Epidemiology of *Paenibacillus larvae*, causative agent of American foulbrood

Barbara Joy Morrissey

PhD

University of York

Biology

September 2015

Abstract

Paenibacillus larvae is the causative agent of American foulbrood (AFB), a damaging disease of honey bees (*Apis mellifera*) with a global distribution. AFB infects the honey bee larvae through spore-contaminated larval food and once infected the larvae will die within twelve days. AFB infection leads to the eventual death of the honey bee colony. In many countries the best control method is thought to be the destruction of the colony, to destroy the hardy, infectious spores. Therefore infection nearly always leads to the death of the colony. In the UK cases of the disease have decreased in recent years due to statutory control methods, however sporadic outbreaks occur each year. Many advances have been made in our knowledge of the mechanisms of infection by *P. larvae* in recent years however, a high resolution, standardised method of strain typing is required to track disease spread and understand outbreak sources. This thesis describes the development of the first multilocus sequence typing (MLST) scheme for *P. larvae*. MLST is the gold standard for pathogenic bacteria typing. It is based on the sequencing of sections of 6-10 housekeeping genes. MLST is standardised, as primer and allele sequences can be made available for other researchers. The new MLST scheme was used to describe previously undetectable patterns of distribution at a global level as well as at a national level. Humans and bees are implicated in the movement of the disease over different spatial scales. Using the MLST scheme a diverse group of isolates were selected for whole genome sequencing. For the first time multiple genomes were compared within and between genotypes. The ability of the MLST scheme to describe relationships amongst sequence types (STs) was tested by comparing phylogenies based on core genes and MLST sequences. I describe the discovery of seven plasmids in four STs previously unknown to harbour plasmids.

Table of Contents

List of Tables

List of Figures

Acknowledgements

My PhD was a BBSRC CASE studentship with Bee disease insurance. The work in Chapters three and four was funded jointly by a grant from BBSRC, Defra, NERC, the Scottish Government and the Wellcome Trust, under the Insect Pollinator Initiative (BB/I000801/1); and a Defra/Fera seedcorn project (PI Budge). Firstly I'd like to thank my supervisors Giles Budge and Thorunn Helgason for their help and support during my PhD. I am grateful for their knowledge and enthusiasm and the advice they gave me during my thesis. I would also like to thank Bernard Diaper for support, interest and commitment to my work. Thanks to Ummey Hany for training me in next generation sequencing, which formed an important part of my PhD. I would like to thank everyone at the National Bee Unit, but especially Jan Bew and Victoria Tomkies for their help in the lab and Ed Haynes and Ben Jones for discussing my work. Peter Ashton, Toby Hodges and Richard Randal-Boggis were a great help whenever I had difficulties with bioinformatics and computer programing.

I would like to thank my friends and family for their help and support throughout the difficult process of writing up my thesis. Ruth, Kath, Clair, Cat and Farida – thanks for the chance to escape and for all the gin. I would also like to thank Ian Cleasby for lots of help during my PhD including helping with analysis, reading chapters and providing comments on everything. I am also grateful for the help that Rufus provided, though he never said much he always cheered me up. Finally, my parents were a great help throughout my PhD and supported me throughout in numerous different ways. I hope this PhD makes them proud.

Author's declaration

Chapter 2 is published as Biogeography of *Paenibacillus larvae*, the causative agent of American foulbrood, using a new Multilocus sequence typing scheme (2015) Barbara J. Morrissey, Thorunn Helgason, Lena Poppinga, Anne Fünfhaus, Elke Genersch, Giles E. Budge. *Environmental Microbiology* **17**: 1414–1424.

All work presented herein is the author's own, except where referenced and with the following exceptions:

Chapter 2.

Sequencing of some samples included in this chapter was carried out at the bee research institute, Germany

Chapter 4 and 5.

Preparation of extracted *P. larvae* DNA for sequencing on the Ion Torrent PGM Sequencer was preformed by Celina Whalley, then of the University of York and Nicola Burns.

This thesis has not been submitted for any other work at this, or any other institution.

1 Introduction

Honey bees of the genus *Apis* have been kept by humans for millennia. There is evidence of managed honey bee colonies from as far back as 2400 BC in Egypt (Crane 1999). As well as being kept for their products honey bees are the most economically valuable pollinators of crops worldwide (McGregor, 1976; Allsopp et al., 2008). They are widely managed and can be easily and conveniently moved to provide pollination services where native species will not visit agricultural fields (Klein et al., 2007). However, honey bees are threatened globally owing to a number of emerging diseases and parasites such as the small hive beetle (Neumann and Ellis, 2008), *Nosema ceranae* (Klee et al., 2007), and the mite *Varroa destructor* which acts as a host for viral diseases (Martin et al., 2012), as well as pesticide usage (McGregor, 1976; Blacquière et al., 2012; Di Prisco et al., 2013). In this thesis I discuss the epidemiology and global distribution of one particularly damaging disease of honey bees, American foulbrood, caused by the pathogen *Paenibacillus larvae*.

*Paenibacillus larva*e is a gram-positive spore forming bacterium, belonging to the Phylum Firmicutes. Many Firmicutes, including *P. larvae* can produce desiccation resistant endospores that allow them to survive in a variety of environments. The genus *Paenibacillus* contains over 100 species, many of which are soil dwelling opportunists (Qin et al., 2006; Dsouza et al., 2014). Because of their phenotypic similarities, *Paenibacilli* were classed within the genus *Bacillus* until 1993 the name *Paenibacillus* means almost *Bacilli* (Ash et al., 1993). *P. larvae* is the causative agent of American foulbrood (AFB) of the honey bee (*Apis mellifera*) one of the most destructive and widespread diseases of honey bee. Left untreated, the disease will lead to the death of the colony. AFB is extremely contagious due to the large numbers of highly resistant spores that are produced.

1.1.1 Pathogenesis of AFB

AFB only affects the brood stage (larvae and pupae) of the domesticated honey bee (*Apis mellifera*). The Asian honeybee has been shown to be vulnerable to infection in laboratory conditions but the hygienic behaviour of this species has been shown to prevent infection at the colony level (Chen et al., 2000). Larvae become infected by ingestion of spore contaminated larval food (glandular secretions and processed honey) (Yue et al., 2008). *P.*

larvae spores infect the larvae in the broodcomb, usually before they are 24 hours old and, once infected they will not recover (Lindström et al., 2008a). Only the spores are infective as the vegetative stages are killed by the secretions of the hypopharyngeal glands of the nurse bees (de Graaf et al., 2001). Infection by *P. larvae* takes place in two stages (Yue et al., 2008). The first stage is non-invasive. Spores germinate in the infected larval midgut where the vegetative bacteria proliferate and live on the food ingested by the host larvae. During this stage *P. larvae* metabolises sugars and sugar derivatives (Neuendorf et al., 2004; Genersch, 2010; Djukic et al., 2014). During the second stage the midgut epithelium is penetrated followed by the destruction of the larval remains (Yue et al., 2008; Genersch, 2010). The bacteria sporulate in the decaying larvae and the remains become characteristically brown and mucilaginous (Lindström et al., 2008a). This stage is one of the diagnostic stages of AFB known as the rope because, when drawn out with a matchstick, the viscous larval remains form a ropy thread (Genersch et al., 2006). The decaying brood dries into hard scales, which tightly adhere to the walls of the brood cell. These scales consist of millions of bacterial spores, the infectious stage of the bacteria (Genersch et al., 2006).

1.1.2 Spread of AFB

AFB is usually spread horizontally but can also be spread vertically, when colonies swarm (Fries et al., 2006). Horizontal transmission of AFB can occur by several means, due to humans through the movement of contaminated honey (Lindström, et al., 2008b) or the movement of contaminated hives or equipment. AFB can also be spread horizontally by bees, either through the movement of adult bees between colonies (drifting) or the behaviour of foragers (robbing). Human movement can spread AFB over much larger distances than by bees (Lindström, et al., 2008b; Pentikäinen et al., 2008). The spores can retain their ability to germinate for many years in honey, old combs kept in store or in derelict hives (Hasemann, 1961, FERA, 2013). It is the robust nature of the spores as well as the production of extremely high numbers in infected colonies that makes control of AFB so difficult (Genersch, 2010).

1.1.3 Control of AFB

Many countries require honey bee colonies infected with *P. larvae* to be burnt to remove the infection risk posed by the highly resistant spores (Genersch, 2010). Some countries

including the USA, Canada, and Argentina allow the use of antibiotics such as Oxytetracycline to keep the disease under control. However antibiotics can only affect the vegetative bacteria, not the infective spores and so repeated treatments are required to keep the disease under control. Left untreated the disease will destroy the colony (Reybroeck et al., 2012). The use of tetracyclines in commercial beekeeping has resulted in the emergence of resistant strains of bacteria (Evans, 2003; Reybroeck et al., 2012). Plasmids containing tetracycline resistance genes have been identified in resistant strains of *P. larvae* from North America, Canada and Argentina (Murray et al., 2007; Alippi et al., 2007, 2014) . Because of the fears of antibiotic resistance alternative treatments such as the breeding of hygienic bees to reduce disease (Spivak and Reuter, 2001) and the use of plant extracts and essential oils (Genersch, 2010; González and Marioli, 2010; Santos et al., 2012) have been developed. However the most effective method of eradication in most countries is considered to be destruction of the infected colony (Genersch, 2010).

1.1.4 AFB control in the UK

In England and Wales AFB is under statutory control by the government-funded apiary inspection programme managed by the National Bee Unit (NBU; http://www.nationalbeeunit.com). The NBU does not allow treatment with antibiotics. Instead the colony is destroyed by burning, and beekeeping equipment is sterilised (Wilkins et al., 2007). Inspectors then maintain vigilance in the area for three years after the last case is reported (Mill et al., 2014). Since the introduction of the inspectorate in 1942 incidence of AFB has decreased from several thousand a year to between 100-200 cases a year (Wilkins et al., 2007). Inspectors look for clinical signs of AFB such as sunken brood cell caps, and pepper pot pattern. Disease is confirmed using the rapid diagnosis field detection kits introduced in 2006 (de Graaf et al., 2013). Prior to this all suspected cases were sent to the NBU lab for diagnosis, and this method is still used for cases that are difficult to identify (Wilkins et al., 2007).

1.1.5 Classification of Paenibacillus larvae

Foulbrood disease of honey bees has been known for centuries but it was not until 1906 that Gilbert White isolated *Bacillus larvae* from infected bee brood that the disease was identified as two separate diseases, European and American foulbrood. The diseases are named after the continents in which they were first described although both are present

globally. It was not until much later that the etiological agent of EFB (*Melissococcus plutonius*) was identified. Phenotypic differences between strains of *P. larvae* originally lead to the bacteria being described as two separate species, *Bacillus larvae* and *Bacillus pulvifacien*s, thought to cause AFB and powdery scale disease respectively. In 1996 a taxonomic study found the high levels of similarity between the species did not support separate classification (Heyndrickx et al., 1996). However, differences in the phenotype and pathology led to the reclassification of the two *Bacillus* species as two subspecies of *Paenibacillus larvae*, *P. larvae larvae* and *P. larvae pulvifaciens* (Heyndrickx et al., 1996). In 2006, Genersch et al. again reclassified the bacteria as one species, *Paenibacillus larvae,* with no subspecies.

1.1.6 ERIC typing of P. larvae

Four genotypes of *P. larvae* have been identified using repetitive element polymerase chain reaction (rep-PCR) fingerprinting with repetitive intergenic consensus (ERIC) primers (Sharples and Lloyd, 1990; Hulton et al., 1991; Genersch and Otten, 2003). The strains previously known as *P. l. larvae* were grouped within ERIC types I and II and, the strains known as *P. l. pulvifaciens* were grouped within ERIC types III and IV (Genersch et al., 2006). The four ERIC types differ in phenotype including colony and spore morphology (Genersch et al., 2006), metabolic capacity (Neuendorf et al., 2004), sporulation (Saville, 2011) and virulence (Genersch et al., 2005, 2006; Rauch et al., 2009; Poppinga et al., 2012). AFB outbreaks caused by ERIC I and ERIC II isolates are found throughout the world, whereas ERIC III is considerably less common. ERIC IV has not been found in field isolates in recent years and only a few strains exist in culture collections (Genersch, 2010; Poppinga et al., 2012). Therefore, the two most practically important genotypes in terms of global apicultural losses are ERIC I and II.

1.1.7 ERIC types I and II

ERIC type II is considered to be more virulent than ERIC type I, as larvae infected with ERIC type II typically die within seven days compared to twelve days in ERIC type I infections (Genersch et al., 2005, 2006). ERIC II type infections are more likely to go undetected as there are fewer clinical symptoms due to bee behaviour (Ashiralieva and Genersch, 2006; Schäfer et al., 2014). Virulence at the level of the individual larvae is negatively correlated with virulence at the colony level (Rauch et al., 2009). Because ERIC

I infections cause the larvae to die more slowly, a greater proportion of ERIC I infected larvae die after cell capping which occurs between 4-11 days after the larvae emerges from the egg (Winston, 1987). They are therefore less likely to be detected and removed by hygienic behaviour of the nurse bees. ERIC II infected larvae are more likely to be detected and removed which reduces spore production by the bacteria, resulting in the disease spreading more slowly throughout the colony (Rauch et al., 2009). Recently a new method to quickly identify *P. larvae* and discriminate amongst ERIC types has been developed based on matrix- assisted laser desorption/ionization time of flight (MALDI-TOF) mass spectrometry (MS) (Schäfer et al., 2014). This new method is rapid, reliable and cheap in comparison to the previous methods of diagnosis, which involved cultivation and subsequent rep-PCR fingerprinting to differentiate into the ERIC genotypes (Genersch et al., 2006). ERIC typing remains the standard method for differentiating amongst *P.larvae* strain types

1.1.8 Strain typing methods for P. larvae *and biogeography*

Knowledge of the distribution of strain types of the causative agent provides information on how the disease is spread as well as how control methods are working. Although the ERIC typing system displays real biological differences and has recently become faster, cheaper and more reliable with the new MALDI-TOF method it does not give enough resolution to be used as a tool for epidemiological studies. Several methods have been used to study the distribution and movement of *P. larvae* strain types in infected colonies but no standard, repeatable method that gives sufficient resolution has previously been developed. Genersch and Often (2003) used three primers (rep-PCR with BOX A1R and MOB REP1) to sub-type *P. larvae* from disease outbreaks across Germany. They identified four sub-types and found that these sub-types were clustered geographically within disease outbreaks (Genersch and Otten, 2003), 2003). Di Pinto et al. (2011) used ERIC primers with non-standard PCR conditions to identify four *P. larvae* types causing disease within Southern Italy. Because the PCR conditions were different from those used by Genersch and Otten (2003) the results are not comparable to the standard ERIC types. This study also did not give any geographical information (Di Pinto et al., 2011). A further study in Italy (Bassi et al., 2015) using the standard ERIC PCR conditions identified that both ERIC I and II are present in Italy but suggested that further study with a scheme that has higher discriminatory power would be necessary for further epidemiological studies (Bassi

et al., 2015). In Finland *Not*I pulsed-field gel electrophoresis (PFGE) was used to detect 52 macrorestriction profiles (MRP) and three biotypes. This study demonstrated movement of the disease across large distances by the beekeeper (Pentikäinen et al., 2008). The established BOX, ERIC and REP primers do not give enough resolution to be used in epidemiological studies and although NotI PFGE gave higher discrimination, the results would be difficult to replicate in different labs. A new typing scheme that gives higher resolution and can be standardised between labs is required for high-resolution study of the disease and its movement. In the UK and Jersey, AFB movement and spread have been studied without strain typing. A Jersey study demonstrated that the disease was spread around the small island by both bees and by keepers over different distances (Datta et al., 2013). AFB has been shown to form disease clusters that reappear year after year in England and Wales. These clusters are eventually wiped out by the inspection regime. However, in some areas clusters can reappear. It has been suggested that this may be due to new sources of infection (Mill et al., 2014) but it is not, with current typing methods, possible to distinguish between these two hypotheses. It was suggested that the importation of honey bees and their products may result in a sporadic supply of AFB spores, which may account for these persistent areas of disease (Mill et al., 2014). A new high resolution, standardised scheme will enable us to type the disease occurring in the UK and understand more about its movement. Such a scheme has recently been published for EFB based on the DNA sequence of housekeeping genes (Haynes et al., 2013).

1.1.9 MultiLocus Sequence Typing (MLST)

In order for a typing scheme to be informative it must have high discrimination power, combined with good inter- and intra-laboratory reproducibility. It should also be easy to set up, use and interpret (Genersch and Otten, 2003). Multilocus Sequence Typing (MLST) meets all of these requirements as it uses unambiguous DNA sequence in order to differentiate between strains, and all that is required to reproduce an MLST scheme is the primer sequences, which can be made available online (Aanensen and Spratt, 2005). The benefit of MLST is that the use of several genes found throughout the genome gives a better overall picture of the rate of evolution between strains (Urwin and Maiden, 2003). MLST schemes have been successfully used to understand the epidemiology of human pathogens in previous studies (Maiden, 2006) including the *Bacillus cerus* group to which *P. larvae* is related (Tourasse et al., 2006). Generally, MLST schemes consist of short

regions of six or seven housekeeping genes (Maiden, 2006). This is because housekeeping genes are highly conserved, and show lower rates of sequence evolution, and at an even rate across strains. As they are involved in essential metabolic tasks any random mutation that has an effect on their function will be selected against (Maiden et al., 1998; Maiden, 2006). Although these genes generally have low variation in comparison to other genes, used in combination with one another they retain the signatures of longer-term evolutionary relationships or clonal stability. Using several loci also prevents skewed results, which may occur with fewer loci (Margos et al., 2008). MLST is a method for globally standardising strain types of bacteria results can be made available on a publically accessed database e.g. PubMLST (Jolley and Maiden, 2010).

1.1.10 **M. plutonius** ***MLST scheme***

An MLST scheme for *M. plutonius*, the causative agent of European foulbrood has recently been developed (Haynes et al., 2013). EFB is a less serious disease of bee brood than AFB. It is treatable using shook swarm or Oxytetracycline and is sometimes referred to as a stress disease (Wilkins et al., 2007). However, like AFB, it is globally widespread and is also under statutory control in the UK where it occurs much more frequently (Wilkins et al., 2007). Using this new scheme researchers have demonstrated cases where movement of the disease over long distances was facilitated by the bee keeper as well as showing local movement (Haynes et al., 2013; Budge et al., 2014). The MLST scheme allowed the authors to identify bacterial strains that were most pathogenic, meaning that treatment can be tailored to tackle strains most effectively (Budge et al., 2014). Using the *M. plutonius* MLST scheme, EFB has also been shown for the first time to have moved between honey bee species (Takamatsu et al., 2014). This thesis describes the creation of a new MLST scheme for *P. larvae* and its uses in furthering our understanding of the epidemiology of AFB.

1.1.11 Bacterial genome sequencing

The first bacterial genome was sequenced in 1995 using Sanger technology (Fleischmann et al., 1995), which took years of work and a six figure budget. Since then next-generation sequencing technologies have radically decreased the amount of time and the cost per base of sequencing. Sequencing a whole bacterial genome now takes hours rather than years (Loman et al., 2012). This development in technology means that bacterial genomes can

now be relatively easily compared using sequencing and bioinformatics (Edwards and Holt, 2013) Comparative genomics helps us to understand the genetic differences among bacterial strains. Whole-bacterial genomes can add an extra dimension to epidemiological studies such as those on Methicillin-resistant *Staphylococcus aureus* (MRSA) (Köser et al., 2012; Harris et al., 2013) and *Escherichia coli* (Mellmann et al., 2011). These studies use genome wide Single Nucleotide Polymorphisms (SNPs) and MLST schemes to determine the movement of bacterial strains causing disease in a short timescale. Whole genome sequence data gives added resolution to the standardised MLST strain type. For species with available MLST schemes whole genome sequence data can be easily and quickly typed using the Bacterial Isolate Genome sequence database (BIGSdb) on the pubmlst database (pubmlst.org; Jolley and Maiden, 2010).

1.1.12 **P. larvae** ***genomics***

Two genomes of *Paenibacillus larvae* have recently been published on Genbank (Accession numbers: DFW00000000 and CP003355-CP003356) in 2014 (Djukic et al., 2014). These genomes were of two different ERIC strains, ERIC I (DSM 25719), consisting of seven contigs, and the complete genome of an ERIC II strain (DSM 25430). The ERIC I genome is around 500,000 bases longer than the ERIC II (4,579,589 compared to 4,056,006bp) and contains nearly a thousand more protein coding genes (4,868 and 3,928) (Djukic et al., 2014). Previously, Qin et al. (2006) uploaded *P. larvae* contigs from a shotgun sequencing project with low coverage (5-6X) (Qin et al., 2006). In 2011 Chan et al. published an updated genome, which decreased the number of contigs from 646 to 388 and increased the coverage (Qin et al., 2006; Chan et al., 2011). However, this genome was incomplete and the authors hypothesised that the fragmentation may be due to areas of long sequence repeats that could not be bridged using shotgun sequencing (Chan et al., 2011). The new genomes confirm the existence of these genomic regions containing repeats (Djukic et al., 2014).

1.1.13 **P. larvae** ***comparative genomics***

P. larvae strains ERIC I and II have different levels of virulence. Although both sequenced strains cause AFB, Djukic et al. (2014) demonstrate that there are many differences between the strains at the genomic level, including length, number of transposases, insertion elements, predicted phage regions, and strain-specific island-like regions (Djukic

et al., 2014). DSM 25430 contains a higher copy number of mutator-type transposases than DSM 275719, however the large number of mobile elements in both strains suggest frequent genome rearrangements and high genome plasticity (Djukic et al., 2014). It is thought that the difference in virulence between ERIC I and II are due to different methods of host infection (Poppinga et al., 2012; Fünfhaus et al., 2013; Djukic et al., 2014). Although ERIC II has been shown to be more virulent at the larval level (Genersch et al., 2005, 2006) DSM 25430 contains five non-functioning toxin genes that were functioning in DSM 25719 (Djukic et al., 2014). The toxins are similar to the family of AB toxin found in other gram-positive, spore forming bacteria such as *Clostridia* and *Bacilli* (Barth et al., 2004; Fünfhaus et al., 2013; Djukic et al., 2014). The toxin loci were present in both of the sequenced strains but in the ERIC II strain DSM 25430, they were non-functional due to point mutations or transposases (Djukic et al., 2014). In contrast, Poppinga et al., (2012) identified a functional s-layer protein in ERIC II isolates that is not functional in ERIC I isolates. The *splA* gene, which codes for a surface layer protein is involved with adhesion to the host's gut wall as well as having protective properties. While the gene is present in both ERIC I and ERIC II genomes, in the ERIC I strains tested, it is non-functional due to a single point mutation causing a frame-shift (Poppinga et al., 2012). Because ERIC I isolates do not express this gene it is likely that they has evolved a different method of infection. Previous genome comparison had only one example of each ERIC strain. In this thesis I compare a number of different strains within and between ERIC I, II and IV.

1.1.14 Plasmids in P. larvae

Djukic et al. (2014) identified plasmids in both genome-sequenced strains. Plasmids are extra chromosomal elements that are able to replicate independently. Plasmids can provide benefits to the host bacterium such as resistance against antibiotics, virulence and additional metabolic capacities. The plasmids; pPLA1_10 found in ERIC I strain DSM25719 and pPLa2_10 found in ERIC II strain DSM 25430 have not been well characterised. Both plasmids are 9.7k long and both contained a putative replication initiation factor (REP) only differing at 49 bases (Djukic et al., 2014). Plasmids of similar length have previously been identified but not characterised (Bodorová-Urgošíková et al., 1992; Neuendorf et al., 2004). Djukic et al. (2014) screened 65 strains of ERIC I and did not identify any further example of the plasmid pPLA1_10, which they suggested was due to strain, but not genotype specificity (Djukic et al., 2014). Tetracycline resistant plasmids

have been discovered in *P. larvae* in areas where AFB is treated with antibiotics (Reybroeck et al., 2012). The plasmid pMA67 identified by Murray et al. (2007) is a short circular plasmid (5kb) that replicates using the rolling circle method. Tetracycline resistance is conferred by the *tetL* gene (Murray and Aronstein, 2006). Alippi et al. (2014) also identified three plasmids in North American strains of *P.larvae*, pPL373, pPL374 and pPL395. These plasmids are highly similar to pMA67 (99%) as well as to plasmids conferring tetracycline resistance in five genera of Gram-positive bacteria found in a variety of ecological niches. This suggests horizontal transfer of tetracycline resistance amongst environmental bacteria in water, soil and food as well as in pathogenic bacteria (Alippi et al., 2014). Alippi et al. (2014) also identified but did not sequence a larger plasmid (8kb) present in two of the three strains. This longer plasmid seemed to facilitate conjugation for the smaller mobilizable plasmids (Alippi et al., 2014).

Much has been learnt about the methods of pathogenesis of *P .larvae* in recent years, and the publishing of *P.larvae* genomes will increased our knowledge of mechanisms in involved in infection by *P. larvae*. However, no standardised typing tool has been produced with sufficiently high resolution, to ascertain broader ecological questions such as the distribution of strain types. In this thesis I describe the development of such a tool and use it for biogeographical studies on both a global and more local scales. I use our new knowledge of sequence types to inform the selection of various diverse STs for whole-genome sequencing and describe genome comparisons, including the discovery of several plasmids in our sequenced isolates.

1.2 This thesis

Chapter two details the creation of a new, high resolution MultiLocus Sequence Typing (MLST) scheme to determine strain types of *Paenibacillus larvae*. This new scheme enables the identification of strain types of *P. larvae* causing the disease around the world. The scheme was used to study the biogeography of *P. larvae* and describe patterns in its global distribution.

Chapter three describes the use of this new MLST scheme to look at the distribution of *P. larvae* strain types causing AFB around the UK. It had previously been determined that AFB outbreaks form clusters of disease, here the STs present in these clusters are compared to outbreaks not in clusters. STs in clusters near to possible risk points are typed and a possible source of *P.larvae* spores, imported honey is examined.

Chapter four. Previous *P. larvae* genome comparisons could only be between ERIC types. However now that 24 MLST sequence types have been described further genome comparisons can be carried out within and between these types. Whole genome sequencing gives further resolution than MLST and describes much more genetic diversity. Here 21 genomes from 14 STs have been sequenced and the results are described. The Phylogenies produced using MLST and WGS are compared to determine how well MLST describes the relationship amongst strains.

Chapter five. For the first time plasmid DNA has been found in P. larvae samples from the UK. Of the 21 Global isolates sequenced using whole genome sequence technology, seven contained plasmids. These plasmids are similar to the pPLA1_10 and pPLA2_10 plasmids described in Djukic et al. (2014). None of the seven plasmids found contain the *tet* genes responsible for tetracycline resistance. Here we describe the seven plasmids and discuss their phylogenetic history.

2 Biogeography of *Paenibacillus larvae*, the causative agent of American foulbrood, using a new MLST scheme

2.1 Abstract

American foulbrood is the most destructive brood disease of honey bees (*Apis mellifera*) globally. The absence of a repeatable, universal typing scheme for the causative bacterium *Paenibacillus larvae* has restricted our understanding of disease epidemiology. We have created the first MultiLocus Sequence-Typing scheme (MLST) for *P. larvae*, which largely confirms the previous ERIC-PCR based typing scheme's divisions whilst providing added resolution and improved repeatability. We have used the new scheme to determine the distribution and biogeography of 294 samples of *P. larvae* from across 6 continents. We found that of the two most epidemiologically important ERIC types, ERIC I was more diverse than ERIC II. Analysis of the fixation index (F_{ST}) by distance suggested a significant relationship between genetic and geographic distance suggesting that population structure exists in populations of *P. larvae*. Interestingly, this effect was only observed within the native range of the host and was absent in areas where international trade has moved honey bees and their disease. Correspondence analysis demonstrated similar sequence type (ST) distributions between native and non-native countries and that ERIC I and II sequence types mainly have differing distributions. The new typing scheme facilitates epidemiological study of this costly disease of a key pollinator.

2.2 Introduction

Paenibacillus larvae, a gram-positive spore-forming bacterium, causes American foulbrood (AFB), which is the most destructive brood disease of the honey bee (*Apis mellifera*). AFB is contagious due to the large numbers of highly resistant spores that are produced and efficiently transmitted by contaminated adult bees within and between colonies (Lindström, et al., 2008a; 2008b). Only the spores are infective and are fed to bee brood by nurse bees in contaminated larval food (glandular secretions and processed honey) (Yue et al., 2008). Once infected larvae usually die within 3 to 12 days (Genersch et al., 2005; Rauch et al., 2009). *P. larvae* spores are able to remain infective for more than 35 years in old hives and are resistant to extremes of temperature (Hasemann, 1961). This makes control of the disease difficult because human activity can spread the disease over

long distances and previously dormant strains may cause an outbreak several years after the original outbreak.

2.2.1 AFB control

Antibiotics only affect the vegetative stage of the bacterium, masking the symptoms of AFB, they have no effect on the infective spores (Genersch and Otten, 2003). In many countries burning infected colonies and hive materials is thought to be the most effective way of preventing the spread of AFB. Therefore, whether AFB is ignored or treated the colony will be killed which leads to considerable economic loss to global apiculture.

2.2.2 P. larvae genotyping

American foulbrood is found on every continent where honey bees are kept (Matheson, 1993). The disease is spread by both humans and bees and it is spread predominantly via horizontal routes although it has been shown to spread vertically (Fries et al., 2006; Lindström, et al., 2008b). Horizontal transmission of AFB can occur by several means, due to humans through the movement of contaminated honey or the movement of contaminated hives or equipment (Genersch, 2010). AFB can also be spread horizontally by bees, either through the movement of adult bees between colonies (drifting) or the behaviour of foragers (robbing) (Lindström *et al.*, 2008b). Using genetic markers to identify outbreaks caused by closely related strains can give important evidence to confirm the source and routes of disease transmission. The shortcomings of phenotypically based typing methods for *P. larvae* (Genersch et al., 2006) have led to the development of molecular typing methods based on the microbial DNA sequence (Alippi and Aguilar, 1998; Genersch and Otten, 2003; Genersch et al., 2006; Antúnez et al., 2007; Pentikäinen et al., 2008). Four genotypes of *P. larvae* have been identified based on repetitive-element PCR (rep-PCR) using enterobacterial repetitive intergenic consensus (ERIC) primers (Genersch et al., 2006). These four genotypes were shown to form two clusters using pulsed-field gel electrophoresis (PFGE) (Genersch et al., 2006). ERIC genotypes differ in phenotype including morphology (Genersch et al., 2006), metabolic capacity (Neuendorf *et al.*, 2004) virulence (Genersch et al., 2005; Rauch et al., 2009) and virulence factors (Poppinga et al., 2012; Fünfhaus et al., 2013). The above typing schemes have been used to categorize crude patterns of *P. larvae* distribution in Europe (Genersch and Otten, 2003; Pentikäinen et al., 2008; Loncaric et al., 2009; Di Pinto et al., 2011), the Americas (Alippi

et al., 2004; Antúnez et al., 2007) and Australasia (Alippi et al., 2004). However, rep-PCR methods have the disadvantage that they are difficult to repeat or to standardize between laboratories and therefore comparisons between different studies is difficult (Rusenova et al., 2013). Most recently it was shown that the four ERIC-genotypes can also be discriminated via matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry (MALDI-ToF) (Schäfer et al., 2014) providing a cost effective, reliable and highly reproducible alternative tool for *P. larvae* ERIC-typing. The advantage of the established ERIC-scheme for *P. larvae* (Genersch et al., 2006) is that it allows grouping into biologically relevant genotypes differing in practically important phenotypic features. However, it does not give enough resolution to be used effectively as an epidemiological tool to study disease spread. Therefore, a state-of-the-art method providing sufficient resolution to be informative to epidemiological studies is urgently needed in order to enhance differentiation beyond the four ERIC-genotypes

2.2.3 MultiLocus Sequence Typing

The utilization of sequence-based typing would allow a single, universal and unambiguous scheme that would help us to better understand the global spread of this damaging disease. Multilocus sequence typing (MLST) can provide a standardized set of strain types that can be used to study bacterial population structure and evolution at both a global and a more local scale. MLST is based on unambiguous DNA sequences and allelic profiles can be shared between laboratories using online databases such as PubMLST (Jolley and Maiden, 2010). MLST schemes have become the gold standard for epidemiological studies, providing insight into the epidemiology of human pathogens such as the *Bacillus cereus* group (Helgason et al., 2004) to which *P. larvae* is related. Generally, MLST schemes consist of short regions of six or seven housekeeping genes, which evolve at a slow even pace across all strains (Maiden, 2006).

Here we report the development of a novel seven gene MLST scheme to enhance differentiation within the species *P. larvae* and we use this scheme to identify global patterns in the population structure of *P. larvae*.

2.3 Methods

2.3.1 MLST development.

Whole genome comparisons were made between two genetically diverse strains of *P. larvae* (Genersch et al., 2006) p6678 and p6993; LMG 16241 and LMG 16247, ERIC I and ERIC IV respectively). Housekeeping genes used in a previous typing scheme for the related *Bacillus cereus* group (Helgason *et al.*, 2004) were found to have no variation when compared between these two isolates, so novel regions with 80-90% similarity were identified. Pairwise comparisons were made using the online program doubleACT and the result visualised using ACT (Carver et al., 2005) and MEGA version 5 (Tamura et al., 2011). Two further genomes were later compared (DSM25719 (ERIC I), NCBI acc. No. ADFW01000002; DSM25430 (ERIC II), NCBI acc. No. NC_023134). This led to the discovery of additional suitable loci that were identified using the *Bacillus cereus* Group Typing Database (University of Oslo, 2012) and by targeting genes likely involved in observed phenotypic differences in sporulation frequency between ERIC types (Saville, 2011).

2.3.2 Primer design.

Primers were designed to candidate loci using primer 3 (v0.4.0) (Rozen and Skaletsky, 2000), and Oligocalc (Kibbe, 2007) with an optimum melting temperature of 55°C (±5°C). In total 31 primer pairs were tested including those for non-coding loci to give a scheme composed of seven coding loci. The primer sets were used to amplify a panel of *P. larvae* isolates of all four ERIC types. Any primer set that did not add extra resolution to the scheme was rejected (Appendix 1).

The PCR reaction were carried out using 2µl template DNA, 12.5µl 2x Fermentas PCR mastermix, and 10 pmol of each primer with a total reaction volume of 25µl. Each reaction was run on an Applied Biosystems 2720 Thermal Cycler. Conditions were as follows: After the initial activation step (3 min, 95 °C), 35 cycles at 95 °C for 30 seconds, 52 °C for 30 seconds, and 72 °C for 1 min were run followed by a final elongation step at 72 °C for 10 min.

2.3.3 Global isolates.

In total, 294 *P. larvae* isolates from existing national and international culture collections were recovered on either blood agar or brain heart infusion agar and PLA (de Graaf et al., 2013). Briefly, spore containing honey was either heated at 90°C for 6 min prior to plating (Genersch and Otten, 2003) or plated directly without heat treatment (Forsgren et al., 2008) because temperatures above 90°C or an incubation time of 10 min could have negatively affected germination of ERIC II strains (Forsgren et al., 2008). Isolates were collected from 38 countries from across six continents: Africa (N= 5), Australasia (N=26), Asia (N=27), Europe (n=199), North America (N= 16) and South America (N=4) as well as some of mixed origin and some from culture collections (N=17) (Appendix 2). DNA was extracted from *P. larvae* cultures using a simple Chelex method. Bacteria were transferred to 300 µl 6% Chelex ®100 and heated to 56°C for 20 minutes followed by boiling for 8 minutes. DNA extracts were stored at -20°C until required. ERIC typing was completed using the method described in Genersch et al. (2006).

2.3.4 Sequencing.

PCR products were purified using Qiagen® PCR purification and sequenced on the ABI 3730xl 96-capillary DNA Analysers. Sequences were aligned using ClustalW in MEGA version 5 (Tamura et al., 2011) and allele types were counted and numbered in order of discovery as described by Aanensen and Spratt (2005). An allele was identified as a sequence or sequences with one or more genuine nucleotide difference from previously assigned sequences. The combination of allele numbers for each of the target genes gives the allelic profile or sequence type (ST) of an isolate. All putative loci were amplified from a panel of *P. larvae* isolates representing all four ERIC types. Loci that failed to add resolution to the scheme were rejected and all isolates were typed using the final MLST scheme.

2.3.5 MLST Analysis.

The ratio of non-synonymous and synonymous substitutions (d_N/d_s) of MLST gene fragments was determined using the modified Nei–Gojobori method (Nei and Gojobori, 1986) in the program START2 (Jolley et al., 2001). Recombination was tested using the index of association with the program START2 (Jolley et al., 2001).

2.3.6 Population structure and biogeography.

PHYLOViZ (Francisco et al., 2012) was used to analyse allelic profiles using the goeBURST algorithm (Francisco et al., 2009; Feil et al., 2004). The program was used to discover clonal complexes and infer founder clones (Francisco et al., 2009). The most parsimonious patterns of descent of all isolates in each clonal complex from the predicted founder(s) were calculated as previously described (Francisco et al., 2009, 2012). A phylogeny was also constructed from an alignment of 2948 sites representing the concatenated sequences of each ST. The sequences were aligned using ClustalW, as implemented in MEGA 5.2 (Tamura et al., 2011). The phylogenetic analysis was then carried out using the Neighbour-Joining method and the Maximum Composite Likelihood model as implemented in MEGA 5.2. (1000 replications).

To test the population structure of *P. larvae* among different countries, pairwise FST was calculated using the haploDiv command in the R package diveRsity (Keenan et al., 2013) and bootstrapped 95% confidence intervals (500 repeats) were calculated. The 40 SNPs identified in the concatenated MLST gene sequences were used to derive pairwise F_{ST} between populations. Populations were taken as all samples from a single country (Appendix 2). Countries where there were samples from fewer than 5 isolates were discounted (Appendix 2) or grouped: Samples from Bangladesh, Japan, China, Singapore, and Mongolia became Asia**.** Geographic Distances were taken as the great circle distance between the centre point location for each country or group of countries. A Mantel test (Mantel, 1967) with 1000 replications was used to determine whether the correlation between physical distance and F_{ST} was significantly different from a random sample of the data. All results were visualised using R (version 2.15.2) (R Core Team, 2012). Correspondence analysis (CA) was also applied to the data. The CA takes into account the STs present in each country, to investigate associations between STs (which types were commonly found together) and patterns in their distribution (which countries are associated with the STs). Finally, rarefaction curves were constructed to compare the sampling efforts between different continents and between countries within Europe. R library vegan (Oksanen et al., 2013) was used to carry out these final two analyses.

2.4 Results

2.4.1 Loci discovery and primer design.

In total, 31 primer pairs, including for non-coding loci, were tested against *P. larvae* isolates representing all four ERIC types. The majority of loci were rejected due to low diversity between test isolates (Appendix 1). Of the remaining loci, seven offered the largest diversity within the 294 isolates of *P. larvae* tested: *clpC* (catabolite control protein A), *ftsA* (cell division protein), *glpF* (glycerol uptake facilitator protein), *glpT* (glycerol-3-phosphate permease), *Natrans* (forward sodium dependant transporter), *sigF* (sporulation sigma factor F) and *rpoB* (RNA polymerase beta subunit) (Table 2.1.). Fragment length and G + C content for the seven selected loci ranged from 271 bp (*glpF*) to 541 bp (*clpC*) and 43.8% to 48.7% respectively (Table 2.2). The percentage of variable sites ranged from 0.65 (*ftsA*) to 2.0 (*Natrans*) resulting in 4-6 alleles per locus (Table 2.2). The d_N/d_S values were all lower than 1 for all genes except, *glpF* which contained a deletion. The ratio of non-synonymous and synonymous

Table 2.1. MLST scheme primer sequences

Gene	Forward primer	Reverse primer	Annealing temp (°C)
clpC	5'TTTGGAAGATTTACTGAACGA3'	5'ATCAGAACCGGGTTATTTTT3'	52
ftsA	5'AAATCGGTGAGGAAGACATT3'	5'TGCCAATACGGTTTACTTTA3'	52
glpF	5'GTCAGCGGGGCTCATTTA3'	5'TGCTTACGATGAGAAATCCT3'	52
glpT	5'GGATTGAAAAACTTGAAACG3'	5'CATGCTGAGAGAAATCTTCC3'	52
Natrans	5'GCTTCGGTAATGGTAACCTA3'	5'TTGAACCCATTGTAAATTCC3'	52
rpoB	5'ATAACGCGAGACATTCCTAA3'	5'GAACGGCATATCTTCTTCAG3'	52
sigF	5'GTCACTGAAGGAATTGGCTA3'	5'TATCTGGTTACGGATGGACT3'	52

Table 2.2 Feature summary of seven loci selected for *P. larvae* typing scheme

Locus	Sequence length (bp)	No. of alleles	Mean G+C content	No. of polymorphic sites*	No. of non-synomymous substitutions	dN/dS ratio
clpC	541	5	48.7	4(0.74)	3	0.4183
ftsA	464	4	46	3(0.65)	2	0.7756
glpF	271	6	45.2	5(1.85)	4	1.4835
glpT	502	4	47.4	9(1.8)	6	0.5569
Natrans	490	6	46.8	10(2.0)	6	0.218
rpoB	339	5	48.7	4(1.5)	2	0.4276
sigF	345	4	43.8	4(1.2)	1	0.1607

* percentage of polymorphic sites in parentheses

substitutions (d_N/d_s) measures the level of selection in a protein coding gene. To ensure consistency in an MLST it is preferable that each locus is under purifying selection. However genes that are under positive selection may give more resolution to the scheme (Maiden, 2006). The ratio of d_N/d_S indicates purifying selection (negative selection) if values are <1, positive selection if values are >1, and neutral evolution if values are close to 1. A value approaching 1 may also indicate a combination of positive and purifying selection.

2.4.2 Index of association

The index of association (I_A) was significantly different from 0 when only one representative of each ST was included in the computation (1.16; P= 0.000), indicating limited recombination events and a clonal population structure in *P. larvae*. The I_A measures the extent of linkage equilibrium within a population by quantifying the amount of recombination among a set of sequences and detecting association between alleles at different loci (Maynard Smith et al., 1993).

2.4.3 Sequence types

Typed isolates included 173 ERIC type I, 92 ERIC type II, 3 ERIC type III and 7 ERIC type IV and 19 isolates where ERIC type was not determined due to a shortage of DNA (Appendix 2). At least one isolate of each ST was ERIC typed. The 7-gene MLST scheme resolved the 294 *P. larvae* isolates into 21 different STs (Figure 2.1; Appendix 2). The allele sequences have been submitted to the EMBL database under Accession numbers HG530076 to HG530109. The entire scheme is available at pubMLST.org/plarvae/ (Jolley and Maiden, 2010). ST designations represented a single ERIC grouping except for a single ERIC III isolate that grouped with ST8 (Figures 2.1 and 2.2). Isolates from ERIC I were separated into 16 STs, whereas ERIC II isolates were only separated into three separate STs. The Chao1 estimates suggest this difference in observed diversity was unlikely to be due to a biased sampling effort. After an initial increase, the mean Chao1 estimate for all geographical regions became relativley level as sample size increased (Figure 2.3), therefore we compared the ST diversity estimates at the highest sample size for each ERIC type (Hughes et al., 2001). Total diversity of ERIC types is significantly different as estimated by Chao1 (Figure 2.3). Chao1 estimates that ERIC I has 18.49 STs (95% CIs 16.37 and 32.91), and ERIC II has 3 STs (CIs 3 and 4.49). The STs were roughly

split following the pattern of the ERIC typing scheme with different ERIC I and II types forming distinct groups with both the eBURST and phylogenetic analysis (Figures 2.1 and 2.2; Appendix 2). ERIC III and IV isolates were less distinct with one isolate (P6266) LMG 16252) ERIC typed as ERIC III but sequence typing as ST8 and grouping with ERIC IV isolates (Figure 2.1, Appendix 2). All isolates within ERIC I and II are linked by single loci variants (SLV) whereas the ERIC III and ERIC IV STs differ from each other at two loci. ERIC I and ERIC II isolates differ from one another by at least four loci (link not shown in figure 2.1) and ERIC III isolates differ from ERIC I isolates by at least six loci (Figure 2.1.). The ERIC III ST 9 is made up of only two isolates found in Chile (Appendix 2).

2.4.4 Population structure and biogeography.

The two most common and widespread types of ERIC I were ST 15 and 5 which were each found causing disease in multiple countries and across 5 continents (Appendix 2). Contrastingly of the two most common STs of ERIC II, ST10 was well distributed and ST 11 was found only in Germany and countries to the East of Germany. In this study no ERIC II isolates were collected from countries to the West of Germany within the native honey bee range.

The relationship between F_{ST} and geographic distance of the global populations of *P.larvae* were not significant (P= 0.996, $r^2 = 5.506 \times 10^{-7}$) suggesting no relationship between genetic and geographic distance when considering the sampled global population of *P. larvae*. However, when the analysis was restricted to isolates collected from within the native range of honey bees (Europe, Africa and Eastern Asia; 226 of 294 isolates (see Appendix 2)) a significant relationship between genetic and geographic distance was detected (P = 0.01, $r^2 = 0.122$; Figure 2.4.).

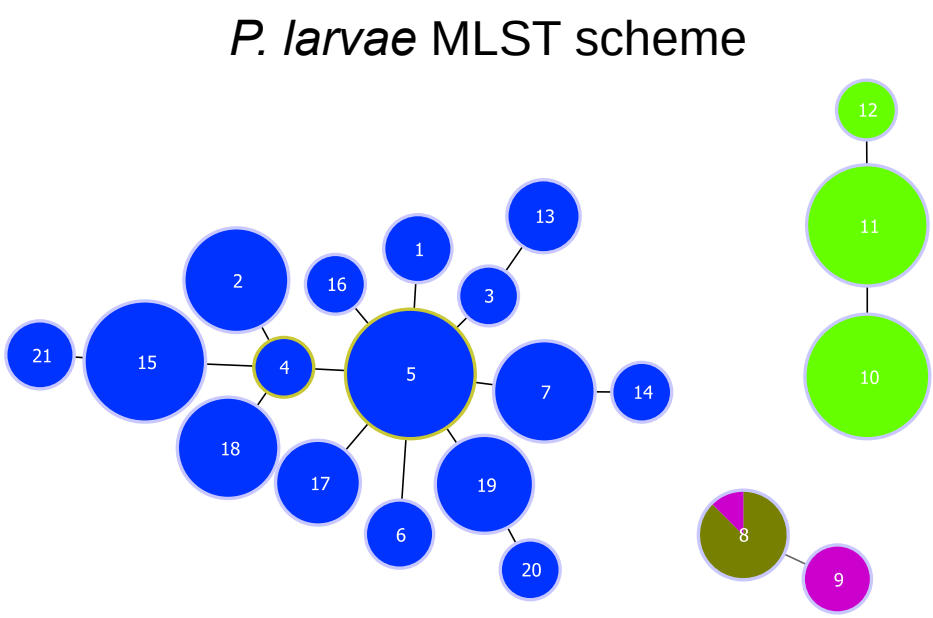

Figure 2.1 eBURST diagram of *P. larvae* MLST scheme. Numbers represent ST of isolates. STs with variations at more than three loci are not connected. Size of circles represents the number of isolates of that type. Blue circles are typed as ERIC I, green as ERIC II, purple as ERIC III, brown ERIC IV

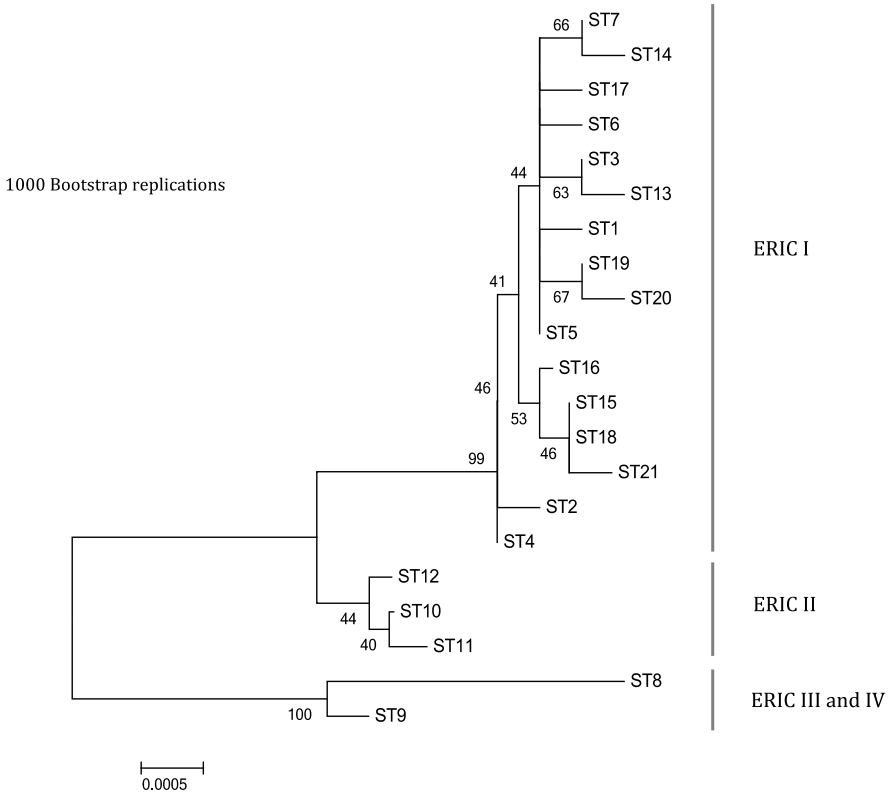

Figure 2.2 Neighbour joining tree inferring the evolutionary history of *P. larvae* STs. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches.

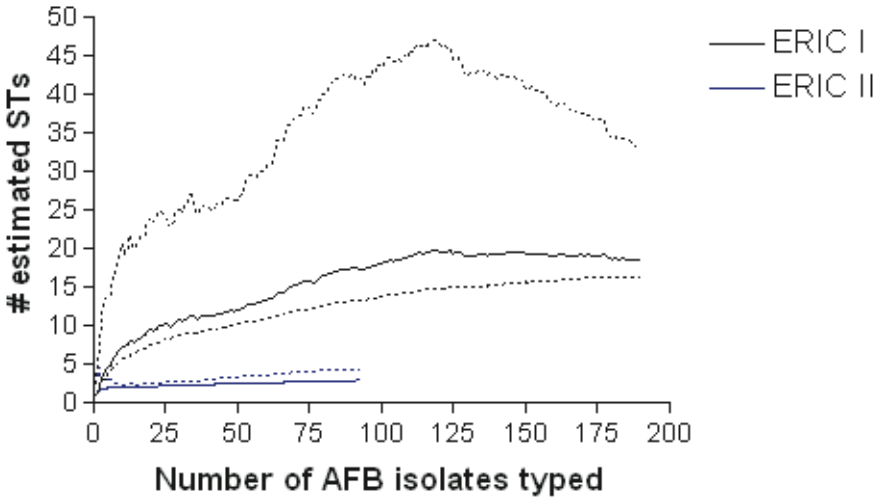

Figure 2.3 Chao1 estimates of ERIC I and ERIC II ST richness as a function of sample size. Dotted lines are 95% CIs and were calculated with the variance formula derived by Chao (1987). The lower solid line represents ERIC II and the upper solid line represents ERIC I

2.4.5 Correspondence Analysis (CA).

The ordination graph (Figure 2.5) describing the results of the CA shows a clear split in the distribution of ERIC I and ERIC II. In addition figure 2.5 shows no split in the distribution of countries, whether they were in the native range of honey bees or not. This suggests that most STs are found in both the native range and countries outside of the native range.

The proportion of variance explained by the first two eigenvectors was 0.4026. In the ordination graph describing the results of the CA (Figure 2.5) STs that have similar distribution are represented by points closer together in space and the proximity of STs to countries indicate that those STs are associated with that country. The first axis highlights the difference in distribution between ERIC I STs and ERIC II STs. The ERIC II STs (ST 10 and ST 11) cluster on the left (negative values) of CA1, and the ERIC I STs on the right, with the exception of ST17. ST10 groups with Arabia, Mount Athos, Kazakhstan, Russia, Sweden and Australia as well as mixed origin isolates from China, Singapore and Japan.

ST11 is associated with Germany and Finland. Native range designation did not appear to influence the distribution of the points on the ordination plot i.e. countries where honey bees have been introduced grouped with countries where honey bees are native.

2.4.6 Rarefaction.

Rarefaction curves comparing the sampling effort for each continent show steep curves for all except Europe. However, when only data from Europe were analysed at country level, it was clear that only Germany and the UK were sampled to adequately describe the resident ST populations. This suggests that further sampling in other countries would yield previously unidentified STs.

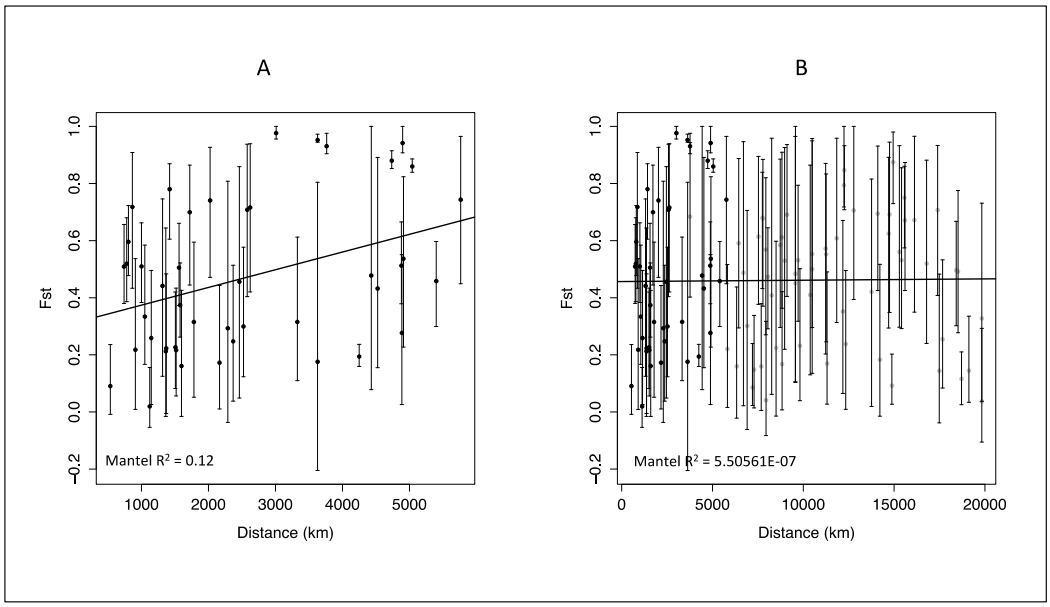

Figure 2.4 FST by distance. Figure 2.4A describes the F_{ST} by distance of *P larvae* populations within the native range of the host (P=0.01). Figure 2.4B describes the F_{ST} by distance of *P. larvae* populations both within (dark grey dots) and outside the host's native range (light grey dots) (P= 0.996).

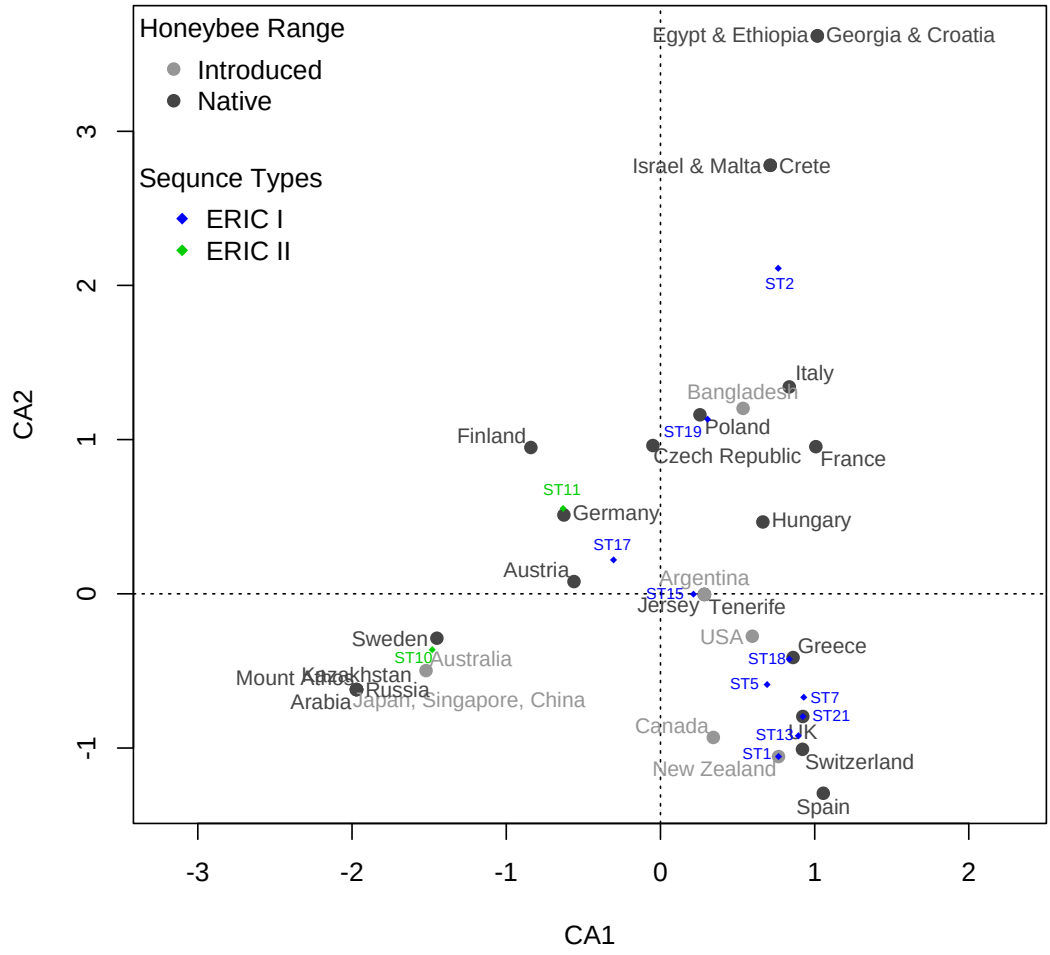

Figure 2.5 Correspondence analysis ordination graph. The CA ordination graph illustrates the association amongst countries of isolate origin and MLST STs. Filled circles represent countries (dark grey represents native host range, light grey represents introduced host range) Diamonds represent STs.

2.5 Discussion

AFB is a serious disease of honeybee brood with near global distribution, however disease epidemiology is poorly understood, in part due to an absence of a repeatable, high-resolution method to discriminate between strain types. We have developed the first MLST scheme, which ratifies and extends the established four group ERIC typing scheme. The new scheme describes 21 different *P.larvae* STs, therefore providing improved resolution when compared to ERIC typing. Whilst other typing schemes, such as PFGE (Pentikäinen et al., 2008) offer greater resolution, i.e. more types, than MLST, these methods are difficult to repeat, preventing comparisons between published studies. MLST is now the gold standard for epidemiological studies (Francisco et al., 2012) and results can be applied to derive epidemiological meaning both locally and internationally.

2.5.1 P. larvae *biogeography*

Our MLST scheme was able to identify a significant relationship between geographic distance and genetic distance amongst *P. larvae* populations within the native range of the honey bee (Figure 2.4A). This relationship is surprising given the history of human movement of bees within Europe in particular (De la Rúa et al., 2009). It might be expected that because of past honey bee movements, mixing of *P.larvae* STs might be equally distributed throughout the native range, however, this is not the case. A detectable link between physical and genetic distance remains despite these large-scale movements and endemic populations of *P. larvae* still appear to exist within the native range. Such a result could suggest that certain *P. larvae* strains are adapted to local honey bee populations. Honey bees are known to differ in their ability to resist pathogens (Jensen et al., 2009; Büchler et al., 2010), and AFB tolerant honey bees have been bred in Argentina (Spivak and Reuter, 2001). However, little is known more generally about the susceptibility of honey bee races to different strains of *P. larvae*. A more comprehensive understanding of inter-race susceptibility of *A. mellifera* is required to understand whether host factors are in part responsible for the significant pathogen population structure within the native range of *A. mellifera*.

The commercial interest in honey bee hive products has lead to the spread of *A. mellifera* far beyond its native geographic range. This global industry has facilitated the spread of

honey bee pests and pathogens such as the ectoparastic mite *Varroa destructor* (Solignac et al., 2005), *Nosema ceranae* (Klee et al., 2007) and Israel acute paralysis virus (IAPV) (Palacios et al., 2008) and it is known that *P.larvae* spores can remain infective in honey (Morse, 1992; Govan et al., 1999; Hansen et al., 2003). Our data indicate that when AFB has been moved outside the native honey bee range, any evidence of significant pathogen population structure breaks down (Figure 2.4B). This finding suggests that the international trade in honey bees and their hive products may have moved *P. larvae* multiple times in a non-systematic manner to infect honey bees beyond the native range. Despite haphazard movements of host and pathogen, there remain some interesting links between historic honey bee movements and ordination graph observations linking STs to locations (Figure 2.5). Some historic honeybee movements are recorded in the literature, and it is possible to trace, for example, honeybee imports into USA and New Zealand back over 400 years to their origins in Europe (Donovan, 1980; Goulson, 2003). It is perhaps unsurprising that the CA (Figure 2.5) groups countries together from within and outwith the native host range group, given the traceable links and potential transmission routes between these countries.

Novel types were sometimes identified only outside the native range of *A. mellifera* with ST3 only found in New Zealand and ST4, ST9, ST16 only identified in the Americas (Appendix 2). There are two plausible explanations for this observation. First, the presence of unique types outside the native range could suggest these types have evolved since becoming isolated from the founder pathogen population. Second and perhaps more likely, our sampling scheme was more intensive in some countries (such as the UK and Germany) and superficial in other countries within the native range (Appendix 2). Therefore it is likely that our sampling scheme was not sufficiently exhaustive to detect the full extent of ST diversity within the native range of *A. mellifera*.

P. larvae isolates classed as ERIC I were more diverse than those classed as ERIC II, containing 16 unique STs compared to only 3 for ERIC II (Figure 2.1). The results of the Chao1 estimate (Figure 2.3) show that although the sampling was uneven with more ERIC I isolates typed than ERIC II, for these data the Chao1 estimator levels off (Figure 2.3) suggesting that the Chao1 estimate is relatively independent of sample size. ERIC I and II isolates differ phenotypically in many ways for example endospore resistance to temperature (Forsgren et al., 2008), rates of sporulation (Saville, 2011), time to host death

(Genersch et al., 2005, 2006; Rauch et al., 2009) and it was suggested that they employ different strategies for killing larvae (Poppinga et al., 2012; Fünfhaus et al., 2013; Djukic et al., 2014). Perhaps these differences account for the disparity in diversity, as some STs may be better able to spread to new areas.

It had previously been assumed that ERIC II is confined to Europe and that it is not rare in Germany or Austria (Genersch, 2010). However, our data suggest that ERIC II STs are much more widely distributed than previously thought, being present in Asia, North America and Australasia as well as Europe although in this study no ERIC II isolates were found in Europe in countries West of Germany. However there is evidence that ERIC II is present in Germany in areas that border France and Belgium (Saarland, see Appendix 2), which might suggest its presence in these countries. This study has split ERIC II into three new STs (Figure 2.1) with ST10 and ST11 being common and ST12 being identified once. Considering the prevalence of ST10, it seems unusual that ST11 remains localised. These differences in prevalence and distribution between ERIC II STs could reflect a difference in phenotype, rather than an artefact of sampling bias, as there was no previous method to discern amongst ERIC II strains. Future work could include genomic and phenotypic comparisons between ERIC II STs, to identify reasons for the observed differences in distribution.

Interestingly ERIC I and II have mainly different distributions as indicated in the ordination graph, ST17 is the only ERIC I isolate on the left side of the graph (Figure 2.5). No ERIC II STs (ST10, ST11) were found in Africa or in the West of Europe in the native range of *A. mellifera* (Appendix 2). The fact that they were not found in Africa and much of the West of Europe may be explained by poor sampling effort. However, the UK was more thoroughly sampled and still no instances of ERIC II STs were found. One suggestion for their differing distribution might be that ERIC II STs are host specific. There are around a dozen subspecies of *Apis mellifera* and they are split into four groups M, W, C and O (African, Western Europe, Eastern Europe and Western Asia respectively). It is thought that *Apis mellifera* originated in Africa and from ther**e** it is likely that there was one expansion into Western Europe and either one or two expansions into the East (Eastern Europe and Asia) (Whitfield et al., 2006). This means that there is a closer relationship between the Western European group of subspecies and African subspecies than between Eastern and Western European subspecies. This may explain the split in the

distribution of ERIC types with STs of ERIC II only being found in Germany and countries to the East within Europe and Asia. The C group subspecies *A. m. carnica* has almost completely replaced the W group *A. m. mellifera* in Central European countries such as Germany, whereas in Poland where no ERIC II was found the majority of bees are still *A. m. mellifer*a (Meixner et al., 2007). The native range of *A. m. mellifera* is from the UK to Scandinavia and from France to Poland. However, Italian (C group) and Carnelian (*carnica*) bees have been transported around Europe in the *A. m. mellifera* range and hybridization has occurred, in fact it is thought that in Scandinavia and the UK that the bees are a mixture of all three sub species (Jensen et al., 2005; De La Rúa et al., 2001) and although both Scandinavia and the UK are thought to have a similar mix of sub species, ERIC II STs were found in Scandinavia**.** As we have no information on the host sub species in our data set it is impossible to determine whether different *P.larvae* STs affect *A. mellifera* subspecies differently. Future work could involve the testing this theory by infecting larvae of the honey bee subspecies with a range of STs of *P. larvae*.

Although local epidemiological observations were not the primary purpose of this study, our data offers some evidence for the hitherto unknown origin of the 2010 AFB outbreak on Jersey, an island in the English Channel. Our scheme matched one Jersey isolate (Appendix 2) with a sequence type that was only found in France (ST6), Jersey's closest neighbour, providing evidence of a potential transmission route and demonstrating the potential power of MLST to inform disease aetiology when coupled to more extensive local sampling efforts.

In summary, we have developed an important new tool for describing the genetic structure of *P. larvae*, which raises unanswered questions about differential host susceptibilities to *P. larvae* STs. Future proofing in an age of rapid advancement in sequencing technologies is an important consideration, and MLST is compatible with methods which could potentially supersede, such as whole genome sequencing (Larsen et al., 2012). National laboratories responsible for the control of AFB can now use this scheme to gather comprehensive data on ST locations and expand the online database pubMLST.org/plarvae/ (Jolley and Maiden, 2010) to build a comprehensive multinational data set to better understand the distribution and transmission networks of AFB on a global scale. Our scheme therefore provides the first universal method for the description of

strains of *P.larvae* and will increase our understanding of the epidemiology of this damaging and costly disease at many spatial scales.

3 Patterns of distribution of the honey bee pathogen *Paenibacillus larvae* in the UK

3.1 Abstract

American foulbrood (AFB) is a globally distributed disease of honey bees (*Apis mellifera*) caused by the bacterium *Paenibacillus larvae*. It is the most damaging brood disease of honey bees in the UK and is under statutory control. AFB has been shown to appear as persistent, perennial disease clusters. We used a newly developed MLST scheme to type disease outbreaks from around the UK between 2006-2013. We typed isolates from disease outbreaks that occurred in clusters as well as those that were not in clusters, these were classed as background outbreaks. 109 honey bee colonies were tested and ten sequence types (STs) were discovered, including two that had not been previously identified. We found that the STs found in English clusters were significantly different from those in the background in Scotland and England and in the Welsh cluster. Clustered outbreaks that were close to risk points were caused by more STs than more isolated clusters. One possible explanation is that risk points represent a diverse infection source, causing repeated multiple infections in nearby clusters. We tested honey that had been imported to the UK for *P. larvae* DNA and found 39% of samples contained traces suggesting a possible source of infection.

3.2 Introduction

Honey bees are globally important economic pollinators of food crops. The annual value of the insect pollination, including honey bees, of these crops was estimated at €1.5 billion worldwide, and £500m in the UK (Fera, 2013). As well as economically important crops, domesticated honey bees (*Apis mellifera*) also pollinate many native plants and wildflowers. It has been widely reported that honey bees are under stress from a variety of pests and diseases (De la Rúa et al., 2009). In the UK two bacterial brood diseases of honey bees are under statutory control, American foulbrood (AFB) and European foulbrood (EFB). Both diseases cause considerable damage to the UK beekeeping industry. While EFB is more widespread than AFB, colonies infected with EFB can be treated with antibiotics or the shook swarm method (Wilkins et al., 2007). In the UK AFB is only controlled by the eradication of infected colonies, antibiotics are not used (Wilkins et al.,

2007). AFB is therefore considered to be the most damaging brood disease of honey bees in the UK (Fera, 2013). AFB is caused by the bacterium *Paenibacillus larvae* and is an extremely damaging disease of the honey bee worldwide (Genersch, 2010).

3.2.1 UK AFB control

Incidence of this disease in the UK has decreased from several thousand cases a year in the 1940s to between 200-100 cases a year (Wilkins et al., 2007). This decrease is due to the UK inspection regime which involves increased scrutiny in an area 3-10 km surrounding a known AFB outbreak or exotic risk point (ports, honey packing plants) and destroying all infected hives. AFB is subject to statutory control within the UK. Any beekeeper who suspects an outbreak of AFB in a colony is legally required to contact the Animal and Plant Health Agency's (APHA), National Bee Unit (NBU), in order to have the colony officially examined by a bee inspector (Wilkins et al., 2007). Bee inspectors check colonies by eye and use lateral flow devices (LFD; De Graaf et al., 2006) to confirm infection. In most countries, including the UK, infected colonies are burned to stop the spread of the disease as antibiotics are not effective against the resistant spores (Genersch and Otten, 2003) however, Oxytetracycline is used to treat the symptoms of AFB in some countries (Reybroeck et al., 2012).

3.2.2 AFB spread

AFB can be spread both by humans and by bees. Humans act as agents for disease spread through the movement of bees and their products or by the use of the same tools at different apiaries (Genersch, 2010). The most common method of disease spread by honey bees is horizontal transmission, although vertical transmission has been recorded (Fries et al., 2006). Horizontal transmission occurs, either through the movement of adult bees between colonies (drifting) or the behaviour of foragers (robbing) (Genersch, 2010). The distance over which bees can spread AFB is around 1-2km (Lindström et al., 2008b). However, studies have shown that apicultural practices can cause the spread of bee diseases over much larger distances (Pentikäinen et al., 2008; Haynes et al., 2013). A recent study in the UK has found that AFB outbreaks can occur in clusters. These clusters occur within areas of 10-30 km and may last from 1-5 years (Mill et al., 2014). AFB is not thought to be endemic in the UK and the sporadic nature of the outbreaks is unusual. Local

spread of diseased material within and between apiaries, whether by honey bees or humans may explain the spatial clustering of disease, but it does not explain the position or timing of the outbreaks (Mill et al., 2014). The majority of the disease clusters are eventually wiped out due to the inspection regime. However, some clusters reoccur and this may be due to repeated infection, either through exotic introductions or from a local source (Mill et al., 2014). *P. larvae* spores remain infective in honey and can survive for up to 35 years on old hives and equipment (Hasemann, 1961). Although there are around 40,000 beekeepers with over 200,000 honey bee colonies in the UK (Temple et al., 2001), the majority of honey used here is imported (BHIPA, 2011). The UK also imports many apiary products such as beeswax, queens and caged bees from European countries (Mill et al., 2014). Therefore, the trade in bees and their products represents a plausible route by which infections arise in the UK.

3.2.3 MLST

MultiLocus Sequence Typing (MLST) is a widely used method that definitively classifies strain types of bacteria (Maiden, 2006) and has been used to study disease outbreaks as well as trace sources of infection in many bacterial species. Here we use a recently developed MLST scheme (Morrissey et al., 2015; Chapter 2) to define and interpret patterns in the distribution of *P. larvae* sequence types (STs) causing disease in the UK. We describe the position and timing of *P.larvae* STs causing disease in the UK for the first time, giving us a clearer picture of the possible sources and movement of bacterial strains. We test whether ST is related to whether the outbreak is part of a cluster or not and whether proximity to risk points has an effect on the number of STs in a cluster.

3.3 Methods

Sampling. In total 109 isolates were sequenced for this study including 63 reported in a previous study (Morrissey et al., 2015; Chapter 2; Table 3.1). We included 72 isolates from 8 clusters in England and Wales as well as 15 isolates from across Scotland and 22 isolates not associated with clusters (background outbreaks) from England and Wales (Table 3.1). Disease clusters identified in Mill et al (2013) were used, cluster E2 was not included in

the final draft of Mill et al due to its small size. Beebase, the National Bee Unit's database of bee disease outbreaks was used to identify the affected colonies, colonies within a 15km radius of the centre of the cluster were included. The disease clusters are temporal as well as spatial and diseased colonies were included from 2006 to 2013 although not all clusters contained outbreaks from each year (Table 3.1). Precise geographic locations of clusters and outbreaks are withheld due to data protection. Bacteria were cultured from LFDs used by bee inspectors in the field between 2006 and 2013. Twenty negative LFDs were tested for the presence of *P. larvae* using the same methods. Disease clusters were assigned from Mill et al. (2014). Background samples were chosen to represent all areas of England and no two samples were taken from the same site/apiary. Clusters defined as close to risk factors were those where at least one outbreak was less than 10km from an exotic risk factor such as ports where honey and bee products were likely to be imported as well as honey packing plants and bee importers. This information was taken from Beebase. Honey samples had been imported into the UK from both EU and non-EU sources and we sampled sources in proportion to actual imports, where possible.

3.3.1 DNA Extraction.

DNA was extracted from LFDs (Vita Europe, Basingstoke, UK). A small piece of the sample pad was removed from an area to the right of the sample port using a sterile scalpel. This was then added to a *P. larvae* agar plate (PLA) (de Graaf et al., 2013) without heat treatment and incubated at 35°C for several days. Single colonies were then spread onto two plates, a brain heart infusion agar plate

Table 3.1 Data for UK *P. larvae* isolates typed in this chapter

Isolate number	ST	SF	Cluster	year	Colonies at apiary	Infected colonies at apiary	Apiary density Within 10km	Risk point within 5km	First reported
p8656	**2**	SF07/8626	Background	2007	1	1	204	airport	Chapter 3
p8672	**2**	SF09/9978	Background	2009	3	1	114	bee importer	Chapter 3
p8664	**2**	SF11/10604	Background	2011	5	4	13	port	Chapter 3
p8657	**2**	SF13/11703	Background	2013	32	3	45	zoo	Chapter 3
p8666	**5**	SF11/10724	Background	2011	1	1	87	No	Chapter 3
UK16/7	**5**	SF07/8872	Background	2007	4	1	78	port	Chapter 3
p8658	**5**	SF13/11685	Background	2013	7	2	191	queen bee importer	Chapter 3
UK12/8	**15**	SF08/9389	Background	2008	3	2	103	airport	Chapter 3
p8660	**15**	SF12/11089	Background	2012	20	1	216	No	Chapter 3
p8663	**15**	SF12/10964	Background	2012	3	2	68	No	Chapter 3
P8606	**15**	SF11/10595	Background	2011	43	2	34	No	Chapter 3
P8607	**15**	SF11/10768	Background	2011	39	2	34	No	Chapter 3
p8659	**17**	SF10/10153	Background	2010	4	1	49	zoo	Chapter 3
UK19/8	**18**	SF08/9246	Background	2008	5	1	125	airport	Chapter 3
p8671	**18**	SF09/9602	Background	2009	4	2	101	bee importer	Chapter 3
UK8/10	**18**	SF10/10316	Background	2010	37	2	126	No	Morrissey et al., 2015
UK9/10	**18**	S10/10060	Background	2010	8	1	125	No	Chapter 3
p8661	**18**	SF12/11131	Background	2012	9	1	196	No	Chapter 3
p8667	**18**	SF11/10828	Background	2011	9	2	123	No	Chapter 3
UK4/10	**18**	SF10/10321	Background	2010	1	1	126	No	Chapter 3
UK4/6	**18**	SF06/8271	Background	2006	10	4	52	port	Chapter 3
p8670	**24**	SF07/8772	Background	2007	16	1	144	No	Chapter 3

UK7/9	**5**	SF09/9494	E1	2009	10	2	43	No	Morrissey et al., 2015
UK5/9	**5**	SF09/9774	E1	2009	2	1	51	No	Morrissey et al., 2015
UK38/9	**5**	SF09/9698	E1	2009	9	1	43	No	Morrissey et al., 2015
UK35/9	**5**	SF09/9647	E1	2009	3	1	51	No	Morrissey et al., 2015
UK27/9	**5**	SF09/9589	E1	2009	4	4	57	No	Morrissey et al., 2015
UK26/9	**5**	SF09/9588	E1	2009	2	2	60	No	Morrissey et al., 2015
UK22/9	**5**	SF09/9496	E1	2009	10	3	43	No	Morrissey et al., 2015
UK22/10	**5**	SF10/10050	E1	2010	2	2	51	No	Morrissey et al., 2015
UK21/10	**5**	SF10/10043	E1	2010	3	1	52	No	Morrissey et al., 2015
UK17/9	**5**	SF09/9588	E1	2009	2	2	60	No	Morrissey et al., 2015
UK17/10	**5**	Sf10/10164	E1	2010	3	1	52	No	Morrissey et al., 2015
UK14/10	**5**	SF10/10050	E1	2010	2	2	51	No	Morrissey et al., 2015
UK12/10	**5**	SF10/10042	E1	2010	1	1	35	No	Morrissey et al., 2015
S2/1	**5**	SF12/10879	E1	2012	1	1	46	No	Morrissey et al., 2015
S1/1	**5**	SF12/10880	E1	2012	3	1	51	No	Morrissey et al., 2015
UK6/9	**5**	SF09/9511	E1	2009	1	1	41	No	Chapter 3
UK24/10	**5**	SF10/10073	E2	2010	17	5	103	No	Morrissey et al., 2015
UK2/10	**5**	SF10/10073	E2	2010	17	10	103	No	Morrissey et al., 2015

UK11/6	**5**	SF06/8348	E2	2006	5	1	107	No	Morrissey et al., 2015
UK33/9	**5**	SF09/9628	E3	2009	2	2	80	Port	Morrissey et al., 2015
UK2/9	**5**	SF09/9915	E3	2009	2	1	80	Port	Morrissey et al., 2015
P8610	**5**	SF12/11480	E3	2012	1	1	78	Port	Chapter 3
P8611	**5**	SF12/11406	E3	2012	3	2	78	Port	Chapter 3
P8604	**5**	SF12/11406	E3	2012	3	2	78	Port	Chapter 3
UK1/8	**15**	SF08/9122	E3	2008	20	3	193	Port	Morrissey et al., 2015
P8597	**15**	SF12/10948	E3	2012	3	1	174	Port	Chapter 3
UK32/9	**15**	SF09/9594	E3	2009	8	1	185	Port	Chapter 3
UK36/9	**15**	SF09/9656	E3	2009	9	2	185	Port	Chapter 3
UK6/7	**21**	SF07/8903	E3	2007	8	1	150	Port	Morrissey et al., 2015
UK25/7	**21**	SF07/8636	E3	2007	9	1	150	Port	Morrissey et al., 2015
UK27/7	**5**	SF07/8656	E5	2007	12	2	44	Honey packer	Morrissey et al., 2015
UK26/7	**5**	SF07/8655	E5	2007	19	1	52	Honey packer	Morrissey et al., 2015
UK22/11	**5**	SF11/10359	E5	2011	5	1	112	Honey packer	Morrissey et al., 2015
UK2/7	**5**	SF07/8656	E5	2007	12	2	44	Honey packer	Morrissey et al., 2015
UK1/10	**5**	SF10/10195	E5	2010	1	1	123	Honey packer	Morrissey et al., 2015
UK3/10	**15**	SF10/9999	E5	2010	3	1	107	Honey packer	Morrissey et al., 2015
UK11/10	**15**	SF10/10081	E5	2010	2	1	123	Honey packer	Morrissey et al., 2015
UK40/6	**18**	SF06/8422	E5	2006	19	3	81	Honey packer	Morrissey et al., 2015

UK27/6	**18**	SF06/8255	E5	2006	16	1	81	Honey packer	Morrissey et al., 2015
UK16/6	**18**	SF06/8422	E5	2006	19	3	81	Honey packer	Morrissey et al., 2015
UK14/6	**18**	SF06/8513	E5	2006	15	2	81	Honey packer	Chapter 3
UK6/10	**2**	SF10/10077	E6	2010	1	1	117	Honey packer	Morrissey et al., 2015
UK7/10	**5**	SF10/10018	E6	2010	4	2	110	Honey packer	Morrissey et al., 2015
UK6/8	**5**	SF08/9063	E6	2008	95	19	67	Honey packer	Morrissey et al., 2015
UK3/8	**5**	SF08/9063	E6	2008	95	19	67	Honey packer	Morrissey et al., 2015
UK21/8	**5**	SF08/9063	E6	2008	95	19	67	Honey packer	Morrissey et al., 2015
UK10/8	**5**	SF08/9240	E6	2008	63	1	67	Honey packer	Chapter 3
UK23/9	**15**	SF09/9505	E6	2009	4	2	110	Honey packer	Morrissey et al., 2015
UK20/10	**5**	SF10/10255	E7	2010	5	1	131	honey bee supplier/queen bee importer	Morrissey et al., 2015
UK16/8	**5**	SF08/9247	E7	2008	11	1	92	honey bee supplier/queen bee importer	Morrissey et al., 2015
UK15/10	**5**	SF10/10259	E7	2010	1	1	136	honey bee supplier/queen bee importer	Morrissey et al., 2015
UK13/9	**15**	SF09/9671	E7	2009	4	1	61	honey bee supplier/queen bee importer	Morrissey et al., 2015

UK3/7	**15**	SF07/8948	E7	2007	4	2	117	honey bee supplier/queen bee importer	Chapter 3
P8634	**7**	SF06/8338	E9	2006	3	2	44	No	Morrissey et al., 2015
P8632	**7**	SF06/8158	E9	2006	6	3	64	No	Morrissey et al., 2015
P8633	**7**	SF07/8563	E9	2007	5	5	143	No	Morrissey et al., 2015
UK38/6	**7**	SF06/8400	E9	2006	6	1	39	No	Morrissey et al., 2015
UK37/6	**7**	SF06/8399	E9	2006	12	1	44	No	Morrissey et al., 2015
UK36/6	**7**	SF06/8398	E9	2006	5	1	53	No	Morrissey et al., 2015
UK34/6	**7**	SF06/8349	E9	2006	5	1	64	No	Morrissey et al., 2015
UK25/6	**7**	SF06/8249	E9	2006	12	3	44	No	Morrissey et al., 2015
UK23/6	**7**	SF06/8158	E9	2006	6	3	64	No	Morrissey et al., 2015
UK21/7	**7**	SF07/8563	E9	2007	5	5	143	No	Morrissey et al., 2015
UK21/6	**7**	SF06/8157	E9	2006	3	3	48	No	Morrissey et al., 2015
UK15/6	**7**	SF06/8398	E9	2006	5	1	53	No	Morrissey et al., 2015
UK10/6	**7**	SF06/8157	E9	2006	3	3	48	No	Morrissey et al., 2015
UK21/11	**15**	SF11/10334	E9	2011	8	5	52	No	Morrissey et al., 2015
P8612	**4**	FS09/123	sco	2009	NA	NA	NA	NA	Chapter 3
P8617	**5**	FS11/60	sco	2011	41	3	84	No	Chapter 3

P8621	**15**	FS12/105	sco	2012	1	1	53	No	Chapter 3
P8622	**15**	FS13/42	sco	2013	12	1	7	No	Chapter 3
P8618	**18**	FS11/88	sco	2011	5	1	12	No	Chapter 3
P8619	**18**	FS12/55	sco	2012	4	1	34	No	Chapter 3
P8620	**18**	FS12/56	sco	2012	4	4	54	No	Chapter 3
P8635	**18**	FS13/98	sco	2013	2	1	24	No	Chapter 3
P8636	**18**	FS13/103	sco	2013	1	1	8	No	Chapter 3
P8637	**18**	FS13/104	sco	2013	3	1	10	No	Chapter 3
P8613	**23**	FS09/151	sco	2009	NA	NA	NA	NA	Chapter 3
P8615	**23**	FS10/15	sco	2010	NA	NA	NA	NA	Chapter 3
P8614	**23**	FS09/158	sco	2009	28	10	7	No	Chapter 3
P8616	**23**	FS10/01	sco	2010	5	1	7	No	Chapter 3
P8623	**23**	FS12/61	sco	2012	6	1	54	No	Chapter 3
UK1/9	**15**	SF09/9912	W3	2009	2	1	71	disused barrel dump	Morrissey et al., 2015
UK9/7	**18**	SF07/8791	W3	2007	4	1	87	disused barrel dump	Morrissey et al., 2015
UK5/8	**18**	SF08/9368	W3	2008	2	1	64	disused barrel dump	Morrissey et al., 2015
UK28/7	**18**	SF07/8767	W3	2007	2	1	59	disused barrel dump	Morrissey et al., 2015
UK14/7	**18**	SF07/8767	W3	2007	2	1	59	disused barrel dump	Morrissey et al., 2015

(BHIA; (de Graaf et al., 2013) and a second PLA plate. DNA was extracted from clonal colonies growing on the BHIA plate after several days in the incubator at 35°C. DNA was extracted using a simple chelex method. Bacteria were transferred to 300 µl 6% Chelex ®100 and heated to 56°C for 20 minutes followed by boiling for 8 minutes. DNA extracts were stored at -20°C until required. Honey samples were warmed to 65°C until liquid. 5ml of honey was removed and added to 5ml sterile distilled water and centrifuged at 12000g for 40 minutes. The pellet was retained and resuspended in 1ml SDW for DNA extraction. The PCR reaction were carried out using 2µl template DNA, 12.5µl 2x Fermentas PCR mastermix, and 10 pmol of each primer with a total reaction volume of 25µl. Each reaction was run on an Applied Biosystems 2720 Thermal Cycler. PCR conditions were as follows: After the initial activation step (3 min, 95 °C), 35 cycles at 95 °C for 30 seconds, 52 °C for 30 seconds, and 72 °C for 1 min were run followed by a final elongation step at 72 °C for 10 min. PCR products were purified using Qiagen® PCR purification and sequenced on a ABI 3730xl 96-capillary DNA Analyser. A published seven-gene MLST scheme was then used to generate sequence information for glpF, glpT, Natrans, sigF, rpoB, ftsA and clpC for all samples (Jolley and Maiden, 2010; Morrissey et al., 2015; pubmlst.org/plarvae). Samples were verified as *P. larvae* using a species-specific real time PCR with Taqman® chemistry (forward: Pl_R24_468F TCC CCG AGC CTT ACC TTT GT, reverse: Pl_R24_583R ACC TAC GAA CTT GAC GCT GTC CT, probe: Pl_R24_489T TGC TCA TAC CCG GTC AGG GAT TCG A). Duplicate real-time PCR reactions were run for each sample on an ABI 7500 real-time PCR machine. Reactions were performed with AmpliTaq Gold ® (Applied Biosystems®), 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 MgCl2, 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 40 cycles of 60 °C for 1 min plus 95 °C for 15 s). *P. larvae* DNA positive controls were included in all qPCR assays, and were always detected. Samples with Ct scores lower than 35 were considered to contain *P. larvae* DNA.

3.3.2 **Analyses**

Non-metric multidimensional scaling (NMDS; Kruskal, 1964a, 1964b) is a tool to visualise and interpret information from multiple dimensions (e.g. multiple clusters) in two dimensions. Here NMDS uses the rank orders of the Bray-Curtis dissimilarity index in

each cluster to display patterns of distribution of STs in clusters. The Bray-Curtis dissimilarity uses the proportion of STs found in each cluster to quantify how dissimilar they are. Using the result of the NMDS we classified outbreaks in two groups. The first group contained all isolates from clusters in England, the second group contained all isolates from Scotland, background and Wales. The Bray-Curtis dissimilarity index was used to compare the groups. The results were analysed with a Mantel test with 1000 iterations to determine significance. To calculate the Sørensen dissimilarity index all samples from England, Wales and Scotland were pooled and presence/absence of STs was compared to Europe and then to the rest of the world excluding Europe. The data for Europe/rest of the world were taken from Morrissey et al. (2015). Sørensen's index was used here because the sampling effort of these groups was different and it was more appropriate to use presence/absence of STs than proportions.

A stepwise binomial generalised linear mixed-effect model (GLMM) was used to test whether any predictors had an effect on the likelihood of infection by ST5. The model with the lowest AIC score was chosen. The following predictors were included: distance from risk point, total number of colonies at apiary, number of infected colonies at apiary, the proportion of infected colonies at the apiary and apiary density within a 10km radius, cluster ID was included as a random effect. We removed three isolates from Scotland from the analysis as the required information was not available for them (Table 3.1.). All analyses were completed using R (2012) packages VEGAN and ECODIST (Goslee et al., 2007; Oksanen et al., 2013). PHYLOVIZ (Francisco et al., 2009) was used to create the eBURST diagram. Data on AFB infection were taken from Beebase, the National Bee Unit's database of honey bee disease.

Table 3.2 MLST allele numbers for UK *P. larvae* STs

ST	***glpF***	***sigF***	***glpT***	***Natrans***	***rpoB***	***ftsA***	***clpC***
1	1	1	1	2	1	4	3
2	1	1	1	2	4	3	2
4	1	1	1	2	4	4	2
5	1	1	1	2	4	4	3
7	1	1	3	2	4	4	3
15	3	1	1	2	4	4	2
17	4	1	1	2	4	4	3
18	3	1	1	6	4	4	2
23	1	1	1	2	6	4	2
24	7	1	1	2	4	4	2

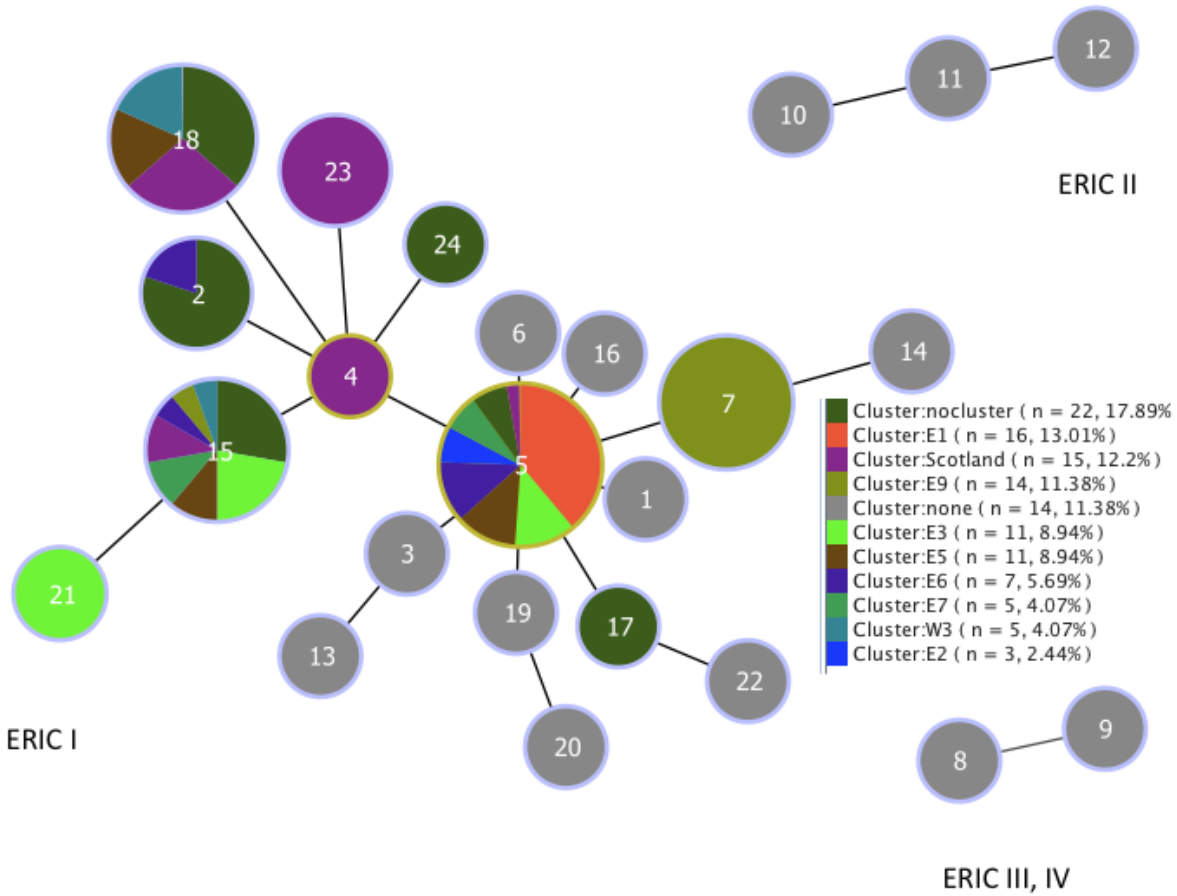

Figure 3.1 eBURST diagram displaying origins of UK STs. Numbers in circles represent STs of isolates. Size of circles represents number of isolates with that ST. Clusters refer to those mapped in figure 3.2. Grey bubbles represent STs that were not found in the UK

3.4 Results

Sequence types.

From 109 isolates we identified ten different STs, all of them grouping within ERIC type I (Figure 3.1.; Genersch et al., 2006), two of these (ST23, ST24) have been identified for the first time in this study (see Table 3.2). ST23 is a single locus variant (SLV) of and ST4 and ST24 is an SLV of STs 15 and 18 (Figure 3.1; Table 3.2). These types have been uploaded to the P. larvae MLST database (pubmlst.org/plarvae). STs 5, 15 and 18 were found in multiple clusters. ST2 was found in one disease cluster but was also found in four background outbreaks. The remaining six STs were either found at a single site (ST4, ST7, ST17, ST21, ST24) or in only one cluster/ country, ST7 was found only in cluster (E9) and ST23 was found exclusively in Scotland (Figure 3.2).

3.4.1 *Clusters and risk points.*

Of eight clusters sampled in England and Wales five were classed as close to risk factors (honey packing plant, barrel sorting facility, port) (E3, E5, E6, E7, W3) and three were further than 10km from these risk factors (E1, E2, E9). Of the 14 samples from cluster E9, thirteen were caused by ST7 in 2006-2007. An outbreak in the same area, four years later, was caused by a different ST (ST15) (Figure 3.2, Table 3.1.).

3.4.2 Clusters and background.

NMDS revealed two groups of disease outbreaks, those in clusters in England (E1-E9) appeared to contain different STs from those in the background of England and Scotland and those in the Welsh cluster (Cluster W3) (Figure 3.3, Table 3.3.). These groups were found to be significantly different when tested with the Bray-Curtis dissimilarity index and Mantel test ($p<0.05$).The proximity to risk factors did not appear to have a strong effect on cluster composition. Using the Chi squared test of association it was found that of the three most common and widespread STs in the UK (STs 5, 15, 18), ST5 is significantly

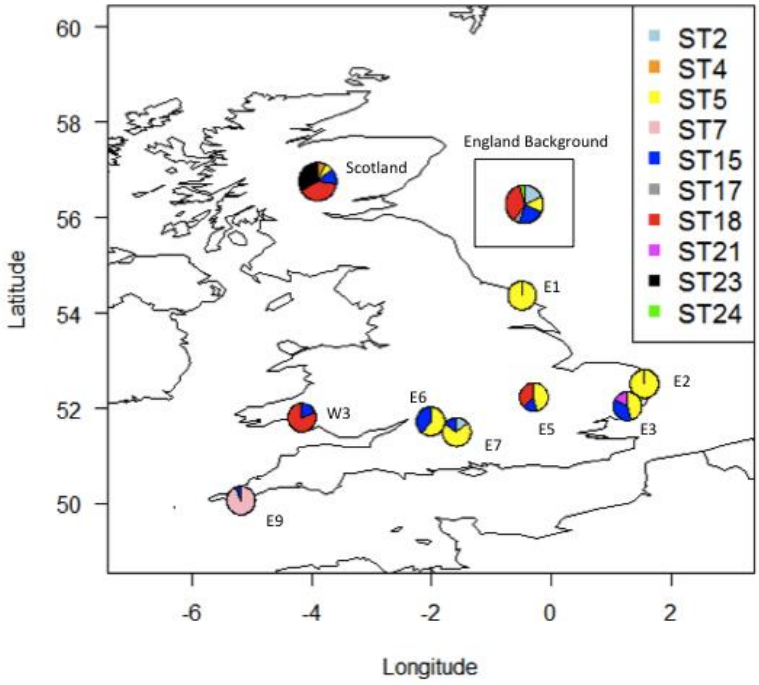

Figure 3.2 Approximate location of AFB disease clusters with pie charts displaying ST composition

Table 3.3 Bray-Curtis test of dissimilarity results

Cluster	E1	E2	E3	E5	E6	E7	E9	W3	Scotland
E2	0.6842105								
E3	0.6296296	0.5714286							
E5	0.6296296	0.5714286	0.3636364						
E6	0.5652174	0.4	0.3333333	0.3333333					
E7	0.7142857	0.25	0.375	0.375	0.3333333				
E9	1	1	0.92	0.92	0.9047619	0.8947368			
W3	1	1	0.875	0.375	0.8333333	0.8	0.8947368		
Scotland	0.9354839	0.8888889	0.7692308	0.4615385	0.7272727	0.7000000	0.9310345	0.5	
Background	0.8378378	0.75	0.5625	0.4375	0.6428571	0.6153846	0.9428571	0.6153846	0.4444444

more likely to be found in a cluster than in background outbreaks, this was true whether or not the large ST5 cluster (Cluster E1) was included ($p<0.001$). ST 18, by contrast was significantly more likely to be found in background outbreaks ($p<0.05$). ST15 was found in the same frequency in both clusters and background. The GLMM did not identify a predictor that had a significant effect on the likelihood of finding ST 5. The model with the lowest AIC contained only the total number of hives at the apiary and total number of infected hives as fixed effects with cluster as a random effect. Between-cluster variance was high ($\sigma^{2=}6.46$) indicating that the likelihood of ST5 being present varied widely across clusters.

3.4.3 *Comparison to global STs*

The Sørensen dissimilarity index between the UK and EU is 0.5 and between the UK and the rest of the World excluding the EU is 0.57 on a scale of 0-1, 1 being completely dissimilar. This means that the STs found in the UK are slightly more similar to those found in the EU than to those found outside Europe.

3.4.4 Honey samples.

Of the 94 honey samples tested for *P.larvae* DNA 76 were of non-EU origin and 18 originated in the EU (see Table 3.4). The majority of the non-EU honey samples originated in China (54/76). In total 33% (25/76) of the non-EU honey and 67% (12/18) of the EU honey contained detectable levels of *P. larvae* (CT <35; Table 3.4.). Further work will include MLST analysis of the honey samples.

3.5 Discussion

American foulbrood is a damaging and costly disease in the UK and although statutory control and eradication has decreased its prevalence (Wilkins et al., 2007), new infections occur each year. It has been shown that some outbreaks of AFB occur in clusters that appear year after year (Mill et al., 2014). This is the first time these clusters have been sequence-typed to determine the strains causing them and assess whether they are caused by re-infection of an apparently

Table 3.4 *P. larvae* occurrence in UK imported honey samples

Country	No. of samples	No. CT score <35	Proportion infected
China	54	20	0.37
Mexico	11	2	0.18
New Zealand	9	1	0.11
Thailand	2	2	1.00
England	1	0	0.00
France	2	0	0.00
Greece	2	1	0.50
Hungary	1	1	1.00
Romania	7	6	0.86
Scotland	1	1	1.00
Spain	4	3	0.75
total	94	37	0.39
EU	18	12	0.67
Non EU	76	25	0.33

persistent ST or by repeated *de novo* infections of new sequence types. Ten STs were identified to be causing disease in the UK between 2006-2013, two of which are new types not previously described (ST23 and ST24). Three of the STs 5, 15 and 18 were found in multiple clusters as well as in background outbreaks. These types are known to be globally widespread: all three are found in Europe, the Americas and Australasia. ST 5 is also found in Asia and ST 15 is found in Africa (Morrissey et al., 2015). In a recent global study of AFB ST distribution, these three types, STs 5, 15 and 18, make up 35%, 20% and 8% respectively of all ERIC I types identified (Morrissey et al., 2015).

The STs found in clusters differed from those found in background (Table 3.3.). The background outbreaks of England and Scotland were composed of more STs than the clusters. Disease clusters are predicted to be dominated by types that then spread locally, either by bees or by keepers (Pentikäinen et al., 2008; Genersch, 2010; Datta et al., 2013). This is consistent with the lower numbers of STs in clusters found here, whereas the background outbreaks are assumed to be caused by unrelated STs. These results also ratify the statistical procedures for determining clusters used by Mill et al., (2014). The NMDS analysis split outbreaks into two groups (Figure 3.2.), those in clusters in England and those in background outbreaks in England and Scotland with the Welsh cluster. These groups are significantly different, most likely due to the difference in abundance of ST5 and 18. ST5 was significantly more likely to be found in all clusters, whereas ST18 was significantly more likely to be found in background outbreaks. In the Welsh cluster four of the five samples are ST18, which explains why it groups with Scotland/English background. Because of the small sample size we are unable to say whether Welsh outbreaks are caused by different STs from English clusters. It is likely that they would be caused by similar STs as Wales is under the same statutory control as England. However, a recent study on the distribution of EFB STs discovered that ST diversity was significantly lower in Wales than England (Budge et al., 2014). ST5 is significantly more common in all clusters than in the background outbreaks. ST 5 was also the most commonly identified ST in a recent study on global *P.larvae* STs (Morrissey et al., 2015). ST5 appears to be a successful pathogen, perhaps due to differences in infection method. For example, it is well documented that different ERIC types of *P.larvae* vary in virulence at colony and individual larvae levels (Rauch et al., 2009) and it has also been shown that different ERIC types produce varying numbers of infectious spores (Saville, 2011). However, until

recently there was no higher resolution strain typing than ERIC type so no work has been done as yet to look into differential virulence amongst STs within ERIC groups. Further work could look into determining the biological reasons for ST5s prevalence, and specifically in clusters. A recent study of EFB STs in the UK identified some STs that are thought to be more virulent than others (Budge et al., 2014) which has ramifications for disease control as the treatment method can be tailored to individual STs or groups of STs.

It has been suggested that the global trade in honey and bee products may have an effect on the strain types of *P. larvae* causing disease in the UK (Mill et al., 2014). In the UK over the last few decades there has been an increase in importation of hive products such as beeswax and honey as well as large numbers of caged bees and queens from European sources and it has been suggested that some clusters may persist due repeated infection from these sources (Mill et al., 2014). *P.larvae* produces extremely resistant spores that can remain infectious in honey as well as in apiary equipment. Five of the clusters included in this study (E3, E5 E6, E7, and W3) are near to likely exotic risk factors including ports, honey packing plants and bee importers. All five of these sites contained outbreaks caused by different STs sometimes within one year, demonstrating multiple infection incidences (see Table 3.1.). Of the three clusters not near a risk point (E1, E2, E9) two contained only isolates of ST5 suggesting fewer infection incidences and local spread of infection. However, cluster E2 has a sample size of only 3, which may mean some STs were missed. Cluster E9 contained thirteen isolates of ST7 from 2006-2007and one isolate of ST15 four years after the previous outbreak in the area. Our results suggest that those clusters nearer to risk points are subject to repeated infections of AFB in comparison to those clusters distant from risk points.

The results showed low similarity between the composition of types in the UK and Europe (0.5 on a scale of 0-1, 0 being identical) and the rest of the world (0.57). However, The UK imports 90% of its honey (BHIPA, 2011) and the largest proportion of this comes from China, Argentina and Mexico (USAID, 2011). The only previous study to type *P. larvae* using the MLST scheme characterised few samples from these regions (Morrissey et al., 2015). Therefore we cannot determine whether the types causing disease in clusters near to exotic risk point are similar to the types found in countries that export large amounts of honey to the UK. We identified *P. larvae* DNA in 39% of the imported honey samples tested. Interestingly a higher proportion of the EU samples we tested contained *P. larvae*

DNA than the non-EU. Previous studies have found that Non-EU sources were more likely to be infected than EU sources (Hansen, 1984; Ritter, 2003). Although trade of bees and their products between European countries is common we found low similarity between the STs found in Europe and those found in the UK. A recent study on another disease of bee-brood, European foulbrood (EFB) found similar results (Haynes et al., 2013). The *P. larvae* population in the UK differs from that of the rest of the world in that no evidence of the three ERIC II STs was found in the UK (Morrissey et al., 2015). The 46 new UK isolates typed for this study were also all ERIC I.

Multiple infections and the persistent nature of outbreaks near risk points as identified by Mill et al. (2014) suggest that these risk points may have an effect on the types of *P. larvae* causing infections. Further work involving typing all AFB outbreaks and *P. larvae* DNA from imported honey could determine whether these outbreaks are related to importations of honey bees and their products. Our results are consistent with an epidemiology of transmission events that occur at different spatial scales due to a combination of the honey bee behaviour (local transmission) and human behaviour (local transmission and longer-distance transmission) that has been demonstrated for EFB (Budge et al., 2014) We see evidence for local transmission in clusters E1 and E9 and long-distance transmission in cluster E3 where new infections appear to develop year after year. Both of these transmission routes have previously been demonstrated in AFB by Pentikäinen et al. (2008).

The *P. larvae* MLST gives us evidence that the UK eradication process is successful and in the case of cluster E9 reoccurrence of disease was most likely due to human processes rather than from a disease reservoir in the landscape. The later infection in 2011 was a different ST (ST15) from the previous cluster in 2006-2007 (ST7) (Table 3.1.). Without the scheme it would not be possible to determine whether the more recent outbreak was related to the previous one. The National Bee Unit began to type all outbreaks of AFB in 2014. Using the scheme all cases of ST18 in this year were linked to a single beekeeper (personal communication, Victoria Tomkies, February 5, 2015).

We demonstrate that disease clusters can be made up of multiple infections or single outbreaks that move locally and that AFB moves at different spatial scales in the UK. The new *P. larvae* MLST scheme allows us to study disease outbreaks in more detail than was

previously possible. We identify honey as a possible source of transmission and demonstrate that exotic risk points may increase the number of STs causing nearby disease outbreaks. The *P. larvae* MLST scheme is a useful tool to determine which strains are causing disease outbreaks. AFB is a less common disease than EFB in the UK but this may be an advantage for tracing sources of infection. Bee inspectors will be able to trace the sale of products once they determine the strain causing an outbreak. In the future a more systematic approach to outbreak typing will give a comprehensive picture of the disease landscape in the UK and this will enable researchers to pinpoint the source of infections and trace the transmission routes.

4 Comparison of *Paenibacillus larvae* population structure determined using multilocus sequence- typing and whole genome sequencing data

4.1 Abstract

Paenibacillus larvae is the causative agent of American foulbrood, a damaging and destructive disease of honey bees (*Apis mellifera*). Recently, a new MultiLocus Sequence Typing (MLST) scheme was published giving greater resolution of the diversity of *P. larvae* than was previously possible and identifying over 20 sequence types (STs). Here, I describe the whole genome sequencing (WGS) and assembly of 19 isolates of *P. larvae* belonging to 13 STs. STs were chosen to represent diversity within and between the standard ERIC genotypes. This is the first time the WGS of multiple STs of *P. larvae* belonging to the same genotype have been compared. The general features of each ST are presented and the discovery of seven plasmids in four STs is reported. I identified 679 core genes present in all strains and compared phylogenies based on these data and MLST data. In general the MLST data accurately describes the relationship between STs. However WGS gives added resolution and can detect relationships not revealed using MLST. In the future WGS will become a more cost efficient method of disease typing in conjunction with MLST to standardise studies.

4.2 Introduction

Since the first bacterial genome was sequenced in 1995 (Fleischman et al., 1995) whole genome sequencing (WGS) has become significantly faster and cheaper. Due to the increase in speed and decrease in cost of WGS (Loman et al., 2012) it has now become a tool available to researchers interested in a variety of topics related to bacterial genetics and evolution (Edwards and Holt, 2013). WGS has a number of applications for the study of pathogenic bacteria, for instance tracking disease outbreaks (Chin et al., 2011; Grad et al., 2012; Köser et al., 2012) and the spread of drug resistance (Holt et al., 2012). WGS gives greater resolution than typing methods such as MultiLocus sequence typing (MLST) but can be used in tandem. Knowledge of the sequence type (ST) is important for linking data to previous and future studies and MLST can be standardised across labs when made available on publically accessible databases, e.g. pubmlst.org (Jolley and Maiden, 2010).

MLST typically requires sequencing of 6-10 fragments of housekeeping genes per isolate (Maiden, 2006). However, the online database BIGSdb (Jolley and Maiden, 2010) and software such as BioNumerics (Applied Maths) allows researchers to use genome data to easily type isolates.

4.2.1 The bacterial genome

A genome refers to the complete set of chromosomes and genes in an individual organism. It was previously thought that the bacterial domain was defined by a single chromosome that was the primary source of all genetic material and was responsible for inheritance (Krawiec and Riley, 1990). However it has since been shown that bacteria can contain more than one replicon, and these can be of different types, plasmids and chromids. Plasmids are secondary replicons which may contain genes that are beneficial to the organism and can be transmitted by horizontal gene transfer (HGT), meaning that beneficial genetic material can be shared by a bacterial population (del Solar et al., 1998). Chromids are large secondary replicons, which may carry genes that are essential for growth of the organism (Harrison et al., 2010). Phages are viruses of bacteria that can also facilitate the movement of DNA within and between bacterial genomes by HGT (Frost et al., 2005). Lysogenic phages are able to incorporate themselves into the host's genome or replicate independently in the cytosol. Phages are important drivers of genomic plasticity in prokaryotes. The genome refers to the complete set of chromosomes and genes in an individual organism including all secondary replicons.

4.2.2 Pan-genome

The term pan-genome was suggested by Tettelin et al., (2005) and comprises the core genome, genes which function as housekeeping genes, present in all strains of the bacterial species, as well as a "dispensable" or accessory genome containing genes that that are present in either a subset of strains or are strain specific (Medini et al., 2005; Tettelin et al., 2008). These accessory genes contribute to the diversity and provides functions that are not necessary to its basic lifestyle but confer selective advantages such as the ability to colonise new hosts, antibiotic resistance and niche adaptation (Tettelin et al., 2008). Non-essential genes are often located on plasmids or genomic islands. The sequence of a single genome does not reflect the true variation of a bacterial species and increased access to whole genome sequencing allows researchers to study the bacterial pan-genome.

4.2.3 Paenibacillus larvae

Paenibacillus larvae is a spore-forming, pathogenic bacterium that causes American foulbrood (AFB) of the honey bee (*Apis mellifera*). AFB is a damaging disease to the global honey trade, if left untreated infected colonies will die. In many countries the best control method is thought to be the burning of infected hives (Genersch, 2010). Antibiotics only affect the vegetative stage of the bacteria, they do not affect the infective spores, therefore if treatment is stopped the disease will return (Reybroeck et al., 2012).

Recently a new multilocus sequence-typing (MLST) scheme has been published and used to describe patterns of distribution of *P. larvae* (Morrissey et al., 2015; Chapter 2, 3). MLST is the gold standard tool for tracking disease outbreaks caused by pathogenic bacteria but a large amount of DNA sequence and gene diversity is not picked up using MLST. Therefore WGS is to needed to give higher resolution to be more effective in the characterisation of outbreak isolates and to strengthen surveillance of pathogens (Sabat et al., 2013).

4.2.4 Genome Sequencing in P. larvae

The first complete genome sequences of *P. larvae* were published in 2014; two strains belonging to different genotypes were added to Genbank (Accession numbers CP003355-CP00336 and ADFW00000000; Djukic et al., 2014). Previous to 2014 two attempts at sequencing the *P. larvae* genome were made using shotgun sequencing (Qin et al., 2006; Chan et al., 2011). However, coverage of both was low and the genome was incomplete. The authors hypothesised that the fragmentation may be due to areas of long sequence repeats that could not be bridged using shotgun sequencing (Chan et al., 2011). The new genomes confirm the existence of these genomic regions containing repeats (Djukic et al., 2014).

4.2.5 Genomic comparison of P. larvae *genotypes*

P. larvae can be classified as belonging to one of four ERIC (Enterobacterial Repetitive Intergenic Consensus sequence) types (Sharples and Lloyd, 1990; Hulton et al., 1991; Genersch et al., 2006). ERIC typing involves using a set of primers to amplify a repetitive region of the genome of varying length. The banding pattern when run out on a gel gives

the ERIC type. The four ERIC types differ in phenotype including colony and spore morphology (Genersch et al., 2006), metabolic capacity (Neuendorf et al., 2004), sporulation (Saville, 2011) and virulence (Genersch et al., 2005, 2006; Rauch et al., 2009, Poppinga et al., 2012). ERIC types I and II are commonly found in infected honey bee colonies whereas ERIC III and IV are rarer, only being discovered a handful of times (Alippi et al., 2004; Poppinga et al., 2012). Djukic et al. (2014) carried out a genome comparison using WGS of two *P.larvae* isolates belonging to ERIC I and II and identified a number of genomic differences that may account for some of the variation in virulence between these two types (Rauch et al., 2009; Genersch, 2010). The ERIC I isolate was shown to have five functional toxin genes involved in breaching the host's epithelial wall that were present in the ERIC II isolate but non-functional. It has also been demonstrated that ERIC II has a functional s-layer protein involved with adhesion to the gut wall as well as providing protective properties that is not functional in ERIC I due to point mutations.

The creation of a new MLST scheme for *P. larvae* that can differentiate amongst over 20 Sequence types (STs) has been reported (Chapter 2; Morrissey et al., 2015; pubmlst.org/plarvae). Using this scheme a diverse group of isolates was selected to sequence using WGS technology. The aim of this Chapter is to sequence and assemble 21 *P. larvae* isolates and report the general features of each. This is the first time multiple isolates belonging to the same ERIC type have been sequenced using WGS technology. Because WGS data gives more resolution than MLST alone, phylogenies made using both types of data will be compared to determine the effectiveness of MLST at describing the *P. larvae* population structure.

4.3 Methods

4.3.1 Bacterial culturing.

Bacterial strains were chosen to represent as wide a diversity as was available in the storage at the National Bee Unit (Table 4.1). Bacteria were recovered from frozen cultures stored at -80°C on Protect microorganism preservation system (ThermoFisher Scientific). Isolates were cultured on brain heart infusion (BHI) agar and incubated at 35°C for several days (de Graaf et al., 2013). DNA was extracted from *P. larvae* either by a simple chelex method or using the MoBio PowerSoil® DNA Isolation Kit. Briefly, the chelex method

was as follows: bacteria were transferred to 300µl 6% Chelex® 100 and heated to 56°C for 20 min followed by boiling for 8 min. DNA extracts were stored at −20°C until required.

4.3.2 Whole genome sequencing

For samples sequenced with Illumina MiSeq technology, DNA was quantified using the Qubit® fluorometer. Samples were prepared for sequencing using the Illumina Nextera XT DNA Library Preparation Kit. Samples sequenced on Ion Torrent technology were quantified on a NanoDrop spectrometer and underwent standard sequencing preparation.

4.3.3 Genome assembly

Resulting reads were assembled into contigs using SPAdes 3.6.2 (Nurk et al., 2013). Contigs were ordered against reference genomes using Mauve (Darling et al., 2010). Isolates of Sequence Types (STs) that group within ERIC I were ordered against the draft genome sequence of *P. larvae* DSM 25719 downloaded from Genbank, accession numbers CP003355-CP00335. Isolates of STs that group within ERIC II were ordered against the complete genome sequence of

Table 4.1 *P. larvae* isolates selected for whole genome sequencing

Isolate	ERIC type	ST	Origin	Platform
1388	i	2	Italy	MiSeq
UK21_8	i	5	England	MiSeq
UK24-10	i	5	England	MISeq
p8625 (J3)	i	6	Jersey	MiSeq
p8632	i	7	England	MiSeq
006 (p8634)	i	7	England	Ion Torrent
007 (p8476)	i	7	France	Ion Torrent
p6260	iv	8	strain collection	MiSeq
p7371	ii	10	Sweden	MiSeq
p7854	ii	11	Germany	MiSeq
p7851	ii	11*	Finland	MiSeq
p7862	i	13	New Zealand	MiSeq
p8624 (J1)	i	14	Jersey	MiSeq
004 (J4)	i	14	Jersey	Ion Torrent
005 (J5)	i	14	Jersey	Ion Torrent
p7370	i	17	Sweden	MiSeq
p8619	i	18	Scotland	MiSeq
p7847	i	18*	Austria	MiSeq
P8488	i	19	Poland	MiSeq
UK25-07	i	21	England	MiSeq
p8615	i	23	Scotland	MiSeq

STs marked with a * were originally mistyped, see appendix 2.

isolate DSM 25430, accession number ADFW00000000 (Djukic et al., 2014; Table 4.1). FastA files were visualised using MAUVE (Darling et al., 2010) and BRIG (Alikhan et al., 2011). The RAST (Rapid Annotations using Subsytems Technology) server (Aziz et al., 2008; Overbeek et al., 2014; Brettin et al., 2015) was used for automatic annotation of all assemblies. A custom BLAST (Basic Local Alignment Search Tool) library was created containing all plasmid sequences available at Genbank and all genomes were BLASTED against it to search for plasmid sequences. The Phage Search Tool (PHAST) was used to search assembled sequences for the presence of phage sequence (Zhou et al., 2011). The online tool, Resfinder (Zankari et al., 2012) was used to search the assembled genomes for acquired antimicrobial resistance genes. The Quality Assessment Tool for genome assemblies (QUAST; Gurevich et al., 2013) was used to evaluate the quality of the genome assemblies.

4.3.4 Core genes and accessory genes.

The core genes were identified using an earlier alignment of each genome against either DSM25719 (ERIC I and IV isolates) or DSM25430 (ERIC II isolates) as a scaffold. All reads were aligned to the scaffold using the BWA-MEM algorithm in BWA (Burrows-Wheeler aligner) (Li and Durbin, 2009). The complete coding sequence of the draft genome of ERIC I strain DSM 25719 was BLASTED against an ERIC II strain to identify genes that are present in both. The list of matches was then filtered to remove any with a sequence length less than 800bp or a percentage identity match lower than 90%. This reduced the list to 1181 genes, which were then blasted against the remaining 18-genome sequences to find only those present in all isolates. The list was then filtered to remove all those shorter than 900 bases and lower than 90% identity match, which resulted in 679 genes present in all 20 strains including DSM25719.

This list of 679 genes in DSM25719 was then BLASTED against each of the assembled genomes and genes that matched were taken to create an alignment. The concatenated core genes were aligned using MAUVE (Darling et al., 2010) and SplitsTree4 (Huson and Bryant, 2006) was used to create a network tree using the Neighbour-net algorithm. Networks trees attempt to provide a more ‘explicit’ representation of evolutionary history than traditional phylogenetic trees such as phylograms. They contain splits in the branches,

which represent ancestral species, the more splits in the branches the more recombination or horizontal gene transfer is likely to have taken place (Underwood et al., 2013).

The program AGEnt (Ozer et al., 2014) was used to determine the number of accessory genes present in each genome. The 679 core genes sequences of DSM25719 were input, the minimum sequence size was set to 300 bases and the minimum overlap with accessory coordinates was set to 95%.

4.4 Results

4.4.1 MLST typing mistakes.

Two of the sequenced isolates had been typed incorrectly using the MLST scheme. When the whole genome sequence was entered at pubmlst.org/plarvae, isolate p7847 was found to be ST18 rather than ST15 and isolate p7851 was found to be ST11 rather than ST12.

4.4.2 Genome features

The general features of all genomes are presented in Table 4.2. ERIC I isolates ranged in length from 4223629-4515833bp ERIC II 3858379-3926568bp and ERIC IV 4342760bp. Two isolates were removed from further analysis, isolate p8615 was of poor quality and isolate p7862 was contaminated with *Enterococcus spp*. and *Melissococcus plutonius*. All isolates contained between 5 and 10 intact phage regions (Table 4.2). All isolates except p7851 contained a vancomycin resistance operon composed of two coding regions. Plasmids were identified in 7 of the 19 genomes tested. These plasmids are similar to pPLA1_10 and pPLA2_10 identified in Djukic et al. (2014). Six of the seven plasmids were identified in ERIC I isolates (Table 4.2). The number of core and accessory open reading frames are presented in Table 4.3. The number of core open reading frames differs from 691 to 797 and range in length from 2938539bp to 3595436bp. The ERIC II isolates have shorter accessory genomes than ERIC I isolates.

Table 4.2 General features of *P. larvae* genomes

Isolate	ST	No. of reads	No. of contigs	N50 contig size	Size (bp)	G+C content (%)	No. of CDS	Plasmid present	Plasmid length	No. of plasmid CDS	No. of phage regions	No. of intact phages	Resistance genes
1388	2	1.3m	298	48617	4335448	44.12	5113	No	NA	NA	19	6	VanE
UK21_8	5	7.84m	351	60541	4448425	44.03	5441	No	NA	NA	15	5	VanE
UK24_10	5	2.64m	342	39740	4369910	44.10	5140	No	NA	NA	17	8	VanE
p8625 (J3)	6	877.82k	271	51901	4430755	44.04	5336	Yes	9390	19	18	9	VanE
p8632	7	1.79m	268	47313	4420656	44.07	5282	Yes	8591	15	21	9	VanE
006 (p8634)	7	1.22m	338	39704	4323692	44.08	5598	Yes	8682	15	19	9	VanE
007 (p8476)	7	1.35m	373	40561	4413320	43.99	5621	No	NA	NA	18	9	VanE
p6260	8	2.69m	270	57387	4342760	44.21	5018	No	NA	NA	18	10	VanE
p7371	10	1.47m	359	24680	3926568	44.73	4505	Yes	9669	23	17	5	VanE
p7854	11	2.66m	333	25559	3858379	44.68	4725	Partial	NA	NA	22	7	VanE
p7851	11*	2.07m	355	25559	3905216	44.61	4785	No	NA	NA	19	5	none
p7862	13	1.5m	NA	NA	NA	NA	NA	NA	NA		NA	NA	NA
p8624 (J1)	14	846.14k	286	50232	4515833	44.02	5379	Yes	8516	23	17	8	VanE
004 (J4)	14	909.47k	327	49651	4331474	44.03	5492	Yes	8666	21	18	8	VanE
005 (J5)	14	2.28m	340	40650	4339667	44.04	5500	Yes	8682	20	17	7	VanE
p7370	17	2.98m	235	54812	4263884	44.14	4952	No	NA	NA	16	5	VanE
p8619	18	2.88m	141	85269	4264001	44.04	5022	No	NA	NA	13	6	VanE
p7847	18*	3.6m	170	77498	4223629	44.13	4955	No	NA	NA	11	7	VanE
P8488	19	1.28m	257	50059	4349350	44.06	5233	No	NA	NA	17	6	VanE
UK25_07	21	3.6m	372	62102	4502918	44.08	5507	No	NA	NA	14	7	VanE
p8615	23	93.11k	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA

Rows coloured light grey are ERIC II isolates, the row coloured dark grey is an ERIC IV isolate.

Chapter 4

Table 4.3 Core and accessory genomes.

Isolate	ST	Core ORFs	Accessory ORFs	Accessory length (bp)
1388	2	698	4213	3402724
UK21_8	5	706	4330	3491220
UK24_10	5	701	4157	3425364
p8625 (J3)	6	710	4391	3513067
006 (p8634)	7	797	4495	3375837
007 (p8476)	7	757	4475	3437094
p8632	7	694	4357	3503053
p6260	8	703	4089	3385325
p7371	10	709	3673	2994674
p7851	11	702	3767	2985293
p7854	11	701	3710	2938539
004 (J4)	14	787	4435	3393530
005 (J5)_	14	766	4425	3391494
p8624 (J1)	14	697	4479	3595436
p7370	17	692	4110	3335135
p7847	18	698	4080	3296537
p8619	18	691	4192	3346411
p8488	19	712	4192	3397143
UK25_7	21	715	4322	3515163

ORF refers to open reading frames longer than 300bp. Rows coloured light grey are ERIC II isolates, the row coloured dark grey is an ERIC IV isolate.

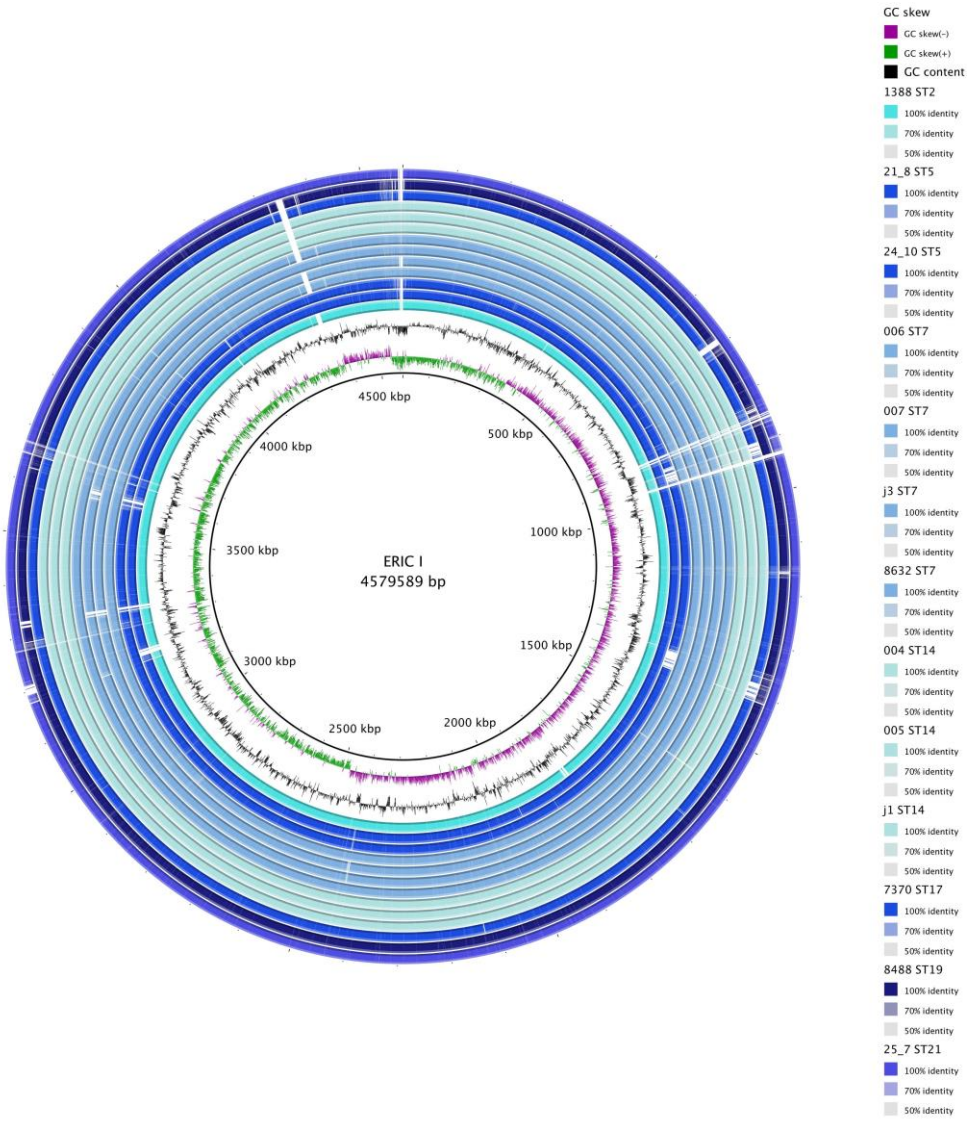

Figure 4.1 Circle plots showing the genome sequences of ERIC I isolates compared to reference sequence DSM 25719.

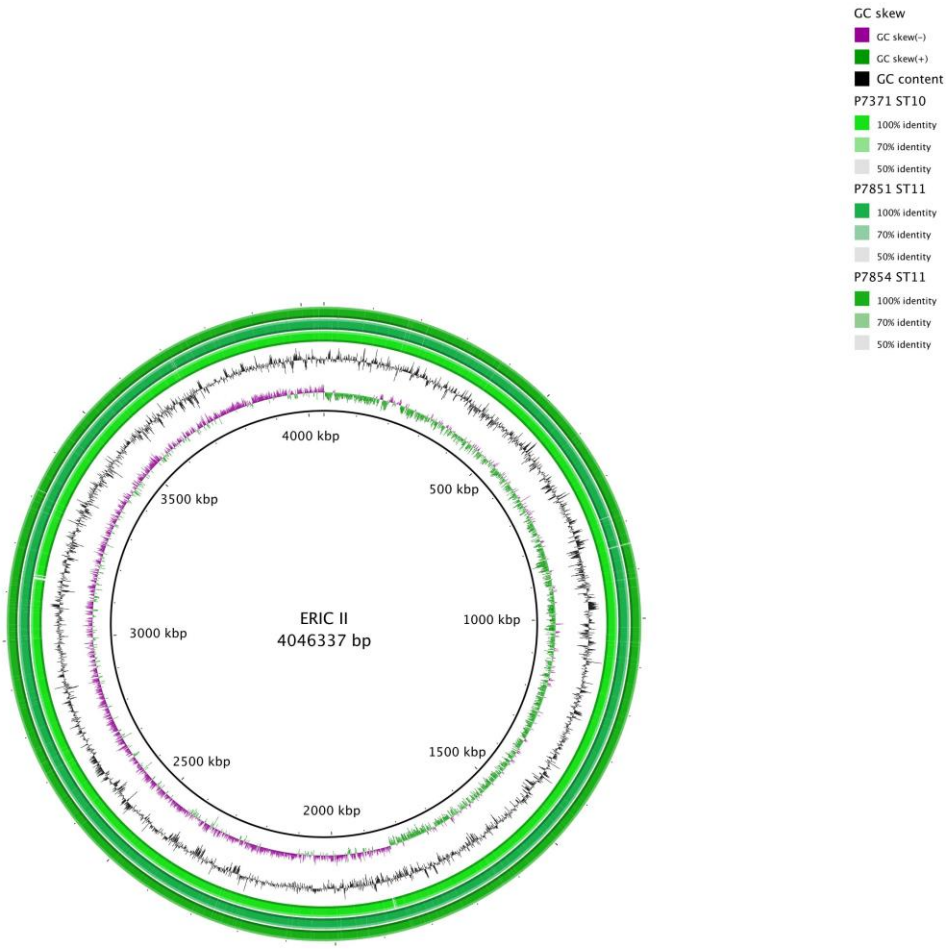

Figure 4.2 Circle plots showing the genome sequences of ERIC II isolates compared to reference sequence DSM 25430**.**

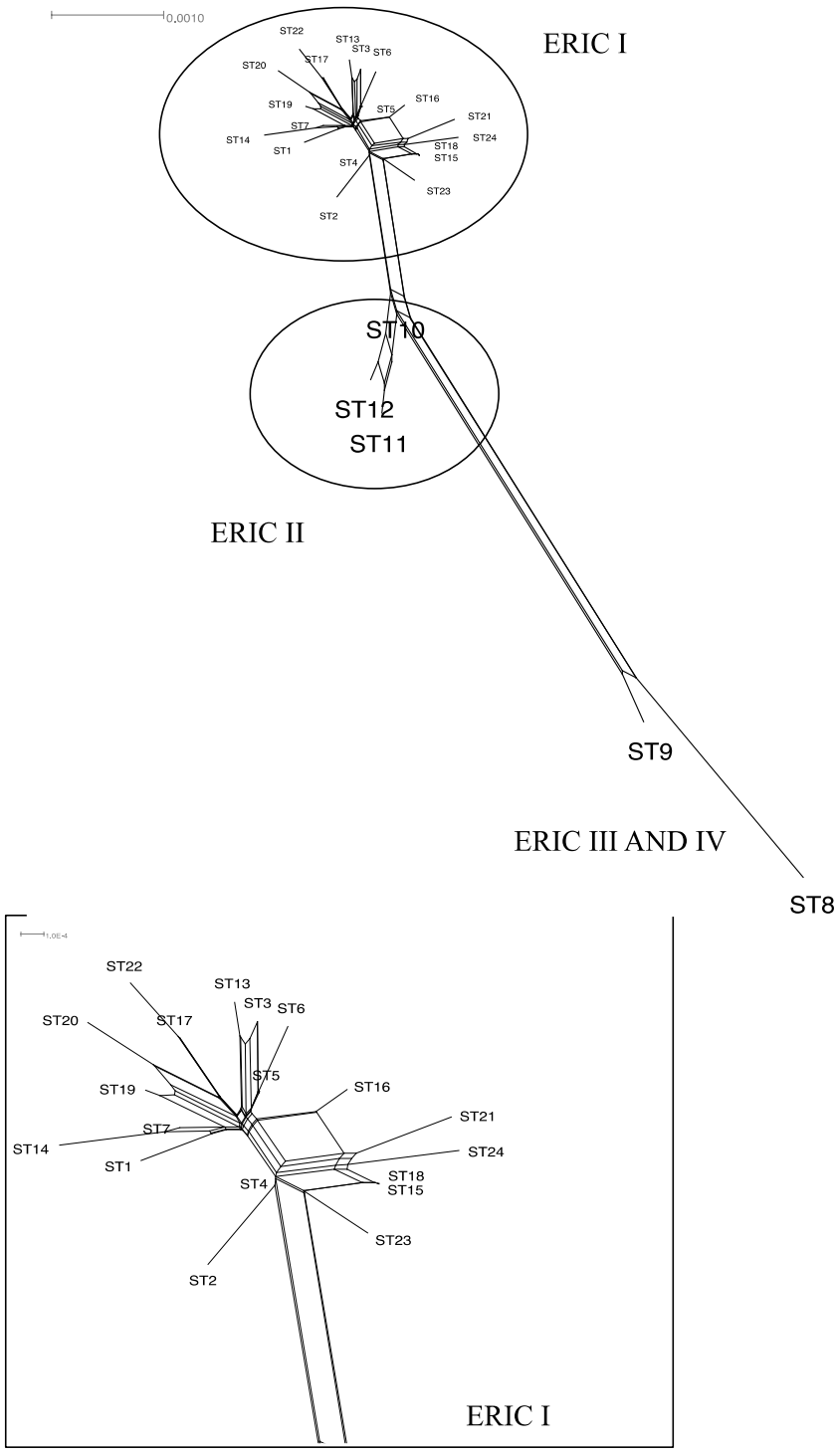

Figure 4.3 Neighbour net tree using concatenated MLST allele sequence data. The tree shows the relationship between all 24 identified STs.

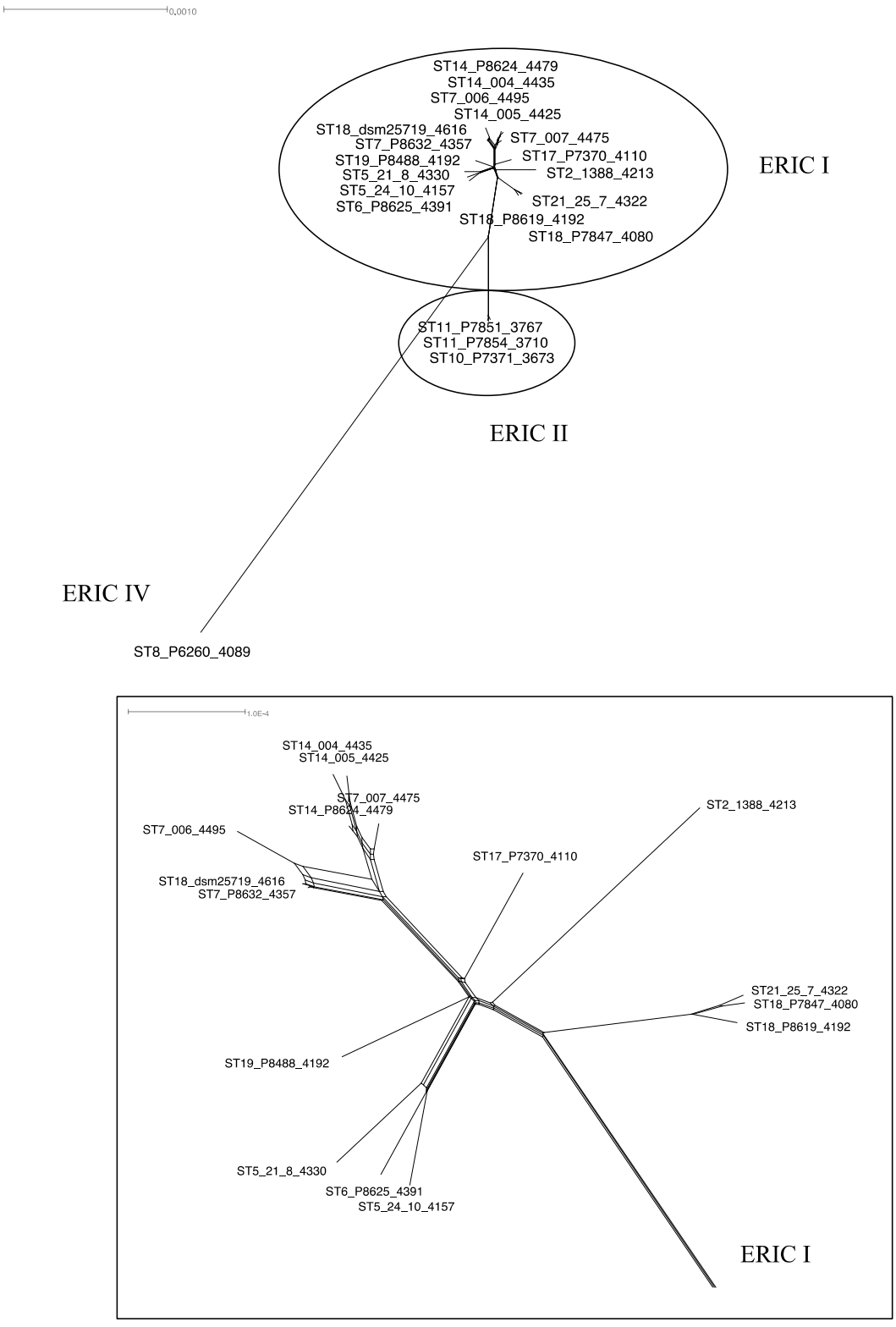

Figure 4.4 Neighbour net tree using core gene sequence data. The tree shows the relationship between the 19 genome-sequenced isolates. The labels contain the ST, Isolate name and number of non-core genes.

4.4.3 Phylogenies.

The 19 assembled genomes tested all contained the 679 core genes. The core gene neighbour network tree demonstrated the same pattern of distribution amongst the ERIC genotypes as the MLST neighbour network, with the ERIC types forming separate groups. Within ERIC I, although all STs are not present in the core gene analysis, the STs form the same groups as in the MLST network (Figures 4.2 and 4.3). ST18 groups with ST21, ST7 groups with ST14 and ST5 groups with ST6. The exception to this is DSM 25719, which is typed as ST18 but groups with ST7 isolates. The core gene tree also gives further differentiation amongst strain types for example ST7 (007) groups with ST14s rather than with the other ST7s and ST6 branches off from one of the two ST5s.

4.5 Discussion

P. larvae is the causative agent of American foulbrood, a destructive disease of honey bees found globally. Due to the creation of the new *P. larvae* MLST scheme this is the first time WGS has been used to compare multiple STs belonging to the same genotype. The new MLST scheme is a useful tool to determine strain types and standardise different studies, however the decreasing cost of WGS means that this technology will soon be regularly used to describe disease outbreak. These results demonstrate that the MLST scheme accurately represents the broad diversity within and between genotypes of *P. larvae* but in finer detail between the STs there are some differences not picked up by the MLST scheme alone.

The MLST typing mistakes in this chapter were due to human error. Typing involves sequencing each isolate seven times. Software is available to type MLST sequence data for example BioNumerics (Applied Maths) can automatically analyse trace sequence files avoiding human error and pubmlst.org/plarvae can also be used to type sequence data including WGS data.

4.5.1 Phylogeny

The core gene tree broadly gave the same pattern of distribution of sequence types to the MLST tree but within the ERIC I group there were some differences. Of the three ST18 isolates included, two group together and form a clade with ST19 and 21 and the other

ST18 (DSM 25719) grouped with two ST7 isolates from England (Figure 4.2). The length of the branch between the two ST18 groups demonstrates that they are diverse. STs 7 and 18 are triple locus variants using the MLST scheme (Morrissey et al., 2015; Chapter 2). The two ST7 isolates (006 and p8632) originate from a single disease cluster in England (Chapter 3; Table 3.1). Both of these isolates contain plasmids similar to the pPLA1_10 plasmid described in DSM25719 (Djukic et al., 2014; Chapter 5). It is likely that these types have had some HGT in the past with DSM 25719. Djukic et al. (2014) propose that the large numbers of mobile genetic elements and prophages in ERIC I and II genomes suggest a high degree of genome plasticity and genome rearrangements (Djukic et al., 2014). ST6 is shown to be close to ST5 in the MLST tree. The core gene tree may demonstrate a split in ST5, with ST6 branching off from one group. ST5 is the most common ST globally (Morrissey et al., 2015) therefore it may be useful in the future to split up the group with an added locus. The core gene tree may be a starting point to finding a split in the ST5 group.

4.5.2 Accessory genes

We have identified the number of genes that are non-core in each isolate (Figure 4.4, Table 4.3) this includes the discovery of plasmids in seven isolates. The number of core genes differs from the 279 identified in DSM25719, this may be due to gene fragmentation. Although genome lengths are similar or shorter than those found in Djukic et al (2014) the total number of ORFs are slightly higher (~ 5000 in ERIC I compared to 4868 in Djukic and ~4000 compared to 3928 in Djukic; Table 4.2) suggesting some gene fragmentation. The number of core genes identified is low in comparison to studies in the closely related *Bacillus cereus* species complex (Tettelin et al., 2005; Fang et al., 2011). Djukic et al. (2014) predict a high number of core genes between ERIC I and II isolates and perhaps the strict method of discovery used in this study has underestimated the number of common genes. It is known that all *P. larvae* strains identified exploit the same niche, infecting honey bee larvae. However, it has been shown that ERIC I and II likely have different methods of infection (Poppinga et al., 2012; Djukic et al., 2014) and further analysis of the functional groups of the accessory genes may give further information on these differences. Djukic et al. (2014) discovered a number of regions which differed between ERIC I and II, these included putative transport proteins and genes that code for toxins that are unique to ERIC I as well as three ERIC II specific regions which included putative

proteins. It is known that some differences in method of infection are not associated with accessory genes but with core genes that have lost function, for example the genes coding for the S-layer protein are present in both ERIC I and ERIC II strains but are only functional in ERIC II due to a frame-shift mutation in ERIC I types (Poppinga et al., 2012). Further work could look into the differences between and among isolates of ERIC I and II to determine further regions that differ and to discover whether these have phenotypic effects.

4.5.3 Phages

The number of complete phages and their positions differed both within and between STs (Table 4.2). Phages are known to alter competition among bacterial strains, and maintain bacterial diversity. They are central to the ecology and evolution of microbial communities (Koskella and Brockhurst, 2014). Phages can also encode virulence factors in pathogenic bacteria, for example *Vibrio cholera* and *Clostridium botulinum* both require phage-encoded toxins to infect hosts (Brüssow et al., 2004). The three ERIC II isolates contained between 5 and 7 intact phage regions, this is in contrast to Djukic et al. (2014). They found that the ERIC II isolate sequenced contained fewer phage regions than the ERIC I isolate, and they reported that there were no complete ERIC II phage regions. Djukic et al. (2014) reported that the ERIC II isolate sequenced contained more mobile genetic elements in total than ERIC I. Future work could compare the total number of mobile genetic elements in each ST.

4.5.4 Plasmids

Accessory genes can have large effects on the phenotype for example in the *Bacillus cereus* species complex *B. cereus, B. thuringiensis* and *B. anthracis* are closely related in terms of core genes however they are phenotypically different, with most of the genes responsible found on plasmids (Helgason et al., 2000; Rasko et al., 2005). The virulent strains *of B. anthracis* carry two plasmids coding for the anthrax virulence factors (Okinaka et al., 1999) and *B. thuringiensis* contains unique genes coding for insecticidal toxins, usually present on plasmids, if these plasmids are lost *B. thuringiensis* can no longer be distinguished from *B. cereu*s (Helgason et al., 2000; Rasko et al., 2005). Plasmids were identified in seven of the 19 *P. larvae* isolates sequenced. Previously, several plasmids have been identified in *P. larvae* (Alippi et al., 2007, 2014; Djukic et al.,

2014) including several that have been shown to confer resistance to antibiotics commonly used in some countries to treat AFB (Reybroeck et al., 2012). The plasmids identified here are similar to those identified by Djukic et al. (2014). Djukic et al. (2014) identified two plasmids pPLA1_10 in the ERIC I strain DSM25719 and pPLA2_10 in the ERIC II strain DSM25430. Six of the seven plasmids identified in this study are found in ERIC I strains. This contrasts with the results of Djukic et al. (2014). 65 ERIC I strains of *P.larvae* were tested for the presence for the plasmid and no other cases were found Djukic et al. (2014). The three ERIC I STs that were found to contain plasmids (ST6, ST7 and, ST14) are all rare STs (Morrissey et al., 2015) and are likely to share a common ancestor due to the geographically small regions in which they originated (South West England and Jersey in the Channel Isles).

The whole genome sequencing of bacterial strains causing outbreaks is likely to be more common in the future. WGS gives more resolution than typing schemes such as MLST and can be used to track disease outbreaks over short timescales where only a few single nucleotide polymorphisms (SNPs) differ amongst strains (Harris et al., 2013). However MLST is a useful tool to use in conjunction with WGS as it enables studies to be standardised. Databases such as BIGSdb (Jolley and Maiden, 2010) make MLST identification within WGS data easy and accurate. This chapter has shown that the *P. larvae* MLST scheme predicts the broad patterns of relatedness amongst STs. WGS gives more resolution and can give further insight into differences between ST, e.g. the presence of plasmids. This is the first time multiple genomes from within genotypes have been compared and it has given further insight into the pan-genome of *P. larva*e with the discovery of intact phage regions in ERIC II and plasmids found in multiple ERIC I isolates. The next step is to link the genetic and phenotypic differences amongst strains to determine whether certain strains are more virulent than others, this could lead to improved control and management of this damaging disease.

5 Description and phylogeny of *Paenibacillus larvae* plasmids discovered in rare strains

5.1 Abstract

Paenibacillus larvae is the causative agent of American foulbrood, a damaging brood disease of the honey bee. Plasmids were identified in seven of 20 *P. larvae* genomes analysed. These plasmids are similar to the recently discovered *P. larvae* plasmids pPLA1_10 and pPLA2_10. Six of the plasmids are found in rare *P. larvae* sequence types from a relatively small geographic area. The plasmids all contain a replication initiation factor and a putative toxin. Maximum likelihood trees demonstrated that plasmids did not group by geographic location or by sequence type unlike the core genome, which grouped, by sequence type. This result suggests horizontal transfer of the plasmids. Plasmids can confer benefits to their host bacteria such as antimicrobial resistance and virulence factors and can move between strains and species via horizontal gene transfer therefore, knowledge and documenting of plasmids is important in epidemiology.

5.2 Introduction

Paenibacillus larvae is a gram-positive bacterium that is the causative agent of AFB, a deadly disease of honey bees found worldwide. The control method favoured by many European countries is the eradication of infected hives and it is therefore a costly and damaging disease. *P.larvae* can be categorised as belonging to one of four groups, known as ERIC types, based on phenotypic and genetic factors (Genersch, 2010). Only ERIC I and II are commonly found in diseased hives and are therefore the most economically important in terms of disease outbreaks (Poppinga et al., 2012). *P. larvae* is a Gram-positive bacterium belonging to the phylum Firmicutes. Firmicutes contains a large number of bacteria pathogenic to animals and humans (Lanza et al., 2015). Plasmids are important drivers of evolution in Firmicutes and can transfer antimicrobial resistance and virulence factors between strain and species via horizontal gene transfer (HGT) (Lanza et al., 2015). I discuss the discovery of plasmids in seven of 20 analysed *P. larvae* genomes belonging to both ERIC I and II.

5.2.1 Plasmid functions

Plasmids are extra-chromosomal elements able to replicate independently, and to move between bacterial strains via horizontal gene transfer (HGT) (Frost et al., 2005). Plasmids can be categorised in two broad classes, conjugative and non-conjugative. Conjugative plasmids contain genes that code for "sex" pili, which allows their transfer to a new host (Petersen, 2011). Non-conjugative and mobilizable plasmids often contain a minimal set of mobility (MOB) genes and require a helper conjugative plasmid to be present in order to transfer by conjugation. These are usually small (15kb) and have a high copy number. Conjugative, self transmissible plasmids tend to be large (30kb) and as well as the MOB region also contain a type four secretion system (T4SS) that allows the use of the mating channel (Garcillán-Barcia et al., 2009). Plasmids can provide benefits to the host bacterium such as resistance against antibiotics, virulence and additional metabolic capabilities.

The *Bacillus cereus* complex contains several closely related species that are genetically similar but that occupy phenotypically different niches. Their diverse functions are thought to have arisen from a common ancestor subjected to HGT and the movement of genetic material due to mobile DNA elements (Auwera et al., 2005). For example *Bacillus anthracis* causes Anthrax due to the presence of two plasmids pX01 and pX02, which contain a number of genes essential for toxin production and survival (Mock and Fouet, 2001). Another member of this group *Bacillus thuringeinsis* can infect a wide range of insects with a disease similar to AFB due to the presence of plasmid genes coding for acquired δ-endotoxin crystals genes (Auwera et al., 2005). It would be reasonable to predict that the closely related *P. larvae* obtained its ability to cause disease in honey bees from plasmids but, to date, plasmids have been detected in only a few strains of *P. larvae*.

5.2.2 Plasmids in **P. larvae**

Several plasmids conferring resistance against tetracycline have been identified in *P. larvae* (Murray and Aronstein, 2006; Murray et al., 2007; Alippi et al., 2014). Oxytetracycline (OTC) is an antibiotic widely used in North and South American countries to prevent and control American foulbrood (AFB). OTC resistance is uncorrelated to bacterial haplotype suggesting HGT of resistance plasmids (Evans, 2003). The plasmid pMA67 identified by Murray et al. (2007) is a short circular plasmid (5kb) that replicates

using the rolling circle method. Tetracycline resistance is conferred by the *tetL* gene (Murray and Aronstein, 2006). Alippi et al. (2014) have identified three new plasmids in three North American strains of *P.larvae* pPL373, pPL374 and pPL395. These new plasmids are highly similar to pMA67 (99%) as well as to plasmids conferring tetracycline resistance in five genera of gram-positive bacteria found in a variety of ecological niches (Alippi et al., 2014). Alippi et al. (2014) also identified but did not sequence two larger plasmids (~8kb) in strains PL373 and PL374, which did not seem to be present in PL395. The authors hypothesised that the larger plasmids, present in strains PL373 and PL374 facilitated conjugation for the smaller mobilizable plasmids (Alippi et al., 2014).

Recently two new circular plasmids of similar length (9kbp) have been identified in *P.larvae*, pPLA1_10 and pPLA2_10 (Djukic et al., 2014). They are respectively found in an ERIC I strain (DSM 25719) and an ERIC II strain (DSM 25430). These plasmids do not contain the *tetL* gene that confers resistance to tetracycline but do contain a putative replication factor (REP) gene. Djukic et al. (2014) screened an international collection of 65 ERIC I strains and 30 ERIC II and found no other strains of ERIC I contained the pPLA1_10. It was hypothesised that the pPLA1_10 plasmid was strain specific but not genotype specific (Djukic et al., 2014). Previously plasmids of similar lengths have been identified in *P. larvae* but not characterised (Bodorová-Urgošíková et al., 1992; Neuendorf et al., 2004).

21 isolates of *P. larvae* were sequenced using whole genome sequencing technology (Ion Torrent and MiSeq; Chapter 4, Table 4.1). The *P. larvae* MLST scheme (Morrissey et al., 2015) was used to select a diverse group of Sequence Types (STs) from the culture collection at Fera, including several isolates from the UK. Of the 20 isolates successfully sequenced, seven contained plasmids. Of these, four were from an outbreak of AFB in Jersey, a small island in the English Channel and two more were identified in a related sequence type from an outbreak in the UK. In this chapter I discuss the functions of these plasmids as well as their phylogeny and possible routes of movement between these sites.

5.3 Methods

P. larvae isolates from the UK and Jersey were cultured from Lateral Flow Devices (LFDs) used by bee inspectors in the field (Table 4.2). Isolates were cultured on brain heart

infusion (BHI) agar and incubated at 35°C for several days (de Graaf et al., 2013). DNA was extracted from *P.larv*ae either by a simple chelex method or using the MoBio PowerSoil® DNA Isolation Kit. Briefly, the chelex method was as follows: bacteria were transferred to 300µl 6% Chelex® 100 and heated to 56°C for 20 min followed by boiling for 8 min. DNA extracts were stored at −20°C until required.

Isolates were sequenced using either Ion Torrent technology or MiSeq as in Chapter 4 (Table 4.1). All reads were assembled de novo using SPAdes 3.6.2 (Nurk et al., 2013). Resulting contigs were ordered against reference genomes using Mauve (Darling et al., 2010). Isolates of Sequence Types (STs) that group within ERIC I were ordered against the draft genome sequence of *P. larvae* DSM 25719 downloaded from Genbank, accession numbers CP003355-CP00335. Isolates of STs that group within ERIC II were ordered against the complete genome sequence of strain DSM 25430, accession number ADFW00000000 (Djukic et al., 2014; Table 4.1)..

The open source tool SAMtools (Li et al., 2009) was then used to sort and index the reads and convert them from SAM to BAM format so that they could be visualised. BAM files were visualised using Tablet (Milne et al., 2013). STs of assembled genomes were checked using the pubMLST *P. larvae* database

Table 5.1 General features of sequenced plasmids

Isolate	ERIC type	ST	Origin	Year (field)	Technology	No. of reads (k)	Length (bp)	GC content	No. of features	pPLA1_10 % Identity, Coverage	pPLA2_10 % Identity, Coverage
J3 (p8625)	i	6	Jersey	2010	MiSeq	6.84	9390	36.86	19	90, 94	88, 89
006 (p8634)	i	7	England	2006	Ion Torrent	16.11	8682	37.5	15	91, 89	90, 85
p8632	i	7	England	2006	MiSeq	65.3	8591	37.51	15	91, 88	87, 84
p7371	ii	10	Sweden	unknown	MiSeq	43.92	9669	37.49	23	89, 93	99, 100
005	i	14	Jersey	2010	Ion Torrent	11.82	8682	37.52	20	91, 89	90, 85
004	i	14	Jersey	2010	Ion Torrent	6.11	8666	37.5	21	91, 89	90, 84
J1 (P8624)	i	14	Jersey	2010	MiSeq	4.66	8516	37.52	23	90, 87	87, 83

(pubmlst.org/plarvae). Whole genome sequences were submitted in fastA format. All assembled sequences were blasted against the Pathosystems resource integration centre PATRIC (Wattam et al., 2013) database of bacterial plasmid DNA to test whether plasmids were present. The genomes that contained plasmids were then reassembled using only either plasmid pPLA1_10 (ERIC I) or pPLA2_10 (ERIC II) as scaffolds using SAMtools (Li et al., 2009). Any read that did not align to the scaffolds were filtered out and removed. All reads that matched the plasmid sequences were then assembled *de novo* using Newbler 2.5 (Roche). This resulted in the complete plasmid sequence available in each sample. The plasmid sequences were then submitted to the RAST (Rapid Annotations using Subsytems Technology) server (Aziz et al., 2008; Overbeek et al., 2014; Brettin et al., 2015) for automatic annotation. The coding sequences were then further analysed using InterProScan (Mitchell et al., 2015) to identify the protein families present. BRIG was used to visualise the plasmid sequences (Alikhan et al., 2011). Core genomes were identified using the method in Chapter 4. Maximum likelihood trees were created using MEGA 6.06 (Tamura et al., 2013) with 1000 bootstrap repetitions.

5.4 Results

Plasmids were identified in seven of the 20 genome sequences analysed using the PATRIC BLAST tool (Wattam et al., 2013). Six plasmids were identified in ERIC I isolates and one was identified in an ERIC II isolate (Table 5.1). The ERIC II plasmid is 99% similar to the previously identified pPLA2_10 (Djukic et al., 2014). The remaining six plasmids are most similar to pPLA1_10 (Table 5.1**)** although pPLA1_10 and pPLA2_10 are highly similar (Djukic et al., 2014). Interestingly, the *P.larvae* isolate DSM 25719, in which pPLA1_10 was discovered in ST18 but the ST18 isolates included in this study (p7847 and p8619; Chapter 4) do not contain plasmids. The ERIC I isolates containing plasmids belong to three rare STs (6, 7 and 14; Morrissey et al., 2015). The six ERIC I plasmids vary in length between 8516bp and 9390bp **(**Table 5.1; figure 5.1**)** whereas the ERIC II plasmid is 9669bp in length, equal to pPLA2_10.

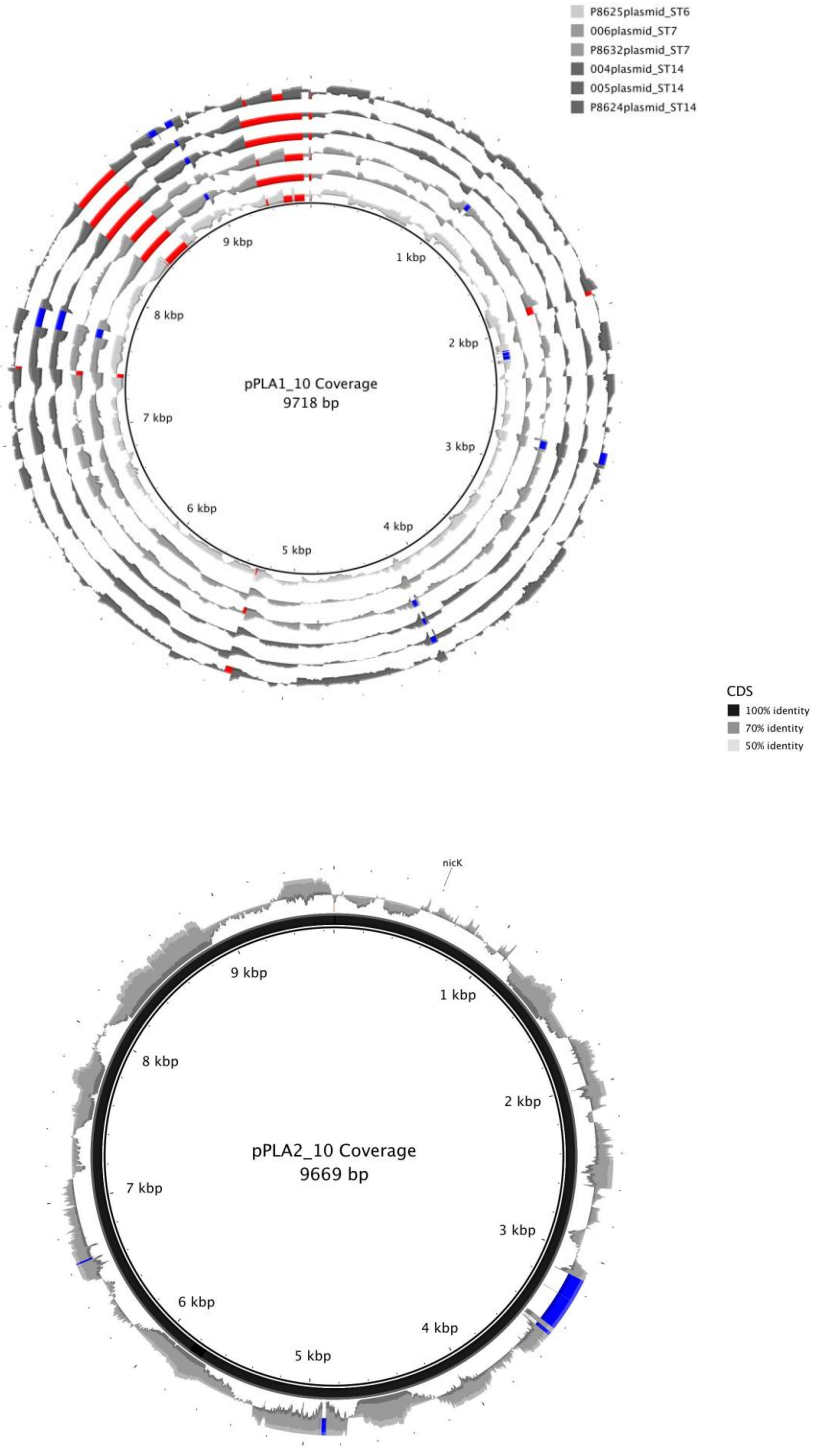

Figure 5.1 Circle diagram of plasmids. ERIC I plasmids above, ERIC II plasmids below. Regions with coverage more than one standard deviation from the mean coverage are coloured blue regions, regions with coverage less than one standard deviation from the mean are coloured red.

5.4.1 Annotation

The RAST server (Aziz et al., 2008) identified several coding regions (CDS) in each plasmid the majority of which were hypothetical proteins. InterProScan (Mitchell et al., 2015) identified family or domain level results for only three CDS for each plasmid. A replication initiation factor was identified in each of the plasmids. This had been previously identified in pPLA1_10 and pPLA2_10 (Djukic et al., 2014). Further CDS were classified as belonging to the Fibronectin type 3 domain and, a Zona occludens Toxin (ZOT). Fibronectin type 3 domain proteins are involved with cell adhesion, cell morphology, thrombosis, cell migration, and embryonic differentiation and is also found in the *Bacillus cereus* group. ZOTs are toxins found in pathogenic bacteria and allow movement through the host's gut wall. There are also closely related proteins present in members of the *Bacillus cereus* complex and it is present in both pPLA1_10 and pPLA2_10 as well as in the genome of *P. larvae* (Marchler-Bauer et al., 2014).

5.4.2 Comparison

All of the new ERIC I plasmids except J3 (ST6) contain a two short deletions 131bp and 246 bp in length, relative to pPLA1_10. Both regions occur in coding sequence. Based on protein BLASTs of aligned pPLA1_10 sequence the shorter is part of a sequence coding for a predicted cell-wall anchored protein and the longer is a section of sequence coding for a major facility transporter (Marchler-Bauer et al., 2014). Plasmids J1, J3 and p8632 (ST14, ST6 and ST7) all contain a 146bp deletion in the sequence coding for the ZOT type protein.

5.4.3 Phylogeny

The maximum likelihood (ML) tree using only plasmid sequences indicated that the ERIC I plasmids discovered in this study are more closely related to one another than to pPLA1_10. They did not cluster together based on ST or by geographic location. The ERIC II plasmid (p7371) formed a separate clade in the

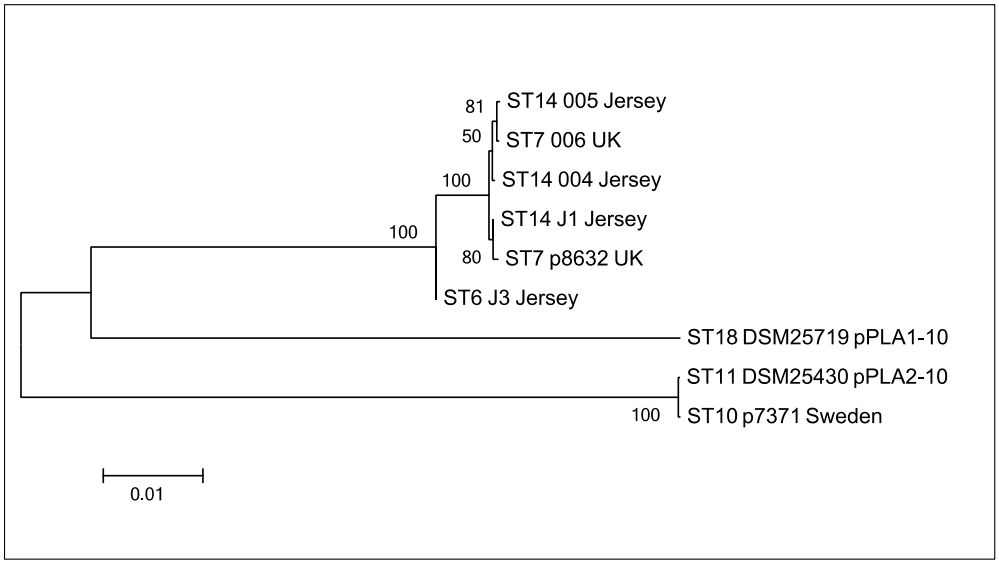

Figure 5.2 Maximum Likelihood tree of plasmid sequences. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches.

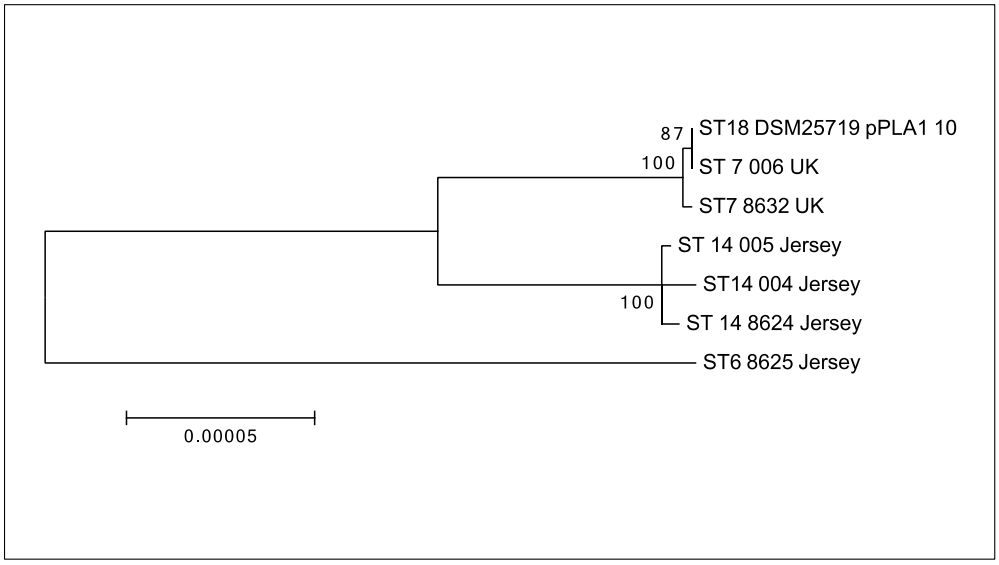

Figure 5.3 Maximum Likelihood tree of isolates containing plasmids using core gene sequences. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches.

tree with pPLA2_10 whereas pPLA1_10 grouped with the ERIC I plasmids but was basal to them.

Isolates group by ST in the core gene ML tree, with the exception of DSM 25719 (ST18). This isolate groups with isolates belonging to ST7 (Chapter 4), which is not congruent with the plasmid tree (Figures 5.2 and 5.3).

5.5 Discussion

We have discovered plasmids in seven out of 20 *P. larvae* isolates that were genome sequenced, including for the first time, in isolates from the UK. Of these plasmids, one, is extremely similar to the previously reported pPLA2_10 and the remaining six were found in STs previously unknown to harbour plasmids. This further contradicts the previous assumption that plasmids are specific to ERIC II (Neuendorf et al., 2004; Djukic et al., 2014).

Our ERIC II plasmid is highly similar to pPLA2_10 and is found in a ST10 isolate. DSM 25430 in which the plasmid was first identified is ST11 (Djukic et al., 2014). We did not find plasmids in the other sequenced ERIC II isolates of different STs. However a plasmid of similar length (pPLL9.4) has previously been identified in ERIC II isolates that are of both ST10 and ST11 (Neuendorf et al., 2004; Djukic et al., 2014; Morrissey et al., 2015) this suggests it is not strain specific.

The six ERIC I plasmids identified in this study are similar to pPLA1_10 and pPLA2_10. All seven contain a replication gene, toxin and Fibronectin genes similar to those identified previously in *Bacillus* species (Marchler-Bauer et al., 2014). BLAST results demonstrated that these plasmids do not appear to be closely related to others in *Bacillus*. This contrasts with the *P. larvae* plasmids conferring tetracycline resistance, which are highly similar to those occurring in several genera of gram-positive bacteria (Alippi et al., 2014) and confer an advantage to their hosts. The tetracycline resistance plasmids identified in *P. larvae* have been shown to replicate by the rolling circle method (Alippi et al., 2014). The plasmids discovered in this study contain a replication initiation factor (REP gene) but none of the other genes required for rolling circle method. It is unsurprising that we did not discover plasmids similar to those identified by Murray and Aronstein (2006) that confer

tetracycline resistance in any of our sequenced European isolates. Oxytetracycline is not commonly used to treat AFB in Europe in contrast to North and South America where these resistance plasmids have been found (Murray and Aronstein, 2006; Murray et al., 2007b; Reybroeck et al., 2012; Alippi et al., 2014).

Breaching the epithelial wall is a crucial step in pathogenesis by *P. larvae* and it has been suggested that this may be accomplished using toxins (Yue et al., 2008; Djukic et al., 2014). We identified a protein sequence associated with the Zona occludens toxin, a toxin associated with helping pathogenic bacteria pass through the host's gut wall (Fasano et al., 1991). Protein BLASTs demonstrated that similar sequences are also found in the genome of *P. larvae*. It is clear that these pPLA-type plasmids are not essential for host gut invasion as it has not been identified in all strains. However possessing multiple copies of toxin producing genes could confer a benefit as in *E. coli* where two copies of a toxin gene allow antigenic heterogeneity of a particular pathogenic strain (Schmitt et al., 1991). It is unclear whether the plasmids discovered in this study provide any benefit or cost to *P. larvae.* Further work is required to determine the costs and benefits for the bacteria of harbouring these plasmids.

5.5.1 Movement and Phylogeny

pPLA type plasmids similar to the ones discovered in this study are not regularly found in ERIC I isolates. Djukic et al. (2014) screened an international collection of 65 ERIC I strains for evidence of the pPLA1_10 plasmid but found no other examples. We identified the ERIC I plasmid in six isolates that are of three rare ST from a relatively small geographic area. Two are from a single disease cluster in the South of England (Chapter 3) and the remaining four are from an outbreak on a small island (14km by 8km) in the English Channel. STs 7 and 14 are known to be closely related to one another (Morrissey et al., 2015; Chapter 4), which suggests that these plasmids have a common ancestor. The ML trees demonstrate that, although STs group together when the bacterial core genes are analysed, the plasmid sequences do not cluster by sequence type or by geographic location. This demonstrates that the plasmid has moved independently of the host bacterium via HGT.

The Jersey samples were collected from an outbreak that occurred on the island in 2010 There is, however there is no previous data on infection in Jersey so it is not known whether these samples were from infections prior to 2010. Although the UK isolate is from an outbreak in 2006, the Jersey isolate ST6 appears to be basal in the ML tree, suggesting the plasmids migrated out from Jersey. All four isolates sequenced from Jersey contained plasmid sequences, suggesting plasmids may be common on the island. Further testing will demonstrate whether this is true.

Plasmids move between host bacterial cells due to horizontal gene transfer (HGT). The gut microbiota of adult honey bees have been shown to be advantageous locations for HGT, for example it has been demonstrated that a single bee can harbour a diverse set of *tet* genes, involved in tetracycline resistance (Tian et al., 2012; Alippi et al., 2014). It is likely that plasmid transfer could occur in the guts of bees in a colony that is then moved by humans to a new location. The import and export of bees to the island of Jersey is likely to have aided in the spread of the plasmid, however it may also be possible for a honey bee colony to fly from France to the island (Datta et al., 2013).

In 2007 the Cornish AFB disease cluster (E9), including at least two plasmid-carrying outbreaks was eradicated (Chapter 3) by UK statutory control methods. These methods involve burning infected hives and increasing inspection in the surrounding area (Wilkins et al., 2007). The MLST scheme demonstrates the effectiveness of the eradication protocol in the UK, which eliminated the possible emerging threat from a new plasmid. After 2007 the next outbreak occurred in 2011 and was of a different ST.

Because of the rarity of plasmids in ERIC I, the rarity of the STs containing the plasmids and the small geographic area in which they were discovered, the six ERIC I plasmids are assumed to have a common ancestor. The plasmids are not MLST ST specific unlike the bacterial core genes. Horizontal gene transfer can occur in bee guts or the environment and the movement of honey bee colonies could have facilitated spread of these plasmids. Plasmids can harbour genes that give resistance to antibiotics, or virulence factors can give bacteria advantages, therefore knowledge and recording of plasmids is important for epidemiology.

6 General Discussion

6.1 Background

American foulbrood is the most destructive brood disease of honey bees (*Apis mellifera*) worldwide and is caused by the Gram-positive, spore-forming bacterium *Paenibacillus larvae. P. larvae* infection will lead to the death of the colony, once infected larvae usually die within 3 to 12 days (Genersch et al., 2005; Rauch et al., 2009). *P. larvae* spores are able to remain infective for more than 35 years in old hives and are resistant to extremes of temperature (Hasemann, 1961). Although antibiotics are used to treat the disease in some countries, treatment must be continuous as it only affects the vegetative bacteria, not the infective spores. In many countries the only effective control measure is to burn the infected colony (Genersch, 2010) therefore, AFB infection nearly always leads to the death of the colony.

In recent years our knowledge of *P. larvae* has hugely increased. In 2006 *P. larvae* was reclassified as a single species after previously being described as two, separate sub species (*P. larvae larvae* and *P. larvae pulvifaciens*) (Genersch et al., 2006). Repetitive-element PCR using enterobacterial repetitive intergenic consensus (ERIC) primers describes four genotypes (Ashiralieva and Genersch, 2006) that differ in phenotype including morphology (Genersch *et al.*, 2006), metabolic capacity (Neuendorf et al., 2004), sporulation (Saville, 2011), virulence (Genersch et al., 2005; Rauch et al., 2009) and, virulence factors (Poppinga et al., 2012; Fünfhaus et al., 2013; Djukic et al., 2014). Only ERIC I and II isolates are regularly found in diseased hives, and are therefore the most economically important genotypes. In 2014 two *P. larvae* genomes belonging to ERIC types I and II were published and a comparison was carried out between them (Djukic et al., 2014). Djukic et al. (2014) demonstrated that there are many differences between these types at the genomic level, including length, number of transposases, insertion elements, predicted phage regions, and strain-specific island-like regions (Djukic et al., 2014).

AFB is a serious disease of honeybee brood with near global distribution and although recent research has improved our understanding of the causative agent and its methods of pathogenesis, disease epidemiology is poorly understood. In part this is due to an absence of a repeatable, high-resolution method to discriminate between strain types. Several

methods have been used to study the distribution and movement of *P. larvae* strain types in infected colonies (Genersch and Otten, 2003; Pentikäinen et al., 2008; Di Pinto et al., 2011; Bassi et al., 2015) but no standard, repeatable method that gives sufficient resolution has previously been developed. In this thesis I discuss the development of a new sequence-based typing scheme for *P. larvae* and its uses in studying the epidemiology of this damaging disease. I use the new scheme as a basis to examine aspect of *P. larvae* from the genomic level to the population level.

6.2 Main findings of thesis

MultiLocus Sequence Typing (MLST) has become a gold-standard for bacterial epidemiology. Chapter 2 describes the creation of the first standardised, sequence-based typing scheme for *P. larvae*. The *P. larvae* MLST scheme has been used to identify 24 global sequence types (STs) to date, and it is hoped that in the future other researchers will add to the pubmlst database (Jolley and Maiden, 2010; pubmlst.org/plarvae). Over 300 *P. larvae* isolates from disease outbreaks around the world have been typed using the scheme. This information was used to determine the global pattern of distribution of *P. larvae* STs. ERIC I was found to be more diverse than ERIC II, containing 19 ST compared to 3 in ERIC II. ERIC I and II STs have differing distributions, no ERIC II isolate was found in the West of Europe, including the UK. However, ERIC II is more widespread than was previously thought (Genersch, 2010). There is a significant relationship between geographic distance and genetic distance of *P. larvae* populations within the native range of the honey bee (Europe, Africa, and Western Asia). This relationship was shown to break down when tested outside the native range of honey bees.

Chapter 3 describes the analysis of AFB outbreaks in the UK for the first time using the new MLST scheme. AFB is not thought to be endemic in the UK and the sporadic nature of outbreaks is unusual. Mill et al., (2014) identified that AFB outbreaks in the UK can occur in clusters. Usually these were eradicated after a few years but some clusters remain persistent. In this chapter the MLST scheme was used to determine the *P. larvae* STs causing these clusters and determine whether they were caused by single STs or by multiple infections from different STs. STs found in English clusters were significantly different from those not in clusters in England and Scotland and in the Welsh cluster. Outbreaks that occurred in clusters near to risk factors were composed of more STs than

those that were not near risk points. This suggests that risk points represent a diverse infection source, causing repeated multiple infections in nearby clusters. Of 94 honey samples imported from both EU and non-EU sources, 39% contained traces of *P. larvae*, suggesting that honey importation is a possible source of AFB infection in the UK.

Chapter 4 describes the whole genome sequencing (WGS) of 21 isolates of *P. larvae* and genome comparisons. Seventeen ERIC I isolates and one ERIC IV isolate were assembled using DSM 25719 as a scaffold (Djukic et al., 2014; accession numbers: P003355-CP003356). Three ERIC II isolated were assembled using DSM 25430 as a scaffold (Djukic et al., 2014; accession number: ADFW00000000). The core genes present in all isolates were used to create a phylogeny. This core gene tree was compared to the tree produced using MLST sequences to confirm that the MLST accurately describes the relationship amongst STs. The core gene tree confirmed the main pattern of distribution amongst ERIC genotypes. However within the ERIC I group there were some differences. DSM 25719 (ST18) grouped with ST7 rather than other ST18s. DSM 25719 contained a plasmid similar to that found in ST7 isolates which suggests HGT. ERIC genotypes were shown to contain differing numbers of phages. All ERIC I isolates contained multiple complete phages, but none of the three ERIC II isolates did. Seven of the 20 isolates analysed contained plasmids similar to those identified in Djukic et al. (2014), including six ERIC I isolates.

Chapter 5 describes the discovery of seven plasmids in the 20 *P. larvae* genome sequences analysed. These plasmids are similar to the pPLA1_10 and pPLA2_10 plasmids described in Djukic et al. (2014). These plasmids contain a replication initiation factor, and a toxin gene. Djukic et al. tested 65 ERIC I isolates and did not identify any further examples of plasmid pPLA1_10. Therefore it is unusual that six were identified in ERIC I isolates in this study. The ERIC I plasmids were all identified in rare STs that are related, in a small geographic area and therefore, it is likely that they share a common ancestor. Maximum likelihood (ML) trees demonstrated that although the core genomes grouped by ST, the plasmids did not group either by ST or by geographic origin, suggesting horizontal gene transfer and movement out of Jersey.

6.3 Direction of future work

Standardised classification of bacterial strains causing disease is fundamental for epidemiological study. MLST schemes have been used to track and study the population structure and evolution of a number of bacterial pathogens of humans and animals (Helgason et al., 2004; Robinson and Enright, 2004; Maiden, 2006; Jolley and Maiden, 2010). Recently, an MLST scheme has been developed for another brood pathogen of honey bees, *Melissococcus plutonius*, the causative agent of European foulbrood (EFB; Haynes et al., 2013). This scheme was used to reveal disease transmission events between beekeepers in the UK (Haynes et al., 2013). The *M. plutonius* scheme along with the *P. larvae* scheme will be used by the National Bee Unit to type all outbreaks of disease in the UK, which will give insight into disease transmission that would be impossible without these methods. For example the *M. plutonius* scheme has been used to determine which STs are more virulent than others and this will inform the control methods in the UK (Budge et al., 2014). Future work on *P. larvae* using the MLST scheme may compare known *P. larvae* distribution and disease severity with genomic data to determine the virulence of STs and use this information to inform control measures.

It has been shown that a balanced gut microbiome is important for honey bee health (Koch and Schmid-Hempel, 2011). Further research could investigate the effects that disease and diet can have on honey bee gut microbiota, and how this may affect susceptibility to certain STs of *P. larvae*. This knowledge could lead to new treatments for honey bee diseases for example, preventative probiotics (Hamdi et al., 2011).

A more holistic use of the *P. larvae* MLST scheme will give a clearer picture of the AFB landscape in the UK, and the scheme has already provided valuable insights after one year, for example in 2014 the *P. larvae* scheme was used to type all AFB outbreaks in England and Wales and it was discovered that all ST18 outbreaks could be traced to a single owner (personal communication, Victoria Tomkies,). Typing of the infected imported honey samples included in Chapter 3 will demonstrate whether the *P. larvae* STs involved are similar to those causing outbreaks clustered near to risk factors. This will clarify whether honey importation is a risk for AFB spread in the UK.

The cost and time required for whole genome sequencing has decreased rapidly in recent years (Loman et al., 2012) and is likely to continue to become more accessible. It may soon be possible to generate WGS data for bacteria more cheaply than MLST data. WGS data gives much more information and finer resolution than MLST data but MLST is still an important tool in standardising strains between studies. WGS data can be easily and accurately typed using online databases such as BIGSdb (Jolley and Maiden, 2010) or software such as BioNumerics (Applied Maths). WGS has been used along with MLST data to give finer resolution in studies, which involve a short-time span and only a few single nucleotide polymorphisms (SNPs) differ between strains for example in MRSA outbreaks in hospitals (Köser et al., 2012; Harris et al., 2013) and to track the source of a Europe-wide *E coli* outbreak (Grad et al., 2012). WGS could be similarly used in studying the spread of *P. larvae* within and between apiaries, this would give insight into short-term spread of AFB that would not be possible with MLST alone. In the near future real-time surveillance and quick detection of outbreaks will become possible (Larsen et al., 2012).

6.3.1 ERIC I and II virulence

P. larvae STs belonging to ERIC I and II are known to have different methods of infection and some of the mechanisms behind this have been discovered (Genersch, 2010; Poppinga et al., 2012; Fünfhaus et al., 2013; Djukic et al., 2014). With the *P. larvae* scheme and WGS data of STs it is possible to compare differences not just between ERIC types but among and within different STs. Plasmids have been shown to have important roles in many bacterial species providing antimicrobial resistance genes e.g. in *P. larvae* (Alippi et al., 2007) and virulence e.g. in *Bacillus* (Helgason et al., 2000). This thesis describes the discovery of plasmids similar to pPLA1_10 in ERIC I isolates. In this thesis three isolates of ST7 were sequences, two with the plasmid and one without. To test the effect of the plasmids on their host, experimental infection could be carried out to compare ST7 isolates with and without the plasmids and the effect of the plasmid on phenotype cost of carrying the plasmid can be investigated (Harrison and Brockhurst, 2012).

Phages can encode virulence factors in pathogenic bacteria (Brüssow et al., 2004). Phage-encoded toxins are required by bacteria such as *Vibrio cholera* and *Clostridium botulinum* in order to infect hosts. Phages can also increase a bacteria's virulence incrementally, a higher number of phage-encoded virulence or fitness factor can increase the fitness of the

pathogen (Brüssow et al., 2004). Phages can also contribute to the diversification of bacterial genomes and can act as anchor points for genome inversions (Brüssow et al., 2004). It has been demonstrated in this thesis (Chapter 4) that the sequenced ERIC I STs all contained multiple complete phages, whereas none of the three ERIC II isolates sequenced did. It is possible that the phages present in ERIC I isolates may increase the fitness of these STs in comparison to ERIC II. ERIC II STs are less widespread and less diverse than ERIC I STs (Morrissey et al., 2015). Future work could look into the functions of the phages identified in the ERIC I isolates to determine their function and whether they have an effect on virulence or fitness on their hosts.

Sporulation rates have been shown differ between ERIC types, with ERIC IV isolates producing many times more spores than ERIC I (Saville, 2011) and this is hypothesised to explain some of the difference in virulence between these strains. ERIC II is known to be more virulent on the individual larval level than ERIC I (Rauch et al., 2009), and it is possible that differences in sporulation may account for some of this variation. Saville (2011) described genetic differences in a number of genes involved in sporulation between ERIC I and IV. However, no comparison between sporulation rates of ERIC I and II or amongst ERIC types have been carried out. Knowledge of the sporulation rates of STs may also impact on future control methods.

6.4 Conclusion

In this thesis I have studied aspects of *P. larvae* from a genomic to biogeographical scale and created a useful tool for further study of this disease causing bacterium The creation of the new *P. larvae* MLST scheme has allowed research into aspects of *P. larvae* that were not previously well characterised. Typing of global strains has allowed the study of the global distribution of *P. larvae* in detail not previously possible. It has also been possible to look in detail at the spread and distribution of strain types at a national level. Whole-genome sequencing and assembly of multiple STs of *P.larvae* opens up many avenues of future research that may lead to better control methods for this damaging disease.

7 Appendices

7.1 Appendix 1 MLST primer sequences tested and rejected

Gene	Forward primer	Reverse primer	Annealing temp (°C)
adk	5'ATGCCTACCTTGCCTAACAT3'	5'ATGCTTCTCCGTTTCGTG3'	**56**
ccpA	5'GTTTCTCGGGTTGTGAATAA3'	5'CCGTCTACCTGTTTTTCAAG3'	**55**
chiA	5'GCGGATGAACTCTAGTGAAC3'	5'ACGATTAAAACGAGCGAAC3'	**55**
gdh	5'TATCATTACAGGGGGTTCC3'	5'CATAGCTAATACGCCTGCTT3'	**55**
gmk	5'AAAGACAATGGAAAGAGAAAGA3'	5'ATTGTTCCATCAGACGAATC3' 5'CTGATGGAACAATATGATTACG3'	**55**
ilvD	5'GATGAACTGTCTGGCTGAAG3' 5'AAGGCGGCATCATAAAAGTC3'	5'CCCTGTTCATTATGTGGATTAT3' 5'AATTCTTCTTCGGATATTTGC3'	**57** **56**
panC	5'TCTCTCCCCAATGAAATATG3' 5'CTTTATCCGGCTGAACAATA3'	5'TTAAAGATGCACAGCAAGTG3' 5'GGAATTAAAAACGGGACATA3'	**55** **55**
sigK	5'GGAGAAGACCTGGAAGATTT3'	5'TTTTGTAAAATTCATGATAAAGC3'	**54**
tpi	5'ATATATTCGCGCACGTTTC3'	5'ACGTTCAAGGGAACCTCTAT3'	**55**

7.2 Appendix 2 Origins of isolates typed in the development of the MLST scheme

Isolate	Country of origin	Region	Continent	ERIC type	Sequence Type
12-8090	Cuba		North America	I	1
11-8014	New Zealand		Australasia	I	1
12-8293	Egypt, Ethiopia		Africa	I	2
12-8296	Egypt, Ethiopia		Africa	I	2
12-8299	Egypt, Ethiopia		Africa	I	2
12-8302	Egypt, Ethiopia		Africa	I	2
12-8338	France		Europe	I	2
12-8343	France		Europe	I	2
p8580 (p143 2012.02.LO)	France		Europe		2
12-8384	Georgia, Croatia		Europe	I	2
12-8387	Georgia, Croatia		Europe	I	2
12-8382	Georgia, Croatia		Europe	I	2
03-159	Germany	Saxony-Anhalt	Europe	I	2
01-391	Germany		Europe	I	2
12-8314	Greece	Crete	Europe	I	2
12-8109 (4)	Israel, Malta		mixed	I	2
1403/1 (P8467)	Italy		Europe		2
1267/1 (P8468)	Italy		Europe		2
1388/1 (P8465)	Italy		Europe		2
UK6/10	UK	England	Europe	I	2
11-8020	New Zealand		Australasia	I	3
11-8032	Mexico, Nicaragua, Argentina, Chile		mixed	I	4
p6678 (LMG 16241)	-			I	5

12-8320	Bangladesh		Asia	I	5
12-8273	Canada		North America	I	5
12-8277	Canada		North America	I	5
12-8278	Canada		North America	I	5
12-8281	Canada		North America	I	5
12-8091 (2)	Cuba		North America	I	5
12-8348	France		Europe	I	5
12-8352	France		Europe	I	5
p8581 (p106 2012.06.LO)	France		Europe		5
P7839 (Thur 99)	Germany	Thuringia	Europe	I	5
p7840 (Thur101)	Germany	Thuringia	Europe	I	5
11-152	Germany	Saxony-Anhalt	Europe	I	5
09-331	Germany	Saxony-Anhalt	Europe	I	5
02-360	Germany		Europe	I	5
03-019	Germany		Europe	I	5
12-367	Germany	Berlin	Europe	I	5
12-8265	Greece		Europe	I	5
12-8268	Greece		Europe	I	5
12-8391	Hungary		Europe	I	5
11-8021	New Zealand		Australasia	I	5
11-8022	New Zealand		Australasia	I	5
11-8023	New Zealand		Australasia	I	5
11-8024	New Zealand		Australasia	I	5
11-8026	New Zealand		Australasia	I	5
11-8027	New Zealand		Australasia	I	5

11-8028	New Zealand		Australasia	I	5
11-8029	New Zealand		Australasia	I	5
11-8010	New Zealand		Australasia	I	5
11-8011	New Zealand		Australasia	I	5
11-8012	New Zealand		Australasia	I	5
AFB7	Spain		Europe		5
12-8242	Switzerland		Europe	I	5
12-8245	Switzerland		Europe	I	5
12-8248	Switzerland		Europe	I	5
12-8251	Switzerland		Europe	I	5
UK1/10	UK	England	Europe	I	5
UK11/6	UK	England	Europe	I	5
UK15/10	UK	England	Europe	I	5
UK16/8	UK	England	Europe	I	5
UK17/10	UK	England	Europe	I	5
UK17/9	UK	England	Europe	I	5
UK3/8	UK	England	Europe	I	5
UK6/8	UK	England	Europe	I	5
UK7/10	UK	England	Europe	I	5
UK12/10	UK	England	Europe	I	5
UK14/10	UK	England	Europe	I	5
UK2/10	UK	England	Europe	I	5
UK2/7	UK	England	Europe	I	5
UK2/9	UK	England	Europe	I	5
UK20/10	UK	England	Europe	I	5
UK5/9	UK	England	Europe	I	5
UK7/9	UK	England	Europe	I	5
UK33/9	UK	England	Europe	I	5
UK38/9	UK	England	Europe	I	5

UK21/8	UK	England	Europe	I	5
UK22/10	UK	England	Europe	I	5
UK26/7	UK	England	Europe	I	5
UK24/10	UK	England	Europe	I	5
UK26/9	UK	England	Europe	I	5
UK27/7	UK	England	Europe	I	5
UK27/9	UK	England	Europe	I	5
UK35/9	UK	England	Europe	I	5
UK22/11	UK	England	Europe	I	5
UK21/10	UK	England	Europe	I	5
UK22/9	UK	England	Europe	I	5
S1/1	UK	England	Europe	I	5
S2/1	UK	England	Europe	I	5
p8582	France		Europe		6
J3/1	Jersey		Europe	I	6
P8476	France		Europe		7
12-8378	Mongolia		Asia	I	7
12-8380	Mongolia		Asia	I	7
UK15/6	UK	England	Europe	I	7
UK20/6	UK	England	Europe	I	7
UK3/6	UK	England	Europe	I	7
UK10/6	UK	England	Europe	I	7
UK10/7	UK	England	Europe	I	7
UK21/6	UK	England	Europe	I	7
UK21/7	UK	England	Europe	I	7
UK23/6	UK	England	Europe	I	7
UK25/6	UK	England	Europe	I	7
UK34/6	UK	England	Europe	I	7
UK36/6	UK	England	Europe	I	7

UK37/6	UK	England	Europe	I	7
UK38/6	UK	England	Europe	I	7
(P6266) LMG 16252	-			III	8
ATCC49483	-			IV	8
DSM 3615	-			IV	8
(P6264) LMG 16247	-			IV	8
LMG6911	-			IV	8
P6993 (LMG 14427)	-			IV	8
P6260 (LMG 16250)	-			IV	8
P6265 (LMG 16249)	-			IV	8
11-8050	Chile		South America	III	9
11-8051	Chile		South America	III	9
12-8120	Arabian Peninsula		Asia	II	10
12-8121	Arabian Peninsula		Asia	II	10
12-8122	Arabian Peninsula		Asia	II	10
12-8123	Arabian Peninsula		Asia	II	10
12-8124	Arabian Peninsula		Asia	II	10
12-8125	Arabian Peninsula		Asia	II	10
12-8126	Arabian Peninsula		Asia	II	10
12-8127	Arabian Peninsula		Asia	II	10
12-8128	Arabian Peninsula		Asia	II	10
12-8119	Arabian Peninsula		Asia	II	10
12-8230	Arabian Peninsula		Asia	II	10
12-8231	Arabian Peninsula		Asia	II	10
12-8232	Arabian Peninsula		Asia	II	10
12-8130	Australia		Australasia	II	10
12-8131	Australia		Australasia	II	10
12-8132	Australia		Australasia	II	10

12-8135	Australia		Australasia	II	10
12-8136	Australia		Australasia	II	10
12-8137	Australia		Australasia	II	10
12-8138	Australia		Australasia	II	10
12-8139	Australia		Australasia	II	10
P7846	Austria		Europe	II	10
12-8272	Canada		North America	II	10
01-649	Germany		Europe	II	10
00-1163	Germany		Europe	II	10
00-0869	Germany		Europe	II	10
12-109	Germany	Veitshöchheim	Europe	II	10
12-510	Germany	Veitshöchheim	Europe	II	10
p7842 (Thur 258)	Germany	Thuringia	Europe	II	10
02-009	Germany	SH / Bad Schwartau	Europe	II	10
03-522	Germany	Brandenburg, Kleinmachnow	Europe	II	10
03-525	Germany	Brandenburg, Kleinmachnow	Europe	II	10
00-897	Germany		Europe	II	10
11-8080	Greece	Mount Athos	Europe	II	10
11-8081	Greece	Mount Athos	Europe	II	10
12-8170	Japan, Singapore, China		Asia	II	10
12-8172	Japan, Singapore, China		Asia	II	10
12-8175	Japan, Singapore, China		Asia	II	10
12-8177	Japan, Singapore, China		Asia	II	10
12-8179	Japan, Singapore, China		Asia	II	10
12-8220	Kazakhstan		Asia	II	10
12-8221	Kazakhstan		Asia	II	10
12-8222	Kazakhstan		Asia	II	10
P7860 (11-8013)	New Zealand		Australasia	II	10
12-8371	Russia		Europe	II	10

12-8369	Russia		Europe	II	10
12-8373	Russia		Europe	II	10
12-8375	Russia		Europe	II	10
26-02	Sweden		Europe	II	10
P7371	Sweden		Europe	II	10
p8583 (AP1141 nr 2)	Austria		Europe		11
P7844 (452|03)	Czech Republic		Europe	II	11
8514-03	Finland		Europe	II	11
8533-03	Finland		Europe	II	11
P7850 (7774|03)	Finland		Europe	II	11
p7852 (8527/03)	Finland		Europe	II	11
01-1714	Germany		Europe	II	11
10-228	Germany	Veitshöchheim	Europe	II	11
10-232	Germany	Veitshöchheim	Europe	II	11
10-654	Germany	Veitshöchheim	Europe	II	11
10-658	Germany	Veitshöchheim	Europe	II	11
10-662	Germany	Veitshöchheim	Europe	II	11
10-678	Germany	Veitshöchheim	Europe	II	11
11-365	Germany	Veitshöchheim	Europe	II	11
11-370	Germany	Veitshöchheim	Europe	II	11
11-381	Germany	Veitshöchheim	Europe	II	11
11-577	Germany	Veitshöchheim	Europe	II	11
11-593	Germany	Veitshöchheim	Europe	II	11
11-599	Germany	Veitshöchheim	Europe	II	11
11-627	Germany	Veitshöchheim	Europe	II	11
12-116	Germany	Veitshöchheim	Europe	II	11
12-128	Germany	Veitshöchheim	Europe	II	11
12-134	Germany	Veitshöchheim	Europe	II	11
12-147	Germany	Veitshöchheim	Europe	II	11

12-490	Germany	Veitshöchheim	Europe	II	11
12-498	Germany	Veitshöchheim	Europe	II	11
12-520	Germany	Veitshöchheim	Europe	II	11
03-098	Germany	Schleswig-Holstein	Europe	II	11
p7853 (646D2920/03)	Germany	Detmold	Europe	II	11
P7854 (647D2920|03)	Germany	Detmold	Europe	II	11
02-334	Germany	Berlin	Europe	II	11
03-194	Germany	Berlin	Europe	II	11
03-200	Germany	Berlin	Europe	II	11
04-309	Germany	Berlin	Europe	II	11
05-085	Germany	Berlin	Europe	II	11
00-0775	Germany		Europe	II	11
11-403	Germany	Veitshöchheim	Europe	II	11
Uni_Saarl.909	Germany	Saarland	Europe	II	11
Uni_Saarl.913	Germany	Saarland	Europe	II	11
Uni_Saarl.914	Germany	Saarland	Europe	II	11
Uni_Saarl.916	Germany	Saarland	Europe	II	11
Uni_Saarl.918	Germany	Saarland	Europe	II	11
P7851 (8501|03)	Finland		Europe	II	12
12-8262	Greece		Europe	I	13
P7862 (11-8025)	New Zealand		Australasia	I	13
AFB2	Spain		Europe		13
J1/1	Jersey		Europe	I	14
B-3650	-				15
11-8061	Argentina		South America	I	15
P7864 (11-8062)	Argentina		South America	I	15
12-8133	Australia		Australasia	I	15
12-8134	Australia		Australasia	I	15

P7847	Austria		Europe	I	15
P7848	Austria		Europe	I	15
P7843 (308|03)	Czech Republic		Europe	I	15
p7841(Thu 170)	Germany	Thuringia	Europe	I	15
01-445	Germany		Europe	I	15
01-358	Germany		Europe	I	15
03-384	Germany		Europe	I	15
01-440	Germany	Saxony-Anhalt	Europe	I	15
782/1	Italy		Europe		15
679/1 (P8466)	Italy		Europe		15
J20/1	Jersey		Europe	I	15
J47/1	Jersey		Europe	I	15
11-8070	Mexico		North America	I	15
11-8030	Mexico, Nicaragua, Argentina, Chile		mixed	I	15
12-8356	Poland		Europe	I	15
12-8361	Poland		Europe	I	15
L1/1	South Africa		Africa	I	15
12-8282	Tenerife		Europe	I	15
12-8285	Tenerife		Europe	I	15
12-8288	Tenerife		Europe	I	15
12-8291	Tenerife		Europe	I	15
UK1/9	UK	Wales	Europe	I	15
UK1/8	UK	England	Europe	I	15
UK11/10	UK	England	Europe	I	15
UK13/9	UK	England	Europe	I	15
UK3/10	UK	England	Europe	I	15
UK23/9	UK	England	Europe	I	15
UK21/11	UK	England	Europe	I	15

P7858 (UK 1991)	UK		Europe	I	15
12-8304	USA		North America	I	15
12-8306	USA		North America	I	15
12-8308	USA		North America	I	15
12-8310	USA		North America	I	15
USA-21	USA		North America	I	15
11-8031	Mexico, Nicaragua, Argentina, Chile		mixed	I	16
ATCC9545	-			I	17
02-129	Germany		Europe	I	17
02-130	Germany		Europe	I	17
12-8366	Poland		Europe	I	17
P7370	Sweden		Europe	I	17
P7857 (UK 1961)	UK		Europe	I	17
P6254	-			I	18
1487/1	Italy		Europe		18
2358/1	Italy		Europe		18
11-8040	New Zealand		Australasia	I	18
UK14/7	UK	Wales	Europe	I	18
UK5/8	UK	Wales	Europe	I	18
UK9/7	UK	Wales	Europe	I	18
UK16/6	UK	England	Europe	I	18
UK8/10	UK	England	Europe	I	18
UK28/7	UK	Wales	Europe	I	18
UK27/6	UK	England	Europe	I	18
UK40/6	UK	England	Europe	I	18

P7856 (UK 1957)	UK		Europe	I	18
08-100	USA		North America	I	18
USA-148	USA		North America	I	18
USA-38	USA		North America	I	18
12-8322	Bangladesh		Asia	I	19
12-8324	Bangladesh		Asia	I	19
12-8326	Bangladesh		Asia	I	19
p7845 (778/03)	Czech Republic		Europe	I	19
00-087	Germany		Europe	I	19
03-125	Germany		Europe	I	19
12-8312	Greece	Crete	Europe	I	19
12-8271	Greece		Europe	I	19
12-8355	Hungary		Europe	I	19
12-8100	Israel, Malta		mixed	I	19
P8477	Poland	Sieradz	Europe		19
p8494	Poland	Radomsko	Europe		19
p8488	Poland	Kutno	Europe		19
P8480	Poland		Europe		19
03-189	Germany	Saxony-Anhalt	Europe	I	20
UK6/7	UK	England	Europe	I	21
UK25/7	UK	England	Europe	I	21

8 References

Aanensen, D.M. and Spratt, B.G. (2005) The multilocus sequence typing network: mlst.net. *Nucleic Acids Res.* **33**: W728–W733.

Alikhan, N.-F., Petty, N.K., Ben Zakour, N.L., and Beatson, S.A. (2011) BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. *BMC Genomics* **12**: 402.

Alippi, A.M. and Aguilar, O.M. (1998) Characterization of Isolates of *Paenibacillus larvae* subsp. *larvae* from Diverse Geographical Origin by the Polymerase Chain Reaction and BOX Primers. *J. Invertebr. Pathol.* **72**: 21–27.

Alippi, A.M., E León, I., and C López, A. (2014) Tetracycline-resistance encoding plasmids from *Paenibacillus larvae*, the causal agent of American foulbrood disease, isolated from commercial honeys. *Int. Microbiol.* **17**: 49–61.

Alippi, A.M., López, A.C., and Aguilar, O.M. (2004) A PCR‐based method that permits specific detection of *Paenibacillus larvae* subsp. *larvae*, the cause of American Foulbrood of honey bees, at the subspecies level. *Lett. Appl. Microbiol.* **39**: 25–33.

Alippi, A.M., López, A.C., Reynaldi, F.J., Grasso, D.H., and Aguilar, O.M. (2007) Evidence for plasmid-mediated tetracycline resistance in *Paenibacillus larvae*, the causal agent of American Foulbrood (AFB) disease in honeybees. *Vet. Microbiol.* **125**: 290–303.

Allsopp, M.H., De Lange, W.J., and Veldtman, R. (2008) Valuing insect pollination services with cost of replacement. *PLoS One* **3**: e3128.

Antúnez, K., Piccini, C., Castro-Sowinski, S., Rosado, A.S., Seldin, L., and Zunino, P. (2007) Phenotypic and genotypic characterization of *Paenibacillus larvae* isolates. *Vet. Microbiol.* **124**: 178–183.

Ash, C., Priest, F.G., and Collins, M.D. (1993) Molecular identification of rRNA group 3 bacilli (Ash, Farrow, Wallbanks and Collins) using a PCR probe test. *Antonie Van Leeuwenhoek* **64**: 253–260.

Ashiralieva, A. and Genersch, E. (2006) Reclassification, genotypes and virulence of *Paenibacillus larvae*, the etiological agent of American foulbrood in honeybees – a review. *Apidologie* **37**: 10.

Auwera, G.A.V. der, Andrup, L., and Mahillon, J. (2005) Conjugative plasmid pAW63 brings new insights into the genesis of the *Bacillus anthracis* virulence plasmid pXO2 and of the *Bacillus thuringiensis* plasmid pBT9727. *BMC Genomics* **6**: 103.

Aziz, R.K. et al. (2008) The RAST Server: Rapid Annotations using Subsystems Technology. *BMC Genomics* **9**: 75.

Bacillus cereus Group Typing Database (University of Oslo, 2012)

Barth, H., Aktories, K., Popoff, M.R., and Stiles, B.G. (2004) Binary Bacterial Toxins: Biochemistry, Biology, and Applications of Common *Clostridium* and *Bacillus* Proteins. *Microbiol. Mol. Biol. Rev.* **68**: 373–402.

Bassi, S., Formato, G., Milito, M., Trevisiol, K., Salogni, C., and Carra, E. (2015) Phenotypic characterization and ERIC–PCR based genotyping of *Paenibacillus larvae* isolates recovered from American foulbrood outbreaks in honey bees from Italy. *Vet. Q.* **35**: 27–32.

BHIPA (British Honey Importers and Packers Association) (2011). Code Of Practice for the Importation, Blending, Packaging and Marketing of Honey. Accessed from: www.honeyassociation.com

Blacquière, T., Smagghe, G., van Gestel, C.A.M., and Mommaerts, V. (2012) Neonicotinoids in bees: a review on concentrations, side-effects and risk assessment. *Ecotoxicology* **21**: 973–992.

Bodorová-Urgošíková, J., Benada, O., and Tichý, P. (1992) Large-scale isolation and partial characterization of plasmid DNA from *B. larvae*. *Folia Microbiol. (Praha)* **37**: 82–86.

Brettin, T. et al. (2015) RASTtk: A modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. *Sci. Rep.* **5**: 8365.

Brüssow, H., Canchaya, C., and Hardt, W.-D. (2004) Phages and the Evolution of Bacterial Pathogens: from Genomic Rearrangements to Lysogenic Conversion. *Microbiol. Mol. Biol. Rev.* **68**: 560–602.

Büchler, R., Berg, S., and Le Conte, Y. (2010) Breeding for resistance to *Varroa destructor* in Europe. *Apidologie* **41**: 393–408.

Budge, G.E. et al. (2014) Molecular epidemiology and population structure of the honey bee brood pathogen *Melissococcus plutonius*. *ISME J.*

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.

Chan, Q.W. et al. (2011) Updated genome assembly and annotation of *Paenibacillus larvae*, the agent of American foulbrood disease of honey bees. *BMC Genomics* **12**: 450–450.

Chen, Y.-W., Wang, C.-H., An, J., and Ho, K.-K. (2000) Susceptibility of the Asian honey bee, *Apis cerana*, to American foulbrood, *Paenibacillus larvae larvae*. *J. Apic. Res.* **39**: 169–175.

Chin, C.-S. et al. (2011) The origin of the Haitian cholera outbreak strain. *N. Engl. J. Med.* **364**: 33–42.

Crane, E., (1999) The world history of beekeeping and honey hunting. 1st ed. Routledge

Darling, A.E., Mau, B., and Perna, N.T. (2010) progressiveMauve: Multiple Genome Alignment with Gene Gain, Loss and Rearrangement. *PLoS ONE* **5**: e11147.

Datta, S., Bull, J.C., Budge, G.E., and Keeling, M.J. (2013) Modelling the spread of American foulbrood in honeybees. *J. R. Soc. Interface* **10**: 20130650–20130650.

Davies, J. and Davies, D. (2010) Origins and Evolution of Antibiotic Resistance. *Microbiol. Mol. Biol. Rev.* **74**: 417–433.

Djukic, M. et al. (2014) How to Kill the Honey Bee Larva: Genomic Potential and Virulence Mechanisms of *Paenibacillus larvae*. *PLoS ONE* **9**: e90914.

Donovan, B.J. (1980) Interactions between native and introduced bees in New Zealand. *N. Z. J. Ecol.* **3**: 104–116.

Dsouza, M., Taylor, M.W., Turner, S.J., and Aislabie, J. (2014) Genome-Based Comparative Analyses of Antarctic and Temperate Species of *Paenibacillus*. *PLoS ONE* **9**: e108009.

Edwards, D.J. and Holt, K.E. (2013) Beginner's guide to comparative bacterial genome analysis using next-generation sequence data. *Microb. Inform. Exp.* **3**: 2.

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

Fang,Y. et al. (2011) A pangenomic study of Bacillus thuringiensis. *J. Genet. Genomics* **38**: 567–576.

Fasano, A. et al. (1991) Vibrio cholerae produces a second enterotoxin, which affects intestinal tight junctions. *Proc. Natl. Acad. Sci.* **88**: 5242–5246.

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. *J. Bacteriol.* **186**: 1518–1530.

Fera (Food and environment research agency) (2013) Foul Brood Disease of Honey Bees and other common brood disorders. Sand Hutton, York accesed at : www.defra.gov.uk/fera

Forsgren, E., Stevanovic, J., and Fries, I. (2008) Variability in germination and in temperature and storage resistance among *Paenibacillus larvae* genotypes. *Vet. Microbiol.* **129**: 342–349.

Francisco, A.P., Bugalho, M., Ramirez, M., and Carriço, 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.

Fries, I., Lindström, A., and Korpela, S. (2006) Vertical transmission of American foulbrood (*Paenibacillus larvae*) in honey bees (*Apis mellifera*). *Vet. Microbiol.* **114**: 269–274.

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

Fünfhaus, A., Poppinga, L., and Genersch, E. (2013) Identification and characterization of two novel toxins expressed by the lethal honey bee pathogen *Paenibacillus larvae,* the causative agent of American foulbrood: Two novel *Paenibacillus larvae* toxins. *Environ. Microbiol.* **15**: 2951–2965.

Garcillán-Barcia, M.P., Francia, M.V., and De La Cruz, F. (2009) The diversity of conjugative relaxases and its application in plasmid classification. *FEMS Microbiol. Rev.* **33**: 657–687.

Genersch, E. (2010) American foulbrood in honeybees and its causative agent, *Paenibacillus larvae*. *J. Invertebr. Pathol.* **103**: S10–S19.

Genersch, E. et al. (2006) Reclassification of *Paenibacillus larvae* subsp. *pulvifaciens* and *Paenibacillus larvae* subsp. *larvae* as *Paenibacillus larvae* without subspecies differentiation. *Int. J. Syst. Evol. Microbiol.* **56**: 501–511.

Genersch, E., Ashiralieva, A., and Fries, I. (2005) Strain- and Genotype-Specific Differences in Virulence of *Paenibacillus larvae* subsp. *larvae*, a Bacterial Pathogen Causing American Foulbrood Disease in Honeybees. *Appl. Environ. Microbiol.* **71**: 7551 –7555.

Genersch, E. and Otten, C. (2003) The use of repetitive element PCR fingerprinting (rep-PCR) for genetic subtyping of German field isolates of *Paenibacillus larvae* subsp. *larvae*. *Apidologie* **34**: 195–206.

González, M.J. and Marioli, J.M. (2010) Antibacterial activity of water extracts and essential oils of various aromatic plants against *Paenibacillus larvae*, the causative agent of American Foulbrood. *J. Invertebr. Pathol.* **104**: 209–213.

Goslee, S.C., Urban, D.L., and others (2007) The ecodist package for dissimilarity-based analysis of ecological data. *J. Stat. Softw.* **22**: 1–19.

Goulson, D. (2003) Effects of Introduced Bees on Native Ecosystems. *Annu. Rev. Ecol. Evol. Syst.* **34**: 1–26.

Govan, V.A., Allsopp, M.H., and Davison, S. (1999) A PCR detection method for rapid identification of *Paenibacillus larvae*. *Appl. Environ. Microbiol.* **65**: 2243–2245.

de Graaf, D.C. et al. (2013) Standard methods for American foulbrood research. *J. Apic. Res.* **52.**:

de Graaf, D., Vandekerchove, D., Dobbelaere, W., Peeters, J., and Jacobs, F. (2001) Influence of the proximity of American foulbrood cases and apicultural management on the prevalence of *Paenibacillus larvae* spores in Belgian honey. *Apidologie* **32**: 587–599.

Grad, Y.H. et al. (2012) Genomic epidemiology of the *Escherichia coli* O104:H4 outbreaks in Europe, 2011. *Proc. Natl. Acad. Sci.* **109**: 3065–3070.

Gurevich,A., Saveliev,V., Vyahhi,N., and Tesler,G. (2013) QUAST: quality assessment tool for genome assemblies. *Bioinformatics* **29**: 1072–1075.

Hamdi, C. et al. (2011) Gut microbiome dysbiosis and honeybee health: Gut microbiome dysbiosis and honeybee health. *J. Appl. Entomol.* **135**: 524–533.

Hansen (1984) The investigation of honey from bee colonies for *Bacillus larvae*. *Dan. J. Plant Soil Sci.* **90**: 81–86.

Hansen, H., Brødsgaard, C.J., Kryger, P., and Nicolaisen, M. (2003) A scientific note on the presence of *Paenibacillus larvae larvae* spores in sub-Saharan African honey. *Apidologie* **34**: 471–472.

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 Microbiol.* **18**: 141–148.

Harrison, E. and Brockhurst, M.A. (2012) Plasmid-mediated horizontal gene transfer is a coevolutionary process. *Trends Microbiol.* **20**: 262–267.

Harris, S.R. et al. (2013) Whole-genome sequencing for analysis of an outbreak of meticillin-resistant *Staphylococcus aureus*: a descriptive study. *Lancet Infect. Dis.* **13**: 130–136.

Hasemann, L. (1961) How long can spores of American foulbrood live? *Am. Bee J.* **101**: 298–299.

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: A typing scheme for *Melissococcus plutonius*. *Environ. Microbiol. Rep.* **5**: 525–529.

Helgason, E. et al. (2000) *Bacillus anthracis*, *Bacillus cereus*, and *Bacillus thuringiensis*—One Species on the Basis of Genetic Evidence. *Appl. Environ. Microbiol.* **66**: 2627–2630.

Helgason, E., Tourasse, N.J., Meisal, R., Caugant, D.A., and Kolsto, A.-B. (2004) Multilocus Sequence Typing Scheme for Bacteria of the *Bacillus cereus* Group. *Appl Env. Microbiol* **70**: 191–201.

Heyndrickx, M. et al. (1996) Reclassification of *Paenibacillus* (formerly *Bacillus*) *pulvifaciens* (Nakamura 1984) Ash et al. 1994, a later subjective synonym of *Paenibacillus* (formerly *Bacillus*) *larvae* (White 1906) Ash et al. 1994, as a subspecies of *P. larvae*, with emended descriptions of *P. larvae* as *P. larvae* subsp. *larvae* and *P. larvae* subsp. *pulvifaciens*. *Int. J. Syst. Bacteriol.* **46**: 270–279.

Holt, K.E. et al. (2012) *Shigella sonnei* genome sequencing and phylogenetic analysis indicate recent global dissemination from Europe. *Nat. Genet.* **44**: 1056–1059.

Hulton, C.S.J., Higgins, C.F., and Sharp, P.M. (1991) ERIC sequences: a novel family of repetitive elements in the genomes of *Escherichia coli*, *Salmonella typhimurium* and other enterobacteria. *Mol. Microbiol.* **5**: 825–834.

Huson, D.H. and Bryant, D. (2006) Application of Phylogenetic Networks in Evolutionary Studies. *Mol. Biol. Evol.* **23**: 254–267.

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. *Mol. Ecol.* **14**: 93–106.

Jensen, A.B., Pedersen, B.V., and Eilenberg, J. (2009) Differential susceptibility across honey bee colonies in larval chalkbrood resistance. *Apidologie* **40**: 524–534.

Jolley, K.A., Feil, E.J., Chan, M.-S., and Maiden, M.C.J. (2001) Sequence type analysis and recombinational tests (START). *Bioinformatics* **17**: 1230–1231.

Jolley, K.A. and Maiden, M.C. (2010) BIGSdb: scalable analysis of bacterial genome variation at the population level. *BMC Bioinformatics* **11**: 595.

Keenan, K., McGinnity, P., Cross, T.F., Crozier, W.W., and Prodöhl, P.A. (2013) diveRsity: An R package for the estimation and exploration of population genetics parameters and their associated errors. *Methods Ecol. Evol.* **4**: 782–788.

Kibbe, W.A. (2007) OligoCalc: an online oligonucleotide properties calculator. *Nucleic Acids Res.* **35**: W43–W46.

Klee, J. et al. (2007) Widespread dispersal of the microsporidian *Nosema ceranae*, an emergent pathogen of the western honey bee, *Apis mellifera*. *J. Invertebr. Pathol.* **96**: 1–10.

Klein, A.-M. et al. (2007) Importance of pollinators in changing landscapes for world crops. *Proc. R. Soc. Lond. B Biol. Sci.* **274**: 303–313.

Koch, H. and Schmid-Hempel, P. (2011) Socially transmitted gut microbiota protect bumble bees against an intestinal parasite. *Proc. Natl. Acad. Sci.* **108**: 19288–19292.

Köser, C.U. et al. (2012) Rapid Whole-Genome Sequencing for Investigation of a Neonatal MRSA Outbreak. *N. Engl. J. Med.* **366**: 2267–2275.

Koskella, B. and Brockhurst, M.A. (2014) Bacteria–phage coevolution as a driver of ecological and evolutionary processes in microbial communities. *FEMS Microbiol. Rev.* **38**: 916–931.

Krawiec, S. and Riley, M. (1990) Organization of the bacterial chromosome. *Microbiol. Rev.* **54**: 502–539.

Kruskal, J.B. (1964a) Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika* **29**: 1–27.

Kruskal, J.B. (1964b) Nonmetric multidimensional scaling: A numerical method. *Psychometrika* **29**: 115–129.

Lanza, V.F., Tedim, A.P., Martínez, J.L., Baquero, F., and Coque, T.M. (2015) The Plasmidome of *Firmicutes*: Impact on the Emergence and the Spread of Resistance to Antimicrobials. *Microbiol. Spectr.* **3**: PLAS–0039–2014.

Larsen, M.V. et al. (2012) Multilocus Sequence Typing of Total-Genome-Sequenced Bacteria. *J. Clin. Microbiol.* **50**: 1355–1361.

Li, H. et al. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**: 2078–2079.

Lindström, A., Korpela, S., and Fries, I. (2008b) Horizontal transmission of *Paenibacillus larvae* spores between honey bee (*Apis mellifera*) colonies through robbing. *Apidologie* **39**: 515–522.

Lindström, A., Korpela, S., and Fries, I. (2008a) The distribution of *Paenibacillus larvae* spores in adult bees and honey and larval mortality, following the addition of American foulbrood diseased brood or spore-contaminated honey in honey bee (*Apis mellifera*) colonies. *J. Invertebr. Pathol.* **99**: 82–86.

Loman, N.J. et al. (2012) High-throughput bacterial genome sequencing: an embarrassment of choice, a world of opportunity. *Nat. Rev. Microbiol.* **10**: 599–606.

Loncaric, I., Derakhshifar, I., Oberlerchner, J.T., Köglberger, H., and Moosbeckhofer, R. (2009) Genetic diversity among isolates of *Paenibacillus larvae* from Austria. *J. Invertebr. Pathol.* **100**: 44–46.

Maiden, M.C.J. et al. (1998) Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc. Natl. Acad. Sci.* **95**: 3140–3145.

Maiden, M.C.J. (2006) Multilocus sequence typing of bacteria. *Annu Rev Microbiol* **60**: 561–588.

Mantel, N. (1967) The Detection of Disease Clustering and a Generalized Regression Approach. *Cancer Res.* **27**: 209–220.

Marchler-Bauer, A. et al. (2014) CDD: NCBI's conserved domain database. *Nucleic Acids Res.* gku1221.

Margos, G. et al. (2008) MLST of housekeeping genes captures geographic population structure and suggests a European origin of *Borrelia burgdorferi*. *Proc. Natl. Acad. Sci.* **105**: 8730–8735.

Martin, S.J. et al. (2012) Global Honey Bee Viral Landscape Altered by a Parasitic Mite. *Science* **336**: 1304–1306.

Matheson, A. (1993) World Bee Health Report. *Bee World* **74**: 176–212.

McGregor, S.E. (1976) Insect pollination of cultivated crop plants Agricultural Research Service, US Department of Agriculture.

Medini, D., Donati, C., Tettelin, H., Masignani, V., and Rappuoli, R. (2005) The microbial pan-genome. *Curr. Opin. Genet. Dev.* **15**: 589–594.

Meixner, M.D., Worobik, M., Wilde, J., Fuchs, S., and Koeniger, N. (2007) *Apis mellifera mellifera* in eastern Europe – morphometric variation and determination of its range limits. *Apidologie* **38**: 191–197.

Mellmann, 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.

Mill, A.C. et al. (2014) Clustering, persistence and control of a pollinator brood disease: epidemiology of American foulbrood: Epidemiology of American foulbrood. *Environ. Microbiol.* **16**: 3753–3763.

Milne, I. et al. (2013) Using Tablet for visual exploration of second-generation sequencing data. *Brief. Bioinform.* **14**: 193–202.

Mock, M. and Fouet, A. (2001) Anthrax. *Annu. Rev. Microbiol.* **55**: 647–671.

Morrissey, B.J., Helgason, T., Poppinga, L., Fünfhaus, A., Genersch, E., and Budge, G.E. (2015) Biogeography of *Paenibacillus larvae,* the causative agent of American foulbrood, using a new multilocus sequence typing scheme: MLST scheme for *Paenibacillus larvae*. *Environ. Microbiol.* **17**: 1414–1424.

Morse, K.S.R. (1992) American foulbrood incidence in some US and Canadian honeys. *J. Apic. Res.* **23**: 497-501.

Murray, D. and Aronstein, K.A. (2006) Oxytetracycline-resistance in the honey bee pathogen *Paenibacillus larvae* is encoded on novel plasmid pMA67. *J. Apic. Res.* **47**: 207–214.

Murray, K.D., Aronstein, K.A., and de León, J.H. (2007) Analysis of pMA67, a predicted rolling-circle replicating, mobilizable, tetracycline-resistance plasmid from the honey bee pathogen, *Paenibacillus larvae*. *Plasmid* **58**: 89–100.

Nei, M. and Gojobori, T. (1986) Simple methods for estimating the numbers of synonymous and non-synonymous nucleotide substitutions. *Mol. Biol. Evol.* **3**: 418–426.

Neuendorf, S., Hedtke, K., Tangen, G., and Genersch, E. (2004) Biochemical characterization of different genotypes of *Paenibacillus larvae* subsp. *larvae,* a honey bee bacterial pathogen. *Microbiology* **150**: 2381 –2390.

Neu, H.C. (1992) The Crisis in Antibiotic Resistance. *Science* **257**: 1064–1073.

Neumann, P. and Ellis, J.D. (2008) The small hive beetle (*Aethina tumida* Murray, *Coleoptera: Nitidulidae*): distribution, biology and control of an invasive species. *J. Apic. Res.* **47**: 181–183.

Nurk,S. et al. (2013) Assembling Genomes and Mini-metagenomes from Highly Chimeric Reads. In, Deng,M., Jiang,R., Sun,F., and Zhang,X. (eds), *Research in Computational Molecular Biology*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 158–170.

Okinaka,R.T. et al. (1999) Sequence and organization of pXO1, the large Bacillus anthracis plasmid harboring the anthrax toxin genes. *J. Bacteriol.* **181**: 6509–6515.

Oksanen, J. et al. (2013) vegan: Community Ecology Package. R package version 2.0-6.

Overbeek, R. et al. (2014) The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Res.* **42**: D206–D214.

Ozer, A., Allen, J.P., and Hauser, A.R. (2014) Characterization of the core and accessory genomes of Pseudomonas aeruginosa using bioinformatic tools Spine and AGEnt. *BMC Genomics* **15**: 737.

Palacios, G. et al. (2008) Genetic Analysis of Israel Acute Paralysis Virus: Distinct Clusters Are Circulating in the United States. *J. Virol.* **82**: 6209–6217.

Paterson, D.L. (2006) Resistance in Gram-Negative Bacteria: *Enterobacteriaceae*. *Am. J. Med.* **119**: S20–S28.

Pentikäinen, J., Kalliainen, E., and Pelkonen, S. (2008) Molecular epidemiology of *Paenibacillus larvae* infection in Finland. *Apidologie* **40**: 73–81.

Petersen, J. (2011) Phylogeny and compatibility: plasmid classification in the genomics era. *Arch. Microbiol.*

Di Pinto, A., Novello, L., Terio, V., and Tantillo, G. (2011) ERIC-PCR Genotyping of *Paenibacillus larvae* in Southern Italian Honey and Brood Combs. *Curr. Microbiol.* **63**: 416–419.

Poppinga, L. et al. (2012) Identification and Functional Analysis of the S-Layer Protein *SplA* of *Paenibacillus larvae*, the Causative Agent of American Foulbrood of Honey Bees. *PLoS Pathog* **8**: e1002716.

Di Prisco, G. et al. (2013) *Neonicotinoid clothianidin* adversely affects insect immunity and promotes replication of a viral pathogen in honey bees. *Proc. Natl. Acad. Sci.* **110**: 18466–18471.

Qin, X., Evans, J.D., Aronstein, K.A., Murray, K.D., and Weinstock, G.M. (2006) Genome sequences of the honey bee pathogens *Paenibacillus larvae* and *Ascosphaera apis*. *Insect Mol. Biol.* **15**: 715–718.

Rauch, S., Ashiralieva, A., Hedtke, K., and Genersch, E. (2009) Negative Correlation between Individual-Insect-Level Virulence and Colony-Level Virulence of *Paenibacillus larvae*, the Etiological Agent of American Foulbrood of Honeybees. *Appl. Environ. Microbiol.* **75**: 3344–3347.

Rasko, D., Altherr, M., Han, C., and Ravel, J. (2005) Genomics of the *Bacillus cereus* group of organisms. *FEMS Microbiol. Rev.* **29**: 303–329.

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

Reybroeck, W., Daeseleire, E., De Brabander, H.F., and Herman, L. (2012) Antimicrobials in beekeeping. *Vet. Microbiol.* **158**: 1–11.

Ritter, W. (2003) Early detection of American Foulbrood by honey and wax analysis. *Apiatica* **38**: 125–130.

Robert D. Fleischmann, Mark D. Adams, Owen White, Rebecca A. Clayton, Ewen F. Kirkness, Anthony R. Kerlavage, Carol J. Bult, et al.(1995) Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* **269**: 496–498+507–512.

Robinson, D.A. and Enright, M.C. (2004) Multilocus sequence typing and the evolution of methicillin-resistant *Staphylococcus aureus*. *Clin. Microbiol. Infect.* **10**: 92–97.

Rozen, S. and Skaletsky, H., J. (2000) Primer3 on the WWW for general users and for biologist programmers. In, *Bionformatics Methods and Protocols: Methods in Molecular Biology*. Humana Press, Totowa, NJ, pp. 365–386.

De La Rúa, P., Galián, J., Serrano, J., and Moritz, R.F.A. (2001) Genetic structure and distinctness of *Apis mellifera L*. populations from the Canary Islands. *Mol. Ecol.* **10**: 1733–1742.

De la Rúa, P., Jaffé, R., Dall'Olio, R., Muñoz, I., and Serrano, J. (2009) Biodiversity, conservation and current threats to European honeybees. *Apidologie* **40**: 263–284.

Rusenova, N., Parvanov, P., and Stanilova, S. (2013) Molecular Typing of *Paenibacillus larvae* Strains Isolated from Bulgarian Apiaries Based on Repetitive Element Polymerase Chain Reaction (Rep-PCR). *Curr. Microbiol.*

Sabat, A.J. et al. (2013) Overview of molecular typing methods for outbreak detection and epidemiological surveillance. *Euro Surveill* **18**: 20380.

Santos, R.C.V. et al. (2012) Antimicrobial activity of Amazonian oils against *Paenibacillus species. J. Invertebr. Pathol.* **109**: 265–268.

Saville, B. (2011) Differentiation of Virulent and Biological Control *Paenibacillus larvae* Strains Associated with American Foulbrood in Bee Hives.

Schäfer, M.O. et al. (2014) Rapid identification of differentially virulent genotypes of *Paenibacillus larvae*, the causative organism of American foulbrood of honey bees, by whole cell MALDI-TOF mass spectrometry. *Vet. Microbiol.* **170**: 291–297.

Schmitt, C.K., McKEE, M.L., and O'Brien, A.D. (1991) Two copies of Shiga-like toxin II-related genes common in enterohemorrhagic *Escherichia coli* strains are responsible for the antigenic heterogeneity of the O157: H-strain E32511. *Infect. Immun.* **59**: 1065–1073.

Sharples, G.J. and Lloyd, R.G. (1990) A novel repeated DNA sequence located in the intergenic regions of bacterial chromosomes. *Nucleic Acids Res.* **18**: 6503–6508.

del Solar, G., Giraldo, R., Ruiz-Echevarría, M.J., Espinosa, M., and Díaz-Orejas, R. (1998) Replication and control of circular bacterial plasmids. *Microbiol. Mol. Biol. Rev.* **62**: 434–464.

Solignac, M. et al. (2005) The invasive Korea and Japan types of *Varroa destructor*, ectoparasitic mites of the Western honeybee (Apis mellifera), are two partly isolated clones. *Proc. R. Soc. B Biol. Sci.* **272**: 411–419.

Spivak, M. and Reuter, G.S. (2001) Resistance to American foulbrood disease by honey bee colonies *Apis mellifera* bred for hygienic behavior. *Apidologie* **32**: 555–565.

Stewart, P.E., Byram, R., Grimm, D., Tilly, K., and Rosa, P.A. (2005) The plasmids of *Borrelia burgdorferi*: essential genetic elements of a pathogen. *Plasmid* **53**: 1–13.

Takamatsu, D., Morinishi, K., Arai, R., Sakamoto, A., Okura, M., and Osaki, M. (2014) Typing of *Melissococcus plutonius* isolated from European and Japanese honeybees suggests spread of sequence types across borders and between different Apis species. *Vet. Microbiol.* **171**: 221–226.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. (2011) MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Mol. Biol. Evol.* **28**: 2731–2739.

Temple, M.L., Emmett, B.J., Scott, P.E., and Crabb, R.J. (2001) ADAS Consulting Limited.

Tettelin, H. et al. (2005) Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: Implications for the microbial "pan-genome." *PNAS* **102**: 13951–13955.

Tettelin, H., Riley, D., Cattuto, C., and Medini, D. (2008) Comparative genomics: the bacterial pan-genome. *Curr. Opin. Microbiol.* **11**: 472–477.

Tian, B., Fadhil, N.H., Powell, J.E., Kwong, W.K., and Moran, N.A. (2012) Long-Term Exposure to Antibiotics Has Caused Accumulation of Resistance Determinants in the Gut Microbiota of Honeybees. *mBio* **3**: e00377–12.

Tourasse, N.J., Helgason, E., Økstad, O.A., Hegna, I.K., and KolstØ, A.B. (2006) The *Bacillus cereus* group: novel aspects of population structure and genome dynamics. *J. Appl. Microbiol.* **101**: 579–593.

Underwood, A.P., Jones, G., Mentasti, M., Fry, N.K., and Harrison, T.G. (2013) Comparison of the *Legionella pneumophila* population structure as determined by sequence-based typing and whole genome sequencing. *BMC Microbiol.* **13**: 302.

Urwin, R. and Maiden, M.C.J. (2003) Multi-locus sequence typing: a tool for global epidemiology. *Trends Microbiol.* **11**: 479–487.

Wattam, A.R. et al. (2013) PATRIC, the bacterial bioinformatics database and analysis resource. *Nucleic Acids Res.* gkt1099.

Whitfield, C.W. 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. (2007) The incidence of honey bee pests and diseases in England and Wales. *Pest Manag. Sci.* **63**: 1062–1068.

Winston, M.L. (1987) The Biology of the Honey Bee. Harvard University Press, Cambridge, Massachusetts.

Yue, D., Nordhoff, M., Wieler, L.H., and Genersch, E. (2008) Fluorescence in situ hybridization (FISH) analysis of the interactions between honeybee larvae and *Paenibacillus larvae*, the causative agent of American foulbrood of honeybees (*Apis mellifera*). *Environ. Microbiol.* **10**: 1612–1620.

Zankari, E. et al. (2012) Identification of acquired antimicrobial resistance genes. *J. Antimicrob. Chemother.* **67**: 2640–2644.

Zhou, Y., Liang, Y., Lynch, K.H., Dennis, J.J., and Wishart, D.S. (2011) PHAST: A Fast Phage Search Tool. *Nucleic Acids Res.* gkr485.