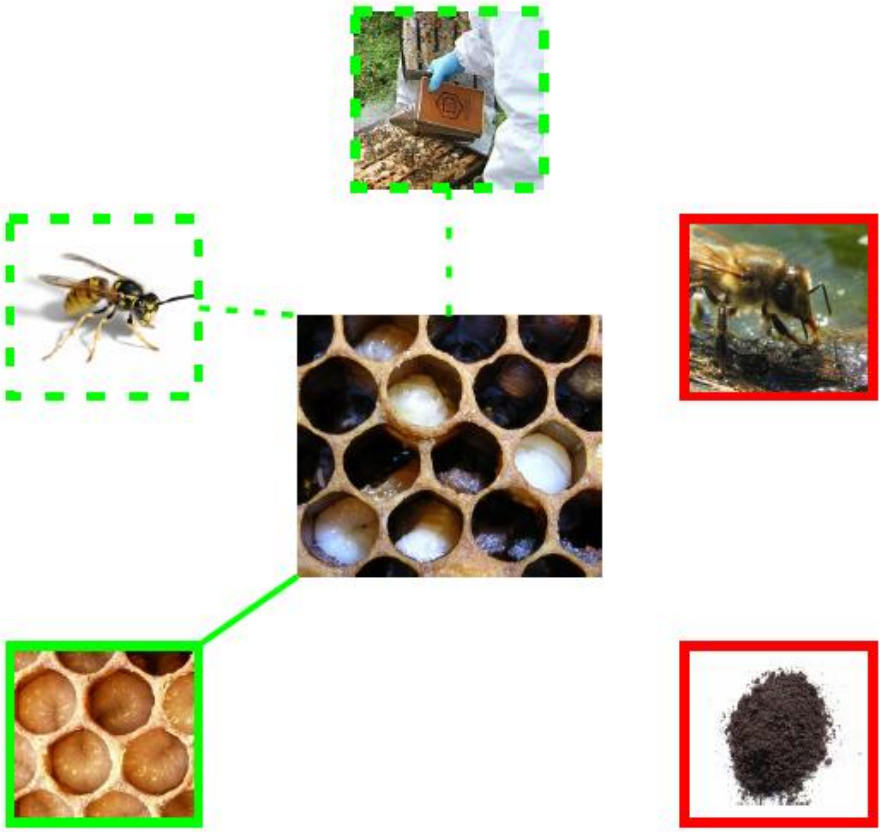


Figure 5.3. Infographic, demonstrating the transmission routes for *M. plutonius* into a honey bee colony identified in this study. Dashed green lines show possible transmission routes suggested by the data, the solid green line shows larvae were confirmed to be asymptomatic carriers, before they developed EFB. No evidence was found of water or soil being reservoirs. Images of symptomatic and asymptomatic larvae, and beekeeping equipment (smoker) courtesy The Food and Environment Research Agency (Fera), Crown Copyright. Images of wasp (courtesy Fir0002/Flagstaffotos) and drinking bee (courtesy Bartosz Kosiorek) are used under GNU Free Documentation License v1.2.

6. General Discussion

European Foulbrood is an important disease of honey bees, which has become more prevalent in recent years in several countries, including the UK (Roetschi et al., 2008; Budge et al., 2010; Dahle et al., 2011). Most molecular epidemiological studies to date on the causative organism *M. plutonius* have focussed on developing pathogen detection methods, and applying them to detect the pathogen in honey bees and their hive products (McKee et al., 2003; Forsgren et al., 2005; Budge et al., 2010). Using molecular techniques for pathogen detection is an important aspect of molecular epidemiology (Foxman and Riley, 2001), and in the context of *M. plutonius* it has revealed sub-clinical infection to be an important pathogen reservoir. At the start of this project, however, this was the extent of the progression of molecular epidemiology for this disease.

Over the last two years, as the scope of next generation sequencing has increased dramatically, two genome sequences for the pathogen have been produced. The genomic analyses in this thesis have revealed that one of these sequences (ATCC 35311 (Okumura et al., 2011)) appears to have been incorrectly arranged, with a large artificial inversion occuring in the middle of the chromosome. Through comparative genomic analysis, and carefully designed molecular assays, the correct orientation of the chromosome was identified. This sequence was then used as a basis for much of the analysis and method development in the rest of the thesis.

This thesis identified three different types of mobile genetic elements in the *M. plutonius* genome; a plasmid, a phage, and a Type I RM system. Several genes involved in either the production of or defence against lantibiotic peptides were identified that may give *M. plutonius* a competitive advantage against co-occurring bacteria. However no antibiotic-resistance genes were identified, something which might have been expected given the continued use of the antibiotic OTC to treat EFB outbreaks (Wilkins et al., 2007; Budge et al., 2010). OTC resistance is known to occur in *P. larvae*, the other bacterial brood pathogen of honey bees, and has been shown to be present on an as yet unidentified mobile element (Evans, 2003). *P. larvae* can occur in the same apiary and rarely in the same colony as *M. plutonius* (Budge, Pers. Comm.), indicating it is a potential source of OTC resistance for *M. plutonius*. Genomics could be fruitfully used to monitor *M. plutonius* populations for acquisition of OTC-resistance; by sequencing the genome of OTC-resistant *P. larvae*, the antibiotic genes (and the mobile elements upon which they travel) could be identified. It would then be simple to design a gene-specific test (such as a PCR or Taqman ® assay) to detect that gene and monitor *M. plutonius* for its introgression, thus eliminating the time and effort required to perform inhibition assays on *M. plutonius* (already a fastidious bacterium to culture) to detect OTC resistance.

Whole genome sequencing allows the greatest possible resolution of the evolutionary relationships between different bacterial isolates. In this thesis, the phylogenetic network produced from the concatenated sequences of all the chromosomal gene sequences clearly differentiates between typical and atypical *M. plutonius*, confirming the separate evolutionary histories of these two groups. There is also no evidence of horizontal gene transfer or recombination between isolates of typical and atypical *M. plutonius*. This is shown both by the single branch linking the atypical isolate 7596 to the typical group in the network, and by the absence of any of the mobile genetic elements seen in the typical isolates in either of the studied atypical isolates (7596 or DAT561). Within typical *M. plutonius* the genome level phylogenetic network is able to distinguish two well-differentiated groups of isolates, one group closely related to the Type Strain, and another group more distantly related.

By comparing the results of this whole genome network with the goeBURST pattern of relatedness calculated for the data from the modified MLST data, we can get an idea of how accurately the MLST reflects the underlying evolutionary relationships. The differentiation between the typical and atypical lineages is obvious from the whole genome analysis and from the MLST (where atypical *M. plutonius* is identified as Clonal Complex 12). Within the typical *M. plutonius* type, the MLST scheme is able to resolve the isolates into the same two groups as the genome scale phylogeny. The cluster identified by the genome sequencing that contains the Type Strain is identified by MLST as Clonal Complex 13, and the cluster more distantly related to the type strain corresponds to Clonal Complex 3 (Figure 6.1).

Within the clonal complexes themselves, the MLST has mixed success at accurately describing the relationships among isolates. The long branch lengths between the Clonal Complex 13 isolates in the phylogenetic network are reflected in the fact that they all differ from each other at two loci in the MLST. However between the Clonal Complex 3 isolates the MLST depicts some relationships less accurately, in some cases overestimating the difference between types (e.g. ST 5 and ST 6) and in some cases underestimating the diversity within a type (e.g. ST 11). Nonetheless, for a four locus scheme to give such a good representation of the diversity and relationships between types whilst using less than 0.2% of the amount of data (four gene fragments out of 1959 chromosomal genes) instils a high degree of confidence in the further use of the scheme. Moreover, future whole genome sequencing of isolates, for example from the widely-found ST 3, may reveal more polymorphic genes which will allow the more common STs to be further subdivided.

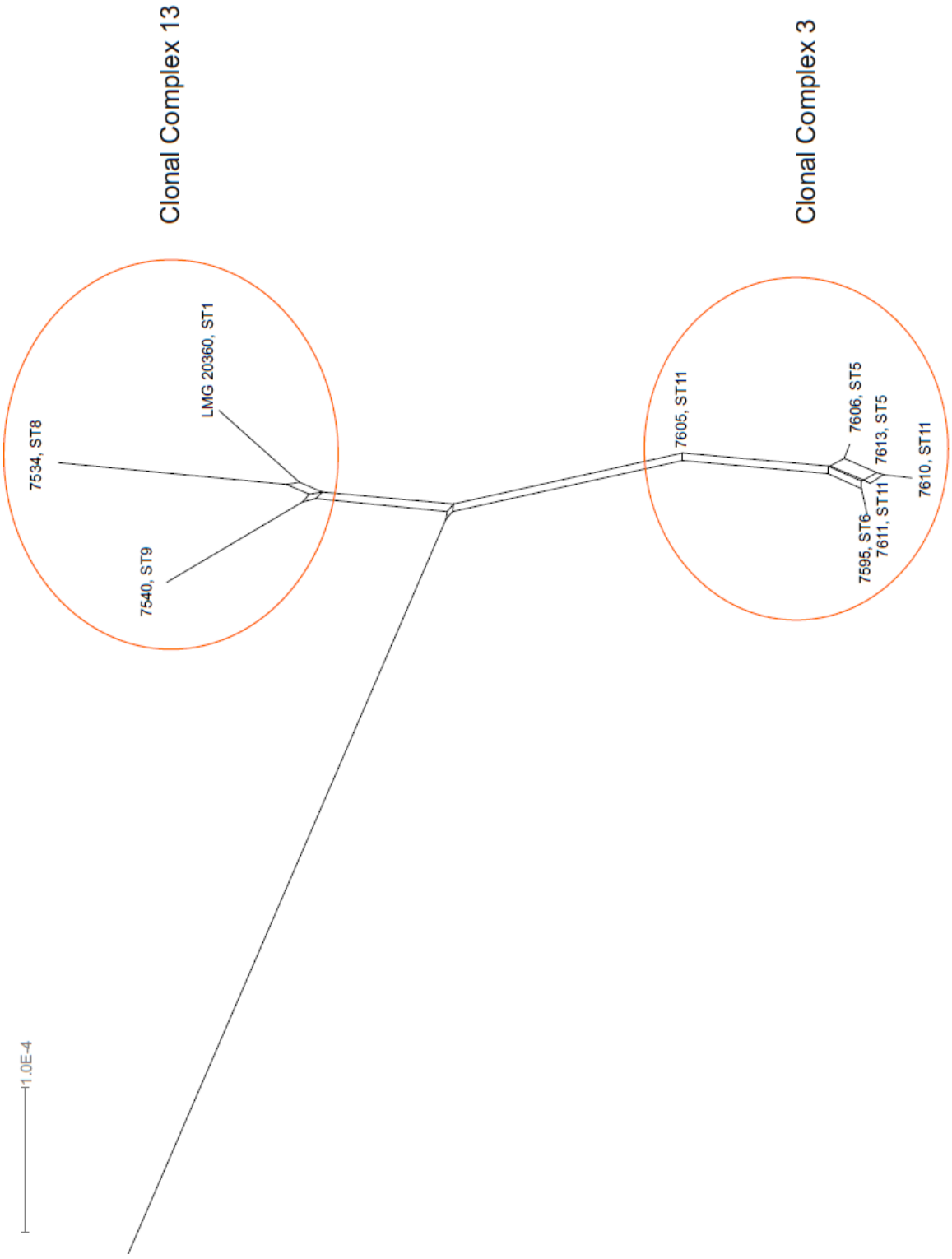


Figure 6.1. Phylogenetic network of typical *M. plutonius*, with the ST and Clonal Complex designations of isolates overlaid. The branch to the left leads to the atypical *M. plutonius* lineage (Figure 2.7).

Having developed the first scheme for strain typing *M. plutonius*, this thesis is the first to show significant patterns of variation across a country. We show disease clustering in the landscape, indicative of local spread, and show that many STs have a restricted geographical range. We can therefore use our knowledge of the current *M. plutonius* distribution in UK as a baseline against which to monitor unusual or unexpected outbreaks. There are examples in the data already that might be break-outs of STs from their home regions (e.g. ST 5 from East Anglia to the North Yorkshire coast, ST 23 from the Welsh Borders to West Yorkshire), and this data could easily be used to monitor future spread within the UK (e.g. if an ST 13 was found outside Norfolk, an initial conclusion would be that it had likely spread from there) or from abroad (like the ST 14 isolate identified in Chapter Three).

The *M. plutonius* populations in Scotland and the rest of Britain are clearly different, with only two possible examples of spread between the two regions (an ST 2 found in Scotland in 2009, and an ST 1 found in England in 2011), even though there is no legislation banning movement of bees across the border. The large EFB-free region between Cumbria and Southern Scotland is likely an important factor preventing natural spread. This is not due to an absence of beekeepers, so there could be some local environmental factor that reduces the prevalence of EFB. On top of this, it is probably the case that migratory beekeepers moving between shared apiaries on heather moors are homogenising the Scottish EFB population (evidence is presented in Chapter Four of *M. plutonius* ST 1 being isolated from Scottish heather honey). The fact that the dominant type in Scotland is ST 1 also allows speculation on the origin of the *M. plutonius* population in the country. The Type Strain (NCDO 2443) also belongs to ST 1, and was isolated by Bailey in England in the 1950s (Bailey, 1957; Bailey Pers. Comm.). It is plausible that the control regime in England has eradicated ST 1 there, but lack of a similar regime in Scotland has allowed that type to persist. Another potential source of the ST 1 in Scotland may be accidental release. In Scotland in the 1960s, research was performed on AFB, and possibly EFB, which may have involved the deliberate inoculation of hives with bacteria (Highet, Pers. Comm.). An obvious source of *M. plutonius* would be the isolate cultured by Bailey several years before, though this is purely circumstantial. A similar phenomenon could be responsible for the presence of ST 1 in Thailand (where this has been confirmed as a genuine outbreak in the environment (Chantawannakul, Pers. Comm.)). This is also difficult to confirm, as any *A. mellifera* present in Thailand will have been ultimately exported from Europe, possibly at a time when ST 1 was more prevalent.

If the control regime in England did effectively remove the ST 1 present and change the pattern of *M. plutonius* in the country, it would indicate that temporal changes in diversity could be an important factor in EFB epidemiology. The IPI project in England and Wales took place over just two years, making it unsuitable for exploring changes in diversity over time. However, there are

sources which could be exploited to that end. For example, all of the Lateral Flow Devices used by Bee Inspectors in England and Wales (since their introduction in 2005) have been stored at the National Bee Unit. *M. plutonius* can be cultured from LFDs (Budge, unpublished data) or a direct MLST PCR could be performed with a DNA extract from the test strip. This would give nine years of samples with which to study changes in diversity. Historical samples of adult honey bees (which have been shown in this thesis and other studies (Roetschi et al., 2008) to be capable of carrying *M. plutonius*) may exist in museums. Strain typing has been performed on bacterial samples that are hundreds of years old (e.g. Taylor et al., 2013), so would be feasible in historical samples if *M. plutonius* could be detected. This could be used to investigate the presence of ST 1 before the control regime was initiated in England, and support or challenge the control regime hypothesis. Another interesting example is that of Switzerland, where the recent increase in *M. plutonius* prevalence began around the year 2000 (Roetschi et al., 2008). If a large enough sample set could be obtained pre-dating the epidemic, then the causes of this increase in prevalence (for example, the introduction of new STs, or changes in frequency of types) could be investigated.

In the course of this thesis, 30 STs of *M. plutonius* have been identified, with the majority of types found in the well-sampled European countries. To identify further global diversity, isolates from *A. cerana* and *A. laboriosa* should be examined. The former should be relatively easy to obtain, given the prevalence of EFB in *A. cerana* (Bailey, 1974; Allen and Ball, 1993), whereas to date *M. plutonius* has only been isolated once from *A. laboriosa* (Allen et al., 1990). A broader sampling of managed bees across Asia would also be welcome. Firstly, as prevalence and diversity of atypical *M. plutonius* seems higher in Japan than non-Asian countries (Arai et al., 2012; Haynes et al., 2013), and secondly because Asia is home to the majority of *Apis* species, which could provide a reservoir for pathogen diversity. The commensal gut microbiota appears frequently shared between honey bee species (Jeyaprakash et al., 2003; Disayathanoowat et al., 2012; Vásquez et al., 2012). This could imply both that limited differentiation of *M. plutonius* will be found between host species, but also that any variation that does exist could be easily transferred among them. Further sampling may also be profitable in Europe, given the long evolutionary history of *A. mellifera* there, and should be undertaken in Africa, from where only one isolate has yet been sequence typed. Africa is also the evolutionary home of *A. mellifera* (Whitfield et al., 2006), and many races of *A. mellifera* absent in Europe are found there, including *A. m. scutellata* and *A. m. capensis*. Limited effort should be expended obtaining samples from outside the native range of the honey bee (Australia, New Zealand and the Americas) unless in an attempt to unravel patterns of trade and import. However, it might be interesting to further sample South and Central America. Unusual types of *M. plutonius* (Allen and Ball, 1993; Haynes et al., 2013) and *P. larvae* (Morrissey, Pers. Comm.) have already been identified from South America. Additionally, it was to Brazil that the original progenitor queens of the rapidly spreading Africanized bees were

introduced from South Africa and Tanzania in 1956 (Winston, 1992), and it is possible that they could have taken African variants of their brood diseases with them. ST 17 is a type of *M. plutonius* whose sole member is a Tanzanian isolate from a culture collection (NCFB 2441). If this type were found in South America, it could possibly have been transported with the African queens.

An important next step with the MLST scheme would be to check the phenotypic characteristics of the STs. Atypical and typical *M. plutonius* are already shown to differ in many respects, with typical showing narrower culture requirements, and a more rapid attenuation of virulence in culture (Arai et al., 2012). However, it will be important to check for any differences in virulence and latency until symptoms show, as this would affect our hypotheses about how virulence affects diversity. There is already some indication that virulence differs between clonal complexes 3 and 13 (Budge et al., 2014; Gauthier, Pers. Comm.) and, to a point, pathogens that are more virulent can also have higher transmissibility (Anderson and May, 1982). Conversely, if there were differences in latency until the appearance of symptoms, those isolate which showed symptoms later would be detected later (under the England and Wales inspection regime (Wilkins et al., 2007)) and may have a greater opportunity to spread. It would also be valuable to sequence the genome of a Scottish ST 1 isolate, to see if there are any mobile genetic elements present, perhaps carrying virulence factors that would make the currently circulating variants more virulent than the Type Strain.

Both strain typing and pathogen identification molecular techniques have contributed to elucidating the spread of *M. plutonius* between colonies and apiaries. There is evidence for a strong anthropogenic component to transmission. This is shown from the beekeeper ownership network (where beekeepers often had the same ST present throughout their operations), from beekeeper to beekeeper sales of bees leading to the transmission of STs, and also from *M. plutonius* being detected on swabs of beekeeper equipment. However, there is also evidence for natural transmission clusters where beekeeper ownership is not a common factor (e.g. ST 2). Additionally, this thesis presents the first evidence of cultivable *M. plutonius* being found on a subject other than a honey bee or honey bee product – a social wasp *Vespula spp.*. This could represent a previously overlooked natural transmission mechanism. Other arthropod vectors of honey bee diseases are known (e.g. *V. destructor* vectoring viruses (Sumpter and Martin, 2004)). Sampling more pests of the honey bee, such as wax moths, hornets or *Varroa* mites, could reveal further vectors of the bacterium.

The next important step for this scheme is to examine the diversity of *M. plutonius* at smaller spatial scales. For example, before sampling a colony to confirm or refute a suspected

transmission event, the number of samples necessary will be informed by the frequency with which different types are found in the same colony, or larva. There is no clear evidence from the genomic work in Chapter Two that co-infection with multiple variants occurs widely. Additionally many whole larval extracts from Scotland and the Netherlands, and several from the US and Poland, have been amplified with and typed by the MLST scheme and not produced multiple alleles (though there is only one dominant type in both Scotland and the Netherlands). Still, a more systematic approach to determining the number of types in a single colony and larva is needed.

If multiple infections are indeed widespread, this would inform hypotheses about drivers of diversity at a country level. It would be unlikely that priority effects, such as the presence of one type in a colony preventing the invasion of a new type, are occurring in non-controlling countries (as suggested in Chapter Four) if bacteria can't exclude each other from the same larval gut. One might expect it to be easier for typical and atypical to exist in the same larva, as bacterial variants that are phylogenetically distant tend to show lower degrees of inhibitory priority effects (Tan et al., 2012), especially given the different substrate utilization patterns of typical (narrow requirements) and atypical (broader substrate utilization) *M. plutonius*. If infection of a colony with multiple types does occur, it might then be possible to utilize that to elucidate the spread of infection around a colony. Typing larvae and nurse bees from across each side of all the frames of the colony would show if types were confined to single side of a comb, or all types were found all over the colony. If the *M. plutonius* population has been homogenized, the pathogen might be being spread by forager bees, feeding nurse bees on multiple frames by trophallaxis (mouth-to-mouth feeding). If types are confined to a surface of a frame, the main mechanism of transmission within the colony may be a circular spread from infected larval faeces or dead brood, spread by the cell cleaning or brood tending behaviours of the nurse bees, or by the movements of newly emerged bees.

This thesis has therefore shown several novel aspects of EFB epidemiology and *M. plutonius* biology. At the genome scale, a correctly orientated chromosome sequence for *M. plutonius* LMG 20360 (the Type Strain) has been produced. This was used to identify mobile genetic elements, carrying important genes for bacterial competition, which were found to have a non-uniform distribution among a set of isolates. These elements were present in isolates of typical *M. plutonius*, but absent in atypical isolates, and the division between these two groups was further highlighted in a phylogenetic network computed from all coding sequences on the chromosome. The chromosome was also used to identify polymorphic loci for the first typing scheme for *M. plutonius*, the results of which correspond closely to those of the phylogenetic network. This scheme highlighted diversity within *M. plutonius* that had been undetectable in earlier studies,

and showed differences in diversity among different countries. These differences are likely to be in part due to the different EFB control regimes in place in the various countries. Within England and Wales the scheme has shown geographical clustering of variants. Some of these clusters were likely driven by beekeeping practices; examples of dissemination of types by sale of bees have been shown, and *M. plutonius* was also detected on beekeeping equipment. Other clusters appear to be natural transmission events. These are likely to be maintained by the bees themselves (*M. plutonius* was frequently detected in asymptomatic larval bees), or by a newly-identified potential vector, social wasps.

It is hoped that some of these advances will help to reduce the burden on beekeepers of this important brood disease. More robust hygiene practises by beekeepers and stricter testing of apparently symptomless material before sale, export or import might reduce the movement of disease between apiaries and beekeepers. At a national level, implementation of the *M. plutonius* MLST scheme on isolates collected by Bee Inspectors would be a useful tool for identifying the geographical or beekeeper source of infections, and implementing appropriate inspection and treatment responses. In the future, if the virulence, latency, or responses to treatment of STs or clonal complexes are found to differ, then there may be scope for differential control options to be initiated dependent upon the infecting variant.

7. Appendices

7.1. Appendix One – Coding Sequences Present on Mobile Genetic Elements

Position of Coding Sequence	Product
575784..575957	hypothetical protein
576000..576485	ADP-ribose pyrophosphatase
578291..580765	Type I restriction-modification system, restriction subunit R
580717..581280	Type I restriction-modification system, restriction subunit R
581323..581949	Type I restriction-modification system, DNA-methyltransferase subunit M
582026..582445	Type I restriction-modification system, DNA-methyltransferase subunit M
582544..582915	Type I restriction-modification system, DNA-methyltransferase subunit M
582912..583184	Type I restriction-modification system, specificity subunit S
583168..584154	Type I restriction-modification system, specificity subunit S
584837..584980	hypothetical protein
586112..586285	hypothetical protein

Coding sequences present on the first mobile genetic element in LMG 20360, and the products they encode.

Position of Coding Sequence	Product
591446..591871	hypothetical protein
592249..592815	hypothetical protein
592941..593525	hypothetical protein
complement 593574..594314	Lantibiotic ABC transporter
complement 594318..595064	lantibiotic ABC transporter permease
complement 595061..595705	Lantibiotic transport ATP-binding protein srtF
596064..596270	hypothetical protein
596292..596750	Two-component response regulator
596719..596979	Two-component response regulator
597149..598378	Two-component sensor histidine kinase, Nisin biosynthesis sensor NisK

Coding sequences present on the second mobile genetic element in LMG 20360, and the products they encode.

Position on chromosome	Product
complement 643642..644046	hypothetical protein
complement 644241..644375	hypothetical protein
complement 644607..645077	Phage integrase site-specific recombinase
complement 645197..645541	hypothetical protein
complement 645638..646309	hypothetical protein
complement 646414..647133	repressor protein
647315..647539	hypothetical protein
647553..648311	Phage antirepressor protein
648324..648503	hypothetical protein
648490..648681	hypothetical protein
648678..648818	hypothetical protein
648873..649181	hypothetical protein
649165..649902	hypothetical protein
649910..650563	hypothetical protein
650671..651357	conserved hypothetical protein
651361..651831	Phage replication initiation
652267..652533	hypothetical protein
652824..653006	Single-stranded DNA-binding protein
653033..653329	Single-stranded DNA-binding protein
653797..654216	hypothetical protein
654223..654969	hypothetical protein
655070..655270	hypothetical protein
655271..655657	hypothetical protein
655647..655913	Phage protein
655927..656346	Transcriptional regulator, RinA family
656660..657199	hypothetical protein
657220..657348	hypothetical protein
657791..658300	Phage terminase, small subunit
658477..658719	Terminase large subunit [Bacteriophage A118]
658768..659175	Terminase large subunit [Bacteriophage A118]
659145..659618	Terminase large subunit [Bacteriophage A118]
659635..661191	Phage minor capsid protein
661194..661406	minor capsid protein
661497..662330	Phage capsid and scaffold
662441..662812	hypothetical protein
662822..663019	hypothetical protein
663031..663900	Major capsid protein (Protein Gp34) (ORF3 protein)
663967..664134	hypothetical protein
664147..664356	hypothetical protein
664353..664703	Uncharacterized protein ORF5 (ORF118)
664704..665057	Phage minor capsid protein
665057..665464	minor capsid protein
665461..665982	minor capsid protein
666288..666593	hypothetical protein
666730..667347	Phage protein
667348..668979	Phage tail length tape-measure protein
669006..669284	Phage tail length tape-measure protein
670086..671624	Phage tail length tape-measure protein
671624..672448	hypothetical protein
672461..673837	hypothetical protein
673837..674085	hypothetical protein
674075..674767	prophage Lp1 protein 54
675248..675622	hypothetical protein

675612..675758	hypothetical protein
675796..676038	hypothetical protein
676040..676234	hypothetical protein
676249..677004	endolysin domain protein
677627..678499	Phage protein
678878..679111	Late competence protein ComGC, access of DNA to ComEA, FIG007487
679108..679512	Late competence protein ComGD, access of DNA to ComEA, FIG038316

Coding sequences present on the third mobile genetic element in LMG 20360, and the products they encode.

Position of Coding Sequence on Plasmid	Product
106606..107958	sugar ABC transporter permease
107964..108803	sugar ABC transporter permease
109436..109978	phenolic acid decarboxylase
110650..112758	hypothetical protein
113359..113712	hypothetical protein
complement 114652..115353	L-serine dehydratase subunit beta
complement 115381..116409	regulatory protein PfoR
complement 116820..118091	seryl-tRNA synthetase
119576..120847	seryl-tRNA synthetase
121258..122286	regulatory protein PfoR
122314..123015	L-serine dehydratase subunit beta
complement 123955..124308	hypothetical protein
complement 124909..127017	hypothetical protein
complement 127689..128231	phenolic acid decarboxylase
complement 128864..129703	maltose/maltodextrin ABC transporter permease MalG
complement 129709..131061	maltose/maltodextrin ABC transporter permease MalF
complement 131209..132315	maltose/maltodextrin ABC transporter substrate binding periplasmic protein MalE
complement 132384..132812	GTPase domain-containing protein
complement 133059..134132	cell surface protein

Coding sequences present on the ATCC 35311 plasmid that do not appear to be present in the contigs from isolates 7534, 7610, 7611, 7613 and 7606.

Product
Maltose/maltodextrin ABC transporter2C substrate binding periplasmic protein MalE
Maltose/maltodextrin ABC transporter2C permease protein MalF
Maltose/maltodextrin ABC transporter2C permease protein MalG
Phenolic acid decarboxylase
hypothetical protein
hypothetical protein
hypothetical protein
L-serine dehydratase2C alpha subunit
L-serine dehydratase2C beta subunit
Transcriptional regulator pfoR
Seryl-tRNA synthetase

Coding sequences present on the small contig found in isolates 7534, 7610, 7611, 7613 and 7606, and not in the ATCC 35311 plasmid or LMG 20360 chromosome.

7.2. Appendix Two – New MLST Allele Sequences

> galK_allele_7

CTACGTTTTTATTCGGAAAAATTTCCACAATTAGGTATTATTCAAAGTAATTTGGATGAATTAGTCTATAAAA
AAGAAGATGATTGGGCAAATTATCCTAAAGGCGTTCTCAAATATTTAAAAGAGAAGTATCCACAATTAACT
TTTGGAATGGATATTTTGTTTTGTGGAGATATTCCAAATGGTGCAGGTCTTTCTTCTTCTGCATCGATTGAAC
TCTTAATGGGTGTTATTGTTGATGATTTGTTTCAAATTGCTATAAAAAGATTATAACTAGTTAAAATAGGTC
AGCAGGTTGAGAATAATTTTATCGGTGTTAACCCTGGAATTATGGATCAATTTGCAATTGGCATGGGCAAA
AAAAACCAAGCGATACTTTTAGATACCAACACTTTAGAGTATAATTATGTACCGGCTTATTTTTCTGATCATC
AAGTCATAATTATGAATACGAATAAACGCCGAGAATTAGCTGATTCAAAATATAATGAAAGAAGAACTGAA
T

> galk_allele_8

CTACGTTTTTATTCGGAAAAATTTCCACAATTAGGTATTATTCAAAGTAATTTGGATGAATTAGTCTATAAAA
AAGAAGATGATTGGGCAAATTATCCTAAAGGCGTTCTCAAATATTTAAAAGAGAAGTATCCACAATTAACT
TTTGGAATGGATATTTTGTTTTGTGGAGATATTCCAAATGGTGCAGGTCTTTCTTCTTCTGCATCGATTGAAC
TCTTAATGGGTGTTATTGTTGACGATTTGTTTTAAATTGCTATAAAAAGATTATAAATAGTTAAAATAGGTC
AGCAGGTTGAGAATAATTTTATCGGTGTTAACCCTGGAATTATGGATCAATTTGCAATTGGCATGGGCAAA
AAAAACCAAGCGATACTTTTAGATACCAACACTTTAGAGTATAATTATGTACCGGCTTATTTTTCTGATCATC
AAGTCATAATTATGAATACGAATAAACGCCGAGAATTAGCTGATTCAAAATATAATGAAAGAAGAACTGAA
T

> galK_allele_9

CTACGTTTTTATTCGGAAAAATTTCCACAATTAGGTATTATTCAAAGTAATTTGGATGAATTAGTCTATAAAA
AAGAAGATGATTGGGCAAATTATCCTAAAGGCGTTCTCAAATATTTAAAAGAGAAGTATCCACAATTAACT
TTTGGAATGGATATTTTGTTTTGTGGAGATATTCCAAATGGTGCAGGTCTTTCTTCTTCTGCATCGATTGAAC
TCTTAATGGGTGTTATTGTTGACGATTTGTTTTAAATTGCTATAAAAAGATTATAAATAGTTAAAATAGGTC
AGCAGGTTGAGAATAATTTTATCGGTGTTAATCCTGGAATTATGGATCAATTTGCAATTGGCATGGGCAAA
AAAAACCAAGCGATACTTTTAGATACCAACACTTTAGAGTATAATTATGTACCGGCTTATTTTTCTGATCATC
AAGTCATAATTATGAATACGAATAAACGCCGAGAATTAGCTGATTCAAAATATAATGAAAGAAGAACTGAA
T

> galK_allele_10

CTACGTTTTTATTCGGAAAAATTTCCACAATTAGGTATTATTCAAAGTAATTTGGATGAATTAGTCTATAAAA
AAGAAGATGATTGGGCAAATTATCCTAAAGGCGTTCTCAAATATTTAAAAGAGAAGTATCCACAATTAACT
TTTGGAATGGATATTTTGTTTTGTGGAGATATTCCAAATGGTGCAGGTCTTTCTTCTTCTGCATCGATTGAAC
TCTTAATGGGTGTTATTGTTGATGATTTGTTTTAAATTGATATAAAAAGATTATAAATAGTTAAAATAGGTC
AGCAGGTTGAGAATAATTTTATCGGTGTTAACCCTGGAATTATGGATCAATTTGCAATTGGCATGGGCAAA
AAAAACCAAGCGATACTTTTAGATACCAACACTTTAGAGTATAATTATGTACCGGCTTATTTTTCTGATCATC
AAGTCATAATTATGAATACGAATAAACGCCGAGAATTAGCTGATTCAAAATATAATGAAAGAAGAACTGAA
T

> gbpB_allele_11

CAAGAAAGTGCTTCTGTTGAAACATCCAAAAATACAGAAAAGAATAAACCTGCTGAAAATAACCAGGCAA
GTGGAACATCTGTAGAACAACCAAAAGAAACACCAAAACAACCTGACGCAGGTCAACAACCGGAACAACC
AAAAGAAACACCAAAACAACCTGACGCAGGTCAACAACCAGAACAACCAAAAGAAACACCAAAACAACCT
GAGGTAGAACAAGTACAACCAGAACAACCAAAAGAAACACCAAAACCTGTTCAACCAGAACAGCCAGTAC
AACACCCTGTTGCCCACACTCCAACGCATAATAATTCAGGAAATGGATCAGCAGCAACTGGCGGCGTATCA
TCAGCAAAACGTGCAGCAGTTAATGCAGCATTGGCTGATGTAGGTAACTCTTATCAAACTGGTTGGAACCA
AC

> gbpB_allele_12

CAAGAAAGTGCTTCTGTTGAAACATCCAAAAATACAGAAAAGAATAAACCTGCTGAAAATAACCAGGCAA
GTGGAACATCTGTAGAACAACCAAAAGAAACACCAAAACAACCTGAGGTAGAACAAGTACAACCAGAACA
ACCAAAAGAAACACCAAAACCTGTTCAACCAGAACAGCCAGTACAACACCCTGTTGCCCACACTCCAACGC
ATAATAATTCAGGAAATGGATCAGCAGCAACTGGCGGCGTATCATCAGCAAAACGTGCAGCAGTTAATGC
AGCATTGGCTGATGTAGGTAACTCTTATCAAACTGGTTGGAACCAAC

> gbpB_allele_13

CAAGAAAGTGCTTCTGTTGAAACATCCAAAAATACAGAAAAGAATAAACCTGCTGAAAATAACCAGGCAA
GTGGAACATCTGTAGAACAACCAAAAGAAACACCAAAACAACCTGACGCAGGTCAACAACCAGAACAACC
AAAAGAAACACCAAAACAACCTGACGCAGGTCAACAACCAGAACAACCAAAAGAAACACCAAAACAACCT
GACGCAGGTCAACAACCAGAACAACCAAAAGAAACACCAAAACAACCTGAGGTAGAACAAGTACAACCAG
AACAACCAAAAGAAACACCAAAACCTGTTCAACCAGAACAGCCAGTACAACACCCTGTTGCCCACACTCCA
ACGCATAATAATTCAGGAAATGGATCAGCAGCAACTGGCGGCGTATCATCAGCAAAACGTGCAGCAGTTA
ATGCAGCATTGGCTGATGTAGGTAACTCTTATCAAACTGGTTGGAACCAAC

> gbpB_allele_14

CAAGAAAGTGCTTCTGTTGAAACATCCAAAAATACAGAAAAGAATAAACCTGCTGAAAATAACCAGGCAA
GTGGAACATCTGTAGAACAACCAAAAGAAACACCAAAACAACCTGACGCAGGTCAACAACCAGAACAACC
AAAAGAAACACCAAAACAACCTGACGCAGGTCAACAACCAGAACAACCAAAAGAAACACCAAAACAACCT
GAGGTAGAACAAGTACAACAAGTACAACCAGAACAACCAAAAGAAACACCAAAACCTGTTCAACCAGAAC
AGCCAGTACAACACCCTGTTGCCCACACTCCAACGCATAATAATTCAGGAAATGGATCAGCAGCAACTGGC
GGCGTATCATCAGCAAAACGTGCAGCAGTTAATGCAGCATTGGCTGATGTAGGTAACTCTTATCAAACTGG
TTGGAACCAAC

> gbpB_allele_15

CAAGAAAGTGCTTCTGTTGAAACATCCAAAAATACAGAAAAGAATAAACCTGCTGAAAATAACCAGGCAA
GTGGAACATCTGTAGAACAACCAAAAGAAACACCAAAACAACCTGACGCAGGTCAACAACCAGAACAACC
AAAAGAAACACCAAAACAACCTGAGGTAGAACAAGTACAACCAGAACAACCAAAAGAAACACCAAAACAA
CCTGACGCAGGTCAACAACCAGAACAACCAAAAGAAACACCAAAACAACCTGAGGTAGAACAAGTACAAC
CAGAACAACCAAAAGAAACACCAAAACCTGTTCAACCAGAACAGCCAGTACAACACCCTGTTGCCCACACT
CCAACGCATAATAATTCAGGAAATGGATCAGCAGCAACTGGCGGCGTATCATCAGCAAAACGTGCAGCAG
TTAATGCAGCATTGGCTGATGTAGGTAACTCTTATCAAACTGGTTGGAACCAAC

8. List of References

2007 No . 506 BEE DISEASES The Bee Diseases and Pests Control (Scotland) Order 2007 (2007).

Abbott,J.C., Aanensen,D.M., Rutherford,K., Butcher,S., and Spratt,B.G. (2005) WebACT--an online companion for the Artemis Comparison Tool. *Bioinformatics* **21**: 3665–3666.

Agroscope (2009) Swiss Bee Research Centre BRC, Project Status 2009.

Ahmed,N., Devi,S.M., Valverde,M. d. l. A., Viyachari,P., Machang'u,R.S., Ellis,W.A., and Hartskeerl,R.A. (2006) Multilocus sequence typing method for identification and genotypic classification of pathogenic Leptospira species. *Annals of Clinical Microbiology and Antimicrobials* **5**: 28.

Aizen,M. and Feinsinger,P. (1994) Habitat fragmentation, native insect pollinators, and feral honey bees in Argentine'Chaco Serrano'. *Ecological Applications* **4**: 378–392.

Allard,M.W., Luo,Y., Strain,E., Pettengill,J., Timme,R., Wang,C., et al. (2013) On the evolutionary history, population genetics and diversity among isolates of Salmonella Enteritidis PFGE pattern JEGX01.0004. *PloS one* **8**: e55254.

Allen,M., Ball,B., and Underwood,B. (1990) An isolate of Melissococcus pluton from Apis laboriosa. *Journal of Invertebrate Pathology* **55**: 439–440.

Allen,M.F. and Ball,B.V. (1993) The cultural characteristics and serological relationships of isolates of Melissococcus pluton. *Journal of Apicultural Research* **32**: 80–88.

Altschul,S., Gish,W., Miller,W., Myers,E., and Lipman,D. (1990) Basic local alignment search tool. *Journal of Molecular Biology* **215**: 403–410.

Anderson,R. and May,R. (1982) Coevolution of hosts and parasites. *Parasitology* **85**: 411–426.

Arai,R., Tominaga,K., Wu,M., Okura,M., Ito,K., Okamura,N., et al. (2012) Diversity of Melissococcus plutonius from Honeybee Larvae in Japan and Experimental Reproduction of European Foulbrood with Cultured Atypical Isolates. *PloS one* **7**: e33708.

Arias,M., Tingek,S., Kelitu,A., and Sheppard,W. (1996) Apis nulensis Tingek, Koeniger and Koeniger, 1996 and its genetic relationship with sympatric species inferred from DNA sequences. *Apidologie* **27**: 415–422.

Arias,M.C. and Sheppard,W.S. (2005) Phylogenetic relationships of honey bees (Hymenoptera:Apinae:Apini) inferred from nuclear and mitochondrial DNA sequence data. *Molecular Phylogenetics and Evolution* **37**: 25–35.

Axelsson,L.T., Chung,T.C., Dobrogosz,W.J. and Lindgren,S.E. (1989) Production of a broad spectrum antimicrobial substance by Lactobacillus reuteri. *Microbial Ecology in Health and Disease* **2**:131-136.

Aziz,R.K., Bartels,D., Best,A.A., DeJongh,M., Disz,T., Edwards,R.A., et al. (2008) The RAST Server: rapid annotations using subsystems technology. *BMC Genomics* **9**: 75.

Bailey,L. (1984) A strain of Melissococcus pluton cultivable on chemically defined media. *FEMS Microbiology Letters* **25**: 139–141.

Bailey,L. (1956) Ætiology of European Foul Brood; a Disease of the Larval Honey-Bee. *Nature* **178**: 1130.

Bailey,L. (1959a) An Improved Method for the Isolation of Streptococcus pluton, and Observations on Its Distribution and Ecology. *Journal of Insect Pathology* **1**: 80–85.

Bailey,L. (1974) An unusual type of Streptococcus pluton from the eastern hive bee. *Journal of Invertebrate Pathology* **23**: 246–247.

Bailey,L. (1981) Honey Bee Pathology 1st ed. Academic Press, London.

Bailey,L. (1983) Melissococcus pluton, the cause of European foulbrood of honey bees (Apis spp.). *Journal of Applied Bacteriology* **55**: 65–69.

Bailey,L. (1959b) Recent research on the natural history of European foulbrood. *Bee World* **40**: 66–70.

Bailey,L. (1957) The isolation and cultural characteristics of Streptococcus pluton and further Observations on Bacterium eurydice. *Journal of general Microbiology* **17**: 39–48.

Bailey,L. and Collins,M.D. (1982) Reclassification of "Streptococcus pluton" (White) in a new genus Melissococcus, as Melissococcus pluton nom. rev.; comb. nov. *Journal of Applied Bacteriology* **53**: 215–217.

Bailey,L. and Gibbs,A. (1962) Cultural characters of Streptococcus pluton and its differentiation from associated enterococci. *Microbiology* **28**: 385.

Bairoch,A., Apweiler,R., Wu,C.H., Barker,W.C., Boeckmann,B., Ferro,S., et al. (2005) The Universal Protein Resource (UniProt). *Nucleic Acids Research* **33**: D154–9.

Balseiro,A., González-Quirós,P., Rodríguez,O., Francisca Copano,M., Merediz,I., de Juan,L., et al. (2013) Spatial relationships between Eurasian badgers (Meles meles) and cattle infected with Mycobacterium bovis in Northern Spain. *The Veterinary Journal* **197**: 739-745

Barbuddhe,S.B., Maier,T., Schwarz,G., Kostrzewa,M., Hof,H., Domann,E., et al. (2008) Rapid identification and typing of listeria species by matrix-assisted laser desorption ionization-time of flight mass spectrometry. *Applied and Environmental Microbiology* **74**: 5402–5407.

Beggs,J., Toft,R., Malham,J., Rees,J., Tilley,J., Moller,H., and Alspach,P. (1998) The difficulty of reducing introduced wasp (Vespula vulgaris) populations for conservation gains. *New Zealand Journal of Ecology* **22**: 55–63.

Behr,T., Koob,C., Schedl,M., Mehlen,A., Meier,H., Knopp,D., et al. (2000) A Nested Array of rRNA Targeted Probes for the Detection and Identification of Enterococci by Reverse Hybridization. *Systematic and Applied Microbiology* **23**: 563–572.

Belloy,L., Imdorf,A., Fries,I., Forsgren,E., Berthoud,H., Kuhn,R., and Charriere,J.-D. (2007) Spatial distribution of Melissococcus plutonius in adult honey bees collected from apiaries and colonies with and without symptoms of European foulbrood. *Apidologie* **38**: 136–140.

Benson,D.A., Karsch-Mizrachi,I., Clark,K., Lipman,D.J., Ostell,J., and Sayers,E.W. (2012) GenBank. *Nucleic Acids Research* **40**: D48–53.

Benson,D.A., Karsch-Mizrachi,I., Lipman,D.J., Ostell,J., and Sayers,E.W. (2010) GenBank. *Nucleic Acids Research* **38**: D46–51.

Benson,D.A., Karsch-Mizrachi,I., Lipman,D.J., Ostell,J., and Wheeler,D.L. (2004) GenBank: update. *Nucleic Acids Research* **32**: D23–26.

Biek,R., O'Hare,A., Wright,D., Mallon,T., McCormick,C., Orton,R.J., et al. (2012) Whole genome sequencing reveals local transmission patterns of Mycobacterium bovis in sympatric cattle and badger populations. *PLoS pathogens* **8**: e1003008.

Biesmeijer,J.C., Roberts,S.P.M., Reemer,M., Ohlemüller,R., Edwards,M., Peeters,T., et al. (2006) Parallel declines in pollinators and insect-pollinated plants in Britain and the Netherlands. *Science* **313**: 351–354.

Blum-Oehler,G., Oswald,S., Eiteljörge,K., Sonnenborn,U., Schulze,J., Kruis,W., and Hacker,J. (2003) Development of strain-specific PCR reactions for the detection of the probiotic Escherichia coli strain Nissle 1917 in fecal samples. *Research in Microbiology* **154**: 59–66.

Boonham,N., Duffy,B., and Frey,J. (2013) Erwinia amylovora loop-mediated isothermal amplification (LAMP) assay for rapid pathogen detection and on-site diagnosis of fire blight. *Journal of Microbiological Methods* **92**: 332–339.

Bremer,H. and Dennis,P. (1996) Modulation of chemical composition and other parameters of the cell by growth rate. *Escherichia coli and Salmonella: cellular and molecular biology* **2**: 1553–1569.

Brul,S. and Coote,P. (1999) Preservative agents in foods: Mode of action and microbial resistance mechanisms. *International journal of food microbiology* **50**: 1-17.

Budge,G.E., Barrett,B., Jones,B., Pietravalle,S., Marris,G., Chantawannakul,P., et al. (2010) The occurrence of Melissococcus plutonius in healthy colonies of Apis mellifera and the efficacy of European foulbrood control measures. *Journal of Invertebrate Pathology* **105**: 164–170.

Budge,G.E., Shaw,M.W., Colyer, A., Pietravalle,S., and Boonham,N. (2009) Molecular tools to investigate Rhizoctonia solani distribution in soil. *Plant Pathology* **58**: 1071–1080.

Budge,G.E., Shirley,M.D.F., Jones,B., Quill,E., Tomkies,V., Feil,E.J., Brown,M.A., and Haynes,E.G. (2014) Molecular epidemiology and population structure of the honey bee brood pathogen *Melissococcus plutonius*. *The ISME Journal* doi:10.1038/ismej.2014.20

Butaye,P., Cloeckaert,A., and Schwarz,S. (2003) Mobile genes coding for efflux-mediated antimicrobial resistance in Gram-positive and Gram-negative bacteria. *International Journal of Antimicrobial Agents* **22**: 205–210.

Bygraves,J., Fox,A., Urwin,R., Gray,S., Russell,J., Feavers,I., and Maiden,M. (1999) Population Genetic and Evolutionary Approaches to Analysis of Neisseria meningiditis Isolates Belonging to the ET-5 Complex. *Journal of Bacteriology* **181**: 5551–5556.

Cai,J. and Collins,M.D. (1994) Evidence for a close phylogenetic relationship between Melissococcus pluton, the causative agent of European Foulbrood Disease, and the genus Enterococcus. *International Journal of Systematic Bacteriology* **44**: 365–367.

Do Carmo,R., Franceschinelli,E., and Silveira,F. (2004) Introduced honeybees (Apis mellifera) reduce pollination success without affecting the floral resource taken by native pollinators. *Biotropica* **36**: 371–376.

Carr,A. (2011) Asian Honeybee Possible Environmental Impacts. Canberra.

Carver,T.J., Rutherford,K.M., Berriman,M., Rajandream,M.-A., Barrell,B.G., and Parkhill,J. (2005) ACT: the Artemis Comparison Tool. *Bioinformatics* **21**: 3422–3423.

Chaudhuri,R.R., Loman,N.J., Snyder,L. a S., Bailey,C.M., Stekel,D.J., and Pallen,M.J. (2008) xBASE2: a comprehensive resource for comparative bacterial genomics. *Nucleic Acids Research* **36**: D543–546.

Cheigh,C.-I. and Pyun,Y.-R. (2005) Nisin Biosynthesis and its Properties. *Biotechnology Letters* **27**: 1641–1648.

Chen,Y., Evans,J., and Feldlaufer,M. (2006) Horizontal and vertical transmission of viruses in the honey bee, Apis mellifera. *Journal of Invertebrate Pathology* **92**: 152–159.

Clapperton,B.K., Alspach,P. a., Moller,H., and Matheson, a. G. (1989) The impact of common and German wasps (Hymenoptera: Vespidae) on the New Zealand beekeeping industry. *New Zealand Journal of Zoology* **16**: 325–332.

Clarke,K.C., McLafferty,S.L., and Tempalski,B.J. (1996) On epidemiology and geographic information systems: a review and discussion of future directions. *Emerging Infectious Diseases* **2**: 85–92.

The European Commission (2013) COMMISSION IMPLEMENTING REGULATION (EU) No 485/2013 of 24 May 2013 amending Implementing Regulation (EU) No 540/2011, as regards the conditions of approval of the active substances clothianidin, thiamethoxam and imidacloprid, and prohibiting the use and s. *Official Journal of the European Union* **11**: 12–26.

Corbet,S., Williams,I., and Osborne,J. (1992) Bees and the pollination of crops and wild flowers in the European Community. *Apiacta* **4**: 47–59.

Cox-Foster,D.L., Conlan,S., Holmes,E.C., Palacios,G., Evans,J.D., Moran,N.A., et al. (2007) A metagenomic survey of microbes in honey bee colony collapse disorder. *Science* **318**: 283–287.

Crane,E. (1999) The World History of Beekeeping and Honey Hunting. 1st ed. Routledge.

Cromley,E.K. (2003) GIS and Disease. *Annual Review of Public Health* **24**: 7–24.

Crowther,J. (1995) ELISA: theory and practice. Humana Press.

Currie,R.W., Pernal,S.F., and Guzmán-Novoa,E. (2010) Honey bee colony losses in Canada. *Journal of Apicultural Research* **49**: 104–106.

Dahle,B., Sørum,H., and Weidemann,J. (2011) European foulbrood in Norway: How to deal with a major outbreak after 30 years absence. In, *Proceedings of the COLOSS Workshop: The future of brood disease research – guidelines, methods and development*. Copenhagen, p. 8.

Dainat,B., Vanengelsdorp,D., and Neumann,P. (2012) Colony collapse disorder in Europe. *Environmental Microbiology Reports* **4**: 123–125.

Danecek,P., Auton,A., Abecasis,G., Albers,C.A., Banks,E., DePristo,M.A., et al. (2011) The variant call format and VCFtools. *Bioinformatics* **27**: 2156–2158.

Darling,A.E., Mau,B., and Perna,N.T. (2010) progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PloS one* **5**: e11147.

Defra (2013) Review of policies on managing and controlling pests and diseases of honey bees - Evidence profile on European Foulbrood (EFB) WWW Document: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/221155/bee-health-consult-efbprofile-20130110.pdf

Desneux,N., Decourtye,A., and Delpuech,J.-M. (2007) The sublethal effects of pesticides on beneficial arthropods. *Annual Review of Entomology* **52**: 81–106.

Disayathanoowat,T., Young,J.P.W., Helgason,T., and Chantawannakul,P. (2012) T-RFLP analysis of bacterial communities in the midguts of Apis mellifera and Apis cerana honey bees in Thailand. *FEMS Microbiology Ecology* **79**: 273–281.

Dixon,P. (2002) Nearest-neighbor contingency table analysis of spatial segregation for several species. *Ecoscience* **9**: 142–151.

Dixon,P. (1994) Testing spatial segregation using a nearest-neighbor contingency table. *Ecology* **75**: 1940–1948.

Djordjevic,S., Noone,K., Smith,L., and Hornitzky,M. (1998) Development of a hemi-nested PCR assay for the specific detection of Melissococcus pluton. *Journal of Apicultural Research* **37**: 165–174.

Djordjevic,S.P., Smith,L.A., Forbes,W.A., and Hornitzky,M.A. (1999) Geographically diverse Australian isolates of Melissococcus pluton exhibit minimal genotypic diversity by restriction endonuclease analysis. *FEMS Microbiology Letters* **173**: 311–318.

Dunn,R.R., Davies,T.J., Harris,N.C., and Gavin,M.C. (2010) Global drivers of human pathogen richness and prevalence. *Proceedings of the Royal Society B: Biological Sciences* **277**: 2587–2595.

Dunn,R.R., Fierer,N., Henley,J.B., Leff,J.W., and Menninger,H.L. (2013) Home Life: Factors Structuring the Bacterial Diversity Found within and between Homes. *PLoS ONE* **8**: e64133.

Dunning Hotopp,J.C., Clark,M.E., Oliveira,D.C.S.G., Foster,J.M., Fischer,P., Muñoz Torres,M.C., et al. (2007) Widespread lateral gene transfer from intracellular bacteria to multicellular eukaryotes. *Science* **317**: 1753–1756.

Eckhart,L., Bach,J., Ban,J., and Tschachler,E. (2000) Melanin binds reversibly to thermostable DNA polymerase and inhibits its activity. *Biochemical and Biophysical Research Communications* **271**: 726–730.

Eisen,J.A., Heidelberg,J.F., White,O., and Salzberg,S.L. (2000) Evidence for symmetric chromosomal inversions around the replication origin in bacteria. *Genome Biology* **1**: 1–0011.

Endo,A. and Salminen, S. (2013) Honeybees and beehives are rich sources for fructophilic lactic acid bacteria. *Systematic and Applied Microbiology* **36**: 444-448.

Engel,M. (1999) The taxonomy of recent and fossil honey bees (Hymenoptera: Apidae; Apis). *Journal of Hymenoptera Research* **8**: 165–196.

Engel,M., Hinojosa-Diaz,I., and Rasnitsyn,A. (2009) A honey bee from the Miocene of Nevada and the biogeography of Apis (Hymenoptera: Apidae: Apini). *Proceedings of the California Academy of Sciences* **60**: 23–38.

Enright,M.C. and Spratt,B.G. (1999) Multilocus sequence typing. *Trends in Microbiology* **7**: 482–487.

Evans,J.D. (2003) Diverse origins of tetracycline resistance in the honey bee bacterial pathogen Paenibacillus larvae. *Journal of Invertebrate Pathology* **83**: 46–50.

Evans,J.D. and Schwarz,R.S. (2011) Bees brought to their knees: microbes affecting honey bee health. *Trends in Microbiology* **19**: 614–620.

Evans,J.D. and Spivak,M. (2010) Socialized medicine: Individual and communal disease barriers in honey bees. *Journal of Invertebrate Pathology* **103**: 62–72.

Feil,E.J., Li,B.C., Aanensen,D.M., Hanage,W.P., and Spratt,B.G. (2004) eBURST : Inferring Patterns of Evolutionary Descent among Clusters of Related Bacterial Genotypes from Multilocus Sequence Typing Data. *Journal of Bacteriology* **186**: 1518–1530.

Fleischmann,R.D., Adams,M.D., White,O., Clayton,R.A., Kirkness,E.F., Kerlavage,A.R., et al. (1995) Whole-genome random sequencing and assembly of Haemophilus influenzae Rd. *Science* **269**: 496–512.

Forsgren,E. (2010) European foulbrood in honey bees. *Journal of Invertebrate Pathology* **103**: S5–9.

Forsgren,E., Budge,G.E., Charrière,J.-D., and Hornitzky,M.A.Z. (2013) Standard methods for European foulbrood research. *Journal of Apicultural Research* **52**: 1–14.

Forsgren,E. and Fries,I. (2010) Comparative virulence of Nosema ceranae and Nosema apis in individual European honey bees. *Veterinary parasitology* **170**: 212–217.

Forsgren,E., Lundhagen,A.C., Imdorf,A., and Fries,I. (2005) Distribution of Melissococcus plutonius in honeybee colonies with and without symptoms of European foulbrood. *Microbial Ecology* **50**: 369–374.

Foxman,B. and Riley,L. (2001) Molecular epidemiology: focus on infection. *American Journal of Epidemiology* **153**: 1135–1141.

Francisco,A.P., Bugalho,M., Ramirez,M., and Carrico,J.A. (2009) Global optimal eBURST analysis of multilocus typing data using a graphic matroid approach. *BMC Bioinformatics* **10**: 152.

Francisco,A.P., Vaz,C., Monteiro,P.T., Melo-Cristino,J., Ramirez,M., and Carriço,J.A. (2012) PHYLOViZ: phylogenetic inference and data visualization for sequence based typing methods. *BMC Bioinformatics* **13**: 87.

Fraser,C.M., Gocayne,J.D., White,O., Adams,M.D., Clayton,R.A., Fleischmann,R.D., et al. (1995) The minimal gene complement of Mycoplasma genitalium. *Science* **270**: 397–404.

Fries,I. (2010) Nosema ceranae in European honey bees (Apis mellifera). *Journal of Invertebrate Pathology* **103**: S73–79.

Frost,L.S., Leplae,R., Summers,A.O., and Toussaint,A. (2005) Mobile genetic elements: the agents of open source evolution. *Nature Reviews Microbiology* **3**: 722–732.

Fukami,T., Martijn Bezemer,T., Mortimer,S.R., and Putten,W.H. (2005) Species divergence and trait convergence in experimental plant community assembly. *Ecology Letters* **8**: 1283–1290.

Gallai,N., Salles,J.-M., Settele,J., and Vaissière,B.E. (2009) Economic valuation of the vulnerability of world agriculture confronted with pollinator decline. *Ecological Economics* **68**: 810–821.

Genersch,E., Forsgren,E., Pentikäinen,J., Ashiralieva,A., Rauch,S., Kilwinski,J., and Fries,I. (2006) Reclassification of Paenibacillus larvae subsp. pulvifaciens and Paenibacillus larvae subsp. larvae as Paenibacillus larvae without subspecies differentiation. *International Journal of Systematic and Evolutionary Microbiology* **56**: 501–511.

Genersch,E., von der Ohe,W., Kaatz,H., Schroeder,A., Otten,C., Büchler,R., et al. (2010) The German bee monitoring project: a long term study to understand periodically high winter losses of honey bee colonies. *Apidologie* **41**: 332–352.

Glenn,T.C. (2011) Field guide to next-generation DNA sequencers. *Molecular Ecology Resources* **11**: 759–769.

Goodchild,A., Watkins,G., Sayers,A., Jones,J., and Clifton-Hadley,R. (2012) Geographical association between the genotype of bovine tuberculosis in found dead badgers and in cattle herds. *Veterinary Record* **170**: 259.

Gopal,R., Goodchild,A., Hewinson,R.G., De La Rua Domenech,R., and Clifton-Hadley,R. (2006) Introduction of bovine tuberculosis to north-east England by bought-in cattle. *Veterinary Record* **159**: 265–271.

Gotelli,N. and Colwell,R. (2001) Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. *Ecology Letters* **4**: 379–391.

Goullet,P. (1977) Relationships between electrophoretic patterns of esterases from Salmonella. *Journal of General Microbiology* **98**: 535–542.

Goulson,D. (2003) Effects of introduced bees on native ecosystems. *Annual Review of Ecology, Evolution, and Systematics* **34**: 1–26.

Govan,V.A., Brozel,V., Allsopp,M.H., and Davison,S. (1998) A PCR Detection Method for Rapid Identification of Melissococcus pluton in Honeybee Larvae. *Applied and Environmental Microbiology* **64**: 1983–1985.

Grad,Y.H., Lipsitch,M., Feldgarden,M., Arachchi,H.M., Cerqueira,G.C., Fitzgerald,M., et al. (2012) Genomic epidemiology of the Escherichia coli O104:H4 outbreaks in Europe, 2011. *Proceedings of the National Academy of Sciences* **109**: 3065–3070.

Grant,J.R. and Stothard,P. (2008) The CGView Server: a comparative genomics tool for circular genomes. *Nucleic Acids Research* **36**: W181–184.

Graystock,P., Yates,K., Evison,S.E.F., Darvill,B., Goulson,D., and Hughes,W.O.H. (2013) The Trojan hives: pollinator pathogens, imported and distributed in bumblebee colonies. *Journal of Applied Ecology*.

Greene,A. (1991) Dolichovespula and Vespula. In, Ross,K. and Matthews,R. (eds), *The Social Biology of Wasps*. Cornell University Press, pp. 263–308.

Greenleaf,S.S. and Kremen,C. (2006) Wild bees enhance honey bees' pollination of hybrid sunflower. *Proceedings of the National Academy of Sciences* **103**: 13890–13895.

Gupta,S. and Maiden,M.C. (2001) Exploring the evolution of diversity in pathogen populations. *Trends in Microbiology* **9**: 181–185.

Guzmán-Novoa,E., Eccles,L., and Calvete,Y. (2010) Varroa destructor is the main culprit for the death and reduced populations of overwintered honey bee (Apis mellifera) colonies in Ontario, Canada. *Apidologie* **41**: 443–450.

Hansen,H. and Brødsgaard,C.J. (1999) American foulbrood: a review of its biology, diagnosis and control. *Bee World* **80**: 5–23.

Hara-Kudo,Y., Yoshino,M., Kojima,T., and Ikedo,M. (2005) Loop-mediated isothermal amplification for the rapid detection of Salmonella. *FEMS Microbiology Letters* **253**: 155–161.

Harris,B.Z. and Singer,M. (1998) Identification and Characterization of the Myxococcus xanthus argE Gene. *J Bacteriol* **180**: 6412–6414.

Harris,S.R., Feil,E.J., Holden,M.T.G., Quail,M. a, Nickerson,E.K., Chantratita,N., et al. (2010) Evolution of MRSA during hospital transmission and intercontinental spread. *Science* **327**: 469–474.

Harrison,P.W., Lower,R.P.J., Kim,N.K.D., and Young,J.P.W. (2010) Introducing the bacterial "chromid": not a chromosome, not a plasmid. *TRENDS in Microbiology* **18**: 141–148.

Hay,S.I., Battle,K.E., Pigott,D.M., Smith,D.L., Moyes,C.L., Bhatt,S., et al. (2013) Global mapping of infectious disease. *Philosophical Transactions of the Royal Society B: Biological Sciences* **368**: 1614.

Haynes,E., Helgason,T., Young,J.P.W., Thwaites,R., and 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* **5**: 525–529.

Henry,M., Béguin,M., Requier,F., Rollin,O., Odoux,J.-F., Aupinel,P., et al. (2012) A common pesticide decreases foraging success and survival in honey bees. *Science* **336**: 348–350.

Héchard,Y. and Sahl,H-G. (2002) Mode of action of modified and unmodified bacteriocins from Gram-positive bacteria. *Biochimie* **84**: 545-557.

Hisashi,F. (2010) Profitable beekeeping with Apis cerana. *Bees for Development* **94**: 8–11.

Hodgetts,J., Tomlinson,J., Boonham,N., Gonzalez-Martin,I., Nikolic,P., Swarbrick,P., et al. (2011) Development of rapid in-field loop-mediated isothermal amplification (LAMP) assays for phytoplasmas. *Bulletin on Insectology* **64**: S41–42.

Hoffmann,S., Otto,C., Kurtz,S., Sharma,C.M., Khaitovich,P., Vogel,J., et al. (2009) Fast mapping of short sequences with mismatches, insertions and deletions using index structures. *PLoS Computational Biology* **5**: e1000502.

Homan,W.L., Tribe,D., Poznanski,S., Li,M., Hogg,G., Spalburg,E., et al. (2002) Multilocus Sequence Typing Scheme for Enterococcus faecium. *Journal of Clinical Microbiology* **40**: 1963–1971.

Horner-Devine,M.C., Lage,M., Hughes,J.B., and Bohannan,B.J.M. (2004) A taxa-area relationship for bacteria. *Nature* **432**: 750–753.

Hornitzky,M. and Smith,L. (1999) Sensitivity of Australian Melissococcus pluton isolates to oxytetracycline hydrochloride. *Australian Journal of Experimental Agriculture* **39**: 881–883.

Hurles,M. (2004) Gene duplication: the genomic trade in spare parts. *PLoS Biology* **2**: E206.

Huson,D.H. and Bryant,D. (2006) Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution* **23**: 254–267.

Huynen,M., Snel,B., Lathe III,W., and Bork,P. (2000) Predicting Protein Function by Genomic Context: Quantitative Evaluation and Qualitative Inferences. *Genome Research* **10**: 1204-1210.

Illumina (2012) FDA Award Summary Project Overview U . S . Food and Drug Administration (FDA) Selects MiSeq ® for Sequencing Foodborne Pathogens. WWW Document: http://www.illuminakk.co.jp/document/pdf/flyer_miseq_fda_award.pdf

Illumina (2013a) illumina Sequencing Technology. WWW Document: http://www.illumina.com/technology/sequencing_technology.ilmn

Illumina (2013b) illumina Paired-End Sequencing. WWW Document: http://www.illumina.com/technology/paired_end_sequencing_assay.ilmn

Illumina (2013c) illumina Mate Pair Sequencing. WWW Document: http://www.illumina.com/technology/mate_pair_sequencing_assay.ilmn

Jackson,P.J., Walthers,E. a, Kalif, a S., Richmond,K.L., Adair,D.M., Hill,K.K., et al. (1997) Characterization of the variable-number tandem repeats in vrrA from different Bacillus anthracis isolates. *Applied and Environmental Microbiology* **63**: 1400–1405.

Janda,J.M. and Abbott,S.L. (2007) 16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: pluses, perils, and pitfalls. *Journal of Clinical Microbiology* **45**: 2761–2764.

Jefferson,T. (1787) Notes on the State of Virginia. Stockdale, London.

Jenkins,H., Woodroffe,R., Donnelly,C., Cox,D., Johnston,W., Bourne,F., et al. (2007) Effects of culling on spatial associations of Mycobacterium bovis infections in badgers and cattle. *Journal of Applied Ecology* **44**: 897–908.

Jensen,A.B., Palmer,K. a, Boomsma,J.J., and Pedersen,B. V (2005) Varying degrees of Apis mellifera ligustica introgression in protected populations of the black honeybee, Apis mellifera mellifera, in northwest Europe. *Molecular Ecology* **14**: 93–106.

Jeyaprakash,A., Hoy,M.A., and Allsopp,M.H. (2003) Bacterial Diversity in Worker Adults of Apis mellifera capensis and Apis mellifera scutellata (Insecta: Hymenoptera) Assessed Using 16S rRNA Sequences. *Journal of Invertebrate Pathology* **84**: 96–103.

Johnson,B.R. (2010) Division of labor in honeybees: form, function, and proximate mechanisms. *Behavioral Ecology and Sociobiology* **64**: 305–316.

Kanbar,G., Engels,W., Nicholson,G.J., Hertle,R., and Winkelmann,G. (2004) Tyramine functions as a toxin in honey bee larvae during Varroa-transmitted infection by Melissococcus pluton. *FEMS Microbiology Letters* **234**: 149–154.

Karlin,S. (2001) Detecting anomalous gene clusters and pathogenicity islands in diverse bacterial genomes. *TRENDS in Microbiology* **9**: 335–343.

Kato,M., Shibata,A., and Yasui,T. (1999) Impact of introduced honeybees, Apis mellifera, upon native bee communities in the Bonin (Ogasawara) Islands. *Researches in Population Ecology* **41**: 217–228.

Kennedy,P.G. and Bruns,T.D. (2005) Priority effects determine the outcome of ectomycorrhizal competition between two Rhizopogon species colonizing Pinus muricata seedlings. *New Phytologist* **166**: 631–638.

Killgore,G., Thompson,A., Johnson,S., Brazier,J., Kuijper,E., Pepin,J., et al. (2008) Comparison of Seven Techniques for Typing International Epidemic Strains of Clostridium difficile: Restriction Endonuclease Analysis, Pulsed-Field Gel Electrophoresis, PCR-Ribotyping, Multilocus Sequence Typing, Multilocus Variable-Number Tandem-Repeat Analysis, Amplified Fragment Length Polymorphism, and Surface Layer Protein A Gene Sequence Typing. *Journal of Clinical Microbiology* **46**: 431–437.

Klein,A.-M., Vaissière,B.E., Cane,J.H., Steffan-Dewenter,I., Cunningham,S. a, Kremen,C., and Tscharntke,T. (2007) Importance of pollinators in changing landscapes for world crops. *Proceedings of the Royal Society B: Biological Sciences* **274**: 303–313.

Kobayashi,I. (2001) Behavior of restriction–modification systems as selfish mobile elements and their impact on genome evolution. *Nucleic Acids Research* **29**: 3742–3756.

Kobayashi,K., Ehrlich,S.D., Albertini,A., Amati,G., Andersen,K.K., Arnaud,M., et al. (2003) Essential Bacillus subtilis genes. *Proceedings of the National Academy of Sciences* **100**: 4678–4683.

Koch,H. and Schmid-Hempel,P. (2011) Bacterial communities in central European bumblebees: low diversity and high specificity. *Microbial Ecology* **62**: 121–133.

Koleff,P., Gaston,K.J., and Lennon,J.J. (2003) Measuring Beta Diversity for Presence–Absence Data. *Journal of Animal Ecology* **72**: 367–382.

Köser,C.U., Holden,M.T.G., Ellington,M.J., Cartwright,E.J.P., Brown,N.M., Ogilvy-Stuart,A.L., et al. (2012) Rapid whole-genome sequencing for investigation of a neonatal MRSA outbreak. *The New England Journal of Medicine* **366**: 2267–2275.

Krawiec,S. and Riley,M. (1990) Organization of the Bacterial Chromosome. *Microbiological Reviews* **54**: 502–539.

Kristin,A. and Miranda,H. (2013) The root microbiota—a fingerprint in the soil? *Plant and Soil* **370**: 671–686.

De la Cruz,M. (2008) Metodos para analizar datos puntuales Asociacion Espanola de Ecologia Terrestre, Universidad Rey Juan Carlos and Caja de Ahorros del Mediterraneo.

De La Rua,P., Jaffe,R., Dall'Olio,R., Munoz,I., and Serrano,J. (2009) Biodiversity, conservation and current threats to European honeybees. *Apidologie* **40**: 263–284.

Lancefield,R. (1933) A serological differentiation of human and other groups of hemolytic streptococci. *The Journal of Experimental Medicine* **57**: 571–595.

Lazcka,O., Del Campo,F.J., and Muñoz,F.X. (2007) Pathogen detection: a perspective of traditional methods and biosensors. *Biosensors and Bioelectronics* **22**: 1205–1217.

Lennon,J.J., Koleff,P., Greenwood,J.J.D., and Gaston,K.J. (2001) The geographical structure of British bird distributions: diversity, spatial turnover and scale. *Journal of Animal Ecology* **70**: 966–979.

Lenski,R.E and May,R.M. (1994) The Evolution of Virulence in Parasites and Pathogens: Reconciliation Between Two Competing Hypotheses. *Journal of Theoretical Biology* **169**: 253-265

Li,H., Handsaker,B., Wysoker,A., Fennell,T., Ruan,J., Homer,N., et al. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**: 2078–2079.

Litvintseva,A.P., Thakur,R., Vilgalys,R., and Mitchell,T.G. (2006) Multilocus Sequence Typing Reveals Three Genetic Subpopulations of Cryptococcus neoformans var. grubii (serotype A), including a Unique Population in Botswana. *Genetics* **172**: 2223–2238.

Lowder,B. V, Cartwright,R.A., Guinane,C.M., Ben Zakour,N.L., Weinert,L.A., Conway-Morris,A., et al. (2009) Recent human-to-poultry host jump, adaptation, and pandemic spread of Staphylococcus aureus. *Proceedings of the National Academy of Sciences* **106**: 19545–19550.

Lwoff,A. (1953) Lysogeny. *Bacteriological Reviews* **17**: 269–337.

Maiden,M., Bygraves,J., Feil,E., Morelli,G., Russell,J., Urwin,R., et al. (1998) Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proceedings of the National Academy of Sciences* **95**: 3140–3145.

Mardis,E.R. (2008) The impact of next-generation sequencing technology on genetics. *Trends in genetics* **24**: 133–141.

Martinson,V.G., Danforth,B.N., Minckley,R.L., Rueppell,O., Tingek,S., and Moran,N. a (2011) A simple and distinctive microbiota associated with honey bees and bumble bees. *Molecular Ecology* **20**: 619–628.

Martiny,J., Eisen,J., Penn,K., Allison,S., and Horner-Devine,M. (2011) Drivers of bacterial β-diversity depend on spatial scale. *Proceedings of the National Academy of Sciences* **108**: 7850–7854.

McKee,B., Djordjevic,S., Goodman,R., and Hornitzky,M. (2003) The detection of Melissococcus pluton in honey bees (Apis mellifera) and their products using a hemi-nested PCR. *Apidologie* **34**: 19–27.

McKee,B., Goodman,R., and Hornitzky,M. (2004) The transmission of European foulbrood (Melissococcus plutonius) to artificially reared honey bee larvae (Apis mellifera). *Journal of Apicultural Research* **43**: 93–100.

McLean,M.J., Wolfe,K.H., and Devine,K.M. (1998) Base composition skews, replication orientation, and gene orientation in 12 prokaryote genomes. *Journal of Molecular Evolution* **47**: 691–696.

Meixner,M.D., Costa,C., Kryger,P., Hatjina,F., Bouga,M., Ivanova,E., and Büchler,R. (2010) Conserving diversity and vitality for honey bee breeding. *Journal of Apicultural Research* **49**: 85–92.

Mellmann,A., Harmsen,D., Cummings,C. a, Zentz,E.B., Leopold,S.R., Rico,A., et al. (2011) Prospective genomic characterization of the German enterohemorrhagic Escherichia coli O104:H4 outbreak by rapid next generation sequencing technology. *PloS one* **6**: e22751.

Milne,I., Bayer,M., Cardle,L., Shaw,P., Stephen,G., Wright,F., and Marshall,D. (2010) Tablet - next generation sequence assembly visualization. *Bioinformatics* **26**: 401–402.

De Miranda,J.R. and Genersch,E. (2010) Deformed wing virus. *Journal of Invertebrate Pathology* **103**: S48–61.

Mishra,A., Afik,O., Cabrera,M.L., Delaplane,K.S., and Mowrer,J.E. (2013) Inorganic nitrogen derived from foraging honey bees could have adaptive benefits for the plants they visit. *PloS one* **8**: e70591.

Moreno-Hagelsieb,G. and Latimer,K. (2008) Choosing BLAST options for better detection of orthologs as reciprocal best hits. *Bioinformatics* **24**: 319–324.

Mori,Y. and Notomi,T. (2009) Loop-mediated isothermal amplification (LAMP): a rapid, accurate, and cost-effective diagnostic method for infectious diseases. *Journal of Infection and Chemotherapy* **15**: 62–69.

Moriya,Y., Itoh,M., Okuda,S., Yoshizawa,A.C., and Kanehisa,M. (2007) KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Research* **35**: W182–185.

Mrázek,J. and Karlin,S. (1998) Strand compositional asymmetry in bacterial and large viral genomes. *Proceedings of the National Academy of Sciences* **95**: 3720–3725.

Mrázek,J., Strosová,L., Fliegerová,K., Kott,T., and Kopecný,J. (2008) Diversity of insect intestinal microflora. *Folia Microbiologica* **53**: 229–233.

Mutinelli,F. (2011) The spread of pathogens through trade in honey bees and their products (including queen bees and semen): overview and recent developments. *Revue Scientifique et Technique-OIE* **30**: 257–271.

Muto,A. and Osawa,S. (1987) The guanine and cytosine content of genomic DNA and bacterial evolution. *Proceedings of the National Academy of Sciences of the United States of America* **84**: 166–9.

Nandi,S., Maurer,J.J., Hofacre,C., and Summers,A.O. (2004) Gram-positive bacteria are a major reservoir of Class 1 antibiotic resistance integrons in poultry litter. *Proceedings of the National Academy of Sciences* **101**: 7118–7122.

NBU (2012a) BeeBase - Inspection Reports. WWW Document: https://secure.fera.defra.gov.uk/beebase/public/BeeDiseases/apiaryReport.cfm

NBU (2012b) BeeBase - Trends of AFB and EFB. WWW Document: https://secure.fera.defra.gov.uk/beebase/public/BeeDiseases/trendDiseaseChart.cfm?id=10

NBU (2013) Live Disease Maps. WWW Document: https://secure.fera.defra.gov.uk/beebase/maps/map.cfm

NBU (2009) Statutory procedures for controlling honey bee pests and diseases. 1–18.

Nelson,K.E., Clayton,R.A., Gill,S.R., Gwinn,M.L., Dodson,R.J., Haft,D.H., et al. (1999) Evidence for lateral gene transfer between Archaea and bacteria from genome sequence of Thermotoga maritima. *Nature* **399**: 323–329.

Ochman,H., Lawrence,J.G., and Groisman,E.A. (2000) Lateral gene transfer and the nature of bacterial innovation. *Nature* **405**: 299-304.

Oertel,E. (1976a) Bicentennial Bees: Early Records of Honey Bees in the Eastern United States. Part I. *American Bee Journal* **116**: 70–71.

Oertel,E. (1976b) Bicentennial Bees: Early Records of Honey Bees in the Eastern United States. Part II. *American Bee Journal* **116**: 128.

Okansen,J., Guillaume Blanchet,F., Kindt,R., Legendre,P., Minchin,P.R., O'Hara,R.B., et al. (2013) vegan: Community Ecology Package.

Okumura,K., Arai,R., Okura,M., Kirikae,T., Takamatsu,D., Osaki,M., and Miyoshi-Akiyama,T. (2011) Complete genome sequence of Melissococcus plutonius ATCC 35311. *Journal of Bacteriology* **193**: 4029–4030.

Okumura,K., Arai,R., Okura,M., Kirikae,T., Takamatsu,D., Osaki,M., and Miyoshi-Akiyama,T. (2012) Complete genome sequence of Melissococcus plutonius DAT561, a strain that shows an unusual growth profile and is representative of an endemic cluster in Japan. *Journal of Bacteriology* **194**: 3014.

Olsen,J.E., Skov,M.N., Threlfall,E.J., and Brown,D.J. (1994) Clonal lines of Salmonella enterica serotype Enteritidis documented by IS200-, ribo-, pulsed-field gel electrophoresis and RFLP typing. *Journal of Medical Microbiology* **40**: 15–22.

Pacific Biosciences (2013a) SMRT TECHNOLOGY. WWW Document: http://www.pacificbiosciences.com/products/smrt-technology/

Pacific Biosciences (2013b) PacBio RSII Brochure. WWW Document: http://files.pacb.com/pdf/PacBio_RS_II_Brochure.pdf

Paini,D.R. (2004) Impact of the introduced honey bee (Apis mellifera) (Hymenoptera: Apidae) on native bees: A review. *Austral Ecology* **29**: 399–407.

Planet,P.J. (2006) Tree disagreement: Measuring and testing incongruence in phylogenies. *Journal of Biomedical Informatics* **39**: 86-102.

Pellet,F. (1924) Native American Bees. *American Bee Journal* **64**: 128–129.

Posthuma-Trumpie,G.A., Korf,J., and van Amerongen,A. (2009) Lateral flow (immuno)assay: its strengths, weaknesses, opportunities and threats. A literature survey. *Analytical and Bioanalytical Chemistry* **393**: 569–582.

Potts,S.G., Biesmeijer,J.C., Kremen,C., Neumann,P., Schweiger,O., and Kunin,W.E. (2010) Global pollinator declines: trends, impacts and drivers. *Trends in Ecology and Evolution* **25**: 345–353.

Potts,S.G., Roberts,S.P.M., Dean,R., Marris,G., Brown,M.A., Jones,R., et al. (2010) Declines of managed honey bees and beekeepers in Europe. *Journal of Apicultural Research* **49**: 15–22.

Purchas,S. (1657) Of American Bees. In, *A Treatise of Political Flying-Insects*.

R Core Team (2013) R: A Language and Environment for Statistical Computing.

Radloff,S., Hepburn,C., and Hepburn,H. (2010) Population structure and classification of Apis cerana. *Apidologie* **41**: 589–601.

Reeson,A.F., Jankovic,T., Kasper,M.L., Rogers,S., and Austin,A.D. (2003) Application of 16S rDNA-DGGE to examine the microbial ecology associated with a social wasp Vespula germanica. *Insect Molecular Biology* **12**: 85–91.

Ribiere,M., Ball,B., and Aubert,M. (2005) Natural History and Geographical Distribution of Honey Bee Viruses. In, *Virology and the Honey Bee*., pp. 15–84.

Rocha,E.P.C. and Danchin,A. (2003) Essentiality, not expressiveness, drives gene-strand bias in bacteria. *Nature Genetics* **34**: 377–378.

Roche (2013) 454 Sequencing, The Technology. WWW Document: 454.com/products/technology.asp

Roetschi,A., Berthoud,H., Kuhn,R., and Imdorf,A. (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* **39**: 362–371.

Rozen,S. and Skaletsky,H. (1999) Primer3 on the WWW for general users and for biologist programmers. In, *Bioinformatics Methods and Protocols*. Humana Press, pp. 365–386.

Ruiz-Garbajosa,P., Bonten,M.J.M., Robinson,D.A., Top,J., Nallapareddy,S.R., Torres,C., et al. (2006) Multilocus Sequence Typing Scheme for Enterococcus faecalis Reveals Hospital-Adapted Genetic Complexes in a Background of High Rates of Recombination. *Journal of Clinical Microbiology* **44**: 2220–2228.

Rutherford,K., Parkhill,J., Crook,J., Horsnell,T., Rice,P., Rajandream,M.-A., and Barrell,B. (2000) Artemis: sequence vizualization and annotation. *Bioinformatics* **16**: 944–945.

de Ruyter,P.G.G.A., Kuipers,O.P., and de Vos,W.M. (1996) Controlled Gene Expression Systems for Lactococcus lactis with the Food-Grade Inducer Nisin. *Applied and Environmental Microbiology* **62**: 3662-3667.

Sakagami,S. (1959) Some interspecific relations between Japanese and European honeybees. *The Journal of Animal Ecology* **28**: 51–68.

Saleh,M., Soliman,H., Sørum,H., Fauske, a K., and El-Matbouli,M. (2012) A novel gold nanoparticles-based assay for rapid detection of Melissococcus plutonius, the causative agent of European foulbrood. *Veterinary Record* **171**: 400.

Sammataro,D., Gerson,U., and Needham,G. (2000) Parasitic mites of honey bees: life history, implications, and impact. *Annual Review of Entomology* **45**: 519–548.

Sanger,F., Nicklen,S., and Coulson,A. (1977) DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences* **74**: 5463–5467.

Schwartz,D.C. and Cantor,C.R. (1984) Separation of yeast chromosome-sized DNAs by pulsed field gradient gel electrophoresis. *Cell* **37**: 67–75.

Shen,M., Yang,X., Cox-Foster,D., and Cui,L. (2005) The role of varroa mites in infections of Kashmir bee virus (KBV) and deformed wing virus (DWV) in honey bees. *Virology* **342**: 141–149.

Shipley,G. (2006) An introduction to real-time PCR. In, *Real-Time PCR*., pp. 1–37.

Smith,N.H., Dale,J., Inwald,J., Palmer,S., Gordon,S. V, Hewinson,R.G., and Smith,J.M. (2003) The population structure of Mycobacterium bovis in Great Britain: clonal expansion. *Proceedings of the National Academy of Sciences* **100**: 15271–15275.

Del Solar,G., Giraldo,R., Ruiz-Echevarria,M.J., Espinosa,M., and Diaz-Orejas,R. (1998) Replication and Control of Circular Bacterial Plasmids. *Microbiology and Molecular Biology Reviews* **62**: 434–464.

Southwick,E. and Southwick,Lj. (1992) Estimating the economic value of honey bees (Hymenoptera: Apidae) as agricultural pollinators in the United States. *Journal of Economic Entomology* **85**: 621–633.

Srivatsan,A., Tehranchi,A., MacAlpine,D.M., and Wang,J.D. (2010) Co-orientation of replication and transcription preserves genome integrity. *PLoS Genetics* **6**: e1000810.

Van der Steen,J. and Blom,M. (2010) Pathogen Transmission in Insect Pollinators. Rapport BIJ-1 onderzoek WP4.

Stipp,R.N., Gonçalves,R.B., Höfling,J.F., Smith,D.J., and Mattos-Graner,R.O. (2008) Transcriptional analysis of gtfB, gtfC, and gbpB and their putative response regulators in several isolates of Streptococcus mutans. *Oral Microbiology Immunology* **23**: 466–473.

Stothard,P. and Wishart,D.S. (2005) Circular genome visualization and exploration using CGView. *Bioinformatics* **21**: 537–539.

Sueoka,N. (1962) On the genetic basis of variation and heterogeneity of DNA base composition. *Proceedings of the National Academy of Sciences* **48**: 582–592.

Suerbaum,S., Lohrengel,M., Sonnevend,A., Ruberg,F., and Kist,M. (2001) Allelic Diversity and Recombination in Campylobacter jejuni. *Journal of Bacteriology* **183**: 2553–2559.

Sullivan,D. and Coleman,D. (1998) Candida dubliniensis: characteristics and identification. *Journal of Clinical Microbiology* **36**: 329–334.

Sumpter,D. and Martin,S. (2004) The dynamics of virus epidemics in Varroa-infested honey bee colonies. *Journal of Animal Ecology* **73**: 51–63.

Sunderland,S., Knox,A., and Highet,F. (2013) The 2012 Scottish Honey Bee Health Surveillance Report.

Takamatsu,D., Arai,R., Miyoshi-Akiyama,T., Okumura,K., Okura,M., Kirikae,T., et al. (2013) Identification of Mutations Involved in the Requirement of Potassium for Growth of Typical Melissococcus plutonius Strains. *Applied and Environmental Microbiology* **79**: 3882–3886.

Takamatsu,D., Okura,M., Yoshiyama,M., Wu,M., Arai,R., and Morinaga,Y. (2012) Detection of atypical Melissococcus plutonius in honeybees. *Veterinary Record* **171**: 630.

Tan,J., Pu,Z., Ryberg,W., and Jiang,L. (2012) Species phylogenetic relatedness, priority effects, and ecosystem functioning. *Ecology* **93**: 1164–1172.

Tarr,H.L.A. (1935) Studies on European foul brood of bees I. A description of strains of Bacillus alvei obtained from different sources, and of another species occurring in larvae affected with this disease. *Annals of Applied Biology* **22**: 709–718.

Tarr,H.L.A. (1937) Studies on European foul brood of bees III. Further experiments on the production of the disease. *Annals of Applied Biology* **24**: 614–627.

Tarr,H.L.A. (1938) Studies on European foul brood of bees IV. On the attempted cultivation of Bacillus pluton, the susceptibility of individual larvae to inoculation with this organism and its localization within its host. *Annals of Applied Biology* **25**: 815–821.

Taylor,G.M., Tucker,K., Butler,R., Pike,A.W.G., Lewis,J., Roffey,S., et al. (2013) Detection and Strain Typing of Ancient Mycobacterium leprae from a Medieval Leprosy Hospital. *PloS one* **8**: e62406.

Life Technologies. (2013) Ion Torrent. WWW Document: http://www.lifetechnologies.com/uk/en/home/brands/ion-torrent.html?cid=fl-iontorrent#

Telford,R., Vandvik,V., and Birks,H. (2006) Dispersal limitations matter for microbial morphospecies. *Science* **312**: 1015.

Teng,F., Kawalec,M., Weinstock,G.M., Hryniewicz,W., and Murray,B.E. (2003) An Enterococcus faecium Secreted Antigen, SagA, Exhibits Broad-Spectrum Binding to Extracellular Matrix Proteins and Appears Essential for E. faecium Growth. *Infection and Immunity* **71**: 5033–5041.

Thomas,C.M. and Nielsen,K.M. (2005) Mechanisms of, and barriers to, horizontal gene transfer between bacteria. *Nature Reviews Microbiology* **3**: 711–721.

Thompson,H.M., Waite,R.J., and Wilkins,S. (2006) Effects of shook swarm and supplementary feeding on oxytetracycline levels in honey extracted from treated colonies. *Apidologie* **37**: 51–57.

Tomkies,V., Flint,J., Johnson,G., Waite,R., Wilkins,S., Danks,C., et al. (2009) Development and validation of a novel field test kit for European foulbrood. *Apidologie* **40**: 63–72.

Truper,H.G. and de' Clari,L. (1998) Taxonomic note: erratum and correction of further specific epithets formed as substantives (nouns) "in apposition." *International Journal of Systematic and Evolutionary Microbiology* **48**: 615.

Tsai,Y.L. and Olson,B.H. (1992) Rapid method for separation of bacterial DNA from humic substances in sediments for polymerase chain reaction. *Applied and Environmental Microbiology* **58**: 2292–2295.

Urwin,R. and Maiden,M. (2003) Multi-locus sequence typing: a tool for global epidemiology. *TRENDS in Microbiology* **11**: 479–487.

Vallenet,D., Labarre,L., Rouy,Z., Barbe,V., Bocs,S., Cruveiller,S., et al. (2006) MaGe: a microbial genome annotation system supported by synteny results. *Nucleic Acids Research* **34**: 53–65.

VanEngelsdorp,D., Evans,J.D., Saegerman,C., Mullin,C., Haubruge,E., Nguyen,B.K., et al. (2009) Colony collapse disorder: a descriptive study. *PloS one* **4**: e6481.

VanEngelsdorp,D., Hayes,J., Underwood,R.M., and Pettis,J. (2008) A survey of honey bee colony losses in the U.S., fall 2007 to spring 2008. *PloS one* **3**: e4071.

VanEngelsdorp,D. and Meixner,M.D. (2010) A historical review of managed honey bee populations in Europe and the United States and the factors that may affect them. *Journal of Invertebrate Pathology* **103**: 80–95.

Vásquez,A., Forsgren,E., Fries,I., Paxton,R.J., Flaberg,E., Szekely,L., and Olofsson,T.C. (2012) Symbionts as Major Modulators of Insect Health: Lactic Acid Bacteria and Honeybees. *PLoS one* **7**: e33188.

Visscher,P., Vetter,R., and Robinson,G. (1995) Alarm pheromone perception in honey bees is decreased by smoke (Hymenoptera: Apidae). *Journal of Insect Behavior* **8**: 11–18.

de Vos,W.M., Kuipers,O.P., van der Meer,J.R., and Siezen,R.J. (1995) Maturation pathway of nisin and other lantibiotics: post-translationally modified antimicrobial peptides exported by Gram-positive bacteria. *Molecular Microbiology* **17**: 427-437.

Waite,R.J., Brown,M.A., Thompson,H.M., and Bew,M. (2003a) Control of American Foulbrood by Eradication of Infected Colonies. *Apiacta* **38**: 134–136.

Waite,R.J., Brown,M.A., Thompson,H.M., and Bew,M.H. (2003b) Controlling European foulbrood with the shook swarm method and oxytetracycline in the UK. *Apidologie* **34**: 569–575.

Waite,R.J., Jackson,S., and Thompson,H.M. (2003) Preliminary investigations into possible resistance to oxytetracycline in Melissococcus plutonius, a pathogen of honeybee larvae. *Letters in Applied Microbiology* **36**: 20–24.

Ward,L., Waite,R., Boonham,N., Fisher,T., Pescod,K., Thompson,H., Chantawannakul,P., and Brown,M. (2007) First detection of Kashmir bee virus in the UK using real-time PCR. *Apidologie* **38**: 181-190.

Watson,J.D. and Crick,F.H.C. (1953) A structure for deoxyribose nucleic acid. *Nature* **171**: 737–738.

White,G.F. (1912) The Cause of European Foul Brood. *Circular of the US Department of Agriculture* **157**: 1–22.

Whitfield,C.W., Behura,S.K., Berlocher,S.H., Clark,A.G., Johnston,J.S., Sheppard,W.S., et al. (2006) Thrice out of Africa: ancient and recent expansions of the honey bee, Apis mellifera. *Science* **314**: 642–645.

Wilkins,S., Brown,M.A., and Cuthbertson,A.G.S. (2007) The incidence of honey bee pests and diseases in England and Wales. *Pest Management Science* **63**: 1062–1068.

Willey,J.M. and van der Donk,W. a (2007) Lantibiotics: peptides of diverse structure and function. *Annual Review of Microbiology* **61**: 477–501.

Williams,K.P. (2002) Integration sites for genetic elements in prokaryotic tRNA and tmRNA genes: sublocation preference of integrase subfamilies. *Nucleic Acids Research* **30**: 866–875.

Wilson,G.G. (1991) Organization of restriction-modification systems. *Nucleic Acids Research* **19**: 2539-2566.

Winston,M. (1992) The Biology and Management of Africanized Honey Bees. *Annual Review of Entomology* **37**: 173–193.

Winston,M. (1987) The Biology of the Honey Bee. Harvard.

Winzeler,E., Shoemaker,D., Astromoff,A., Liang,H., Anderson,K., Andre,B., et al. (1999) Functional Characterization of the S. cerevisiae Genome by Gene Deletion and Parallel Analysis. *Science* **285**: 901–906.

Woodhall,J.W., Webb,K.M., Giltrap,P.M., Adams,I.P., Peters,J.C., Budge,G.E., and Boonham,N. (2012) A new large scale soil DNA extraction procedure and real-time PCR assay for the detection of Sclerotium cepivorum in soil. *European Journal of Plant Pathology* **134**: 467–473.

Woodroffe,R., Donnelly,C.A., Cox,D.R., Gilks,P., Jenkins,H.E., Johnston,W.T., et al. (2009) Bovine tuberculosis in cattle and badgers in localized culling areas. *Journal of Wildlife Diseases* **45**: 128–143.

Wu,M., Sugimura,Y., Takaya,N., Takamatsu,D., Kobayashi,M., Taylor,D., and Yoshiyama,M. (2013) Characterization of bifidobacteria in the digestive tract of the Japanese honeybee, Apis cerana japonica. *Journal of Invertebrate Pathology* **112**: 88–93.

Yadav,S.P. (2007) The Wholeness in Suffix -omics, -omes, and the Word Om. *Journal of Biomolecular Techniques* **18**: 227.

Young,J.P.W., Crossman,L.C., Johnston,A.W.B., Thomson,N.R., Ghazoui,Z.F., Hull,K.H., et al. (2006) The genome of Rhizobium leguminosarum has recognizable core and accessory components. *Genome Biology* **7**: R34.

Zdobnov,E. and Apweiler,R. (2001) InterProScan–an integration platform for the signature-recognition methods in InterPro. *Bioinformatics* **17**: 847–848.

Van der Zee,R. and Pisa,L. (2011) Monitor Bijensterfte Nederland 2009-2010.

Zufelato,M.S., Bitondi,M.M., Simões,Z.L., and Hartfelder,K. (2000) The juvenile hormone analog pyriproxyfen affects ecdysteroid-dependent cuticle melanization and shifts the pupal ecdysteroid peak in the honey bee (Apis mellifera). *Arthropod Structure & Development* **29**: 111–119.