Evaluation of population genetic structure and horizontal gene transfer for the human pathogen *Burkholderia pseudomallei*

Thesis

How to cite:


For guidance on citations see FAQs.

© 2008 The Author

Version: Version of Record

oro.open.ac.uk
Evaluation of population genetic structure and horizontal gene transfer for the human pathogen

*Burkholderia pseudomallei*

Sarinna Tumapa
BSc, MSc

Thesis submitted for PhD examination
(Life and Biomolecular Science) through the Open University
Sponsoring Establishment Mahidol-Oxford Tropical Medicine Research Unit (MORU)
Supervisors: Dr Sharon Peacock & Professor Nicholas Day
November 2007
Acknowledgements

It is with pleasure and satisfaction that I acknowledge the help of the following persons whose invaluable contribution helped me in completion of this thesis.

First, I wish to acknowledge a number of people for their help, advice and encouragement throughout this study. I would like to thank my supervisor, Dr. Sharon J. Peacock, who generously offered support and gave me the opportunity to undertake PhD. She also provided invaluable guidance, understanding, comments, suggestions and encouragement throughout this thesis.

I would like to extend my appreciation to co-advisors, Prof. Nicholas J. White and Prof. Nicholas PJ. Day, who kindly gave me support, advice and valuable comments.

Sincere thanks and appreciation go to my third party monitor, Asst. Prof. Kesinee Chotivanich, and senior colleagues, Mrs. Piengchan Sonthayanon, Dr. Wirongrong Chierakul, Dr. Narisara Chantratita, Mrs. Vanaporn Wuthickanun, Dr. Direk Limmathurotsakul, Dr. Rachaneeporn Tiyavisutsri and Dr. Khaemaporn Boonbumrung for their cooperation, help and support in technical exchanges and informatics strategies, statistical and clinical knowledge. Special thanks to Dr. Mongkol Vesaratchavest, for his great practical advice and genuine help which allow me to overcome many difficult situations.

I would like to thank the Wellcome Trust of Great Britain who provided the funding for this study. Without the support and facilities of the Mahidol-Oxford Tropical Medicine Research Unit, these studies would not have been possible. I am also indebted to my microbiology laboratory colleagues, Ms. Premchit Amornchai, Ms. Aunchalee Tanwisai and Mr. Sayan Langla for their kind help in bacterial
isolation and stock maintenance. Thanks for cooperative help and support also go to the doctors, nurses and technicians at Sappasithiprasong hospital, Ubon Ratchathani province.

Last, but not least, I would like to dedicate all my gratitude and my heartfelt thanks to my beloved family for their support and encouragement. Finally, special thanks for my daughter, Genie, born during this study, for keep me walking, fulfils my life and tells me what I mean to her.
Chapter 1. Introduction

1.1 General background to *Burkholderia pseudomallei* ........................................ 1

1.1.1 Historical aspects ................................................................. 1

1.1.2 Classification and description .............................................. 1

1.2 *B. pseudomallei* in the environment ........................................ 6

1.3 Geographical distribution of *B. pseudomallei* .................................... 9

1.3.1 Global epidemiology ........................................................... 9

1.3.1.1 Australia – Pacific region ............................................... 11

1.3.1.2 Asia ................................................................. 12

1.3.1.3 Areas outside Asia ...................................................... 15

1.4 *B. pseudomallei* infection ..................................................... 15

1.4.1 Melioidosis in animals ....................................................... 15

1.4.2 Human melioidosis ............................................................ 16

1.4.2.1 Risk factors ............................................................. 16
1.7 General background to *Burkholderia thailandensis* .............................................. 44

1.8 Aims of this dissertation ........................................................................................ 48

Chapter 2. Materials and Methods

2.1 Chemicals and reagents.......................................................................................... 49

2.2 Bacterial culture and storage conditions ................................................................ 49

2.3 Laboratory facilities ............................................................................................... 49

2.4 *B. pseudomallei* identification ............................................................................... 50

2.5 Bacterial isolates .................................................................................................... 51

2.5.1 Isolates used for genomic island analysis ....................................................... 51

2.5.2 Isolates used for MLST ................................................................................... 51

2.6 Extraction of genomic DNA .................................................................................. 51

2.7 Polymerase chain reaction (PCR) to detect genomic islands ................................. 52

2.7.1 Introduction ..................................................................................................... 52

2.7.2 PCR to detect the presence of genomic islands .............................................. 54

2.7.3 PCR to confirm absence of genomic islands .................................................. 55

2.7.4 Analysis of genomic island data ..................................................................... 59

2.8 MLST of *B. pseudomallei* ................................................................................... 59

2.8.1 Primers ............................................................................................................ 59

2.8.2 PCR ................................................................................................................. 59

2.8.3 Clean-up of PCR products .............................................................................. 60

2.8.4 Sequencing ...................................................................................................... 60
Chapter 3. Results I: *Burkholderia pseudomallei* genome plasticity: genomic island variation in relation to virulence and origins

3.1 Chapter content ...................................................................................................... 65

3.2 Bacterial isolates .................................................................................................... 66

3.3 Genomic island identification ................................................................................ 67

3.4 Clinical definitions ................................................................................................. 67

3.5 Analysis .................................................................................................................. 68

3.6 Results and discussion ........................................................................................... 68

3.6.1 Overview of five genomic islands in *B. pseudomallei* strain K96243 .......... 68

3.6.2 Sampling framework for *B. pseudomallei* isolates ......................................... 70

3.6.3 Presence of five genomic islands in *B. pseudomallei* ..................................... 74

3.6.4 Comparison of the presence of genomic islands in soil and invasive isolates 74

3.6.5 Relationship between clinical features and genomic islands in disease-associated isolates .................................................................................................... 75

3.6.6 Variability in genomic islands between clones of *B. pseudomallei* .......... 79

3.6.7 *in silico* analysis of genomic islands ............................................................... 81

3.6.8 Chapter summary ............................................................................................ 87

Chapter 4. Results II: Distribution of *B. pseudomallei* clones in relation to geographical location and virulence

4.1 Chapter content ...................................................................................................... 88
Chapter 5. Results III: Multilocus sequence typing of *Burkholderia thailandensis*

5.1 Chapter content ................................................................. 101

5.2 Bacterial strains ................................................................. 101

5.3 Results ................................................................. 103

5.3.1 Allelic distribution and population structure of *B. thailandensis* .......... 103

5.3.2 Diversity of *B. thailandensis* Thai isolates ................................ 111

5.3.3 Comparison between Thai isolates of *B. thailandensis* and *B. pseudomallei* ................................................................. 113

5.4 Discussion ................................................................. 115

5.5 Chapter Summary ................................................................. 116
Chapter 6. Concluding comments

References.................................................................................................................. 118

Appendices................................................................................................................. 142

Appendix I: Composition and preparation of bacterial medium ....................... 142

Appendix II: Composition and preparation of Agarose gel electrophoresis........ 144

Appendix III: Gene contents of five genomic islands ....................................... 145

Appendix IV: Sequence type and presence of genomic islands for 186 B. pseudomallei isolates ................................................................. 148

Appendix V: Sequence type of 79 B. pseudomallei parotitis isolates .......... 152

Appendix VI: Sequence type and strain information of all 92 B. thailandensis isolates ................................................................. 154
Tables

Table 2.1 Description of 5 genomic islands of *B. pseudomallei* examined in this study

Table 2.2 Primer sequences and grouping of primers used during multiplex PCR

Table 2.3 Primer sequences used to amplify region across putative insertion site of genomic islands

Table 2.4 Primer pairs and PCR cycling conditions used for MLST of *B. pseudomallei*

Table 2.5 Primer pairs used for sequencing of amplification products

Table 3.1 Overview of gene content within five genomic islands of *B. pseudomallei* K96243

Table 3.2 Presence of five genomic islands in 186 *B. pseudomallei* isolates

Table 3.3 Cumulative presence of genomic islands in 186 *B. pseudomallei* isolates

Table 3.4 Relationship between clinical factors and the presence of five genomic islands

Table 3.5 Presence of genomic islands in the three largest bacterial clones as defined by MLST

Table 3.6 Distribution of islands in sequenced *B. pseudomallei* strains

Table 4.1 Allele frequency and diversity for 266 Thai isolates of *B. pseudomallei*

Table 4.2 Sources of *B. pseudomallei* isolates in the 10 largest MLST clones for northeast Thai strains

Table 5.1 Data on 14 *B. thailandensis* isolates obtained from the MLST database

Table 5.2 Allelic data for *B. thailandensis* isolates

Table 5.3 Allelic diversity for 86 *B. thailandensis* and 266 *B. pseudomallei* from Thailand
Figures

Chapter 1

Figure 1.1 B. pseudomallei in culture .................................................................4
Figure 1.2 Worldwide distribution of melioidosis .................................................10
Figure 1.3 Overview of MLST ............................................................................42

Chapter 2

Figure 2.1 PCR strategy to confirm lack on genomic island ...............................57

Chapter 3

Figure 3.1 Geographic origin of 83 B. pseudomallei isolates cultured from the environment in Ubon Ratchathani, northeast Thailand ..................72
Figure 3.2 Geographic origin of 103 B. pseudomallei isolates cultured from patients with melioidosis in Ubon Ratchathani, northeast Thailand ....................73
Figure 3.3 Comparison of GI 11 in B. pseudomallei strains K96243, 1106a and 1106b. ...........................................................................................................84
Figure 3.4 Comparison of GI 16 in B. pseudomallei strains 1655, K96243 and 1710b. ...........................................................................................................86

Chapter 4

Figure 4.1 eBURST of 266 B. pseudomallei isolates obtained from the environment (n=83) or associated with human melioidosis (n=183), together with 158 isolates from Northern Australia which were mainly disease associated .................................92
Figure 4.2 Neighbor-joining tree using the concatenated sequences of all seven loci for Thai and Australian isolates (n=266 and n=158, respectively) ..............98
Chapter 5

Figure 5.1 UPGMA dendrogram of 92 *B. thailandensis* isolates ......................... 105

Figure 5.2 eBURST for 92 *B. thailandensis* isolates............................................ 106

Figure 5.3 Neighbor-joining tree using concatenated sequence of all 7 loci for *B. thailandensis* isolates. ........................................................................................................... 108

Figure 5.4 Neighbor-joining tree using concatenated sequence of all 7 loci for *B. thailandensis*, *B. pseudomallei*, *B. mallei*, *B. oklahomensis* and *Burkholderia* spp. 109

Figure 5.5 eBURST for 86 *B. thailandensis* isolates ............................................. 112
# Glossary of terms

<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACT</td>
<td>Artemis Comparison Tool</td>
</tr>
<tr>
<td>ASA</td>
<td>Ashdown selective agar</td>
</tr>
<tr>
<td>ASM</td>
<td>The American Society for Microbiology</td>
</tr>
<tr>
<td>bipD</td>
<td>a cell invasion protein (in TTS)</td>
</tr>
<tr>
<td>bp</td>
<td>base pair</td>
</tr>
<tr>
<td>BSL3</td>
<td>Biosafety containment level three</td>
</tr>
<tr>
<td>BURST</td>
<td>based upon related sequence types</td>
</tr>
<tr>
<td>CC</td>
<td>clonal complex</td>
</tr>
<tr>
<td>CDS</td>
<td>coding sequence</td>
</tr>
<tr>
<td>cfu</td>
<td>colony forming unit</td>
</tr>
<tr>
<td>CI</td>
<td>Classification index</td>
</tr>
<tr>
<td>cm</td>
<td>centimeter</td>
</tr>
<tr>
<td>CRP</td>
<td>C-reactive protein</td>
</tr>
<tr>
<td>CT</td>
<td>computed tomography</td>
</tr>
<tr>
<td>DLV</td>
<td>double locus varience</td>
</tr>
<tr>
<td>DNA</td>
<td>deoxyribonucleic acid</td>
</tr>
<tr>
<td>dNTP</td>
<td>deoxy nucleotide triphosphate</td>
</tr>
<tr>
<td>DR</td>
<td>directed repeat</td>
</tr>
<tr>
<td>BURST</td>
<td>based upon related sequence types</td>
</tr>
<tr>
<td>ELISA</td>
<td>Enzyme-linked immunosorbent assay</td>
</tr>
<tr>
<td>EHEC</td>
<td>enterohaemorrhagic E. coli</td>
</tr>
<tr>
<td>EPEC</td>
<td>enteropathogenic E. coli</td>
</tr>
<tr>
<td>fliC</td>
<td>flagellin gene</td>
</tr>
<tr>
<td>G-CSF</td>
<td>granulocyte colony-stimulating factor</td>
</tr>
<tr>
<td>GI</td>
<td>genomic island</td>
</tr>
<tr>
<td>HR</td>
<td>hazard ratio</td>
</tr>
<tr>
<td>IFA</td>
<td>indirect fluorescent assay</td>
</tr>
<tr>
<td>IHA</td>
<td>indirect hemagglutination assay</td>
</tr>
<tr>
<td>IS</td>
<td>insertion sequence</td>
</tr>
<tr>
<td>Kb</td>
<td>kilobase</td>
</tr>
<tr>
<td>kDa</td>
<td>kilodalton</td>
</tr>
<tr>
<td>LB</td>
<td>Luria-Bertani broth</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>LEE</td>
<td>locus of enterocyte effacement</td>
</tr>
<tr>
<td>LPS</td>
<td>lipo-polysaccharide</td>
</tr>
<tr>
<td>Mb</td>
<td>megabase</td>
</tr>
<tr>
<td>MLEE</td>
<td>multilocus enzyme electrophoresis</td>
</tr>
<tr>
<td>MLST</td>
<td>multilocus sequence typing</td>
</tr>
<tr>
<td>m²</td>
<td>meter squared</td>
</tr>
<tr>
<td>MIC</td>
<td>minimum inhibitory concentration</td>
</tr>
<tr>
<td>min</td>
<td>minute</td>
</tr>
<tr>
<td>mmHg</td>
<td>millimeters of mercury</td>
</tr>
<tr>
<td>mL</td>
<td>milliliter</td>
</tr>
<tr>
<td>OD</td>
<td>optical density</td>
</tr>
<tr>
<td>PBS</td>
<td>phosphate-buffered saline</td>
</tr>
<tr>
<td>PCR</td>
<td>polymerase chain reaction</td>
</tr>
<tr>
<td>PFGE</td>
<td>pulsed field gel electrophoresis</td>
</tr>
<tr>
<td>PAI</td>
<td>pathogenicity island</td>
</tr>
<tr>
<td>RNA</td>
<td>ribonucleic acid</td>
</tr>
<tr>
<td>rpsU</td>
<td>ribosomal protein subunit</td>
</tr>
<tr>
<td>rRNA</td>
<td>ribosomal RNA</td>
</tr>
<tr>
<td>tRNA</td>
<td>transfer RNA</td>
</tr>
<tr>
<td>sec</td>
<td>second</td>
</tr>
<tr>
<td>SLV</td>
<td>single locus variant</td>
</tr>
<tr>
<td>ST</td>
<td>sequence type</td>
</tr>
<tr>
<td>TMP-SMX</td>
<td>trimethoprim-sulfamethoxazole</td>
</tr>
<tr>
<td>TSB</td>
<td>triptcise soy broth</td>
</tr>
<tr>
<td>TTS</td>
<td>type three secretion system</td>
</tr>
<tr>
<td>µg</td>
<td>microgram</td>
</tr>
<tr>
<td>UPGMA</td>
<td>unweighted pair group method with arithmetic mean</td>
</tr>
<tr>
<td>UV</td>
<td>ultraviolet</td>
</tr>
<tr>
<td>VNTR</td>
<td>variable number tandem repeat</td>
</tr>
</tbody>
</table>


Abstract

Burkholderia pseudomallei is a soil-dwelling saprophyte and the cause of melioidosis. Multilocus sequence typing (MLST) of 266 Thai B. pseudomallei isolates (83 soil and 183 invasive) defined 123 sequence types (STs). Invasive isolates were over-represented in the 10 largest clones, and there was a significant difference in the classification index between environmental and disease isolates, confirming that genotypes were not distributed randomly between the two samples. MLST profiles for 158 predominantly invasive isolates from northern Australia contained a similar number of STs (96) as the Thai invasive isolates, but no ST was found in both populations. This analysis revealed strong genetic differentiation on the basis of geographical isolation, and a significant differentiation on the basis of virulence potential.

The presence and distribution of five genomic islands described for B. pseudomallei K96243 was investigated in natural populations. In silico analysis of 10 B. pseudomallei genome sequences demonstrated variable presence of these regions, together with evidence for micro-evolutionary changes that generates GI diversity. Their presence was assessed in environmental (83) and invasive (103) B. pseudomallei isolates using PCR. Positivity ranged from 12% for a prophage-like island (GI 9), to 76% for a metabolic island (GI 16). The presence of each of the five GIs did not differ between environmental and disease-associated isolates. There was no reproducible association between the individual or cumulative presence of five GIs and clinical features in 103 patients with melioidosis.

Burkholderia thailandensis is a non-pathogenic soil saprophyte which is highly related to B. pseudomallei. MLST of a collection of B. thailandensis isolates demonstrated two clusters. One cluster contained three isolates (one each from...
France, the U.S. and Cambodia), and the second larger cluster contained isolates from Thailand, Vietnam and Laos. eBURST analysis indicated that the larger cluster formed a single clonal complex, while all three strains of the small cluster were singletons.
Chapter 1. Introduction

1.1 General background to *Burkholderia pseudomallei*

1.1.1 Historical aspects

*Burkholderia pseudomallei* is the cause of a serious human infection termed melioidosis. This organism was originally described in 1912 by Whitmore and Krishnaswami who were working as pathologists in Rangoon, Burma (now Yangon, Myanmar). The term "melioidosis" was introduced in 1921 by Stanton and Fletcher to describe a disease of horses that was similar to, but not the same as glanders (which is caused by the highly related *Burkholderia mallei*). This word is derived from Greek; 'melis' means a distemper of asses and 'eidos' means resembles glanders. Melioidosis is clinically and pathologically similar to glanders, but the ecology and epidemiology of melioidosis are markedly different. *B. pseudomallei* is now regarded as a major tropical pathogen, and has been classified by the United States Centers for Disease Control as a category B biothreat agent. This organism is classified as a biological hazard group 3 agent in Europe and many other countries of the world and must be handled in BSL3 containment facilities.

1.1.2 Classification and description

*B. pseudomallei* has been given a variety of names including *Bacillus pseudomallei*, *Bacillus whitmorii*, *Pfefferella whitmorii*, *Actinobacillus pseudomallei*, *Loefflerella whitmorii*, *Malleomyces pseudomallei*, *Pfefferella pseudomallei* and *Pseudomonas pseudomallei*. In 1973, it was grouped within the genus *Pseudomonas* as a *Pseudomonas* RNA homology group II. In 1992, members of this group were transferred to the new genus, *Burkholderia*, based on 16S rRNA sequence, DNA-
DNA homology, cellular lipid and fatty acid composition and phenotypic characteristics.\textsuperscript{276}

\textit{B. mallei}, the causative agent of glanders, and \textit{B. thailandensis}, a normally non-virulent organism, are genetically closely related to \textit{B. pseudomallei}. These species can be differentiated from \textit{B. pseudomallei} by the characteristics of motility and arabinose assimilation. \textit{B. mallei} is non-motile due to its inability to express flagella,\textsuperscript{132} compared with \textit{B. pseudomallei} and \textit{B. thailandensis} which are motile. \textit{B. thailandensis} assimilates arabinose while \textit{B. pseudomallei} and \textit{B. mallei} do not. Arabinose utilization is determined by growth on minimal salt agar containing 0.2\% L-arabinose.\textsuperscript{215} Both \textit{B. pseudomallei} and \textit{B. mallei} have a deleted operon encoding arabinose fermentation.\textsuperscript{162}

\textit{B. pseudomallei} is a small (~0.6 x 3 \textmu m) motile, Gram-negative bacterium. It is non-spore forming, oxidase-positive and grows aerobically but can survive under anaerobic conditions in the presence of nitrate or arginine.\textsuperscript{276} Optimum temperature for growth of \textit{B. pseudomallei in vitro} is 37 – 42°C; the organism is unable to grow at temperatures less than 21°C.\textsuperscript{180} However, some strains appear to be able to survive for prolonged periods (greater than 190 days) at 5°C under laboratory conditions.\textsuperscript{277} \textit{B. pseudomallei} is positive for catalase and indophenoloxidase. The bacterium usually exhibits bipolar staining (likened to a safety pin) on Gram stain because of intracellular deposits of \textbeta-hydroxybutyric acid. The accumulation of prominent granules of \textbeta-hydroxybutyric acid as an energy store reflects a metabolism adapted to long-term survival.\textsuperscript{104} \textit{B. pseudomallei} in clinical specimens may not have bipolar staining, and may rarely be observed on Gram stain as long filaments.

The colony appearance of \textit{B. pseudomallei} on solid agar differs depending on the medium used. On blood agar the colonies are smooth and shiny. Ashdown's
selective medium is commonly used to isolate the organism from clinical specimens containing mixed flora. This medium contains gentamicin, crystal violet and glycerol. The typical colony morphology after 48 hours or more of incubation in air at 37°C is a wrinkled, purple colony which has been likened to a cornflower head (Figure 1.1). Less common morphologies have been described on Ashdown’s agar, including a smooth colony type. Colonies growing on agar are described as having an earthy odour, although sniffing of plates is prohibited for safety reasons. In vitro subculture on ferric citrate-containing medium leads to alteration in colony morphology on Ashdown’s agar. It can change into either a larger, glistening, mucoid, dark purple colony type or a smaller, flat, light purple colony type.

When B. pseudomallei is grown in broth culture, a dry pellicle with a matt surface often forms. This pellicle forms at the broth-air interface, possibly because of positive aerotaxis of B. pseudomallei.
B. pseudomallei is intrinsically resistant to many antibiotics, including first, second, and third-generation cephalosporins, aminoglycosides, penicillins and polymyxin. The organism is usually susceptible to ceftazidime, the carbapenem antibiotics, amoxicillin-clavulanate, piperacillin-tazobactam, doxycycline and trimethoprim-sulfamethoxazole (TMP-SMX). Carbapenem antibiotics have the greatest activity in vitro, while fluoroquinolones have poor activity in vitro. Rare spectrum cephalosporins due to the production of a beta-lactamase. Phenotypic characterization of the beta-lactamase produced by B. pseudomallei has shown it to be a weakly inducible, membrane associated chromosomally encoded cephalosporinase which is slowly active against carbenicillin, ceftazidime and cefotaxime and inactivated by clavulanic acid. Sequence analysis of the beta-lactamase gene *penA* (or *blaA*) of B. pseudomallei revealed conserved motifs typical of class A beta-lactamases and a relationship to *penA* (in *B. cepacia*) and *blat* (in *Yersinia enterocolitica*) lineages. Macrolide and aminoglycoside resistance is mediated via a multidrug efflux pump.

The genome of B. pseudomallei strain K96243 (originating from Thailand) has been sequenced and comprises two chromosomes of 4.97 Mb and 3.17 Mb. The large chromosome carries many genes associated with core functions such as cell growth and metabolism, while the smaller chromosome carries more genes encoding...
**B. pseudomallei** is intrinsically resistant to many antibiotics, including first, second and third-generation cephalosporins, aminoglycosides, penicillins and polymyxin. The organism is usually susceptible to ceftazidime, the carbapenem antibiotics, amoxicillin-clavulanate, piperacillin-tazobactam, doxycycline and trimethoprim-sulfamethoxazole (TMP-SMX). Carbapenem antibiotics have the greatest activity *in vitro*, while fluoroquinolones have poor activity *in vitro*. Rare clinical isolates are susceptible to gentamicin and macrolides, occurring in the order of approximately 1 in every 1000 isolates. The beta-lactams and ciprofloxacin are bactericidal *in vitro*, whereas the agents conventionally used to treat melioidosis during the oral phase of treatment (doxycycline and TMP-SMX) have bacteriostatic activity only. A detailed summary of *in vitro* activities of antibiotics against *B. pseudomallei* can be found in a review by Cheng and Currie, 2005.

*B. pseudomallei* is inherently resistant to ampicillin and broad- and expanded-spectrum cephalosporins due to the production of a beta-lactamase. Phenotypic characterization of the beta-lactamase produced by *B. pseudomallei* has shown it to be a weakly inducible, membrane associated chromosomal cephalosporinase which is strongly active against carbenicillin, cefotaxime and cefuroxime, and inactivated by clavulanic acid. Sequence analysis of the beta-lactamase gene penA (or blal; Ambler class A) revealed conserved motifs typical of class A beta-lactamases and a relationship to penA (in *B. cepacia*) and blal (in *Yersinia enterocolitica*) lineages. Macrolide and aminoglycoside resistance is mediated via a multidrug efflux pump.

The genome of *B. pseudomallei* strain K96243 (originating from Thailand) has been sequenced and comprises two chromosomes of 4.07 Mb and 3.17 Mb. The large chromosome carries many genes associated with core functions such as cell growth and metabolism, while the smaller chromosome carries more genes encoding...
accessory functions that could be associated with adaptation and survival in different environments. Around 6% of the genome is made up of putative genomic islands that have probably been acquired via horizontal gene transfer. These are mostly absent from the *B. thailandensis* genome (and are absent from the *B. mallei* genome); it is unclear whether these regions have a role in disease pathogenesis. The Institute of Genomic Research (Rockville, MD, USA) is currently sequencing nine further *B. pseudomallei* isolates and 25 *B. pseudomallei* bacteriophage genomes originating from various sources. In addition, sequencing of the *B. thailandensis* (strain E264) genome is now completed (www.tigr.org) and published.\textsuperscript{128}

1.2 *B. pseudomallei* in the environment

*B. pseudomallei* is a saprophytic organism which can be found in soil or surface water across much of southeast Asia and Northern Australia. Variables that may affect the distribution of the organism in soil include climatic factors such as temperature, pH, moisture (rainfall), sunlight (UV), and soil composition (physical, chemical and biological).\textsuperscript{117} Well-drained, light, sandy soils are less able to support the prolonged persistence of *B. pseudomallei* than waterlogged, heavy clay soils.\textsuperscript{111} Water content in soil has been reported to determine the duration of *B. pseudomallei* survival in the laboratory. This bacterium was able to survive for 70 days if the water content in soil was less than 10%, and was prolonged to 726 days if there was more than 40% water content.\textsuperscript{244,273}

In melioidosis-endemic regions, environmental conditions have been shown to play a role in growth and persistence of *B. pseudomallei* in soil. It was reported that a higher *B. pseudomallei* count was present in superficial clay soil compared to deeper samples during the rainy season.\textsuperscript{240} However, the overall isolation rate was higher
during the dry season than during the wet season. This could be the result of better aeration of soil.\textsuperscript{272} The organism can resist drying, since \textit{B. pseudomallei} can be cultured from dry soil 4 weeks after inoculation in the laboratory, provided suitable resuscitation methods using serial subculture on 5\% horse blood agar were used in the initial stages of bacterial isolation.\textsuperscript{115} \textit{B. pseudomallei} has also been reported to survive in water at room temperature for up to 8 weeks, in muddy water for up to 7 months, in soil in laboratory conditions for up to 30 months\textsuperscript{239} and in sterile triple-distilled water without any addition of nutrients for more than 3 years.\textsuperscript{273}

Wet rice fields have been reported to reach a temperature of 40-43°C,\textsuperscript{229} a temperature that is permissive for \textit{B. pseudomallei} growth. Soil temperature is usually greater in cleared land than in forested areas, which might account for the greater yield of \textit{B. pseudomallei} in rice fields.\textsuperscript{163,217} Evidence for persistence in the soil within non-tropical climates includes survival of \textit{B. pseudomallei} for several years in the environment during the French outbreak in the 1970’s.\textsuperscript{160} This affected a zoo in the Jardin des Plantes, equestrian clubs, and human contacts associated with extensive environmental contamination.\textsuperscript{64}

pH is another factor that can influence the survival of \textit{B. pseudomallei}. The organism can be isolated from soil and water at a wide range of pH (2.0 – 9.0),\textsuperscript{84} but the optimum pH range for \textit{B. pseudomallei} has been reported to be between pH 5 and 8. Under laboratory conditions, a pH below 4.5 has been reported to lead to rapid bacterial inactivation which may represent a viable but non-culturable form.\textsuperscript{244} The soil from which \textit{B. pseudomallei} was most commonly isolated in the highly endemic region of Thailand was reported to be acidic.\textsuperscript{123} The survival of \textit{B. pseudomallei} may be favored by the relatively acidic environment of rice paddy fields, which are pH 4.4-7.7 in northeast Thailand.\textsuperscript{123} It has also been reported that \textit{B. pseudomallei} was
more acid tolerant and adaptable to starvation than *Burkholderia cepacia* and *Pseudomonas aeruginosa* under laboratory conditions.\(^{70}\)

Other factors that may effect the survival of this organism are aeration and UV exposure. During the rainy season, in Thailand the isolation rate increased with increasing depth, but remained uniform between 30 – 60 cm depth during the dry season.\(^{272}\) The mean *B. pseudomallei* count from soil sampled from Northeast Thailand was 20-fold higher than that from Central of Thailand.\(^{217}\) In Australia, one study reported that samples from clay soils collected at 25 – 45 cm depth were most likely to be positive and surface soil was rarely culture positive.\(^{240}\) *B. pseudomallei* could be found in clay at depths of 120 cm.\(^{29}\) Tong *et al* (1996)\(^{244}\) reported that under laboratory conditions, *B. pseudomallei* was more easily killed by UV light than other soil bacteria. However, at present, there is little evidence to suggest that *B. pseudomallei* is significantly different from other bacterial species in the ability to resist UV light or repair the damage that UV light induces.\(^{117}\) The effect of osmotic stress has been examined using variable sodium chloride (NaCl) concentrations. *B. pseudomallei* can survive in NaCl at less than 2.5% wt/vol.\(^{106}\) Little is known about survival of this organism in sea water.\(^{117}\)

Survival of *B. pseudomallei* following internalization into amoebic cysts has been reported.\(^{116}\) Persistence in amoebic cysts may represent a means of surviving stressful environmental conditions, although this has not yet been proven in the natural environment. Amoebic species in which *B. pseudomallei* can survive include *Acanthamoeba astronyxis*, *Acanthamoeba castellanii*, *Acanthamoeba palestiniensis* and *Acanthamoeba polyphaga*.\(^{116}\) There are several similarities between *B. pseudomallei* and *Legionella* spp., both of which are environmental saprophytes capable of intracellular parasitism. *B. pseudomallei* employs the coiling phagocytosis
mechanism to penetrate free-living amoeba. Bacillus-filled amoebic vacuole size is close to that of aerosolized particles that can enter the human respiratory tract. As in the case of Legionella, the infective particle for melioidosis and the normal intracellular habitat may be the same entity.\textsuperscript{28} It is known that virulence of Legionella is enhanced on amoebic passage,\textsuperscript{53} but this has not been studied for \textit{B. pseudomallei}.

Additional putative environments for \textit{B. pseudomallei} are reported to be the cytoplasm of arbuscular mycorrhizal fungi, legumes (nitrogen-fixing plants) and other plants such as rice, rubber and cassava.\textsuperscript{116,141,163,175,234} It had been proposed that \textit{B. pseudomallei} can reduce nitrate in order to grow in an anaerobic environment in association with legumes, and the use of nitrate fertilizers might thus contribute to its proliferation in agricultural land. These relationships with other organisms may not only play an important role in bacterial survival, but in potential ecological niches and the distribution of \textit{B. pseudomallei}.

1.3 Geographical distribution of \textit{B. pseudomallei}

1.3.1 Global epidemiology

\textit{B. pseudomallei} is present in the environment between latitudes 20°N and 20°S of tropical or sub-tropical climates. South East Asia and northern Australia are notable endemic areas for melioidosis. Most reported cases of melioidosis are associated with agricultural occupations (particularly rice farming in Thailand), or other activities leading to exposure to soil and water. The numbers of cases in endemic areas are highest during the rainy season. In non-endemic areas, cases of melioidosis involve infection in returning travelers and in transported animals. The worldwide distribution of melioidosis as it is currently understood is summarized in Figure 1.2.
The first report of melioidosis within this region involved a diabetic patient in Townsville in 1950,[13,14] one year after an outbreak in sheep in Winton, Northern Queensland.[19] Most reported cases in Australia occur in the northern territories. *B. pseudomallei* has since been isolated from the environment in Northern Australia, and it is estimated that the incidence among the general population is between 1 in 4,000 and 1 in 10,000.[20,21] The first case report of melioidosis was in a serviceman stationed in PNG during World War II.[22] Thereafter, a small number of documented cases from Port Moresby, Central PNG,[23,24] and other areas of PNG have been reported.[25] A series of confirmed cases of melioidosis have been described in Balimo, Western PNG.[26] Two small serological studies failed to detect antibody to *B. pseudomallei*,[27,28] but a larger serological study conducted in Balimo reported that the seroprevalence and estimate annual incidence rate (6.3–13.4% and 20 per 100,000) were similar to the Northern Territory of Australia (5–13% and 16.5–34.9 per 100,000).[29,30]

Figure 1.2 Worldwide distribution of melioidosis (reproduce from[35]).
1.3.1.1 Australia – Pacific region

The first report of human melioidosis within this region involved a diabetic patient in Townsville in 1950,\textsuperscript{54,197} one year after an outbreak in sheep in Winton, Northern Queensland.\textsuperscript{54} Most reported cases in Australia occur in the northern territories. \textit{B. pseudomallei} has since been isolated from the environment in Northern Australia, including paddocks, moist soil and mud, vegetable matter and water reservoirs.\textsuperscript{29,61,111,239,240}

Other regions outside of northern Australia including the southwest, central and eastern regions have also reported indigenous melioidosis.\textsuperscript{57,91,209} Infection has also occurred in relation to a contaminated drinking supply and pipes in Western and Northern Australia.\textsuperscript{63,113,114} Serosurveys of populations throughout Australia that defined the presence of antibodies to \textit{B. pseudomallei} have demonstrated that seropositivity in the Northern Territory is higher than Queensland,\textsuperscript{59} a finding that is consistent with disease incidence.

Papua New Guinea (PNG), particularly the western region, has been reported to be a melioidosis endemic region.\textsuperscript{59,261} The first case report of melioidosis was in a serviceman stationed in PNG during World War II.\textsuperscript{129} Thereafter, a small number of human cases and first bacterial isolation (from a tree climbing kangaroo)\textsuperscript{76} were documented from Port Moresby, Central PNG.\textsuperscript{55,134,200} A series of confirmed cases of melioidosis have been described in Balimo, Western PNG.\textsuperscript{260} Two small serological studies failed to detect antibody to \textit{B. pseudomallei},\textsuperscript{22,200} but a larger serological study conducted in Balimo reported that the seroprevalence and estimate annual incidence rate (6.2 – 10.4% and 20 per 100,000) were similar to the Northern Territory of Australia (5 – 13% and 16.5 – 34.5 per 100,000).\textsuperscript{59,261}
Six cases of melioidosis have been diagnosed between 1999 and 2004 in New Caledonia, which is located in Oceania, east of Queensland, Australia. All were from the Northern Province. Molecular typing of two of the associated isolates suggested a link to an Australian strain.¹³³

1.3.1.2 Asia

Thailand: Thailand is considered to be the world epicentre of melioidosis, reporting more than 1,000 cases each year from the northeast of the country alone. Yet the first case in an indigenous Thai patient was not described until 1955,⁵⁰ and the second report only appeared in 1974.¹⁷⁶ Since then there has been rapid economic development, and microbiology facilities in provincial hospitals have improved considerably. By 1985, Thai investigators were able to review a detailed clinical experience of over 800 cases of melioidosis.¹³⁸ This illustrates how closely the apparent epidemiology of melioidosis (and many other tropical infections) is linked with the availability of microbiology facilities and trained personnel.²⁴¹ Melioidosis is now recognized as the most common cause of severe community-acquired bacteraemia in northeast Thailand where it accounts for 20% of cases and 40% of sepsis-related deaths.³⁶ The calculated incidence rate in northeast Thailand of 4.4/100,000²³² is likely to be a significant underestimate because of continued ascertainment difficulty.⁵⁹,²⁶² Around three quarters of cases present during the rainy season, with a strong association with rainfall.²³²

PDR Laos: The first ever introduction of a blood culture system at Mahosot Hospital in Vientiane, the capital of Lao PDR, has led to the description of a significant burden of disease from melioidosis.¹⁷⁹ B. pseudomallei has also been isolated from soil taken
from areas surrounding the capital city. The number of melioidosis cases in other areas of Laos is currently unknown.

**Vietnam:** This country came under the spotlight as a region endemic for melioidosis during WWII and then during the conflict with the US. Between 1965 until the withdrawal of U.S. forces in 1972, the Melioidosis registry of the United State Army Office of the Surgeon-General (OTSG-MR) received reports of 343 cases, with 15 deaths, and a further 21 deaths for which melioidosis was a contributor. Serological studies indicated that between 25,000 and 225,000 army personnel were exposed.

A significant proportion of the infections reported by the military presented were sub-acute or chronic pulmonary infections (10% of the cases in American soldiers were misdiagnosed initially as pulmonary tuberculosis). In some the latency between exposure and disease was months or years (the disease has been labelled the "Vietnamese time-bomb"). A soil survey has demonstrated the presence of *B. pseudomallei* in soil sampled around Ho Chi Minh City.

**Singapore, Malaysia and Indonesia:** Melioidosis has been recognized as an important human infectious disease in Singapore since the 1990’s. The first reported case occurred in 1989. The estimated annual incidence in Singapore is 1.7 per 100,000 population.

In Malaysia in the early 1910s, melioidosis caused an outbreak among laboratory animals in the Institute for Medical Research of the Federated Malay States. Seven patients were diagnosed with acute septicaemic melioidosis in Kuala Lumpur between 1976 to 1979, and another 10 cases were diagnosed between 1981 and 1986. Melioidosis became increasingly recognised thereafter, with a reported
85 cases from both Peninsular and east Malaysia, and 50 cases from Kuala Lumpur. A total of 135 cases diagnosed between 2000 and 2003 were reported from Pahang, of whom 54% died.

Indonesia is likely to be endemic for melioidosis, although reported cases are limited to infection in returning travelers to Australia, United Kingdom and United States, and to cases that occurred in the aftermath of the tsunami of 2004.

China, Hong Kong and Taiwan: Melioidosis was first reported in Taiwan in 1984, in which a previously healthy patient presented with multilobar pneumonia after near drowning in the Philippines. Several cases have since been described, together with an outbreak. Forty cases were reported after a typhoon in 2005. This confirms that melioidosis is endemic to Taiwan.

Small numbers of cases have been reported in Hong Kong since 1983, while infection in marine mammals in the Hong Kong oceanarium is an on-going problem. In China, 4.2% of environmental soil and water samples were reported to be positive for \textit{B. pseudomallei}. Seroprevalence in Southern China's Hainan Island ranges from 8.8 to 66.7%, with the highest antibody titers found in farmers in areas of Xinglong. A seropositivity rate of 34% has also been reported among rice farmers, but the incidence of melioidosis is currently unknown.

Other parts of Asia: Occasional patients have been reported from India, Sri Lanka, Guam, and the Philippines.
1.3.1.3 Areas outside Asia

Sporadic case reports have been documented in Central and South America, West and East Africa, Europe and the Caribbean, encompassing such widely divergent areas as Brazil, Panama, Mexico, Puerto Rico, Ecuador and other countries of northern South America.\(^{52,66,159}\) \(B.\) pseudomallei has also been reported in Northern America (California).\(^3\) Cases of melioidosis have been reported from Turkey, Iran, Madagascar, Kenya, Central West Africa, South Africa and Europe,\(^65\) but these were predominantly in citizens or ex-servicemen who had returned home after traveling within or emigrating from an endemic area.

1.4 \(B.\) pseudomallei infection

1.4.1 Melioidosis in animals

\(B.\) pseudomallei has an extremely broad host range. Among domestic animals, melioidosis is most commonly reported in sheep, goats and swine. Sporadic cases have been reported in a range of other animals including the monkey, gibbon, orang-utan, kangaroo, wallaby, deer, buffalo, cow, camel, llama, zebra, koala, dog, cat, horse, mule, parrot, certain bird species, rat, hamster, rabbit, guinea pig, ground squirrel, snake, fish, seal, dolphin and crocodile.\(^{224}\)

Outbreaks affecting animals have been reported to occur following importation from endemic areas. Fournier (1965)\(^{86}\) considered that this accounted for the introduction of melioidosis into Aruba, and also into Australia after World War II. The gift of an infected panda from China, the so called “Affaire du Jardin des Plantes” may have precipitated an outbreak in France in 1975.\(^{160}\) Infection subsequently spread to other zoos in Paris and Mulhouse and throughout France, and led to the death or slaughter of an unknown number of animals.\(^{72}\) An outbreak in the United Kingdom
has also been reported to have occurred in primates imported from the Philippines and Indonesia.\textsuperscript{68}

Studies in animal models have been performed including inbred mouse strains, hamsters, chickens, rats and guinea pigs.\textsuperscript{254,255,263} BALB/c mice represent the most common animal model used for the study of \textit{B. pseudomallei} virulence.\textsuperscript{250}

1.4.2 Human melioidosis

1.4.2.1 Risk factors

In northeast Thailand, Suputtamongkol \textit{et al.}, 1994\textsuperscript{232} noted that 81\% of cases of melioidosis occurred in rice farmers or their families. In northern Australia, 72.5\% of patients with melioidosis were reported to be male, and 49\% were Aborigines who have a lifestyle involving frequent exposure to tropical soils and surface water through daily activities and hunting.\textsuperscript{62} Merianos \textit{et al.} (1993)\textsuperscript{155} and Suputtamongkol \textit{et al.} (1999)\textsuperscript{231} reported that the risk of melioidosis was approximately 10-fold higher in those with heavy occupational soil exposure compared with indoor workers, and in Aborigines irrespective of their employment.

Underlying conditions such as diabetes mellitus, renal failure or dysfunction and alcoholism are important risk factors for melioidosis.\textsuperscript{34,62,138,214} Diabetes mellitus is the most commonly recognized risk factor for melioidosis in the major melioidosis-endemic countries including Thailand,\textsuperscript{36,231} Malaysia,\textsuperscript{185} Singapore and Australia.\textsuperscript{60,138,231} In Thailand, the rate of diabetes mellitus in melioidosis cases is reported to be 60\%.\textsuperscript{231} Renal disease is present in 20-27\% of melioidosis patients in Thailand,\textsuperscript{36,231} respectively) compared with 9\% in northern Australia.\textsuperscript{62} Malignancy, chronic lung disease, connective tissue disease and steroid therapy are additional risk factors. Chronic lung disease may be important in circumstances involving infection
following inhalation of *B. pseudomallei*. Alcoholism is a documented risk factor. In Thailand, 12% of melioidosis cases had a history of heavy alcohol consumption, compared with 37% in northern Australia. Alcohol excess appears to be less common as a risk factor in Malaysia and Singapore. This may reflect the differences in alcohol consumption by different populations. Studies using animal models show that acute alcohol intoxication results in significant immune dysfunction. Other risk factors include thalassemia, which is present 7% in Thai patients, iron overload and tuberculosis, although their roles as independent risk factors are yet to be confirmed. HIV does not appear to be a risk factor for melioidosis.

A study of risk factors in northern Australia showed that the absence of risk factors predicted for survival, and that the only individual risk factor associated with mortality was chronic renal disease. The higher mortality rate in those with renal disease was associated with a higher rate of bacteraemia (85%).

### Disease acquisition and incubation period

Melioidosis is limited to individuals exposed to environments containing *B. pseudomallei*. Acquisition of melioidosis may occur via three major routes, *i.e.* inhalation, ingestion and inoculation. Contamination of existing wounds or new penetrating wounds with *B. pseudomallei* present in soil and water is likely to represent the major route of infection in farmers in northeast Thailand. Evidence for inhalation as a route of infection comes from cases of melioidosis in US soldiers involved in the war with Vietnam, when helicopter crews and servicemen were infected after inhalation of particles created by helicopters. Melioidosis may follow aspiration of contaminated water during near drowning, and occurred
following near drowning in several people involved in the tsunami of December 2004. An outbreak associated with contaminated drinking water indicates that ingestion is an additional possible route, although this is probably rare. Other possible modes of infection include laboratory accidents.

Melioidosis is seasonal in the tropics where most cases occur during the rainy season. This can be explained by increased contact with the organism. Rice farmers plant at the start of the monsoon and work in flooded rice paddies until harvest. Thai farmers rarely wear protective footwear and their feet often show signs of repeated trauma and injuries. Extreme weather may be associated with a shift in the mode of acquisition of infection. Aerosols are created during heavy rain, and this may result in repeated inhalation of the organism. Heavy rainfall and winds consistently result in a shift towards more pneumonia in patients presenting with melioidosis in Northern Australia. Severe or penetrating injury and near-drowning are known risk factors for melioidosis, as highlighted by a study of a cluster of melioidosis cases in Southern Thailand following the 2004 tsunami.

Limited person-to-person transmission of \( B. \) pseudomallei has been reported. Transmission may occur in utero, and melioidosis in two infants in Northern Australia was related to breast-feeding by mothers with mastitis caused by \( B. \) pseudomallei. The wife of a Vietnam veteran with chronic prostatitis caused by \( B. \) pseudomallei was shown to have developed an antibody response to the organism in the absence of clinical manifestations of melioidosis. Person-to-person transmission may have occurred between a diabetic brother and sister living in endemic northeast Thailand, and between 2 siblings with cystic fibrosis. Nosocomial melioidosis from a suspected environmental source has been reported from an endemic area, but cross infection between hospital in-patients has not been reported. Vector-borne
transmission via mosquito (*Aedes aegypti*) and rat flea (*Xenopsylla cheopis*) has been described.\(^{167}\)

The period between *B. pseudomallei* exposure and onset of clinical manifestations is highly variable and often difficult to define. In one study, only 25\% of cases reported a specific inoculation event from which an incubation period of 1-21 days (mean 9 days) was derived.\(^{59}\) Aspiration may be associated with a large bacterial inoculum, and the incubation period may be very short.\(^{49}\) The incubation period may also be very prolonged,\(^{51,152}\) the maximum recorded being 62 years.\(^{166}\) Factors that provoke the reactivation of latent *B. pseudomallei in vivo* include stress and reduced immunity.\(^{120,241}\)

The time from onset of disease to clinical presentation is highly variable, reflecting the broad range in disease severity. In Northern Australia, 13\% of patients presenting for the first time had symptoms for more than two months.\(^{59}\) In the Wellcome Unit Melioid Lab in Ubon Ratchathani, northeast Thailand, around a third of patients had symptoms for less than 7 days, one half report being unwell for 7-28 days and the remainder had symptoms for more than 28 days.\(^{270}\)

### 1.4.2.3 Clinical features

Manifestations of disease are extremely broad ranging. Melioidosis has been called the great mimicker because of its protean clinical features.\(^{282}\) These range from asymptomatic seroconversion, acute fulminant infection and chronic disease which can mimic tuberculosis or malignancy.\(^{247,262}\) A common clinical picture is that of a septicemic illness, often associated with bacterial dissemination to distant sites such that concomitant pneumonia and hepatic and splenic abscesses are common. Pneumonia occurs in around 50\% of patients. Infection may also occur in bone, joints,
skin (superficial pustules and cutaneous abscesses), soft tissue (pyomyositis) and prostate.\textsuperscript{59,185,262} A specific syndrome of meningoencephalitis with flaccid paraparesis or peripheral motor weakness occurs in 4% of cases in Northern Australia,\textsuperscript{60} and cerebral abscess is well described.\textsuperscript{31} Involvement of the vascular tree is recognized but unusual. Acute parotitis accounts for one third of childhood cases in Thailand,\textsuperscript{67,262} but is unusual in Thai adults and is rarely seen in any age group outside of Thailand.

\subsubsection*{1.4.2.4 Diagnosis}

Isolation and identification of the organism represents the diagnostic gold standard. \textit{B. pseudomallei} is not thought to exist as a member of the normal flora. This is based on studies of the presence of the organism on throat swab taken from patients with or without melioidosis. In one study, 36\% of 1,011 patients with melioidosis were throat swab positive compared with none of 3,524 healthy subjects or patients with other diseases.\textsuperscript{274} In a second study, 8 of 17 children with melioidosis were throat swab positive, but no swabs were positive for 1000 children admitted to the same ward with other conditions.\textsuperscript{124} Patients admitted to Sappasithiprasong Hospital in Ubon Ratchathani hospital over the last 20 years who were culture positive for \textit{B. pseudomallei} invariably have features consistent with melioidosis (S. Peacock, personal communication). Individuals in endemic areas with contaminated wounds but no signs or symptoms of infection have not undergone systematic study, and the possibility of wound colonization in the absence of disease has not been excluded. Several cases reported to ProMED following the tsunami of 2004 were culture positive for \textit{B. pseudomallei} but did not have symptoms consistent with melioidosis. Full clinical information for these cases has not been published to date. In
the meantime, it is prudent to suspect melioidosis in anyone with a positive culture from any site, and to investigate accordingly.

1.4.2.4.1 Bacterial identification

Samples of blood, urine and respiratory secretions should be obtained for culture from all patients, together with pus and wound swabs where relevant. Early discussion with senior members of the clinical microbiology laboratory is important during investigation of suspected cases. This will raise awareness for the presence of a significant pathogen in a mixed culture. Furthermore, *B. pseudomallei* is classified as a hazard group 3 biological agent and should be cultured using the appropriate containment level. ASM have produced sentinel laboratory guidelines for suspected *Burkholderia mallei* and *B. pseudomallei*. These are useful reference documents for laboratories that are not familiar with culture and identification of *B. pseudomallei*.²

Samples of pus, respiratory secretions, urine and throat swab are plated directly onto non-selective (blood agar) and selective media (MacConkey agar and Ashdown's medium). If Ashdown's medium is not available, *B. cepacia* medium is a good alternative.⁸⁷ Additional steps can be taken to increase the yield of positive *B. pseudomallei* cultures from most samples, as follows. Urine should be plated neat, then spun (3000rpm for 5 min) and the entire pellet plated. Quantitative count of *B. pseudomallei* does not appear to have the same significance in terms of defining urinary tract sepsis, and bacteria in urine are likely to reflect systemic disease rather than isolated urinary sepsis.¹⁴² In view of this, even a single colony in urine should be interpreted as clinically significant. Yield from pus samples will be improved if a 48-hour enrichment step in liquid medium such as TSB is used. Yield from respiratory secretions, throat and wound swabs that contain mixed flora will be enhanced by 48
hours pre-incubation in a selective liquid medium. All media are incubated at 37°C in air. Colonies are usually visible on agar after 24 hours incubation, but may be very small on selective media. All plates should be re-examined after 48 hours incubation.

Additional information can be derived from blood culture. Time to blood culture positivity has prognostic significance, with a mortality rate of 73.7% for those with a positive culture within 24 hours compared with 40.9% in patients with blood cultures positive after 24 hours. The quantitative count of B. pseudomallei is related to outcome; around one quarter of 108 patients examined in northeast Thailand had >100 colony forming units (cfu)/mL blood. Counts of ≤ 1 cfu/mL were associated with a mortality of 42%, compared with a mortality of 96% in those with counts of > 100 cfu/mL.

Bacterial colonies on Ashdown's medium are usually purple, dry and wrinkled ('cornflower head') after 48 hours incubation. Gram stain shows Gram-negative rods; bipolar (safety pin) staining has been described, but this is variably present and not unique to this species. The organism is motile, indole negative, oxidase positive, and resistant to colistin and gentamicin, features that aid identification. Laboratories in endemic settings describe the use of direct immunofluorescent microscopy and latex agglutination methodology for rapid bacterial identification, but these are in-house reagents and are not widely available. Commercial biochemical profile kits such as API20NE have been shown to be highly reliable for identification of strains in Thailand, but a recent study of isolates from a diverse geographical area demonstrated a much lower accuracy (60%). If commercial kits do not identify a suspected B. pseudomallei isolate, then confirmatory identification can be carried out by a reference laboratory using standard biochemical testing and/or PCR amplification and sequencing. Use of conventional and real-time PCR to identify B.
**pseudomallei** direct from blood or from bacterial colonies has been described, but this is not part of current routine identification strategies. Susceptibility testing to ceftazidime, imipenem or meropenem, co-amoxiclav, doxycycline and trimethoprim-sulfamethoxazole should be performed using relevant guidelines. Susceptibility to trimethoprim-sulfamethoxazole should be assessed using one of the established MIC methods; E-test is reliable and easy to perform.

1.4.2.4.2 Serology

Most healthy individuals (>80%) living in northeast Thailand show evidence of seroconversion during childhood (indirect hemagglutination (IHA) titre of ≥1:10), and some healthy individuals have very high titres (≥1:160), but the overwhelming majority will not present with melioidosis during their lifetime. Thus serology has low diagnostic utility in melioidosis-endemic regions. Interpretation of serology in the non-endemic setting requires knowledge of travel history and place of normal residence. Paired sera demonstrating a rising antibody titre to *B. pseudomallei* taken from an individual who does not normally reside in an endemic area supports the diagnosis of melioidosis in the presence of clinical features of disease. Seroconversion in an asymptomatic individual implies exposure, but it is unclear whether one should automatically infer that the patient is either actively infected, or will present with melioidosis after a variable incubation period.

A range of serological tests has been evaluated, most of which are not commercially available and have not been validated in the non-endemic setting. The IHA test has been studied extensively in endemic areas where it is clear that background seropositivity results in poor test specificity when used for diagnostic purposes. Increasing the IHA cut-off leads to a rise in specificity, but a
predictable loss in sensitivity. Evaluation of the IHA in Singapore (where disease is sporadic) indicates that this test is more useful in the low endemic or non-endemic setting. Using a cut-off titre of 1:16, less than 1% of the asymptomatic population had a positive titre, while all 20 patients tested with melioidiosis had titres of 1:16 or greater. The diagnostic cut-off used to define a positive result varies between studies; ≥ 1:10 has been used to define seroconversion in Thailand and titres of ≥ 1:16, 1:40 and 1:160 have all been used to define a reactive or diagnostic titre.

Several other serological tests have been described, including an indirect immunofluorescent assay (IFA), enzyme-linked immunosorbent assay (ELISA) using culture filtrate or affinity purified antigen, and dot immunoassay. Data are not available on their utility and performance compared with IHA in non-endemic areas. An immunochromatographic test kit has been developed for the rapid detection of IgG and IgM antibodies to B. pseudomallei, and this has been evaluated in Thailand and Australia. Based on results for IgG, this kit represents an alternative to the IHA but has not been released onto the market.

1.4.2.4.3 Molecular diagnostics

Real-time PCR has been developed for the rapid detection of B. pseudomallei DNA, including assays that target genes encoding 16S rRNA, flagellin (FliC), ribosomal protein subunit S21 (rpsU), type III secretion systems (TTS1 and TTS2), and two sequences unique to B. pseudomallei designated 8653 and 9438. A prospective clinical evaluation conducted in Darwin, Australia of a real-time PCR assay targeting the TTS1 type III secretion system gene cluster included 33 individuals with culture-confirmed melioidosis; sensitivity and specificity for patient
diagnosis were 91% and 95%, respectively. A retrospective study evaluated the real-time PCR targets 8653 and 9438, using samples collected in northeast Thailand from 28 patients with culture-confirmed melioidosis and 17 patients with bacteremia caused by other pathogens. The sensitivity was 71% and 54% for 8653 and 9438 respectively, and specificity was 82% and 88%, respectively. These findings suggest that detection of *B. pseudomallei* in clinical samples using molecular approaches may have utility for the early diagnosis of melioidosis.

### 1.4.2.5 Treatment

Appropriate antimicrobial agents should be commenced immediately on suspicion of the diagnosis of melioidosis. Ideally, culture should be performed prior to administration of antibiotics but treatment should not be delayed if culture cannot be performed rapidly, as in the case of aspiration of pus under imaging or operative intervention. Collections of pus should be drained where feasible (including operative washout for infected joints). Patients with severe melioidosis associated with septic shock, respiratory failure associated with severe infection or adult respiratory distress syndrome, acute renal failure, and other manifestations of a severe septic illness require intensive care management where this is available. The availability of ICU in different geographical regions is an important contributor to the marked difference in outcome between Australia and Thailand. A predictive scoring system for outcome has been devised based on age at diagnosis, presence of pneumonia, serum urea, serum bilirubin, lymphocyte count, and serum bicarbonate. A number was assigned from 0 to 2 based on the degree of abnormality and a score formed from the sum. A score of $\leq 3$ was associated with a mortality of 8.6%, while $\geq 4$ was associated with a mortality of 44.6%. APACHE II score is an independent predictor of mortality in
melioidosis, and has the advantage of being a routine part of intensive care management.

Routine laboratory tests are useful to detect the onset of acute renal failure, abnormal liver function tests and anemia, all of which are well recognized during severe melioidosis. Arterial blood gases should be taken in patients with lung involvement and/or any evidence of respiratory impairment. One small study indicated that serial serum C-reactive protein levels aid patient management; CRP in five culture-positive patients with complicated disease remained elevated until the infection was brought under control. However, a recent review of CRP in 175 patients found that admission CRP may be normal or only mildly elevated, including patients with severe sepsis, fatal cases, and in relapsed melioidosis.

Chest radiographs should be taken in all patients with suspected melioidosis. Common radiographic patterns include localized patchy alveolar infiltrate, focal, multifocal or lobar consolidation, diffuse interstitial shadowing considered consistent with blood borne spread of infection, pleural effusion, and upper-lobe involvement which may include cavitation. The radiographic pattern may mimic tuberculosis. The development of empyema and/or lung abscess are well recognized, and repeat chest radiographs may be clinically indicated for patients with respiratory involvement. Abdominal ultrasound or CT scan should be performed to exclude the presence of abscesses in liver and spleen. Clinical evidence of prostatic involvement requires appropriate imaging (trans-rectal ultrasound or CT scan). The need for other imaging will depend on clinical features and organ involvement.

Fever clearance is often slow, (median fever clearance time of 9 days), and without evidence of clinical deterioration is not normally sufficient to indicate the need for a change in therapy. Sputum and draining abscess cultures may remain
positive for several weeks in a patient who is otherwise responding to treatment. A patient who has clinical deterioration or persistently positive blood cultures should be viewed as failing treatment, at which stage the need for imaging, drainage of collections and change in antimicrobial therapy should be considered.

Treatment recommendations are based on clinical evidence, supported by a Cochrane review published in 2001 and updated in 2002. Treatment is required to complete 12-20 weeks, or longer if clinical indicated. This is divided into intravenous and oral phases. There are some differences in treatment protocols developed at Sappasithiprasong Hospital, Ubon Ratchathani, Thailand, and the Royal Darwin Hospital, Australia.

Initial parenteral therapy is given for 10-14 days or until clinical response is seen (which ever is the longer). Ceftazidime or a carbapenem antibiotic is the treatment of choice. Ceftazidime is used as first line therapy in Thailand, with a switch to a carbapenem antibiotic in the event of treatment failure on ceftazidime. In Thailand, intravenous amoxicillin-clavulanate is used to treat children and pregnant women and is second line empiric treatment for adults, but this preparation is not available in Australia. Parenteral treatment at the Royal Darwin Hospital consists of ceftazidime, or meropenem plus Granulocyte colony-stimulating factor (G-CSF) if the patient has septic shock. The routine addition of TMP-SMX to ceftazidime or meropenem during the initial intensive therapy phase has recently been discontinued in Darwin, although this drug is still used to treat patients with neurological or prostatic melioidosis in view of its excellent penetration into CSF and prostate.

Oral treatment to complete 12-20 weeks of therapy consists of TMP-SMX alone (Australia) or together with doxycycline (adults in Thailand). First-line oral
treatment for pregnant women and children is amoxicillin-clavulanate; this is also an alternative for adults who cannot tolerate TMP-SMX.

G-CSF was adopted for use in patients with septic shock due to melioidosis at the Royal Darwin Hospital, northern Australia in December 1998. This was associated with a decrease in mortality from 95% to 10%. Risk factors, the duration of illness before presentation, and the severity of illness were similar in patients before and after G-CSF introduction, but improvements in intensive care management also occurred during the study period. A randomized placebo-controlled trial of G-CSF for severe melioidosis has been completed at Sappasithiprasong Hospital, northeast Thailand. This was not shown to improve outcome. A placebo-controlled trial of an intravenous PAF receptor antagonist (lexipafant) in 131 adult Thai patients with suspected severe sepsis included 36 patients with melioidosis; no benefit was detected. Other interventions shown to be of benefit in critically ill septic patients such as goal-directed therapy, intensive glycemic control and activated protein C have not been evaluated in patients with melioidosis.

1.4.2.6 Relapse

Disease relapse is common; recurrent infection occurred in 13% of survivors over 10 years in Northern Australia, with a mean time from initial diagnosis to recurrence of 8 months. This mirrors recurrence rates in Thailand. Choice and duration of oral antimicrobial therapy are the most important determinants of relapse, followed by positive blood culture (HR 1.86; 95%CI, 1.18-2.92) and multifocal distribution (HR 1.95; 95%CI, 1.03-3.67). Patients treated with an appropriate oral antibiotic regimen for 12 to 16 weeks have a 90% decreased risk of relapse (HR 0.10; 95%CI, 0.02-0.44) compared with patients treated for ≤ 8 weeks. Three quarters of
recurrent cases are due to relapse, the remainder representing re-infection by a different strain. Culture and susceptibility testing of the causative isolate is central to further management. Bacterial genotyping is now a necessity for treatment trials that use recurrence rates as an outcome measure, allowing treatment failure to be distinguished from re-infection.

1.4.2.7 Prevention

Exposure to *B. pseudomallei* is very difficult to prevent, especially in rice-farming areas and occupational-related soil exposure. Persons with diabetes and skin lesions should avoid contact with soil and standing water in endemic areas. Protective clothing such as rubber boots and gloves during agricultural work can prevent infection through the feet and hands.

There is no effective vaccine available that protects against *B. pseudomallei* infection. Current approaches under evaluation include conjugate-, DNA-, attenuated- and heterologous vaccines. Attenuated mutants that are invasive but have a reduced ability to multiply in phagocytes have been identified using transposon mutagenesis, and these induce a high degree of protective immunity in mice. A mutant of *B. pseudomallei* that is auxotrophic for branched-chain amino acids induced a protective response against a subsequent challenge with an otherwise lethal dose of wild-type *B. pseudomallei* in an animal model. Mice inoculated with a *B. pseudomallei bipD* mutant were partially protected against subsequent challenge with wild-type *B. pseudomallei*, although immunization with purified BipD protein was not protective. However, it seems unlikely that live attenuated vaccination will be feasible for use in humans. Antibodies raised against *B. pseudomallei* flagellin markedly reduced bacterial motility and provided passive protection against
experimentally induced *B. pseudomallei* infection.\textsuperscript{27} Evaluation of LPS and capsular polysaccharide as subunit vaccines against experimental melioidosis demonstrated partial protection in a mouse model.\textsuperscript{164} Guinea pig inoculation with *B. thailandensis* provided partial protection from a subsequent challenge with virulent *B. pseudomallei*.\textsuperscript{110}

It was recently postulated that the T-cell response to primary *B. pseudomallei* infection is biphasic, with an early cytokine – most notably IFN\(\gamma\) - induced phase in which T cells appear to be functionally redundant for initial bacterial clearance, followed by a later antigen-induced phase in which *B. pseudomallei*-specific T-cells - in particular CD4\(+\) T-cells are important for host resistance.\textsuperscript{77} In view of the fact that *B. pseudomallei* infection generates a rapid and potent IFN\(\gamma\) response from NK T- and conventional T cells, it is has been suggested that an effective subunit vaccine against *B. pseudomallei* should target the generation of IFN\(\gamma\)-secreting T cells.\textsuperscript{98} This was further underscored by a recent report suggesting that any potential vaccine would need to stimulate both cell-mediated and humoral immunity.\textsuperscript{100} In this study, dendritic cells were utilized as a vaccine delivery vector to induce cell-mediated immune responses to *B. pseudomallei*. Purified dendritic cells were pulsed with heat-killed whole-cell *B. pseudomallei* and used to immunize syngeneic mice. Strong cellular immune responses were elicited, although antibody responses were low. Subsequently, booster immunizations of either a second dose of dendritic cells or heat-killed *B. pseudomallei* were administered to increase the immune response. Immunized animals were challenged with fully virulent *B. pseudomallei*, and protection was demonstrated in those with both a strong humoral and cell-mediated immunity.\textsuperscript{100} The possible utility of these observations for vaccine development awaits further investigation.
1.5 Horizontal Gene Transfer

The significance of horizontal gene transfer between bacteria was first recognized in the 1950's with the emergence of multi-drug resistance. The ability of different bacteria to develop resistance to the same spectrum of antibiotics suggested that these traits were being transfer between species. Horizontal or lateral gene transfer has since been recognized as the basis for genetic variability and phenotypic plasticity for many bacterial species, and represents a mechanism by which bacteria can evolve and adapt to new environments. Unlike vertical gene transfer which occurs between organisms during reproduction, horizontal gene transfer occurs in the absence of reproduction. A significant proportion of bacterial genetic diversity can be attributed to the acquisition of sequences from other organisms, contrasting with eukaryotes which evolve principally through the modification of existing genetic information.

1.5.1 Identification of horizontal gene transfer in bacterial genomes

DNA sequence information has been used to define features that indicate horizontal gene transfer. For example, DNA segments gained through horizontal gene transfer may display a restricted phylogenetic distribution among related species. In addition, these regions may show small levels of DNA or protein sequence similarity that is absent from closely related organisms. Another important marker of gene acquisition is the G+C content. The average G+C content of a given bacterial species can be calculated, allowing variation from the average to be determined. Such regions will highlight gene acquisition from species with a different G+C content, but will fail to recognize transfer between species with similar G+C contents. In the latter case,
more subtle markers such as base composition, frequency and patterns of codon usage can be determined. A different approach used is to search for genes with functions that are often associated with horizontal gene transfer, such as mobility genes, integrases, transposases or phage-related genes.

1.5.2 Genomic islands

Genomic islands are regions of DNA of variable length (between 10 to 200 kilobases) that are acquired through horizontal gene transfer.\textsuperscript{125} These regions often differ from the recipient genome with respect to G+C content and codon usage. Genomic islands commonly contain genes associated with mobilization, and are often located adjacent to tRNA genes (phage attachment sites) and/or insertion sequence (IS) elements. Directed repeats (DR), which are defined as DNA sequences of 16 to 45 bp with a perfect or nearly perfect sequence repetition, frequently flank the genomic islands. Islands may demonstrate a mosaic structure, a reflection of repeated integration events.\textsuperscript{93}

Genomic islands first came to prominence when it became clear that they could carry genes that were involved in disease pathogenesis; these were termed 'pathogenicity islands'. However, islands that appear to contribute to other aspects of bacterial existence such as fitness in the environment due to the presence of genes encoding metabolic functions have since been described. The division of fitness islands into different subtypes is not based on their intrinsic genetic composition, but on their effects in a specific niche and within a particular organism. In other words, the same fitness island may act as an ecological island when the bacterial recipient resides outside of a host, but become a pathogenicity island when the bacterium enters a host.
1.5.3 Mechanisms of horizontal gene transfer

Horizontal gene transfer occurs between the same or related species. The prerequisites for this process are that: (i) there is a means by which donor DNA can be delivered into the recipient cell; (ii) the acquired sequences must be incorporated into the recipient's genome or into an autonomous replicating element such as plasmid; and (iii) the incorporated genes must be expressed in a manner that is appropriate to the recipient environment. The third step depends on compatibility between the transferred genes and the transcriptional and translational machinery of the host organism. The known mechanisms by which horizontal gene transfer occurs are transformation, transduction and conjugation.

1.5.3.1 Transformation

This process involves transfer of DNA present in solution after cell death/lysis or following release of DNA into the environment. Bacteria that are able to take up free DNA are termed 'competent'. Some bacterial species are always in a state to accept DNA, whereas others become competent at specific time points or as a result of changes in environmental conditions, such as in the presence of calcium chloride. Transformation has the potential to transfer DNA between distantly related organisms. Efficient transformation in certain bacterial species such as Neisseria gonorrhoeae and Haemophilus influenzae requires the presence of specific recognition sequences which are present in their respective genomes. Although the presence of specific uptake sequences enhances the transformation efficiency between related species, many of the naturally competent bacterial species do not display sequence preference and are capable of high levels of transformation. However, transformation
proficiency does not necessarily translate into correspondingly high rates of interspecies gene transfer.\textsuperscript{173}

\textbf{1.5.3.2 Transduction}

This is the process of DNA transfer through independently replicating viruses, called bacteriophage, that infect a range of bacteria. This does not require the donor and recipient cells to be present at the same time or place. Phage-encoded proteins mediate the delivery of double-stranded DNA into the recipient cytoplasm, and can promote the integration of DNA into the chromosome and protect the transferred sequences from degradation by host restriction endonucleases. The genomes of phage can range in size from a few to several 100 kb, the amount of DNA that can be transferred in a single event being limited by the size of the phage capsid. Essential bacteriophage genes include specific replicase genes, genes encoding phage components and genes encoding the proteins that package DNA in a protein coat.

Bacteriophages may replicate vigorously and lead to lysis of host bacteria, or may become integrated into the genome when they are non-lytic and termed prophage. Integrated phage is replicated along with the host chromosome and passed on to the daughter cells, but can be triggered by environmental stimuli such as DNA damaging agents to enter into the lytic stage. During cell lysis, host cell DNA can be packaged and later transferred to a new bacterial strain. New genetic material can be introduced into a recipient by donor DNA fragments or the donor genes adjacent to the phage attachment site. Although phages are prevalent in the environment, the spectrum of microorganisms that can be transduced depends upon the presence of specific receptors which are recognized by bacteriophage. As a result, this type of gene transfer is limited to related bacterial species.
1.5.3.3 Conjugation

Conjugation is the process of gene transfer via mobile self-replicating elements termed plasmids. These encode proteins required for conjugation and excision from the donor, and some other factors that may be required for survival in a restricted environment. These genetic elements encode proteins that facilitate their own transfer from the donor plasmid-carrying cell to a recipient cell that lacks the plasmid. Preparation of the DNA prior to transfer is similar among conjugative systems but the mechanism of mating-pair formation varies between organisms. Actively transmissible conjugative elements range in size from approximately 40 to 250 kb and are of low copy number. The conjugative or transferable genetic elements establish a stable mating pair and trigger DNA transport from the donor to the recipient cell through a specialized transfer pore called a conjugation bridge. This bridge provides protection to the DNA from the action of environmental nucleases.

There are several types of conjugative mechanisms, the greater distinctions being between those of Gram-positive compared with Gram-negative bacteria. The effects of conjugation for Gram-negative bacteria include the acquisition of antibiotic resistance. This is a particular problem within hospital settings where antibiotic usage is high, such as in an intensive care setting. Since the recipient cell becomes a donor after transfer of a plasmid, it is easy to see why an antibiotic resistant gene carried on a plasmid can quickly convert a sensitive population of organisms into a resistant one. In addition, prophages can undergo recombination with other mobile genetic elements that reside outside or within the same bacterium, and this can contribute to a “mosaic structure” of genomic islands.
Despite the diversity of mechanisms mediating genetic exchange among prokaryotes, the introduction of DNA into a recipient cell's cytoplasm does not ensure successful gene transfer unless the transferred sequences are stably maintained in the recipient microorganism. In addition to the process of acquiring sequences, expression of the acquired gene(s) at a significant level is required. The latter step depends on the compatibility of the transferred genes with the transcriptional and translational machinery of the host organism.\textsuperscript{75,173} Stable incorporation of foreign DNA into bacterial genomes can be mediated by any of the following processes:

1. Persistence of the mobile genetic element as an episome, which requires selection to avoid loss.
2. Homologous recombination (usually takes places among closely related taxa).
3. Integration mediated by bacteriophage integrases or mobile element transposases.
4. Illegitimate incorporation through chance double-strand break repair, as postulated for the integration of mitochondrial sequences into the yeast genome.\textsuperscript{196}

The type and number of genes transferred seems to be limited only by the selective pressures of the host or niche, rather than by any characteristic of the core replicative genes of the mobile genetic elements.

1.5.4 Classification of genomic islands

1.5.4.1 Pathogenicity islands (PAI)

The term Pathogenicity Island (PAI) was coined by Hacker \textit{et al.} in 1990. Currently, this term is commonly used to describe regions in the genomes of certain
pathogens that are absent in the non-pathogenic strains of the same or closely related species. These regions may contain continuous blocks of virulence genes. PAIs have recently been characterized in a wide range of bacterial pathogens. This has led to the identification of many virulence factors, and has increased our understanding of the evolution of bacterial pathogenicity.

PAIs were first described in *Escherichia coli*\(^9\), but have recently been found in the genomes of various pathogens for humans, animals and plants. PAIs are genetically unstable elements, many of which carry genes encoding adhesins, toxins or other virulence factors.\(^{126}\) PAIs in enteropathogenic and enterohaemorrhagic strains of *E. coli*, which are important human and veterinary pathogens, are well described. In brief, enteropathogenic *E. coli* (EPEC) is an important cause of diarrhea in infants and travelers in underdeveloped countries or regions with poor sanitation. Disease manifestations vary from minor gastroenteritis to a severe cholera-like syndrome. EPEC are acquired by ingestion of contaminated food and water, and adults in endemic areas appear to develop immunity.\(^{165}\) Enterohaemorrhagic *E. coli* (EHEC) can cause a spectrum of diseases ranging from uncomplicated watery diarrhoea to bloody diarrhea and hemorrhagic colitis occasionally with haemolytic–uraemic syndrome. EHEC produce a cytotoxin (Shiga-like toxin or Vero toxin), which is believed to be important in the pathogenesis of the bloody diarrhea.\(^{171}\) These pathogenic *E. coli* are closely related to each other and there is some evidence that a clone of EHEC is derived from EPEC. Both EPEC and EHEC appear to have evolved over time from non-pathogenic strains of *E. coli* by accumulating a variety of virulence factors, such as type III secretion systems (TTS3), the secreted translocator proteins and the secreted effector proteins. These functionally different modules have been termed the locus of enterocyte effacement (LEE) PAI. Both *E. coli* strains,
EPEC and EHEC, share the same LEE but EHEC contains 13 more ORFs within LEE than EPEC. Hypothetically, the remaining ORFs belong to phage-encoded Shiga toxins and have probably been acquired at a later time point. On this basis, EHEC may have evolved from EPEC via acquisition of a putative prophage element.\textsuperscript{178,195}

1.5.4.2 Metabolic islands

Acquisition of genes that alter metabolic properties of bacteria allow recipient organisms to explore new environments and endows selective advantages under specific environmental conditions. Molecular genetic analyses have established that species-specific metabolic traits can be attributed either to the acquisition of genes through horizontal gene transfer or to the loss of ancestral genes from a given lineage. Mobilization of complex metabolic traits depends on transfer of physically clustered genes in a single step. As a result, horizontal inheritance will select for gene clusters that act together (operons), which can be expressed in recipient cells by a host promoter at the site of insertion. For instance, some bacteria of the \textit{Pseudomonas} group that are part of the plant rhizosphere carry genomic islands that encode enzymes involved in degradation of phenolic compounds.\textsuperscript{190-192} These can help in enhancement of the capacity of bacteria to grow and disseminate in the soil. In addition, the symbiosis islands of the family \textit{Rhizobiaceae} carry nitrogen fixation genes whose products are necessary for the interactions of these bacteria with plant cells.

The discovery of genomic islands has accelerated with the advent of whole genome sequencing. There are currently more than 380 completed bacterial genomes. This has led to the observation that many bacteria carry genomic islands and may have multiple different islands.
Genomic islands can be used as molecular markers for the purposes of bacterial identification. These methods can also be applied for estimation of its pathogenic potential and even its antibiotic resistance pattern. This can be achieved by high sensitivity techniques such as Polymerase Chain Reaction (PCR) and Microarray DNA-Chip analysis.

### 1.6 B. pseudomallei multilocus sequence typing (MLST)

#### 1.6.1 General background

Clinical microbiologists are often asked to determine whether bacterial isolates are related to, or distinct from each other. Consequently, numerous typing schemes have been developed, including bacteriophage typing, serotyping and biotyping (metabolic capabilities of cell). For *B. pseudomallei*, various strain typing methods have been developed. These include ribotyping,\textsuperscript{181,212,245} PCR-based methods,\textsuperscript{256} random amplified polymorphic DNA (RAPD) analysis,\textsuperscript{139,187} variable-number tandem repeats (VNTR),\textsuperscript{144} and pulsed-field gel electrophoresis (PFGE).\textsuperscript{63} Many of these methods suffer from lack of reproducibility both within and between laboratories. Multilocus sequence typing (MLST), a system based on multilocus enzyme electrophoresis (MLEE), overcomes this problem. This has become the preferred typing method for epidemiologic and phylogenetic studies for a range of organisms because of its portability and unambiguous output data.\textsuperscript{249}

#### 1.6.2 Basis of MLST methodology

MLST is a nucleotide sequence-based approach that compares the sequences of housekeeping gene fragments, now aided by comparison with sequence held in several databases that can be accessed via the internet.\textsuperscript{149} This technique was
developed using a similar approach to that used by MLEE, a typing scheme that analyses the phenotype of housekeeping enzymes.

MLEE was first described for the study of mammals in 1971.\textsuperscript{210} It was subsequently applied to the study of \textit{E. coli} in 1973.\textsuperscript{157} This protein-based method determines the mobility of a selected set of metabolic enzymes across a starch gel matrix. This involves 20 or more loci, and the alleles at each locus go towards defining an electrophoretic type (ET). An assumption is made that differences in the mobility of enzymes are the result of molecular mass, electrical charge and conformation. Protein migration is directly related to its amino acid sequence. Charge differences due to amino acid substitutions in the polypeptide sequence reflect changes in the DNA encoding the polypeptide. Not all mutations at the DNA level have consequences for the amino acid sequence, and post-translational modifications are not usually taken into account in the interpretation of the results. Thus, strains showing identical migration patterns for one enzyme will not be distinguished by MLEE but may have different genetic alleles at that locus. Thus, MLEE has its limitations, and in addition is difficult to perform and produces results that are difficult to compare between laboratories.

MLST was first described in 1998 for the human pathogen \textit{Neisseria meningitidis}.\textsuperscript{149} MLST defines alleles at each of 7 housekeeping gene loci on the basis of nucleotide sequence of 400 to 600 bp internal fragments. The seven loci are chosen so as to be well separated and scattered on the chromosome in order to assess the contribution of recombination to the variation within each species.\textsuperscript{96} There are three main factors to consider for the design of a new MLST scheme: (i) the choice of isolates to be used in the initial evaluation; (ii) the choice of genetic loci to be
characterized; and (iii) the design of primers for gene amplification and nucleotide sequence determination.251

Once sequence is obtained for a given locus, this is checked with the complementary strand and then compared with all of the previously identified sequences at that locus. Allele numbers are then assigned at each of the seven loci, a new number (the next available ascending number) being assigned to a new allele (defined as any sequence with one or more base pair differences from other known sequences at that locus). The combination of the seven allelic numbers defines the allelic profile, and each profile (ex. 2-2-1-3-4-2-1) is assigned a sequence type (ST). The process by which numbers are assigned to alleles and sequence types means that these are uninformative in terms of inter-allele or ST relationships.148 An overview of the process is shown in Figure 1.3.

MLST has several advantages over MLEE. A greater degree of variation can be detected by MLST, since many more alleles are assigned per locus compared with MLEE. MLST is based on the nucleotide sequence while MLEE defines a phenotypic property. Sequence data can be compared readily between laboratories, such that MLST is fully portable, and data stored in a single expanding central multilocus sequence database can be interrogated electronically via the internet to produce a powerful resource for global epidemiology without exchanging strains. MLST also has the advantage that it can be applied directly to clinical specimens, reducing the need for bacterial culture.149
Chromosomal DNA

Amplify ≈ 450-bp fragments

Sequence the seven gene fragments

Each different sequence at a locus is given a unique allele number

The allele numbers at the seven MLST loci gives the allelic profile of the isolate

Compare the allelic profile of isolate to those of all isolates within a central database on the web via the internet

**Figure 1.3 Overview of MLST** (reproduced from\textsuperscript{225}).
1.6.3 Use of MLST databases

At the time of writing, there are 41 species represented in MLST databases and published schemes. There are three separate MLST sites, each consisting of several databases and related software. The MLST home page at http://www.mlst.net includes links to two pages, namely the PubMLST home page at http://pubmlst.org and MLST databases at the MPI für Infektionsbiologie http://web.mpiib-berlin.mpg.de/mlst/. These MLST databases are hosted at the University of Oxford, UK; Imperial College, London, UK; and the Max Planck Institute for Infection Biology, Berlin, Germany, respectively.

MLST users can access information on primer sequences and PCR conditions, sequencing protocols, and can use the databases to assign allelic numbers and STs. Web databases also provide downloadable submitted data of isolates and allelic profiles. Web-based programs for MLST data analysis (some available for free-download) are also provided. Sequence from an uncharacterized strain is submitted to the appropriate website, and alleles at each locus or multiple loci can be obtained either separately or together through the enquiry procedure. Each MLST website holds the sequences of all known alleles at each of the seven loci, and all known allelic profiles, and through the curator assigns new allele numbers and STs.

1.6.4 MLST scheme for B. pseudomallei

The B. pseudomallei MLST scheme was established in 2003 by Godoy and colleagues for epidemiological studies of melioidosis. This characterized B. pseudomallei and the closely related organisms, B. mallei and B. thailandensis. MLST data can be accessed via the B. pseudomallei MLST website (http://bpseudomallei.mlst.net/). Candidate housekeeping genes involved in essential
metabolic processes were selected by using BLASTX with sequence contigs of *B. pseudomallei* strain K96243. Complete genome sequencing of K96243 revealed that all 7 genes were located on chromosome 1. This study demonstrated that MLST was able to discriminate between *B. pseudomallei*, *B. mallei* and *B. thailandensis*. *B. mallei* appeared to represent a single clone that had recently diverged from *B. pseudomallei*.

1.7 General background to *Burkholderia thailandensis*

*B. thailandensis* is a Gram-negative, motile saprophyte that has been isolated from soil in central and northeast Thailand, Laos, Cambodia and Vietnam. Early reports described the organism as a non-virulent biotype of *B. pseudomallei*, but a clear distinction was made in 1998 when the organism was recognized as a separate species. 16S rRNA gene sequencing, DNA-DNA hybridization and MLST profile are consistent with the designation of *B. thailandensis* as a unique species. This bacterium is genetically closely related to *B. pseudomallei*, with which it shares several antigenic and morphologic similarities. Both species grow on the same selective media and have similar biochemical profiles, but a key difference between the two is the ability of *B. thailandensis* to assimilate L-arabinose. *B. pseudomallei* is unable to do so as a result of a gene deletion event involving the arabinose operon, and this distinction is used widely in the laboratory as a simple way to distinguish between the two. *B. pseudomallei* and *B. thailandensis* can also be differentiated by colony morphology when grown on modified Ashdown's selective media (Ashdown's selective media plus 100 μg/ml streptomycin and 15 μg/ml gentamicin; MASM). *B. thailandensis* colonies are smooth and glossy with a pink pigmentation, while *B. pseudomallei* colonies are
rough and wrinkled with a dark purple pigmentation. It is also possible to discriminate between *B. pseudomallei* and *B. thailandensis* using species-specific PCR or Real-Time PCR.

*B. thailandensis* is generally non pathogenic for animals and humans, and disease due to *B. thailandensis* is extremely rare. A report from Thailand described the isolation of *B. thailandensis* from purulent material obtained from an amputated knee following a motorcycle accident in Thailand, although it is not clear whether this was the cause of the infection or a wound coloniser. A case report from the U.S. described a boy who was involved in a road traffic accident and near drowning, who subsequently developed pneumonia and blood cultures positive for *B. thailandensis*. Given the frequency of exposure to this organism and the low attack rate, it is clear that infection only occurs under exceptional circumstances. This is consistent with its low virulence potential in experimental animal models. In one study, the mean 50% lethal dose of *B. thailandensis* in mice was $10^9$ CFU, compared with 182 CFU for *B. pseudomallei*.

Whole genome sequencing of *B. thailandensis* E264, a strain isolated from a rice field in central Thailand, has been performed. This is comprised of two chromosomes of 3.8 and 2.9 Mb, respectively, with a total of 5,645 predicted protein-encoding ORFs (3282 on Chromosome 1 and 2363 on Chromosome 2, respectively). The genome sequence and annotation has been deposited in the GenBank database (accession numbers CP000086 and CP000085 for chromosomes I and II, respectively). Comparison of the *B. pseudomallei* and *B. thailandensis* genomes has revealed broad similarities, including two highly syntenic chromosomes with comparable numbers of genes and density of single sequence repeat (SSR)
Contents. 168 B. thailandensis contains an eight-gene arabinose assimilation operon (BTH_II1626–1633) on Chromosome 2 that is absent in B. pseudomallei. Conservation of core genes involved in amino acid metabolism, cofactor and carrier synthesis, nucleotide and protein biosynthesis between the two species is consistent with the ability of B. pseudomallei and B. thailandensis to occupy similar environmental niches. 215

One of the objectives behind whole genome sequencing of B. thailandensis was to define the difference in the putative virulence determinant component between this and B. pseudomallei. Subtractive hybridization was used as a forerunner to whole genome sequencing to identify virulence genes that were present in B. pseudomallei but absent in B. thailandensis. 193 Differences have been defined in a range of genes, including a capsular polysaccharide gene cluster that was unique to B. pseudomallei. Virulence related genes, in particular members of the Type III secretion complex, were collectively more divergent between B. pseudomallei and B. thailandensis compared to the rest of the genome, possibly contributing towards the ability of B. pseudomallei to infect mammalian hosts. 283 71% of B. pseudomallei genes were also present as orthologs in B. thailandensis at an average similarity of greater than 80%, including TTS2 and TTS3, antibiotic resistance genes, type IV pili-generating proteins, hemolysin-related genes, and several adhesion factors and proteases. Although a considerable amount of information has been generated to date, the precise factor(s) that are absent in B. thailandensis but present in B. pseudomallei that confer virulence have not been defined.

Despite the importance of B. thailandensis in relation to B. pseudomallei, little is known about the population genetic structure of the former. An MLST scheme developed for B. pseudomallei was found to give sequence at the 7 housekeeping
loci. The alleles at all seven loci in the *B. thailandensis* isolates were different from those at the seven loci in all *B. pseudomallei* isolates, and *B. thailandensis* formed a distinct group. By using the concatenated sequences, the average sequence diversity among both the *B. pseudomallei* and the *B. thailandensis* STs was 0.2%, whereas the average divergence between the STs of the two species was 3.1%. A minimum-evolution tree showed that the isolates cultured from the farmer in Oklahoma (see above) were distantly related to all isolates of both *B. pseudomallei* (5.2% divergence) and *B. thailandensis* (4.7% divergence). This study included 10 *B. thailandensis* isolates, and no other studies have been published to date on the population genetic structure of this species.
1.8 Aims of this dissertation

The work contained in this dissertation describes a series of studies that explore the population genetic structure of *B. pseudomallei* and the closely related organism, *B. thailandensis*. This was achieved through the use of MLST and an evaluation of genomic islands. Specific aims were to:

1. Determine the presence of putative genomic islands in soil and invasive isolates of *B. pseudomallei*, to compare the frequency between the two groups and to examine whether a relationship existed between their presence and clinical features/outcome from melioidosis. Genomic islands were also considered in relation to the population genetic structure of *B. pseudomallei* as defined by MLST.

2. Explore the population genetic structure of *B. pseudomallei* using MLST, to compare isolates from the environment with those from invasive disease, and to compare isolates from Thailand with those from Australia. Isolates associated with a single, well-defined clinical manifestation (parotid infection) were examined to determine whether this represented a more restricted population.

3. Undertake typing of an expanded number of *B. thailandensis* isolates using MLST. This was used as a basis for defining the relatedness of *B. thailandensis* and its closely related species, and to define the utility of MLST as a typing tool for this species.
Chapter 2. Materials and Methods

2.1 Chemicals and reagents

All chemicals and reagents were obtained from Sigma-Aldrich Company Ltd. or Merck Ltd. Bacterial culture media were obtained from Oxoid. Sterile plasticware was obtained from Falcon, Becton Dickinson Ltd. and Terumo.

2.2 Bacterial culture and storage conditions

(a) Liquid media: *B. pseudomallei* and *B. thailandensis* were grown in Trypticase soy broth (TSB) under static conditions at 37°C in air. Cultures were routinely grown in 3 ml of media contained in 15 mL universal containers.

(b) Solid media: All agar plates used in this study were prepared as recommended by the manufacturer (Appendix I). *B. pseudomallei* and *B. thailandensis* were routinely cultured on Ashdown’s Selective Agar (ASA) at 37°C in air (Appendix I).

(c) Bacterial storage: *B. pseudomallei* and *B. thailandensis* were suspended in TSB containing glycerol (15%, v/v) and stored at −80°C.

2.3 Laboratory facilities

All experimental work involving viable bacterial cells was performed in a Category 2 containment facility under negative pressure of −50 pascals. All infectious work was carried out within a class II biological safety cabinet.
2.4 *B. pseudomallei* identification

The *B. pseudomallei* isolate collection used during this work was maintained in locked \(-80^\circ\text{C}\) freezers at the Wellcome Unit, Faculty of Tropical Medicine, Mahidol University. All isolates were identified in the clinical diagnostic laboratory in Ubon Ratchathani as *B. pseudomallei* using the following criteria:

(i) Growth on Ashdown’s agar with typical colony morphology after \(\geq 48\) hours of incubation at \(37^\circ\text{C}\) in air.

(ii) Oxidase test positive.

(iii) Resistant to Gentamicin (disc strength 10\(\mu\text{g}\)) and Colistin (disc strength 10\(\mu\text{g}\)) after overnight incubation at \(37^\circ\text{C}\) in air on Columbia agar (National Committee for Clinical Laboratory Standards, 2000).

(iv) Bacterial colonies on agar plates were confirmed as *B. pseudomallei* using a latex agglutination reaction. In brief, a small amount of colony is emulsified using a sterile toothpick into a solution containing latex particles coated with a mixture of three monoclonal antibodies specific for a 200-kD surface antigen of *B. pseudomallei*. The reaction is read within 1-2 min. The presence of agglutination is taken as a positive reaction. *E. coli* is used as a negative control.

(v) API 20NE biochemical strips were used as recommended by the manufacturer (bio-Merieux, Inc.), if any uncertainly remained about the bacterial species. Typical API profiles are 1156577, 1556577 or 1156576 by APILAB Plus software, which gives an automated interpretation on a computer workstation (bioMerieux, 2005).

To exclude the possibility that some isolates were *B. thailandensis*, all isolates were tested for their inability to assimilate arabinose. This was performed in batches at the end of the rainy season of each year in the Wellcome Unit, Bangkok. *B.
pseudomallei is negative and B. thailandensis is positive. Isolates were point inoculated onto agar plates containing arabinose as the only source of nutrition (Appendix I). B. pseudomallei fails to grow on this medium after 48 hours incubation at 37°C in air, while B. thailandensis shows growth.

2.5 Bacterial isolates

2.5.1 Isolates used for genomic island analysis

A total of 186 B. pseudomallei isolates were obtained from northeast Thailand. Of these, 83 were isolated from the environment in the province of Ubon Ratchathani, northeast Thailand between 1990 and 2003.

2.5.2 Isolates used for MLST

A total of 266 B. pseudomallei isolates were characterized by MLST. Of these, 83 were the environmental and 103 the clinical isolates described above. A further 79 isolates were associated with a single clinical syndrome (acute suppurative parotitis in children). These were cultured from consecutive patients presenting to Sappasithiprasong hospital with parotitis since 1986. A total of 77 B. thailandensis isolates were obtained from soil collected from central and northeast Thailand.

2.6 Extraction of genomic DNA

Bacterial isolates were streaked from the freezer vial onto Ashdown's agar. A single colony was inoculated into TSB and incubated overnight in air at 37°C, after which genomic DNA was extracted using the Wizard Genomic DNA purification kit (Promega). DNA concentration was determined using the spectrophotometric method. Stock DNA was maintained at -80 °C.
2.7 Polymerase chain reaction (PCR) to detect genomic islands

2.7.1 Introduction

The presence or absence of 5 genomic islands was defined by PCR (see Table 2.1). These were selected from the 16 possible islands so as to represent the range of island types described for the sequenced *B. pseudomallei* strain K96243 (prophage, prophage-like, putative integrated plasmid and metabolic island).\textsuperscript{104} The target genes were predominantly putative hypothetical proteins of unknown function; genes that encoded functions necessary for the transfer and maintenance of the island, such as integrases, were avoided as similar sequences may be found on other unrelated islands.
Table 2.1 Description of 5 genomic islands of *B. pseudomallei* examined in this study

<table>
<thead>
<tr>
<th>Island</th>
<th>Size (kb)</th>
<th>Number of CDSs</th>
<th>CDS Coordinates</th>
<th>Integrase</th>
<th>GC%</th>
<th>Functional note</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Chromosome 1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GI 2</td>
<td>36.2</td>
<td>49</td>
<td>BPSL0129-BPSL0176</td>
<td>1</td>
<td>- (65.4)</td>
<td>Prophage φK96243</td>
</tr>
<tr>
<td>GI 6</td>
<td>15.0</td>
<td>21</td>
<td>BPSL1137-BPSL1157</td>
<td>1</td>
<td>+ (58.8)</td>
<td>Prophage-like</td>
</tr>
<tr>
<td>GI 9</td>
<td>9.8</td>
<td>19</td>
<td>BPSL2568-BPSL2586</td>
<td>1</td>
<td>- (64.0)</td>
<td>Prophage-like</td>
</tr>
<tr>
<td>GI 11</td>
<td>15.3</td>
<td>14</td>
<td>BPSL3257-BPSL3270</td>
<td>0</td>
<td>+ (55.9)</td>
<td>Putative integrated plasmid or integrated conjugative element contains recombinase,conjugal plasmid transfer and replication proteins</td>
</tr>
<tr>
<td><strong>Chromosome 2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GI 16</td>
<td>61.8</td>
<td>41</td>
<td>BPSS2051-BPSS2090</td>
<td>0</td>
<td>+ (59.3)</td>
<td>Metabolic island</td>
</tr>
</tbody>
</table>

The size and number of CDSs contained within the putative genomic islands are indicated as defined by Holden *et al.* 2004.104 The numbers of putative integrases contained within islands are given. Anomalies in the GC% content of DNA within islands as compared to the surrounding DNA is indicated by +. The value of the G + C% for an island is given in brackets in the GC% column.
2.7.2 PCR to detect the presence of genomic islands

Two target genes were selected for each genomic island, as follows: BPSL0130 and BPSL0135 (island 2), BPSL1138 and BPSL1155 (island 6), BPSL2578 and BPSL2579 (island 9), BPSL3258 and BPSL3260 (island 11), and BPSS2053 and BPSS2061 (island 16). Primer design was based on the whole genome sequence of *B. pseudomallei* K96243 using PrimerSelect (Lasergene V6.1, DNAstar, USA). The primer pairs used to amplify genes situated within genomic islands are shown in Table 2.2. The specificity of primer sequences for the chosen gene was confirmed through BLASTN analysis using the K96243 whole genome sequence (http://www.ncbi.nlm.nih.gov/BLAST/), accession number NC_006350 and NC_006351, for Chromosome I and chromosome II, respectively.).

PCR amplifications were performed using a PTC-0200 DNA engine (MJ Research, Cambridge, USA). Primers were multiplexed in three reactions (Table 2.2). Each 10 µl of PCR mixture contained 1X reaction buffer, 1.5 mM MgCl₂, 0.7 µM of each primer, 1 µl of working template DNA, 200 µM of deoxynucleotide triphosphate (dNTP) and 0.5 U of Taq DNA polymerase (Promega, USA). Reaction mixtures were pre-heated to 95°C for 2 minutes followed by 35 cycles of 95°C for 30 seconds, specific annealing temperature for 30 seconds and 72°C for 1 minute. A final extension at 72°C for 5 minutes was followed by cooling to 4°C. A positive control (*B. pseudomallei* K96243) and a negative control (reaction mixture without DNA) were included in each PCR run. Aliquots of reaction mixtures were analyzed by 2% agarose gel electrophoresis. Bands were detected by staining with ethidium bromide and visualization under UV light using the Gel Doc 2000 System (Bio-Rad, USA). A single product of the predicted size was interpreted as a positive result for a given
gene. The presence of one or both genes in a given island was interpreted to mean that the given island was present.

2.7.3 PCR to confirm absence of genomic islands

For isolates negative for a given island, a process of further verification was undertaken in which a PCR reaction was performed to amplify across the putative insertion site. Primers were designed in flanking genes, and the extension time set to amplify the region across the site in the absence of an island. If an island was present, this reaction would fail to give a product. This strategy is summarized in Figure 2.1.

Primer pairs were designed using PrimerSelect (Lasergene V6.1, DNAsStar, USA). Specificity of each primer pair was verified as before. Primers and cycling conditions are shown in Table 2.3. Amplification across genomic island 9 had an extension time of 1 minute 30 seconds, while 1 minute was used for other genomic islands.

Two positive products for each insertion site were randomly selected and verified by sequencing. Products were purified and sequenced as described in 2.9.3 and 2.9.4. Sequences were compared with the genome sequence of K96243 using SeqMan II (Lasergene V6.1, DNAsStar, USA) and BLASTN search (http://www.ncbi.nlm.nih.gov/BLAST/).9
Table 2.2 Primer sequences and grouping of primers used during multiplex PCR

<table>
<thead>
<tr>
<th>Multiplex No.</th>
<th>Gene ID</th>
<th>Forward Primer (5'-3')</th>
<th>Reverse Primer (5'-3')</th>
<th>Annealing temp.</th>
<th>Product size (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>I.</td>
<td>BPSL3258</td>
<td>AGCGCATGCCGCCGACAAAAAGAAGG</td>
<td>CGGCCGCGGAGATCAATGTAAT</td>
<td>63°C</td>
<td>899</td>
</tr>
<tr>
<td></td>
<td>BPSS2053</td>
<td>GGCCTGTTCTGTCGTCTAG</td>
<td>GGCGCGGCTAGTTTCAG</td>
<td>534</td>
<td></td>
</tr>
<tr>
<td></td>
<td>BPSL2578</td>
<td>GCCGTCCCGGCGGAGCTATTC</td>
<td>GACAGGGCTCAGCAGCTACCC</td>
<td>151</td>
<td></td>
</tr>
<tr>
<td>II.</td>
<td>BPSL0130</td>
<td>GCGCCGCTCGACTTCCTCTCTCT</td>
<td>GAGGGGCGCAGACTGACTTACAC</td>
<td>63°C</td>
<td>1110</td>
</tr>
<tr>
<td></td>
<td>BPSL1155</td>
<td>TATGCCTCGCTCGGAAACAGAT</td>
<td>GCACGTGCGCGCGCTCTAC</td>
<td>148</td>
<td></td>
</tr>
<tr>
<td></td>
<td>BPSL2579</td>
<td>CGCCGTGCAACCGGCTTTCA</td>
<td>GCTCGTCACCTGCTAAAACC</td>
<td>482</td>
<td></td>
</tr>
<tr>
<td></td>
<td>BPSL3260</td>
<td>TGTCGTTGCTCTGCTGCTG</td>
<td>TACCTGTTGCTCTCTCGGTGTC</td>
<td>238</td>
<td></td>
</tr>
<tr>
<td>III.</td>
<td>BPSL0135</td>
<td>CGTCAAGGCTGCACCGCAAG</td>
<td>CTGCCTCGCCTGGCTGC</td>
<td>65°C</td>
<td>136</td>
</tr>
<tr>
<td></td>
<td>BPSL1138</td>
<td>GATTGGTGTGGCTCGGTGTT</td>
<td>CGACCTTGCGCGGAATATGAG</td>
<td>424</td>
<td></td>
</tr>
<tr>
<td></td>
<td>BPSS2061</td>
<td>AACGCTCGGCGCTTTAC</td>
<td>AATGCCCTCAGATTTATG</td>
<td>239</td>
<td></td>
</tr>
</tbody>
</table>
Figure 2.1 PCR strategy to confirm lack on genomic island
Table 2.3 Primer sequences used to amplify region across putative insertion site of genomic islands

<table>
<thead>
<tr>
<th>Genomic island</th>
<th>Forward Primer (5'-3')</th>
<th>Reverse Primer (5'-3')</th>
<th>annealing temperature (°C)</th>
<th>MgCl₂ (mM)</th>
<th>Product size (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>GGGCGAGCGCAACGCACGTGAC</td>
<td>CCTGCCGCGCGGAAAAACATAC</td>
<td>48; 3.0</td>
<td>1,007</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>GATCAGCGGCCAGGACTTCTC</td>
<td>GGCAAGGGCCAGGTAAGACATC</td>
<td>55; 1.5</td>
<td>1,152</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>CTGGCCGCAGCTCCTCAA</td>
<td>GGGCGCCGTGTAATCTG</td>
<td>55; 1.5</td>
<td>1,441</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>GCCGGCGGCGCATCATAAC</td>
<td>GTGCCTCGGTCGTCTAGTG</td>
<td>63; 1.5</td>
<td>962</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>GCCGGCGCGGGGTGATCT</td>
<td>CCTGCTCGCGGGTTGGTC</td>
<td>56; 3.0</td>
<td>183</td>
<td></td>
</tr>
</tbody>
</table>
2.7.4 Analysis of genomic island data

An island was classified as present if either or both target genes were positive, and negative if no amplification products were obtained for either target gene, together with a PCR product across the putative insertion site. Comparison of genome sequences for GIs was performed using the ACT (Artemis Comparison Tool) suite of programs (www.sanger.ac.uk Software ACT).30 Statistical tests were performed using the statistical program STATA/SE, version 9.0 (StataCorp LP, College Station, Tx.). Proportions were compared using the Chi-square test or Fisher’s exact test, as appropriate. Comparison of continuous data was performed using the Student’s t test. Unifactorial analysis was used to examine the association between presence of each GI, cumulative number of GIs, and clinical factors.

2.8 MLST of B. pseudomallei

2.8.1 Primers

The primer pairs used to amplify housekeeping gene fragments on chromosome 1 are shown in Table 2.4.

2.8.2 PCR

PCR was performed using a PTC-0200 DNA engine (MJ Research, Cambridge, USA). Reactions were carried out in a total volume of 50 µl. Each reaction contained 1X reaction buffer, 1.5 mM MgCl₂, 0.7 µM of each primer, 5 µl of working template DNA, 200 µM of deoxynucleotide triphosphate (dNTP) and 2.5 U of Taq DNA polymerase (Promega, USA). Cycling conditions were an initial denaturation at 95°C for 4 min followed by 34 cycles of denaturation at 95°C for 30 sec, the annealing temperature shown in Table 2.4 for 30 sec, and extension at 72°C
for 60 sec. The samples were then maintained at 72°C for 10 min and cooled to 4°C. The amplification product was visualized using 2% agarose gel electrophoresis which was stained with ethidium bromide and visualized using a Gel Doc 2000 system (Bio-Rad, USA).

2.8.3 Clean-up of PCR products

The amplified DNA fragments (25 µl) were precipitated by mixing with 60 µl of 20% polyethylene glycol 8000/2.5M NaCl for 1 h at room temperature or 4°C overnight. Following centrifugation at 13,000 rpm for 40 min, the supernatant was discarded and the pellet was washed with 600 µl of 70% ethanol. The tube was centrifuged again as before. The supernatant was then removed and the DNA pellet dried at 37°C for 1 h and resuspended with 10 µl of sterile distilled water.

2.8.4 Sequencing

The DNA fragments for each strand were sequenced using a DYEnamic ET Dye Terminator Kit (Amersham) and a MegaBACE 500 sequencer with nested primer pairs shown in Table 2.5.
Table 2.4: Primer pairs and PCR cycling conditions used for MLST of *B. pseudomallei*

<table>
<thead>
<tr>
<th>Primer name</th>
<th>Primer Sequence (5'-3')</th>
<th>Annealing Temp. (°C)</th>
<th>Product size (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outer_aceFwd</td>
<td>GCCGCTCGGCGCTTCTCAAA</td>
<td>60</td>
<td>767</td>
</tr>
<tr>
<td>Outer_aceRev</td>
<td>AGCCGCCGTTTCAGCGAAAAATC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Outer_lepAFwd</td>
<td>GCATCAGACGCGCGACGTAGTTGT</td>
<td>60</td>
<td>724</td>
</tr>
<tr>
<td>Outer_lepARev</td>
<td>CTGGCGGGCCTTTTGCCTAAA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Outer_lipAFwd</td>
<td>ACGATCCGACCGCCAAGCAGAAGG</td>
<td>60</td>
<td>807</td>
</tr>
<tr>
<td>Outer_lipARev</td>
<td>ACGTACTCGGCACCCGGCAGATGG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Outer_narKFwd</td>
<td>GCCGTACGCTGAGCGGCTGCT</td>
<td>60</td>
<td>767</td>
</tr>
<tr>
<td>Outer_narKRev</td>
<td>AGCCCCGCTTCTGAACCACA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Outer_ndhFwd</td>
<td>TATCCCCGCGATCAGAAACAGTCC</td>
<td>60</td>
<td>719</td>
</tr>
<tr>
<td>Outer_ndhRev</td>
<td>GCCGGGTCTCGCCTCGTC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Outer_gltBFwd</td>
<td>TGTCGCGGCCCCTTTCATCT</td>
<td>63</td>
<td>727</td>
</tr>
<tr>
<td>Outer_gltBRev</td>
<td>ATCAGCACCAGGCCGCTACGAC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Outer_gmhDFwd</td>
<td>TCGCCAGGGCGACGCAGTT</td>
<td>65</td>
<td>705</td>
</tr>
<tr>
<td>Outer_gmhDRev</td>
<td>GGTGCGGACCCTGAGACC</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 2.5 Primer pairs used for sequencing of amplification products

<table>
<thead>
<tr>
<th>Primer name</th>
<th>Primer Sequence (5’-3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nested_aceF</td>
<td>GCTCGGC\text{CG}CTCTCAT\text{AAA}</td>
</tr>
<tr>
<td>Nested_aceR</td>
<td>ATGT\text{CCG}TGCCGAT\text{TAGC}</td>
</tr>
<tr>
<td>Nested_lepAF</td>
<td>TTT\text{CGCT}TG\text{ACG}CG\text{ACTG}</td>
</tr>
<tr>
<td>Nested_lepAR</td>
<td>CG\text{AACCACGAATCGAT}\text{GTA}</td>
</tr>
<tr>
<td>Nested_lipAF</td>
<td>CATACGG\text{TGTCGAGGAAGC}</td>
</tr>
<tr>
<td>Nested_lipAR</td>
<td>GC\text{AGGATCTC}CGTG\text{GTCG}</td>
</tr>
<tr>
<td>Nested_narKF</td>
<td>GC\text{CACCCTCGCGG}\text{CTGACG}</td>
</tr>
<tr>
<td>Nested_narKR</td>
<td>AG\text{CCCGTTCTGCAACCACA}</td>
</tr>
<tr>
<td>Nested_ndhF</td>
<td>GCAG\text{TTCGTCGG}\text{ACTATC}</td>
</tr>
<tr>
<td>Nested_ndhR</td>
<td>GG\text{CAGCGCATGAA}\text{GCTC}</td>
</tr>
<tr>
<td>Nested_gltBF</td>
<td>GG\text{CGGCAAGTCGAACACG}</td>
</tr>
<tr>
<td>Nested_gltBR</td>
<td>GC\text{AGGCGTTCAGACG}</td>
</tr>
<tr>
<td>Nested_gmhDF</td>
<td>TC\text{GCAGGCAGAC}\text{GCAGTT}</td>
</tr>
<tr>
<td>Nested_gmhDR</td>
<td>GC\text{TCAAGGACCGCG\text{GCTGAG}}</td>
</tr>
</tbody>
</table>
2.8.5 MLST analysis

The alleles at each of the loci were assigned using the *B. pseudomallei* MLST website (http://bpseudomallei.mlst.net). Novel sequences were repeated, and assigned new alleles and deposited in the MLST allele database. Each allele was given an allelic number, and the allelic profile used to define sequence type (ST). The previously published MLST profiles for 158 strains isolated by Professor Bart Currie in Northern Australia were downloaded from the MLST web site for comparison (http://www.mlst.net).

Neighbor-joining trees were constructed using the Kimura-2-parameter method of distance estimation as implemented in MEGA version 2.1. UPGMA dendrograms were constructed using the START package available from http://www.pubmlst.org. Maximum likelihood congruence analysis of gene trees was carried out using PAUP* ver.4b following the method of Feil et al., 2001 except that all trees were reconstructed and scored on the basis of the HKY85+G+Γ model of DNA substitution. eBURST v3 (http://eburst.mlst.net) was used to demonstrate relationships between closely related STs (those differing at only a single locus). Estimates of synonymous and non-synonymous changes (dS/dN) were calculated using the method of Nei and Gojobori as implemented in MEGA version 2.1. Mean heterozygosity per locus (H) was calculated using LIAN version 3.1 (http://adenine.biz.fh-weihenstephan.de/lian/). Comparisons of allelic and nucleotide divergence were calculated using the program BLAND. Comparisons between isolates from Thailand and Australia, and between Thai isolates recovered from soil and cases of invasive disease, were carried out using the classification index (CI) proposed by Jolley et al., 2005. This parameter is similar to F_{ST}, but is more sensitive in cases where more than two alleles are present, as it directly considers the
frequencies of each allele. Significant differences in the frequency of STs, and the frequency of alleles at specific loci, were identified by comparing the observed classification index with those obtained from 10,000 randomized trials.
Chapter 3. Results I: *Burkholderia pseudomallei* genome plasticity: genomic island variation in relation to virulence and origins

3.1 Chapter content

The time from onset of symptoms of melioidosis to hospital presentation varies from hours to months. Disease manifestations are extremely wide ranging and vary from acute, fulminant sepsis to localized disease. The basis for marked variability in clinical presentation and disease severity is unknown. The role of bacterial factors in determining disease variability and severity is undefined.

The whole genome sequence of *B. pseudomallei* K96243 has been published.\(^{104}\) The 7.25Mb genome is divided into two chromosomes, the smaller of which (chromosome 2) may have evolved from a megaplasmid. Considering both chromosomes, sixteen genomic islands (GIs) have been identified, which encompass distinct types of mobile genetic elements. The presence of eleven GIs was found to be variable in 40 natural isolates of *B. pseudomallei*.\(^{104}\) This is consistent with a key role for lateral gene transfer in the microevolution of this species. It was proposed that there is considerable variability in the presence of genomic islands within the wider *B. pseudomallei* population, leading to diversity in gene complement within this species. Genomic islands in other bacterial species encode many different functions, and selection may favor the maintenance of islands that increase fitness in a specialized environmental niche. Genomic islands may also play a pivotal role in virulence of a large number of bacterial pathogens, carrying clusters of virulence genes encoding a wide range of functions, including iron uptake systems, adhesins, superantigens, and genes that alter the antibiotic resistance phenotype.\(^{208}\)
The aim of this study was to examine the utility of a range of approaches in defining the presence and significance of genomic variation in natural populations of *B. pseudomallei*. This was applied to a representative sample of five GIs previously described for *B. pseudomallei* K96243, the choice being based on inclusion of the range of different island types. Comparisons were drawn between isolates obtained from the environment and from patients with melioidosis in order to examine the role of genomic islands on virulence. An *in silico* analysis of eight complete genome sequences revealed the micro-evolutionary processes responsible for the short-term diversification of these regions. The rapid loss and/or acquisition of gene islands were observed by assaying the presence of islands within isolates belonging to single clones (as defined by MLST data).

### 3.2 Bacterial isolates

A total of 186 *B. pseudomallei* isolates were obtained from northeast Thailand. Of these, 83 were isolated from the environment in the province of Ubon Ratchathani, northeast Thailand between 1990 and 2003. All sites were flooded rice paddies, the majority of which were sampled after ploughing but before planting. These isolates were processed for the presence of *B. pseudomallei* as previously described by Smith *et al.*, 1995.217 Briefly, soil samples were collected from approximately 30 cm depth. 100 g of soil was added to 100 ml of sterile distilled water then mixed and left overnight. The surface liquid of each sample (neat and serial dilutions) was spread plated onto ASA and inoculated into selective or enrichment broth. *B. pseudomallei* isolates were identified as described previously in chapter 2. Individual freezer vials represented several colonies from each positive site. A further 103 isolates were cultured from consecutive patients presenting with melioidosis during 2001 to
Sappasithiprasong hospital, Ubon Ratchathani. A single isolate was used from each patient. Clinical manifestations of infection were varied; 63 patients had positive blood cultures with or without involvement of one or more organs or tissues, and 40 patients had negative blood cultures but one or more organs or tissues involved. All isolates were maintained at -70°C in TSB with 15% glycerol.

3.3 Genomic island identification

The genomic islands were previously identified during whole genome sequence analysis of \textit{B. pseudomallei} K96243 (accession number BX571965 and BX571966).\textsuperscript{104} Five out of sixteen genomic islands were selected as being representative of these. These are GI 2, GI 6, GI 9, GI 11 and GI16, which correspond to putative prophage, prophage-like, prophage-like, putative integrated plasmid and putative metabolic islands, respectively.

3.4 Clinical definitions

Multiple organ involvement was defined when there was >1 non-contiguous focus of infection, not including blood. Pneumonia was defined as the presence of clinical features plus radiographic changes and/or sputum culture positive for \textit{B. pseudomallei}. Impaired renal function was defined as an estimated glomerular filtration rate below 60 mL/min/1.73 m\textsuperscript{2} during admission. Glomerular filtration rate was estimated using an abbreviated form of the Modification of Diet in Renal Disease study equation.\textsuperscript{1} Impaired hepatic function was defined as an elevation of aminotransferase more than 5 times the upper normal limit or the presence of jaundice during admission. Hypotension was defined as a systolic BP <90mmHg on or during admission.
3.5 Analysis

Comparison of genome sequences was performed with ACT (Artemis Comparison Tool (www.sanger.ac.uk Software ACT)\(^{30}\) using BLASTN and TBLASTX \(^{11}\). Comparisons were made using 3 additional complete \textit{B. pseudomallei} genome sequences 1106a (accession numbers CP000572 and CP000573), 1710b (accession numbers CP000124 and CP000125) and 668 (accession numbers CP000570 and CP000571), 6 further \textit{B. pseudomallei} strains currently undergoing whole genome sequencing (406e, 1106b, 1710a, Pasteur 6068, S13, and 1655) by The Institute of Genome Research (TIGR; http://www.tigr.org), and \textit{B. thailandensis} E264 (accession numbers CP000086 and CP000085).\(^{283}\)

Statistical tests were performed using the statistical program STATA/SE, version 9.0 (StataCorp LP, College Station, Tx.). Proportions were compared using the Chi-square test or Fisher’s exact test, as appropriate. Comparison of continuous data was performed using the Student’s t test. Unifactorial analysis was used to examine the association between presence of each GI, cumulative number of GIs, and clinical factors.

3.6 Results and discussion

3.6.1 Overview of five genomic islands in \textit{B. pseudomallei} strain K96243

The five genomic islands studied here include examples of prophage (GI 2; K96243) which have been shown to be mobile in this strain,\(^{104}\) two prophage-like islands (GI 6 and GI 9), a putative integrated plasmid (GI 11), and a putative metabolic island (GI 16) that lacks any obvious genes for mobilisation or integration. The criteria used to identify these island regions were wide; as with many genomes, no single characteristic typical of genomic islands is shared by all of these regions.
For example, some of the islands contain integrase genes, some are located next to a tRNA gene and are flanked by direct repeats, and others have anomalous base composition. Atypical base composition is indicative of fairly recent acquisition of the island through horizontal gene transfer by *B. pseudomallei*, while islands with typical base composition for the genome are likely to be more ancient and stable integrations into the genome. By assaying the presence of these islands in a large strain collection within the population framework provided by the MLST data, it is possible to investigate the frequency of island acquisition and loss, as well as their possible role in disease.

A summary of the gene content of each of the five GIs in *B. pseudomallei* K96243 is provided in Table 3.1, and a comprehensive list is provided in Appendix III. The overwhelming majority of genes that are not phage or plasmid related are of unknown function. Four GIs (2, 6, 9 and 11) do not contain genes with homology to those with known metabolic or putative virulence functions. However, GIs 2, 6 and 11 contain low %GC regions relative to the rest of the island, indicating that the islands are mosaic and some genes may have been acquired very recently. GI 16 contains several CDSs with similarity to known virulence determinants, including a putative haemagglutinin and processing protein, and genes involved in the acquisition and utilization of nutrients. The abundance of CDSs on this island that encode functions that potentially broaden the metabolic repertoire of K96243 (for example, possible sugar utilization and amino acid catabolism gene clusters and accompanying regulators), has led to this island being previously designated a putative metabolic island.104 CDSs in this island region exhibited varying levels of similarity to proteins from a range of taxonomically diverse organisms, and it was not possible to speculate on the likely source of this island region from similarity searches.
3.6.2 Sampling framework for *B. pseudomallei* isolates

Google Earth software was used to map the geographical site of isolation for environmental isolates (Figure 3.1), and the home address of patients with melioidosis grouped according to Amphoe (second level administrative districts) (Figure 3.2). There was considerable overlap in location of environmental isolates and patient address, and on this basis it is suggested that soil and invasive isolates do not correspond to distinct sub-populations.
Table 3.1 Overview of gene content within five genomic islands of

*B. pseudomallei* K96243

<table>
<thead>
<tr>
<th>Island</th>
<th>Gene classification</th>
<th>No. CDSs</th>
<th>Functional Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>GI 2</td>
<td>unknown function</td>
<td>27</td>
<td>3’ region of the prophage contains a predicted membrane protein with no database matches in a low %GC region: BPSL0176 (55.9% GC)</td>
</tr>
<tr>
<td></td>
<td>bacteriophage structural and</td>
<td>22</td>
<td></td>
</tr>
<tr>
<td></td>
<td>regulatory proteins</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GI 6</td>
<td>unknown function</td>
<td>16</td>
<td>5’ region of the island contains two CDSs in a low %GC region: BPSL1137 (45.7% GC) and BPSL1138 (46.5% GC) that encode a putative nucleotide binding protein and conserved hypothetical protein with a RelA/SpoT conserved domain.</td>
</tr>
<tr>
<td></td>
<td>bacteriophage-related proteins</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>additional functional proteins</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>GI 9</td>
<td>unknown function</td>
<td>17</td>
<td>Bacteriophage-like proteins include an integrase and a phage-like regulatory protein</td>
</tr>
<tr>
<td></td>
<td>bacteriophage-related proteins</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>GI 11</td>
<td>unknown function</td>
<td>7</td>
<td>Plasmid-like proteins include conjugal transfer proteins and plasmid replication protein</td>
</tr>
<tr>
<td></td>
<td>plasmid-related proteins</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>GI 16</td>
<td>unknown function</td>
<td>11</td>
<td>This island encodes several metabolic and transport functions that may be important in the generation and acquisition of nutrients.</td>
</tr>
<tr>
<td></td>
<td>transport proteins</td>
<td>8</td>
<td>Theis includes two ATP Binding Cassette transporter type systems, and a L-asparaginase, fatty aldehyde dehydrogenase, and putative alpha-galactosidase. The island also includes a putative cell surface haemaglutinin protein and two-partner secretion system protein that may have a role in niche colonization or virulence.</td>
</tr>
<tr>
<td></td>
<td>metabolic proteins</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td></td>
<td>regulatory proteins</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>transposase proteins</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>misc. functional proteins</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td></td>
<td>pseudogenes and gene remnants</td>
<td>4</td>
<td></td>
</tr>
</tbody>
</table>
Figure 3.1 Geographic origin of 83 *B. pseudomallei* isolates cultured from the environment in Ubon Ratchathani, northeast Thailand. Red dots demonstrate sites of isolation of environmental isolates.
3.6.3 Presence of five genomic islands in *B. pseudomallei*

Presence or absence of the five GIs was defined for 186 Thai isolates obtained from the environment (n=83) and from patients with melioidosis (n=103). The hypotheses that GIs play a role in the ability to cause human disease was examined by testing whether their presence was significantly different in clinical samples. The number of GIs per isolate was defined for the 186 Thai isolates; this ranged from 0 to 5 (median 2, IQR: 1 to 3) (Table 3.3). The distribution of cumulative GI number did not differ between environmental and disease-associated isolates (p=0.27) (Table 3.3). These data

Figure 3.2 Geographic origin of 103 *B. pseudomallei* isolates cultured from patients with melioidosis in Ubon Ratchathani, northeast Thailand. Green dots show the areas of residence of patients with invasive melioidosis, grouped into Amphoe (second level administrative districts). The size of the dot is proportional to number of cases, with larger dots representing a larger number of cases.
3.6.3 Presence of five genomic islands in \textit{B. pseudomallei}

Presence or absence of the five GIs was defined for 186 Thai isolates obtained from the environment (n=83) and from patients with melioidosis (n=103). The proportion of all isolates positive by PCR varied between GIs, and ranged from 12% for a prophage-like island (GI 9) to 76% for a metabolic island (GI 16) (Table 3.2). These proportions are consistent with previous findings reported for 40 Thai isolates,\textsuperscript{104} with the exception that GI 2 and GI 6 were more frequently detected in the larger bacterial population examined here (72% versus 53% for GI 2, and 23% versus 8% for GI 6, p=0.02 & p=0.03, respectively). Isolates in both collections were from northeast Thailand. A very small proportion of strains were positive for one or two gene targets for a given GI, and gave an amplification product for PCR across the putative insertion site. One explanation for this is that GIs may have alternative insertion sites. This explanation is shown to be true for some islands in the sequenced strains (see analysis of the sequenced strain GIs below).

3.6.4 Comparison of the presence of genomic islands in soil and invasive isolates

The hypothesis that GIs play a role in the ability to cause human disease was examined by testing whether their presence was significantly different in clinical versus environmental isolates. There was no significant difference in the presence of any of the GIs between these two groups (p>0.05 for all five islands) (Table 3.2). The possibility remained that combinations of GIs could alter the ability to cause disease through either a cumulative or synergistic effect. The cumulative number of GIs per isolate was defined for the 186 Thai isolates; this ranged from 0 to 5 (median 2, IQR 1 to 3) (Table 3.3). The distribution of cumulative GI number did not differ between environmental and disease-associated isolates (p=0.27) (Table 3.3). These data
indicate that none of the islands examined were numerically associated either individually or on a cumulative basis with the ability to cause melioidosis.

Table 3.2 Presence of five genomic islands in 186 *B. pseudomallei* isolates

<table>
<thead>
<tr>
<th>Genomic island</th>
<th>Positive (%)</th>
<th>Positive (%)</th>
<th>P°</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All isolates (n=186)</td>
<td>Soil (n=83)</td>
<td>Invasive (n=103)</td>
</tr>
<tr>
<td>2</td>
<td>133 (72%)</td>
<td>62 (75%)</td>
<td>71 (69%)</td>
</tr>
<tr>
<td>6</td>
<td>43 (23%)</td>
<td>22 (27%)</td>
<td>21 (20%)</td>
</tr>
<tr>
<td>9</td>
<td>22 (12%)</td>
<td>11 (13%)</td>
<td>11 (11%)</td>
</tr>
<tr>
<td>11</td>
<td>39 (20%)</td>
<td>15 (18%)</td>
<td>24 (23%)</td>
</tr>
<tr>
<td>16</td>
<td>142 (76%)</td>
<td>66 (80%)</td>
<td>76 (74%)</td>
</tr>
</tbody>
</table>

° Chi-square

Table 3.3 Cumulative presence of genomic islands in 186 *B. pseudomallei* isolates

<table>
<thead>
<tr>
<th>Cumulative number of genomic islands</th>
<th>All isolates</th>
<th>Soil</th>
<th>Invasive</th>
<th>P°</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>5 (3%)</td>
<td>1 (1%)</td>
<td>4 (4%)</td>
<td>0.27</td>
</tr>
<tr>
<td>1</td>
<td>46 (25%)</td>
<td>20 (24%)</td>
<td>26 (25%)</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>84 (45%)</td>
<td>36 (43%)</td>
<td>48 (47%)</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>40 (22%)</td>
<td>20 (24%)</td>
<td>20 (19%)</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>10 (5%)</td>
<td>6 (7%)</td>
<td>4 (4%)</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>1 (0.5%)</td>
<td>0</td>
<td>1 (1.0%)</td>
<td></td>
</tr>
</tbody>
</table>

° Student's T test
3.6.5 Relationship between clinical features and genomic islands in disease-associated isolates

The lack of association between the presence of the five GIs and the ability to cause human disease does not exclude the possibility that these GIs are associated with specific clinical features or outcome. To study this, a detailed statistical analysis was undertaken of the relationship between a range of clinical features in 103 patients with melioidosis and the presence of GIs in their associated *B. pseudomallei* isolates. Univariate analysis was used to compare the presence of each GI and their cumulative number with the presence of blood cultures positive for *B. pseudomallei*, disease distribution, presence of ultrasound confirmed abscess(es) in liver and/or spleen, lung involvement as defined by abnormal chest radiograph consistent with infection, the presence on admission or subsequent development of hypotension, impaired renal function and/or impaired liver function, and death during hospital admission. GI 6 was negatively associated with positive blood culture, and GI 11 was negatively associated with impaired renal function (Table 3.4). Cumulative number of GIs was negatively associated with impaired renal function.
Table 3.4 Relationship between Univariate analysis of clinical factors and the presence of genomic islands

<table>
<thead>
<tr>
<th>Clinical factors</th>
<th>Number of patients (%)</th>
<th>GI 2</th>
<th>GI 6</th>
<th>GI 9</th>
<th>GI 11</th>
<th>GI 16</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N=103</td>
<td>Odds ratio</td>
<td>Odds ratio</td>
<td>Odds ratio</td>
<td>Odds ratio</td>
<td>Odds ratio</td>
<td>P</td>
</tr>
<tr>
<td>Blood culture positive</td>
<td>63 (61%)</td>
<td>1.12</td>
<td>0.30</td>
<td>1.79</td>
<td>0.88</td>
<td>0.90</td>
<td>0.82</td>
</tr>
<tr>
<td></td>
<td>(0.48-2.62)</td>
<td>0.30</td>
<td>0.11-0.82</td>
<td>0.45-7.21</td>
<td>0.42-2.76</td>
<td>(0.36-2.23)</td>
<td></td>
</tr>
<tr>
<td>Multiple organ</td>
<td>21 (20%)</td>
<td>0.67</td>
<td>0.35</td>
<td>0.85</td>
<td>0.73</td>
<td>0.61</td>
<td>4.17</td>
</tr>
<tr>
<td>involvement</td>
<td>(0.25-1.83)</td>
<td>0.35</td>
<td>0.07-1.64</td>
<td>0.17-4.29</td>
<td>0.22-2.42</td>
<td>(0.90-19.3)</td>
<td>0.07</td>
</tr>
<tr>
<td>Pneumonia</td>
<td>34 (33%)</td>
<td>1.73</td>
<td>1.33</td>
<td>1.18</td>
<td>1.02</td>
<td>0.97</td>
<td>2.04</td>
</tr>
<tr>
<td></td>
<td>(0.68-4.41)</td>
<td>1.33</td>
<td>0.49-3.59</td>
<td>0.32-4.35</td>
<td>0.39-2.69</td>
<td>(0.74-5.66)</td>
<td>0.17</td>
</tr>
<tr>
<td>Liver and/or</td>
<td>25 (24%)</td>
<td>0.94</td>
<td>0.68</td>
<td>0.67</td>
<td>1.05</td>
<td>3.26</td>
<td>0.89</td>
</tr>
<tr>
<td>splenic abscess</td>
<td>(0.36-2.49)</td>
<td>0.68</td>
<td>0.21-2.26</td>
<td>0.13-3.31</td>
<td>(0.37-3.03)</td>
<td>(11.9)</td>
<td></td>
</tr>
<tr>
<td>Impaired</td>
<td>32 (31%)</td>
<td>1.22</td>
<td>1.14</td>
<td>0.46</td>
<td>0.89</td>
<td>0.82</td>
<td>2.42</td>
</tr>
<tr>
<td></td>
<td>(0.36-2.49)</td>
<td>0.67</td>
<td>0.21-2.26</td>
<td>0.13-3.31</td>
<td>(0.37-3.03)</td>
<td>(11.9)</td>
<td>0.11</td>
</tr>
<tr>
<td>hepatic function</td>
<td>impaired renal function</td>
<td>Hypotension</td>
<td>Died during admission</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------------</td>
<td>-------------------------</td>
<td>-------------</td>
<td>----------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>54 (52%)</td>
<td>0.55</td>
<td>0.79</td>
<td>1 (40%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.49-3.06)</td>
<td>(0.23-1.29)</td>
<td>(0.27-2.38)</td>
<td>(0.34-2.45)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.73</td>
<td>2.15</td>
<td>0.79</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.33-2.42)</td>
<td>(0.85-5.43)</td>
<td>(0.34-1.84)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.36</td>
<td>1.47</td>
<td>0.86</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.14-0.94)</td>
<td>(0.46-4.69)</td>
<td>(0.23-3.11)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.63</td>
<td>0.32</td>
<td>0.81</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.21-2.57)</td>
<td>(0.49-8.85)</td>
<td>(0.85-5.43)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.33</td>
<td>0.14</td>
<td>0.85</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.03-1.65)</td>
<td>(0.34-2.45)</td>
<td>(0.34-1.84)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.17</td>
<td>0.21</td>
<td>0.91</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.23-1.62)</td>
<td>(0.34-2.45)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.62</td>
<td>0.68</td>
<td>0.58</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.21-2.57)</td>
<td>(0.34-2.45)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.34</td>
<td>0.68</td>
<td>0.58</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.15-1.28)</td>
<td>(0.34-1.84)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.69</td>
<td>0.79</td>
<td>0.79</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.28-1.67)</td>
<td>(0.34-2.45)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Culture positive for *B. pseudomallei*
These significant associations could be a chance finding related to multiple comparisons. To test this hypothesis, the presence of GIs were determined for an independent set of 255 invasive isolates obtained from patients with melioidosis presenting to Sappasithiprasong Hospital between 2002-2003. There was no significant association between GI 6 and positive blood cultures (OR 1.12; 95%CI 0.51-2.43, p=0.78), or GI 11 and impaired renal function (OR 0.67; 95%CI 0.34-1.32, p=0.25). Cumulative number of GIs was not negatively associated with impaired renal function. It can be concluded that there were no significant associations between clinical features and any of the five GIs examined.

3.6.6 Variability in genomic islands between clones of *B. pseudomallei*

The mobility of the five GIs was explored by examining whether the presence of gene islands reflects the underlying population structure, such that closely related isolates might have more similar GI repertoires than distantly related isolates. This was possible as all of the isolates had been characterized by MLST, and these data revealed the presence of three clonal lineages consisting of very closely related isolates: ST 70 (n=15 isolates), ST 54 (n=11), and ST 167 (n=9). The variation in GI repertoire was compared within and between these clones. There was considerable variability in presence of the five GIs between the three clones (Table 3.5). For example, GI 6 was more common in isolates belonging to ST 70 than the other two clones (p=0.04), GI 16 was ubiquitous in ST 54 isolates but present in three quarters or less in the other two ST (p=0.04), and GI 11 was only detected in 4 isolates belonging to ST 70.
Table 3.5 Presence of genomic islands in the three largest bacterial clones as defined by MLST

<table>
<thead>
<tr>
<th>Genomic island</th>
<th>All isolates</th>
<th>Sequence type</th>
<th></th>
<th></th>
<th></th>
<th>$P^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>ST 70</td>
<td>ST 54</td>
<td>ST 167</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of isolates</td>
<td>186</td>
<td>15</td>
<td>11</td>
<td>9</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>GI 2</td>
<td>133 (72%)</td>
<td>11 (73%)</td>
<td>6 (55%)</td>
<td>6 (67%)</td>
<td>0.60</td>
<td></td>
</tr>
<tr>
<td>GI 6</td>
<td>43 (23%)</td>
<td>8 (53%)</td>
<td>1 (9%)</td>
<td>1 (11%)</td>
<td>0.04</td>
<td></td>
</tr>
<tr>
<td>GI 9</td>
<td>22 (12%)</td>
<td>0</td>
<td>1 (9%)</td>
<td>0</td>
<td>0.57</td>
<td></td>
</tr>
<tr>
<td>GI 11</td>
<td>39 (21%)</td>
<td>4 (27%)</td>
<td>0</td>
<td>0</td>
<td>0.08</td>
<td></td>
</tr>
<tr>
<td>GI 16</td>
<td>143 (77%)</td>
<td>10 (67%)</td>
<td>11 (100%)</td>
<td>5 (56%)</td>
<td>0.04</td>
<td></td>
</tr>
</tbody>
</table>

$^a$ Fisher's exact test
This suggests that horizontal gene transfer can alter the gene repertoire ("accessory genome") of this species very rapidly, i.e. before any observable changes in the vertically inherited "core" genome, and is consistent with the hypothesis that isolates that are identical by MLST might exploit different ecological niches.\textsuperscript{249} The MLST data for \textit{B. pseudomallei} has revealed a very high frequency of homologous recombination resulting in limited clustering and low levels of linkage disequilibrium.\textsuperscript{40,90,97} These results also suggest that the loss and acquisition of genomic islands by horizontal gene transfer significantly contributes to the dynamism of the genome. However, it should be noted that these two processes are distinct, and that a high rate of island transfer does not necessarily predict a high rate of homologous recombination or \textit{vice versa}. For example, MLST data for the Gram positive pathogen \textit{Staphylococcus aureus} has revealed only modest rates of homologous recombination, yet the "accessory" genome of this species is known to change very rapidly.\textsuperscript{79,103,143,177,249} \textit{B. pseudomallei} represents an alternative scenario, the data indicating both high rates of homologous recombination together with rapid changes in island repertoire.

\textbf{3.6.7 in silico analysis of genomic islands}

An \textit{in silico} analysis of the five GIs was undertaken to explore the degree of variability in presence, structure and chromosomal insertion site. Nine isolates that are currently undergoing whole genome sequencing by TIGR were examined and compared with strain K96243, which has been sequenced by the Wellcome Trust Sanger Institute\textsuperscript{104} (Table 3.6). Seven TIGR isolates and K96243 are unique strains, but strains 1106b and 1710b were recovered from patients with relapsing melioidosis, and are identical by MLST to strains 1106a and 1710a respectively.

The \textit{in silico} structure, composition and insertion site of the five GIs was
examined using the ten available genome sequences (complete and assemblies from unfinished genomes). GIs 2, 6, 9 and 11 all had flanking repeats, integrases, and mobile element-like genes in the other strains, consistent with the suggestion that these were recently acquired mobile genetic elements. There were two instances where the GI was absent at the predicted site of insertion, but DNA with extended similarity to the island was detected elsewhere in the genome (one example each for GI 2 and GI 6). This may be due to genomic rearrangement in the same strain over time, or may reflect the presence of multiple potential insertion sites for these GIs in the *B. pseudomallei* genome.

Comparison of GI 11 identified an alternative island inserted at orthologous sites in the 1106a and 1106b genomes (Table 3.6, Figure 3.3). This novel island is larger than the K96243 GI 11 (~35 kb). Comparison of the islands in 1106a and 1106b demonstrated that they are virtually identical except for a small internal indel region that is present in 1106a but absent in 1106b. The indel region is between two identical IS elements in 1106a. In 1106b there is a single IS element, suggesting that recombination between the IS elements may have lead to the deletion in this region in 1106b. Although the 1106a and 1106b islands appear to be inserted at the same site as the K96243 GI 11, the attachment sites are different. In the case of K96243, the attachment site is an Ala tRNA, generating 45mer perfect repeats that flank the island. The GI 11 regions in 1106a and 1106b are next to the Ala tRNA, but the flanking DNA in both islands is a 13mer sequence that does not appear to be part of the Ala tRNA, or surrounding sequence in K96243. The alternative GI 11 in 1106a and 1106b contains a putative filamentous haemagglutinin and processing protein that are similar to proteins found in *B. thailandensis* (BTH_I2723 and BTH_I2721 respectively). The homologues of the filamentous haemagglutinin and processing protein in *B. thailandensis* strain E264 are not in an orthologous position relative to the *B.
pseudomallei GI 11 flanking DNA. In the genome of this closely related species, an alternative GI 11 is present (BTH_I3130 to BTH_I3143), which is integrated at the orthologous attachment site (Ala tRNA) to the GI 11 in B. pseudomallei K96243.

Table 3.6 Distribution of islands in sequenced B. pseudomallei strains

<table>
<thead>
<tr>
<th>BP strain</th>
<th>Origin</th>
<th>GI 2</th>
<th>GI 6</th>
<th>GI 9</th>
<th>GI 11</th>
<th>GI 16</th>
</tr>
</thead>
<tbody>
<tr>
<td>K96243</td>
<td>Thailand</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>406e</td>
<td>Thailand</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>NC</td>
</tr>
<tr>
<td>1106a</td>
<td>Thailand</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>*</td>
<td>Alt</td>
</tr>
<tr>
<td>1106b</td>
<td>Thailand</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>*</td>
<td>Alt</td>
</tr>
<tr>
<td>1710a</td>
<td>Thailand</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+/- (BPSS2057 to BPSS2076)</td>
</tr>
<tr>
<td>1710b</td>
<td>Thailand</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+/- (BPSS2057 to BPSS2076)</td>
</tr>
<tr>
<td>Pasteur 6068</td>
<td>Vietnam</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+/- (BPSS2057 to BPSS2076)</td>
</tr>
<tr>
<td>S13</td>
<td>Singapore</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+/- (BPSS2057 to BPSS2076)</td>
</tr>
<tr>
<td>668</td>
<td>Australia</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1655</td>
<td>Australia</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
</tbody>
</table>

Strain pairs 1106a & 1106b, and 1710a & 1710b were obtained from two patients who each relapsed (from a persistent focus) 3 years after the primary episode of melioidosis. ‘a’ denotes primary isolate and ‘b’ denotes relapse isolate.

- Island absent; + Island present; +/- Island partial present (CDSs missing); -* Island absent at the site of insertion but similar region present elsewhere in the genome; +* Island present and contains extra sequence; NC No contig covering this region entirely; Alt Alternative island present at this site (no sequence similarity)
Figure 3.3 Comparison of GI 11 in *B. pseudomallei* strains K96243, 1106a and 1106b. The results of a TBLASTX comparison of the GI 11 region from K96243 (top), 1106a (middle) and 1106b (bottom) is displayed using the Artemis Comparison Tool (ACT). A plot of the % GC content of each sequence is displayed. The colored bars separating each genome (red and blue) represent similarity matches identified by BLASTN analysis. Red lines link matches in the same orientation; blue lines link matches in the reverse orientation.
Micro-evolutionary diversification was also observed in GI 16, in which an internal deletion encompassing CDSs BPSS2057 to BPSS2076 was noted in four unique strains and both relapse isolates (Table 3.6). One other strain (1655) contained the complete GI 16 plus additional sequence, both within the island and downstream of it. Figure 3.4 shows a comparison of GI 16 regions in strains K96243, 1655 and 1710b (which has the internal deletion). The lack of integrases and other mobile genetic element genes, and high variability in gene content, suggests that GI 16 is distinct from the other GIs considered here which have structures more typical of mobile genetic elements. GI 16 has a mosaic structure consistent with a long history of multiple insertion and deletion events. The role of horizontal gene transfer in the evolution of this region is supported by anomalous GC composition (Figure 3.4). One of the genes used during PCR to detect GI 16 lies within the deleted region (BPSS 2061); 58% of the Thai isolate collection were positive for GI 16 using primers to this target, compared with a combined result of 76% for both GI 16 gene targets. The internal deletion may account for the discrepancy between these figures.
Figure 3.4 Comparison of GI 16 in *B. pseudomallei* strains 1655, K96243 and 1710b. The results of a BLASTN comparison of the GI 16 regions from the 1655 (top), K96243 (middle) and 1710b (bottom) is displayed using the Artemis Comparison Tool (ACT). A plot of the % GC content of each sequence is displayed. The colored bars separating each genome (red and blue) represent similarity matches identified by BLASTN analysis. Red lines link matches in the same orientation; blue lines link matches in the reverse orientation.
This survey of divergent putative GIs in a well defined and clinically significant *B. pseudomallei* population has demonstrated that genetic exchange is widespread. The *in silico* analysis provided evidence that GI 16 may represent a composite region rather than a genomic island, and defined a novel GI at the predicted insertion site for GI 11. A detailed analysis of the relationship between five GIs and both the ability to cause disease and a range of clinical features failed to show an association between the two. This study confirms the utility of a range of approaches in defining the presence and significance of genomic variation in natural populations of *B. pseudomallei*. Further studies are required to determine the full extent of variability in GIs, and their relationship to biological fitness in the environment and to disease pathogenesis in the human host.
Chapter 4. Results II: Distribution of \textit{B. pseudomallei} clones in relation to geographical location and virulence

4.1 Chapter content

MLST has been used to characterize the population structure and diversity of a number of important human pathogens, most of which are associated with colonization and human-to-human transmission. The free movement of colonized individuals between countries and continents leads to worldwide dispersal of bacterial genotypes with little geographic structure apparent in many species. In contrast, \textit{B. pseudomallei} is an environmental saprophyte which does not colonize humans or other mammals and is not transmitted between them. Inter-continental migration of soil microorganisms may potentially arise by long-range wind dispersal, but the survival of non-spore forming bacteria may be limited to shorter distances. If so, allopatric populations should arise corresponding to distinct endemic genotypes ("geotypes"). \textit{B. pseudomallei} provides an ideal species to address the issue of migration and its effect on population structure. Furthermore, the ability to assign a country or geographic region as the likely source of a deliberately released strain may be of value.

The MLST data currently available for \textit{B. pseudomallei} has revealed a low level of sequence diversity but a relatively large number of multilocus genotypes (STs) and limited clustering. Analyses of these data, which are mostly restricted to isolates from Northern Australia, suggest a high rate of recombination. However, there is currently little evidence concerning variation in virulence potential between isolates, nor have comparisons been drawn between isolates from different
geographical origins to examine the extent of large-scale migration. The objective of this study was to use MLST to compare *B. pseudomallei* isolates recovered from disease and from the environment within a single geographical region (northeast Thailand), and compare these results to pre-existing data to examine the extent of migration between Northern Australia and South East Asia.

4.2 Bacterial strains

A total of 266 *B. pseudomallei* isolates were characterized by MLST. Of these, 83 were isolated from the environment as described in section 3.2. 183 isolates were from patients presenting to Sappasithiprasong Hospital in Ubon Ratchathani between 1989 and 2002, of which 79 were associated with a single clinical syndrome (acute suppurative parotitis in children). The remaining 104 invasive isolates were from consecutive adult patients diagnosed with melioidosis as described in section 3.4. A single isolate was used from each patient. The previously published MLST profiles for 158 strains isolated by Prof. Bart Currie, in Northern Australia were downloaded from the MLST web site for comparison (httn://www.mlst.net).

4.3 Results and Discussion

4.3.1 Diversity of the *B. pseudomallei* population from northeast Thailand

The 266 clinical and environmental isolates of *B. pseudomallei* from Thailand characterized by MLST corresponded to 123 STs. The four most commonly recovered STs were ST 70 (*n*=21), ST 167 (*n*=15), ST 54 (*n*=12) and ST 58 (*n*=11). Of the remainder, 37 (13.9%) STs were represented by between 2 to 8 isolates, and 82 (66.7%) of the STs were represented by a single isolate. A greater degree of allelic diversity was noted in *gmhD* and *nark* than in the other five gene loci; 10 and 13
alleles were found in these two loci respectively, whereas only 3 to 6 alleles were noted at the other five loci (Table 4.1). Despite the large number of STs detected, overall nucleotide sequence diversity was very low; when all unique alleles identified in the current study were considered, pairwise comparisons revealed an average of only 0.26-0.48% nucleotide divergence over all seven loci (Table 4.1). Estimates of synonymous and non-synonymous changes are shown in Table 4.1.

eBURST was used to identify groups of closely related STs (clonal complexes) and to examine the extent to which the data corresponded to a simple model of clonal expansion (Figure 4.1). The majority of Thai isolates corresponded to a single large clonal complex (CC48) which did not conform to a simple pattern of radial expansion; the bootstrap support for the founder of this group was 75% and therefore it could not be assigned with a high degree of confidence. Because of the relatively high level of diversity at narK and gmhD, many of the single locus variant links shown in figure 4.1 correspond to allelic differences at these two loci. A second clonal complex (CC70) was also identified which much more closely corresponded to a simple radial expansion from a founding genotype, ST70 (bootstrap support 99%). In contrast, a much higher degree of variation (a greater proportion of singleton or unlinked isolates) was observed for the Australian population (Figure 4.1).
Table 4.1. Allele frequency and diversity for 266 Thai isolates of *B. pseudomallei*

<table>
<thead>
<tr>
<th>Locus</th>
<th>Number of alleles</th>
<th>Average pair-wise diversity (%)</th>
<th>$dS/dN$</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>ace</em></td>
<td>3</td>
<td>0.26</td>
<td>$\alpha$</td>
</tr>
<tr>
<td><em>gltB</em></td>
<td>5</td>
<td>0.31</td>
<td>$\alpha$</td>
</tr>
<tr>
<td><em>gmhD</em></td>
<td>10</td>
<td>0.47</td>
<td>8.48</td>
</tr>
<tr>
<td><em>lepA</em></td>
<td>5</td>
<td>0.33</td>
<td>$\alpha$</td>
</tr>
<tr>
<td><em>lipA</em></td>
<td>4</td>
<td>0.41</td>
<td>$\alpha$</td>
</tr>
<tr>
<td><em>nark</em></td>
<td>13</td>
<td>0.5</td>
<td>14.15</td>
</tr>
<tr>
<td><em>ndh</em></td>
<td>6</td>
<td>0.38</td>
<td>3.65</td>
</tr>
</tbody>
</table>
Figure 4.1 eBURST of 266 \textit{B. pseudomallei} isolates obtained from the environment (n=83) or associated with human melioidosis (n=183), together with 158 isolates from Northern Australia which were mainly disease associated. Isolates from Thailand are labeled in blue, and those from Australia are labeled in black. The majority of Thai isolates correspond to a single large clonal complex (CC48) which does not conform to a simple pattern of radial expansion. A second clonal complex (CC70) much more closely corresponds to a simple radial expansion from a founding genotype, ST70. A much higher degree of variation is observed for the Australian population.
4.3.2 Comparisons within the Thai bacterial population

It is unclear whether strains of *B. pseudomallei* differ substantially in their ability to cause human infection or whether strains are equally likely to cause disease in susceptible hosts. This question was addressed by comparing the genotypes of isolates from invasive disease and those cultured from soil or water in northeast Thailand. 59 STs were identified among the 83 environmental isolates (0.71 STs per isolate), and 88 STs among the 183 invasive disease isolates (0.48 STs per isolate). The origin (environmental or invasive) of isolates corresponding to the ten most commonly observed STs are shown in Table 4.2. Invasive isolates predominated within these common STs; 43.7% of all invasive isolates belonged to these 10 STs compared to only 25.3% of soil isolates (p<0.0001). There was a significant difference in the classification index (CI) between environmental and invasive isolates (p<0.001), confirming that genotypes were not distributed randomly between the two samples.

ST70 was the most frequently recovered strain from invasive disease in northeast Thailand and was noted 17 times among the 183 disease isolates (9.3%), but only four times in the 83 environmental isolates (4.8%). ST70 and a single-locus variant (SLV) of ST70 (ST32), have been identified as the cause of melioidosis in captive birds, marine mammals and humans in Hong Kong. Other STs also appeared to be associated with invasive disease in northeast Thailand (e.g. STs 34, 58, 167 and 221), whilst others were under-represented amongst the disease isolates (e.g. STs 54 and 60).
Table 4.2 Sources of *B. pseudomallei* isolates in the 10 largest MLST clones for northeast Thai strains

<table>
<thead>
<tr>
<th>ST</th>
<th>Invasive isolates (n=183)</th>
<th>Soil isolates (n=83)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ST 70</td>
<td>17 (9.3%)</td>
<td>4 (4.8%)</td>
</tr>
<tr>
<td>ST 167</td>
<td>13 (7.1%)</td>
<td>2 (2.4%)</td>
</tr>
<tr>
<td>ST 54</td>
<td>6 (3.3%)</td>
<td>7 (8.5%)</td>
</tr>
<tr>
<td>ST 58</td>
<td>11 (6.0%)</td>
<td>0</td>
</tr>
<tr>
<td>ST 34</td>
<td>7 (3.8%)</td>
<td>1 (1.2%)</td>
</tr>
<tr>
<td>ST 60</td>
<td>4 (2.2%)</td>
<td>4 (4.8%)</td>
</tr>
<tr>
<td>ST 221</td>
<td>7 (3.8%)</td>
<td>0</td>
</tr>
<tr>
<td>ST 174</td>
<td>6 (3.3%)</td>
<td>0</td>
</tr>
<tr>
<td>ST 33</td>
<td>5 (2.7%)</td>
<td>1 (1.2%)</td>
</tr>
<tr>
<td>ST 48</td>
<td>4 (2.2%)</td>
<td>2 (2.4%)</td>
</tr>
<tr>
<td>TOTAL</td>
<td>80 (43.7%)</td>
<td>21 (25.3%)</td>
</tr>
</tbody>
</table>
A classification index proposed by Jolley et al.\textsuperscript{122} can also be used to detect population differences in allele frequency at each of the MLST loci.\textsuperscript{122} This approach is likely to be less sensitive in detecting differences between populations than when STs are examined, particularly in highly recombinogenic species. When the environmental and invasive isolates were compared on a locus-by-locus basis, only \textit{gmhD} (p=0.01) and \textit{narK} (p=0.05) showed a non-random distribution of alleles. For example, \textit{gmhD} allele 5 was noted 13 times in the 183 invasive isolates, but was absent from the 83 environmental strains (p=0.01). Similarly, \textit{narK} allele 29 was noted 7 times in the invasive isolates but was absent from the environmental strains, whilst \textit{narK} allele 2 was noted 20 times in both cases. A relatively high allelic diversity (and hence discriminatory power, Table 4.1) may account for why differences were only detected in two genes. These alleles may not have any direct bearing on virulence potential, but it is possible that they are linked to nearby alleles or genomic islands that influence pathogenicity. Such close linkage may also help to explain the high level of diversity at these loci.

To explore further the putative role of bacterial factors in disease, Thai isolates were examined from cases associated with acute suppurative parotitis, a single defined clinical manifestation with a relatively good prognosis that is rare in northern Australia, but is the presenting feature in one third of Thai childhood cases.\textsuperscript{67} There were 45 STs among the 79 isolates associated with parotitis (0.56 STs per strain) and 63 STs among the other 104 invasive disease isolates (0.6 STs per strain), indicating that isolates causing parotid infection were as diverse as those associated with a range of other clinical presentations. There were no significant differences in the CI for STs associated with parotid disease and other invasive isolates, but a significant difference was found in the distribution of \textit{ace} alleles (p=0.03). Only three alleles were
observed at this locus for all Thai isolates, one of which (allele 4) was rare and occurred five times. *ace* allele 1 was represented in 46/79 (58.2%) of the isolates associated with parotitis, and 41/103 (39.8%) of the other invasive isolates. These figures are inverted for allele 3, being noted in 31/79 (39.2%) of the parotid isolates and 60/103 (58.2%) of the other invasive isolates.

### 4.3.3 Comparison between Thai and Australian bacterial populations

To detect geographical structuring on an inter-continental scale, the 88 STs from invasive isolates obtained from Thailand were compared with 96 STs identified previously among 158 isolates (mainly associated with disease) from Northern Australia. No STs were common to both Thailand and Australia (p<0.0001). This complete differentiation was also found when the environmental isolates from Thailand were included in the analysis. Comparisons of CI also revealed that allele frequencies at each gene were non-randomly distributed between the Thai and Australian populations (p=0.0001). There was also a non-random distribution of single polymorphic sites; of the 67 polymorphisms present in the Australian data, only 19 (28.3%) were also present in the Thai data. Individual 2x2 chi squared tests revealed highly significant differences in the frequencies of 12 of these 19 polymorphisms (p<0.0001). Together, these analyses strongly suggest very limited gene flow and that the populations of *B. pseudomallei* have been diverging independently in Australia and Thailand.

To explore further the separation of Australian and Thai isolates, a neighbor-joining tree using the concatenated sequences of all seven loci was constructed (Figure 4.2). This dendrogram confirmed that the Australian and Thai isolates were distinct, with some exceptions. The tree suggested four separate clades; one consisted
of Thai isolates alone, another consisted of the majority of Australian isolates, and the remaining two clades were mixed, with some evidence for separation of the two populations within the clades. The tree was poorly supported by bootstrap scores (not shown), hence the topological details are not likely to be indicative of the true phylogeny. Despite the poor phylogenetic signal, however, the tree supports a broad distinction between the Thai and Australian populations.

In addition to being distinct on the basis of ST, allele, and polymorphic site frequency, there were other notable differences between the Thai and Australian samples. A total of 46 different alleles were noted in the 183 invasive Thai isolates, whereas 98 distinct alleles were noted in the 158 Australian isolates, confirming the suggestion from Figure 4.1 that the Australian population is more diverse. The mean heterozygosity per locus \((H)\) was higher for the Australian (0.67 +/- 0.07) than for the Thai invasive isolates (0.57 +/- 0.07), although this difference was not significant. There was a striking difference in the distribution of allele frequencies between the two populations. Of the 98 alleles in the Australian data, 38 (38.7%) were found in only one isolate, corresponding to an average of 0.24 unique alleles per isolate. In contrast, the number of alleles occurring once in Thai isolates was 9/46 (19.6%), corresponding to an average of 0.05 unique alleles per isolate (p= 0.02). Australian strains containing one or more unique alleles were no more diverged from each other (average pairwise allelic mismatch = 4.7, mean \(H = 0.694 +/- 0.0976\) ) than Australian isolates overall (average pairwise allelic mismatch = 4.6; mean \(H 0.67 +/- 0.07\) ), confirming that this difference is not due to the inclusion of a few very distinct genotypes in the Australian sample. Alternative possibilities include sampling artifacts (the Australian isolates were recovered over a wider geographic area),
differences in mutation or recombination rates, or an Australian origin of *B. pseudomallei* in Australia and subsequent spread to South East Asia.

The Thai data revealed a small number of alleles at each locus but a relatively large number of STs, suggesting that alleles in the population may frequently recombine. This is consistent with a previous analysis for the entire MLST dataset. The most direct method to evaluate the evidence of recombination is to observe the nature of the changes that occurred in the ancestral STs and the effect on ST variability. Applying this approach to the data from this study revealed a low incidence of recombination. The low levels, when detected, were associated with an extra variable character or allele that may have been introduced more recently into the population. Despite the evidence for recombination, the majority of isolates from both regions clustered into major clade complexes, where again the great majority of these changes can be explained as the result of recombination.

An examination of phylogenetic congruence based on the method of Peltier et al. 2003 was also performed. A NJ tree was constructed from the allele profiles of the isolates using the START package. Fifty divergent STs were plotted from this tree, and Maximum Likelihood trees constructed for each gene. For all 42

Figure 4.2 Neighbor-joining tree using the concatenated sequences of all seven loci for Thai and Australian isolates (n=266 and n=158, respectively). Isolates from Thailand are labeled in red; those from Australia are labeled in blue. The majority of Australian isolates appear to belong to a single clade, although a small number of Thai isolates also cluster within this group, possibly reflecting recent migration events. A clade composed entirely of Thai isolates is also apparent.
4.3.4 Impact of homologous recombination on diversification of the *B. pseudomallei* population

The Thai data revealed a small number of alleles at each locus but a relatively large number of STs, suggesting that alleles in the population may frequently recombine. This is consistent with a previous analysis for the entire MLST dataset.\(^97\) The most direct method to evaluate the contribution of recombination is to observe the nature of the changes that occur as an ancestral ST diversifies to form SLVs. Applying this approach to the ST70 clonal complex showed that the altered alleles in 9 of the 10 SLVs differed from the allele in ST70 at more than one nucleotide site. The low level of sequence diversity would be expected to lead to an underestimate of the extent of recombination, as alleles introduced by recombination may frequently differ at a single site and could be scored as a point mutation. The fact that only 1/10 allelic changes among the SLVs involved a single nucleotide change suggests that the rate of recombination is very high, relative to mutation. This is supported by an examination of the SLVs of the subgroup founders\(^82,226\) in the major clonal complex, where again the great majority of allelic changes can be assigned as the result of recombination.

An examination of phylogenetic congruence based on the method of Feil *et al.*, 2001\(^80\) was also performed. A UPGMA tree was constructed from the allelic profiles of all the Thai isolates using the START package. 30 diverse STs were picked from this tree, and Maximum Likelihood trees constructed for each gene. For all 42 pairwise comparisons, the maximum likelihood trees constructed for each gene from the Thai dataset were no more similar to the other gene trees than to trees of random topology. This total lack of congruence is consistent with high rates of recombination, but the paucity of informative sites might limit the ability to detect significant
congruence between the loci. The lack of phylogenetic consistency within the Thai data contrasts with the broad differentiation between the Thai and Australian data evident from Figure 4.2.

4.4 Chapter summary

This study has demonstrated the utility of MLST in defining inter-continental geographical segregation of an important soil-dwelling pathogen. It is likely that the differences between the Australian and Thai populations of *B. pseudomallei* reflect a historical pattern of migration and concomitant genetic drift, possibly in association with niche adaptation. The clear distinction between the endemic populations of Thailand and Australia shown here has important implications for tracking the source of outbreaks or deliberate release. From the study of isolates from Thailand, it was apparent that the most common STs are over-represented in the isolate population associated with disease. Putative differences in virulence potential in a given ST may reflect rapid acquisition and loss of mobile genetic elements, and strain differentiation may underlie differences in disease presentation between Thailand and Australia. Many cases of melioidosis occur in individuals with risk factors such as diabetes mellitus and renal impairment. Further study is required to define whether at-risk individuals become infected by the same bacterial population as those without definable risk factors, or whether only a subset of strains are able to cause disease in the latter group.
Chapter 5. Results III: Multilocus sequence typing of *Burkholderia thailandensis*

5.1 Chapter content

*B. thailandensis* is a saprophytic organism that is closely related to *B. pseudomallei*. The genetic and biochemical profiles of the two organisms are similar, but *B. thailandensis* is avirulent with the exception of very rare case reports.\(^{271}\) MLST has been used previously to define a small number of *B. thailandensis* isolates. The purpose of this study was to:

1. Extend the MLST analysis to a collection of *B. thailandensis* isolates obtained from the environment in Thailand.
2. Compare the genetic relatedness of *B. pseudomallei* and *B. thailandensis* as defined by MLST.

5.2 Bacterial strains

A total of 92 *B. thailandensis* isolates were included in this study. Seventy-seven isolates were obtained from soil collected from central and northeast Thailand between 1990 and 2002, and one isolate was obtained from soil in Seam Reap, Cambodia in 2005. The remaining 14 isolates represent strains present in the *B. pseudomallei* MLST database ([http://bpseudomallei.mlst.net](http://bpseudomallei.mlst.net)), details for which are shown in Table 5.1. Overall, 86 strains were isolated from Thailand.
Table 5.1 Data on 14 *B. thailandensis* isolates obtained from the MLST database

<table>
<thead>
<tr>
<th>No.</th>
<th>ST</th>
<th>Strain</th>
<th>original number</th>
<th>Country</th>
<th>Year of isolation</th>
<th>Source</th>
<th>Sender</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>73</td>
<td>82172</td>
<td>34</td>
<td>France</td>
<td>1982</td>
<td>Chicken</td>
<td>Dr. Tyrone L. Pitt</td>
</tr>
<tr>
<td>2</td>
<td>74</td>
<td>E27</td>
<td>A3</td>
<td>Thailand</td>
<td>1990</td>
<td>Environment</td>
<td>Dr. Andrew J. Simpson</td>
</tr>
<tr>
<td>3</td>
<td>75</td>
<td>VN 534b</td>
<td>A5</td>
<td>Vietnam</td>
<td>1997</td>
<td>Environment</td>
<td>Dr. Andrew J. Simpson</td>
</tr>
<tr>
<td>4</td>
<td>76</td>
<td>LE1</td>
<td>A16</td>
<td>Laos PDR</td>
<td>1999</td>
<td>Environment</td>
<td>Dr. Andrew J. Simpson</td>
</tr>
<tr>
<td>5</td>
<td>77</td>
<td>E125</td>
<td>E125 (ara+)</td>
<td>Thailand</td>
<td>1991</td>
<td>Environment</td>
<td>Dr. Tyrone L. Pitt</td>
</tr>
<tr>
<td>6</td>
<td>77</td>
<td>E327</td>
<td>A9</td>
<td>Thailand</td>
<td>1998</td>
<td>Environment</td>
<td>Dr. Andrew J. Simpson</td>
</tr>
<tr>
<td>7</td>
<td>79</td>
<td>E294</td>
<td>A14</td>
<td>Thailand</td>
<td>1994</td>
<td>Environment</td>
<td>Dr. Andrew J. Simpson</td>
</tr>
<tr>
<td>8</td>
<td>80</td>
<td>E111</td>
<td>E111 (ara+)</td>
<td>Thailand</td>
<td>1991</td>
<td>Environment</td>
<td>Dr. Tyrone L. Pitt</td>
</tr>
<tr>
<td>9</td>
<td>80</td>
<td>E216</td>
<td>E216 (ara+)</td>
<td>Thailand</td>
<td>1992</td>
<td>Environment</td>
<td>Dr. Tyrone L. Pitt</td>
</tr>
<tr>
<td>10</td>
<td>80</td>
<td>G32</td>
<td>G32 (ara+)</td>
<td>Thailand</td>
<td>1993</td>
<td>Environment</td>
<td>Dr. Tyrone L. Pitt</td>
</tr>
<tr>
<td>11</td>
<td>80</td>
<td>E264</td>
<td>ATCC 700388</td>
<td>Thailand</td>
<td>1993</td>
<td>Environment</td>
<td>Dr. Jay E. Gee</td>
</tr>
<tr>
<td>12</td>
<td>101</td>
<td>200301589</td>
<td>-</td>
<td>USA</td>
<td>2003</td>
<td>Human</td>
<td>Dr. Jay E. Gee</td>
</tr>
<tr>
<td>13</td>
<td>345</td>
<td>E254</td>
<td>-</td>
<td>Thailand</td>
<td>1992</td>
<td>Environment</td>
<td>Dr. Jay E. Gee</td>
</tr>
<tr>
<td>14</td>
<td>352</td>
<td>132/02</td>
<td>PHLS E082</td>
<td>Thailand</td>
<td>1990</td>
<td>Rice Field</td>
<td>Dr. Tyrone L. Pitt</td>
</tr>
</tbody>
</table>
5.3 Results

5.3.1 Allelic distribution and population structure of *B. thailandensis*

The existing *B. pseudomallei* MLST scheme was used to characterize the *B. thailandensis* isolate collection. Twenty STs were defined for the 92 *B. thailandensis* isolates (0.22 ST per isolate), the MLST profiles for which are shown in Appendix VI. The three largest clones were ST80, ST345 and ST77, which contained 21 (22.8%), 16 (17.4%) and 12 (13.0%) isolates, respectively.

Table 5.2 shows the number of alleles found in the 92 isolates. This ranged from 2 (*ace*) to 6 (*lipA*). None of the *B. thailandensis* alleles were present in the *B. pseudomallei* database.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Number of alleles</th>
<th>allele number</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>ace</em></td>
<td>2</td>
<td>5, 6</td>
</tr>
<tr>
<td><em>gltB</em></td>
<td>5</td>
<td>5, 8, 9, 10, 31</td>
</tr>
<tr>
<td><em>gmhD</em></td>
<td>3</td>
<td>9, 15, 16</td>
</tr>
<tr>
<td><em>lepA</em></td>
<td>5</td>
<td>5, 8, 11, 29, 30</td>
</tr>
<tr>
<td><em>lipA</em></td>
<td>6</td>
<td>7, 9, 10, 11, 14, 24</td>
</tr>
<tr>
<td><em>nark</em></td>
<td>3</td>
<td>7, 14, 20</td>
</tr>
<tr>
<td><em>ndh</em></td>
<td>4</td>
<td>5, 8, 9, 14</td>
</tr>
</tbody>
</table>
A UPGMA dendrogram based on allelic data for *B. thailandensis* is shown in Figure 5.1. Two distantly related branches can be observed, with the majority of isolates in one cluster, and three STs in a second cluster (ST73, ST101 and ST1000); the latter were isolated in France, the US and Cambodia, respectively. The allelic data were further represented using eBURST (Figure 5.2). This clearly shows separation of ST73, ST101 and ST1000 (which are singletons), from the remainder which form a single clonal complex.
Figure 5.1 UPGMA dendrogram of 92 B. thailandensis isolates
Figure 5.2 eBURST for 92 *B. thailandensis* isolates. Numbers represent STs. Linking lines between STs represent single locus variants (sharing six of seven alleles). The predicted group founder is labeled in blue, while the sub-group founders are labeled in yellow. The number of isolates in each ST is proportional to the size of the dot.
Neighbor-joining trees using concatenated sequence of all seven MLST genes were constructed in order to compare the relatedness between the population of *B. thailandensis* isolates (Figure 5.3), and between the closely-related species *B. thailandensis*, *B. mallei* and *B. pseudomallei* (Figure 5.4). The *B. thailandensis* tree was consistent with eBURST and the UPGMA dendrogram and showed two branches, one of which contained ST73, 101 and 1000 (Figure 5.3). The tree constructed using data for all three species clearly demonstrated that *B. thailandensis* is diverse from *B. pseudomallei*, while *B. mallei* clustered with *B. pseudomallei*. *B. oklahomensis* is genetically distant and is shown as an out-group. A population snapshot of all three species using eBURST (Figure 5.5) showed a clonal complex of *B. thailandensis* as a separate group.
Figure 5.3 Neighbor-joining tree using concatenated sequence of all 7 loci for *B. thailandensis* isolates.
Figure 5.4 Neighbor-joining tree using concatenated sequence of all 7 loci for *B. thailandensis* (▲), *B. pseudomallei* (▲), *B. mallei* (▲), *B. oklahomensis* (▲) and *Burkholderia spp.*(▲). Data for the last three were downloaded from the MLST database.
Figure 5.5 eBURST displays population snapshot of *B. thailandensis* (labeled in pink) compared with all *B. pseudomallei* and closely related species. The green label is specified for ST1000 which is not submitted to the MLST database.
5.3.2 Diversity of *B. thailandensis* Thai isolates

The 86 *B. thailandensis* isolates from Thailand corresponded to sixteen STs. ST 80 was the largest clone (20 isolates; 23.3%) and 6 STs were found only once (ST 74, ST 356, ST 357, ST 358, ST 359 and ST 363). STs were considered according to site of isolation, in which strains were separated into those isolated in the northeast (n=67) and central region of Thailand (n=19) (Table 5.3). Six STs were found in both regions (pink digit), including the three largest clones (ST 80, ST 345 and ST 77), while another 9 STs were found only in northeast Thailand (black digits), and one ST (ST 363) was unique to the central region (green digit) (Figure 5.5).

Some alleles were restricted to northeast Thailand (for example, *gltB* 8 and 31, *lepA* 29 and 30 and *lipA* 24), but no alleles were only found in isolates from the central region.
Figure 5.5 eBURST for 86 *B. thailandensis* isolates. The number of isolates in each ST is proportional to the size of the dot. Black digit means isolates found only in northeast Thailand, green means ST found only in central Thailand, and pink means STs found in both regions.
5.3.3 Comparison between Thai isolates of *B. thailandensis* and *B. pseudomallei*

Eighty-six *B. thailandensis* isolates from Thailand were compared with the 266 Thai *B. pseudomallei* isolates described previously. STs per isolate were calculated to be 0.46 for *B. pseudomallei* compared with 0.19 ST per isolate for *B. thailandensis*, indicating a greater diversity for *B. pseudomallei*. The allelic diversity of these two species was defined and compared (Table 5.3). The average pairwise diversity and number of polymorphic nucleotide sites for *B. thailandensis* (0.20% and 0-3 sites) was considerably lower than that for *B. pseudomallei* (0.38% and 2-9 sites). Estimates of synonymous and non-synonymous changes for both species are shown in Table 5.5. The mean heterozygosity per locus (*H*) within the Thai population of *B. thailandensis* and *B. pseudomallei* were 0.24 ± 0.08 and 0.58 ± 0.07, respectively. These findings imply that the *B. pseudomallei* population is considerably more diverse than the *B. thailandensis* population.

A second comparison was made between the 86 *B. thailandensis* isolates from Thailand and the 83 *B. pseudomallei* isolates obtained from the environment. This confirmed the initial observation that *B. pseudomallei* isolated from the environment is more diverse than the saprophytic *B. thailandensis*. The mean *H* of the *B. pseudomallei* Thai environmental isolates (0.57 ± 0.06) was similar to the overall Thai population (0.58 ± 0.07), with an average pairwise diversity of 0.37% and 0.38%, respectively.
Table 5.3 Allelic diversity for 86 \textit{B. thailandensis} and 266 \textit{B. pseudomallei} from Thailand

<table>
<thead>
<tr>
<th>Locus</th>
<th>86 \textit{B. thailandensis} isolates</th>
<th>266 \textit{B. pseudomallei} isolates</th>
<th>Total different nucleotide sites</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No. of alleles</td>
<td>polymorphic nucleotide sites</td>
<td>Avg. pairwise diversity (%)</td>
</tr>
<tr>
<td>ace</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>gltB</td>
<td>3</td>
<td>3</td>
<td>0.32</td>
</tr>
<tr>
<td>gmspD</td>
<td>2</td>
<td>1</td>
<td>0.21</td>
</tr>
<tr>
<td>lepA</td>
<td>3</td>
<td>2</td>
<td>0.27</td>
</tr>
<tr>
<td>lipA</td>
<td>4</td>
<td>3</td>
<td>0.37</td>
</tr>
<tr>
<td>mnrk</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ndh</td>
<td>2</td>
<td>1</td>
<td>0.23</td>
</tr>
</tbody>
</table>
5.4 Discussion

This chapter describes the application of the *B. pseudomallei* MLST scheme to the closely related *B. thailandensis*. Consistent with a previous report, MLST demonstrated a clear genetic distinction between *B. thailandensis* and *B. pseudomallei*. No alleles were shared between the two species. The *B. thailandensis* population was resolved into two phylogenetic groups, with 3 strains distantly related to the remainder of the collection but closely related to each other. These 3 strains were from France, the U.S. and Cambodia. The strain from Cambodia was recently isolated from soil in Seam Reap, the strain from the U.S. was associated with pneumonia and septicaemia following near drowning in a child, and the strain from France was isolated from a chicken in whom it was considered a pathogen, thus all three can be accurately placed. It is unclear whether more strains would reside in this phylogenetic cluster, since the collection tested was strongly biased towards strains from Thailand (a function of availability).

The population structure of *B. thailandensis* and *B. pseudomallei* were compared. The population of *B pseudomallei* was more diverse than that for *B. thailandensis*. This was clearly demonstrated using eBURST, and was reflected in the higher value of ST per isolate and the greater degree of allelic variation for *B. pseudomallei*. This difference was observed when comparing similar strain numbers for each species cultured from the soil in Thailand. The basis for this difference is not clear. One explanation is that *B. pseudomallei* is ancestral to *B. thailandensis*, with a shorter evolutionary time of the latter over which to diverge. This is consistent with the finding that *B. mallei*, which appears to have recently evolved from *B. pseudomallei*, is clonal. An alternative explanation for tighter clustering of *B. thailandensis* is that the population is more constrained by its ecological niche.
compared with *B. pseudomallei* which has a range of hosts in addition to existence in the environment. It is also possible that the efficiency of recombination is lower in *B. thailandensis*, perhaps due to differences in mechanisms that facilitate horizontal gene transfer such as susceptibility to phage and their distribution in the environment.

Although *B. pseudomallei* and *B. thailandensis* are considered to be closely related, the finding that no housekeeping gene loci were shared between the two species indicates that there is considerable genetic diversity between the two. It is possible that the complete lack of shared alleles reflects a specific barrier to genetic exchange, the most likely explanation being phage susceptibility.

Evaluation of Thai isolates cultured from the northeast and central regions of Thailand demonstrated that some STs were shared between the two regions, and some were unique. The small number of isolates from the central regions combined with the relatively high genetic diversity of *B. thailandensis* indicates that any interpretation should be made with caution. Although it is possible that some STs are region specific, a larger study would be required before this could be accepted as a robust conclusion.

### 5.5 Chapter Summary

This study represents the first evaluation of the population genetic structure of *B. thailandensis*. The population was more clonal than that for *B. pseudomallei* isolated from the same geographical region. The implications of the phylogenetic split observed for *B. thailandensis* is unclear and warrants further study with a larger strain collection isolated from a more diverse geographical area.
Chapter 6. Concluding comments

1. MLST was applied to a collection of *B. pseudomallei* isolated in Thailand, and these data were compared to that for a collection of *B. pseudomallei* isolated in Northern Australia. Isolates associated with a single disease entity (parotitis) in Thailand were as diverse as Thai isolates associated with a range of clinical manifestations. A striking finding was the geographical segregation observed between isolates from Thailand and Australia. This is consistent with geographic segregation and independent evolution, possibly following land mass separation of Asia from Australia.

2. An evaluation was conducted of the presence and significance of 5 specific genomic islands in a population of *B. pseudomallei* isolated in Thailand. There was considerable variability, both within and between the three clones examined, in the presence of the five GIs, suggesting that horizontal gene transfer can rapidly alter the gene repertoire of this species. This study also demonstrated the utility of a range of approaches in defining the presence and significance of genomic variation in natural populations of *B. pseudomallei*.

3. The *B. pseudomallei* MLST scheme was applied to a population of *B. thailandensis* isolates cultured from the environment in Thailand, and existing MLST data was downloaded from the mlst database. This demonstrated that *B. thailandensis* fell into two phylogenetically distinct clusters. The smaller cluster contained isolates from outside of Thailand, and it is possible that this clustering represents geographical variation in population although conclusions are limited by the number of isolates obtained from outside of Thailand. *B. thailandensis* was more clonally restricted. The basis and effects of this require further study.
References


63. Currie, B. J., M. Mayo, N. M. Anstey, P. Donohoe, A. Haase, and D. J. Kemp. 2001. A cluster of melioidosis cases from an endemic region is
clonal and is linked to the water supply using molecular typing of

Rev. 4:52-60.

74:115-119.


67. Dance, D. A., T. M. Davis, Y. Wattanagoon, W. Chaowagul, P. Salphan,
suppurative parotitis caused by Pseudomonas pseudomallei in children. J.
Infect. Dis. 159:654-660.

Rec. 130:525-529.

and immunity in mice and hamsters. II. Studies with avirulent strains of

70. Dejsirilert, S., E. Kondo, D. Chiewsilp, and K. Kanai. 1991. Growth and

1999. Phylogenetic analysis of Ara+ and Ara- Burkholderia pseudomallei
isolates and development of a multiplex PCR procedure for rapid

infectious disease, melioidosis, in temperate countries. Arch. Inst. Pasteur
Tunis 63:69-73.

Winstanley. 2006. Use of a variable amplicon typing scheme reveals
considerable variation in the accessory genomes of isolates of Burkholderia

244.

diversity. J. Biosci. 27:27-33.

39:243-244.


identification of capsular polysaccharide of *Burkholderia pseudomallei* as a major virulence determinant. Infect. Immun. 69:34-44.


PAGES NOT SCANNED AT THE REQUEST OF THE UNIVERSITY

SEE ORIGINAL COPY OF THE THESIS FOR THIS MATERIAL