Assessing patterns of genetic and antigenic diversity in Calliphoridae (blowflies)

Submitted by

Laura Marie McDonagh

To the University of Exeter as a thesis for the degree of Doctor of Philosophy in Biosciences, September 2009.

This thesis is available for Library use on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

I certify that all material in this thesis which is not my own work has been identified and that no material has previously been submitted and approved for the award of a degree by this or any other University.

Signature Date
I would firstly like to thank my supervisors Drs Jamie Stevens and Chris Thornton for their guidance, continual support, and encouragement and throughout my PhD.

Thanks must also go to Dr Kate Heesom and Dr Will Mawby at the University of Bristol Proteomics Facility, for her help and advice with 2D gel electrophoresis and MS analyses, and to Pete Splatt and Gavin Wakely, at the University of Exeter Bioimaging Centre, for help with electron microscopy work. For the provision of sample material I would like to thank Dr Rene García (IAEA Consultant, Mexico), Dr Martin Hall (Natural History Museum, London), Professor Richard Wall (University of Bristol, UK), Dr James Wallman (University of Wollongong, NSW), and Dr Jeff Wells (West Virginia University, USA).

I would also like to thank everyone in the School of Biosciences for making the past few years as enjoyable as they have been. In particular, I thank Dr Lucie Evans for technical assistance, and for her blowfly work from which much of my own research has continued, and to Dr Patrick Hamilton for all his advice concerning phylogenetic methods, and for generally being a constant fountain of knowledge. I would also like to take this time to thank the following; Jemma Clancy, Okhyun Lee, Jill (Ashton) Goodwin, Marta Soffker, Viv Fowler, Alvine Mehinto, Dr Angela Pountney, Kate Ball, Toby Coe, all past and present members of Lab C4, and last but not least my fellow GTA’s, for their friendship and support over the years, and particularly over the past few months. I would also like to specially thank Dr Max Frenzel for his friendship, support, and encouragement, and more specifically for his roast dinners and cakes!

Finally, I would like to thank my family for their continual support and encouragement in everything I do.

This work was funded by the School of Biosciences, under the Graduate Teaching Assistant program.
Abstract

The blowflies (Diptera: Calliphoridae) include some of the world’s most economically significant parasites of livestock. The defining characteristic of blowflies is the need for their larval stages to feed on a proteinaceous substrate, often including the tissues of a living vertebrate host, a process known as myiasis. While the evolution of myiasis has been linked to the development of key adaptations in behaviour and physiology (Stevens et al., 2006), patterns of blowfly evolution suggest that parasitism evolved independently in different blowfly groups after periods of geographic isolation (Stevens et al., 2006).

However, understanding the origin and evolution of myiasis in Calliphoridae is restricted by a lack of agreed theories of evolutionary relationships and taxonomic classification (Stevens, 2003). Mitochondrial genes are some of the most widely used molecular markers in insect systematics, yet most studies have utilised only single genes, with few having systematically assessed which if any are best suited for studying particular insect orders. Accordingly, this thesis presents a comprehensive analysis of 62 hexapod mitochondrial genomes, including 55 from Insecta, and assesses the ability of mitochondrial genes to recover currently recognised insect orders as monophyletic groupings. The greatest amount of phylogenetic signal was recovered when all mitochondrial genes were analysed together, regardless of optimality criterion used (PhyML, RaxML, MrBayes). Of the single-gene analyses, COX1 out-performed all other genes, even performing as well as a combined-gene analysis under Bayesian inference. In view of this finding, nucleotide sequence data from COX1 (mitochondrial protein-coding), EF-1α (nuclear protein-coding gene), and 28S (nuclear rRNA) were combined to present one of the most comprehensive multi-gene phylogenetic studies of Calliphoridae to date, resolving many ambiguous relationships, and also including several taxa that have not previously been analysed in molecular phylogenetic studies.

Within Calliphoridae, Cochliomyia hominivorax (New World screwworm fly), is widely considered one of the most destructive insect parasites of livestock in the Western hemisphere. While successful eradication programmes using sterile insect technique (SIT) have been completed in North and Central America, and on some Caribbean islands, in some areas SIT has failed. It has been hypothesized that failure of SIT may be related to genetic differentiation between populations of C. hominivorax. Consequently, intra-specific variation using nucleotide sequence data from both mitochondrial (COX1 and 12S) and nuclear (EF-1α) markers, was explored. Phylogenetic analysis of these data confirmed some population sub-structuring and suggested a South American origin to all Caribbean island populations, with the exception of Cuba. In agreement with previous studies, Cuban populations appeared distinct from all other Caribbean populations; however, our findings do not support a North American origin for Cuba, as has previously been suggested.

Finally, this thesis attempted to explore the relationship between antigenic proteins expressed in larvae from species displaying different forms of parasitism, and in doing so assessed the utility of such target proteins as potential candidates for species-specific vaccines and diagnostic tools. However, while this work discovered distinct antigenic profiles for different blowfly species, the ability to characterize specific antigens was fundamentally limited by an apparent lack of homologous proteins in current databases.
Chapter One: Introduction

1.1 Parasitism in Diptera (true flies) ................................................................. 15
1.2 The evolution of myiasis in Diptera .............................................................. 15
  1.2.1 Calliphoridae (blowflies) ........................................................................... 16
  1.2.2 Sarcophagidae (flesh flies) ....................................................................... 18
  1.2.3 Oestridae (bot and warble flies) ................................................................. 19
1.3 Myiasis in wild and domesticated animal populations .................................. 20
1.4 The economic significance of myiasis ......................................................... 23
  1.5 The medical, veterinary and economic importance of Calliphoridae .......... 23
    1.5.1 Ecological significance ............................................................................. 23
    1.5.2 Forensic entomology ............................................................................... 24
    1.5.3 ‘Maggot therapy’ ..................................................................................... 24
    1.5.4 ‘Sheep strike’ ......................................................................................... 25
    1.5.5 New and Old World screwworm flies .................................................... 28
1.6 Methods of controlling blowfly myiasis ......................................................... 30
  1.6.1 Insecticides ................................................................................................. 30
  1.6.2 Mulesing ..................................................................................................... 31
  1.6.3 Genetic approaches to improving sheep strike resistance ....................... 33
  1.6.4 Grazing management practices .................................................................. 34
  1.6.5 Sterile Insect Technique ............................................................................ 35
  1.6.6 Genetic sexing systems ............................................................................. 37
  1.6.7 Vaccine development ................................................................................ 39
  1.6.8 Baited traps ................................................................................................. 42
1.7 Diagnostic techniques .................................................................................... 43
  1.7.1 Morphology ............................................................................................... 43
  1.7.2 DNA-based analyses .................................................................................. 43
  1.7.3 Rapid diagnostic tests ................................................................................ 45
1.8 Methods of phylogenetic inference ............................................................... 46
## 1.8 Substitution models
- Substitution models .................................................. 46
- Distance methods .......................................................... 48
- Maximum parsimony ...................................................... 48
- Maximum likelihood ...................................................... 49
- Bayesian inference .......................................................... 49

### 1.9 Choice of molecular marker
- Mitochondrial DNA ......................................................... 50
- Nuclear DNA ................................................................. 51
- Nuclear ribosomal DNA .................................................. 52

### 1.10 Present study
- Aims and objectives ....................................................... 53

## Chapter Two: Materials and Methods 55

### 2.1 Molecular methods
- Sample collection and storage for DNA analysis .................. 56
- DNA extraction ............................................................... 56
- PCR amplification .......................................................... 58
- Gel electrophoresis ........................................................ 62
- Automated sequencing ................................................... 62
- Polyclonal antibody (PAb) production ................................ 62
- SDS PAGE gel electrophoresis ........................................ 63
- Western blotting ............................................................ 63
- N-terminal sequencing .................................................... 64
- Two-dimensional gel electrophoresis ................................ 65

### 2.2 Numerical methods
- Multiple sequence alignment .......................................... 66
- Nucleotide substitution model selection ............................ 66
- PAUP* .......................................................... 66
- MrBayes ................................................................. 67
- PhyML ................................................................. 68
- RaxML ................................................................. 68
- Sliding windows analysis ................................................ 69

### 2.3 Analytical methods
- Mass spectrometry ......................................................... 69
- Peptide de novo sequencing from MS/MS spectrum ............ 69
- Electron microscopy gold labelling .................................. 70
Chapter Three: Which mitochondrial gene (if any) is best for reconstructing insect phylogenetic relationships?  

3.1 Introduction ........................................................................................................................................... 74

3.2 Materials and Methods .......................................................................................................................... 77

3.2.1 Phylogenetic reconstruction using selected mtDNA genomes ......................................................... 77

3.2.1.1 mtDNA genome selection ............................................................................................................ 77

3.2.1.2 Multiple sequence alignment and substitution model selection ................................................. 77

3.2.1.3 Phylogenetic analysis .................................................................................................................. 78

3.2.1.4 Quantifying the phylogenetic utility of each mtDNA gene ......................................................... 80

3.2.2 Quantifying nucleotide diversity using all available mtDNA genomes ........................................... 80

3.2.2.1 mtDNA genome sequences and alignment .................................................................................. 80

3.2.2.2 Calculating nucleotide diversity by sliding window analysis ..................................................... 81

3.3 Results .................................................................................................................................................. 83

3.3.1 Phylogenetic Reconstruction ............................................................................................................. 83

3.3.1.1 Nucleotide versus amino acid sequence data ............................................................................. 83

3.3.1.2 Measuring clade support .............................................................................................................. 84

3.3.1.3 Single-gene analyses .................................................................................................................... 87

3.3.1.4 Concatenated-gene analyses ....................................................................................................... 90

3.3.1.5 Comparing phylogenetic methods .............................................................................................. 90

3.3.2 Nucleotide diversity .......................................................................................................................... 91

3.3.2.1 Single-gene levels of nucleotide diversity .................................................................................... 91

3.3.2.2 Levels of nucleotide diversity across the mtDNA genome ....................................................... 95

3.3.3 The importance of gene length ......................................................................................................... 96

3.4 Discussion ......................................................................................................................................... 98

3.4.1 Should mtDNA markers be used to resolve deep level relationships? ........................................... 98

3.4.1.1 Detecting phylogenetic ‘signal’ .................................................................................................... 98

3.4.2 Which mitochondrial gene (if any) is best? ....................................................................................... 100

3.4.2.1 Detecting ‘noise’ in the data ....................................................................................................... 101

Chapter Four: Molecular systematics of blowflies (Diptera: Calliphoridae):  
A multi-gene approach. ......................................................................................................................... 103

4.1 Introduction ....................................................................................................................................... 104

4.1.1 Calliphoridae (Diptera: Oestroidea) ............................................................................................... 104

4.1.1.1 Auchmeromyiinae ...................................................................................................................... 105

4.1.1.2 Bengaliinae ............................................................................................................................... 106

4.1.1.3 Calliphorinae ............................................................................................................................ 106

4.1.1.4 Chrysomyinae .......................................................................................................................... 108

4.1.1.5 Luciliinae .................................................................................................................................... 108
4.1.2 The evolution of myiasis in Calliphoridae .................................................. 113
4.1.3 Choice of molecular marker ........................................................................ 114
4.1.4 Aims ................................................................................................................. 116

4.2 Methods and Materials .................................................................................. 120
4.2.1 Sample collection ............................................................................................. 120
  4.2.1.1 Lucilia bufonivora specimens ................................................................. 120
  4.2.1.2 Lucilia cuprina hybrid specimen ............................................................ 120
4.2.2 DNA extraction ............................................................................................... 121
4.2.3 PCR amplification ........................................................................................... 121
  4.2.3.1 Cytochrome oxidase 1 .............................................................................. 121
  4.2.3.2 Elongation factor-1α ............................................................................... 121
  4.2.3.3 28S ............................................................................................................. 121
4.2.4 PCR product purification and sequencing ...................................................... 122
4.2.5 Sequence proofreading ................................................................................... 122
4.2.6 Multiple sequence alignment .......................................................................... 122
4.2.7 Nucleotide substitution model selection ....................................................... 122
4.2.8 Bayesian phylogenetic inference .................................................................... 122
4.2.9 Partition homogeneity test ............................................................................. 123

4.3 Results ............................................................................................................... 127
4.3.1 Substitution models ......................................................................................... 127
4.3.2 Congruence between gene trees ..................................................................... 127
4.3.3 Auchmeromyiinae ........................................................................................... 127
4.3.4 Bengaliiinae .................................................................................................... 128
4.3.5 Calliphoridae .................................................................................................... 128
4.3.6 Chrysomyinae .................................................................................................. 129
4.3.7 Lucilinae ......................................................................................................... 129
4.3.8 Larval feeding habits within Calliphoridae .................................................... 131

4.4 Discussion ......................................................................................................... 136
4.4.1 Taxonomic review of Calliphoridae ............................................................... 136
4.4.2 Hybridisation and introgression ..................................................................... 137
4.4.3 Evolution of parasitism within Calliphoridae ................................................. 139
4.4.4 Future work ..................................................................................................... 141
Chapter Five: Phylogenetic analysis of the New World screwworm fly, Cochliomyia hominivora

5.1 Introduction ............................................................................................................. 145
5.2 Materials and methods .......................................................................................... 150
  5.2.1 Samples and DNA extraction ........................................................................... 150
  5.2.1.1 ‘Historical’ North American samples and DNA extraction ....................... 154
  5.2.2 DNA sequences ............................................................................................... 154
  5.2.3 Sequence alignment and phylogenetic analysis .............................................. 155
5.3 Results ..................................................................................................................... 158
  5.3.1 Phylogenetic analysis: EF-1α ........................................................................... 158
  5.3.2 Phylogenetic analysis: 12S .............................................................................. 158
  5.3.3 Phylogenetic analysis: COX1 .......................................................................... 159
  5.3.4 Combined dataset ............................................................................................. 159
5.4 Discussion ............................................................................................................... 163
  5.4.1 Regional genetic variability of NWS fly populations .................................... 163
  5.4.2 Mitochondrial vs. nuclear genealogy .............................................................. 164
  5.4.3 General findings of the FAO/IAEA research project ..................................... 165

Chapter Six: An investigation into the diversity of antigenic larval proteins, concentrating on four key blowfly species, Calliphora vicina, Lucilia caesar, Lucilia cuprina and Lucilia sericata.

6.1 Introduction ............................................................................................................. 171
  6.1.1 The evolution of myiasis within Calliphoridae .............................................. 171
  6.1.2 The cycle of myiasis in sheep ....................................................................... 172
  6.1.3 Immunomodulation of host responses ......................................................... 173
  6.1.4 Levels of resistance to sheep strike ............................................................... 174
  6.1.5 Methods for control of myiasis .................................................................... 175
  6.1.6 Immunisation trials and vaccine development ............................................. 175
  6.1.7 DNA-based species identification of blowfly species ................................. 177
  6.1.8 Aims ................................................................................................................ 181
6.2 Methods and Materials ........................................................................................... 182
  6.2.1 Taxonomic sampling ...................................................................................... 182
  6.2.2 Sample collection ........................................................................................... 182
  6.2.3 Polyclonal antibody (PAb) production .......................................................... 182
  6.2.4 Gel electrophoresis and Western blotting ..................................................... 183
  6.2.5 N-terminal sequencing .................................................................................. 183
  6.2.6 Two-dimensional gel electrophoresis ........................................................... 183
  6.2.7 Mass spectrometry (MS) .............................................................................. 183
6.2.7.1 Peptide de novo sequencing from MS/MS spectrum ..................... 184
6.2.8 Electron microscopy (EM) gold labelling .................................... 184
6.3 Results ......................................................................................... 185
  6.3.1 Sample collection ..................................................................... 185
  6.3.2 Rabbit polyclonal antiserum .................................................. 185
  6.3.3 Gel electrophoresis and Western blotting ................................. 186
    6.3.3.1 Biogeographic diversity .................................................. 189
  6.3.4 Two dimensional gel electrophoresis ....................................... 190
  6.3.5 Mass Spectrometry .................................................................. 192
    6.3.5.1 MALDI-TOF/TOF .......................................................... 192
    6.3.5.2 De novo peptide sequencing ........................................... 192
  6.3.6 Localization of antibody binding by EM gold labelling ............... 198
6.4 Discussion .................................................................................... 204
  6.4.1 Future work ............................................................................ 208

Chapter Seven: General discussion .................................................. 209
  7.1 Introduction ............................................................................... 210
  7.2 Phylogenetic signal within the mitochondrial genome ..................... 210
  7.3 Molecular systematics of Calliphoridae ....................................... 213
  7.4 Intra-specific diversity of Cochliomyia hominivorax ......................... 215
  7.5 Antigenic diversity within Calliphoridae ....................................... 217
  7.6 Future work ............................................................................... 220

References ....................................................................................... 222
Appendix A ......................................................................................... 267
List of Figures

Chapter One
Figure 1: Schematic illustration of interactions between blowfly larvae and the host immune system, representing both innate (left) and adaptive (right) immune response pathways induced in the host by larval antigens on the wound surface (Otranto, 2001).

Chapter Two
Figure 2: Schematic representation of the genes used in this study, indicating position of primers used for PCR amplification of the gene sequence.

Chapter Three
Figure 3: Phylogenetic relationships within Insecta (in group), and representatives of Myriapoda, Collembolla, and Crustacea (outgroups), based on a partitioned Bayesian analysis of the concatenated mtDNA nucleotide data set, using 14 partitions (12 protein-coding genes, and two rRNA genes).
Figure 4: Number of taxonomic groups recovered as monophyletic (with clade support of BP > 75% or PP > 0.95) for all single-gene and concatenated phylogenetic reconstructions using PhyML, RaxML and MrBayes.
Figure 5: Sliding window analyses of the mtDNA genome alignments for Hemiptera, Lepidoptera, Diptera, and Insecta. Nucleotide diversity (π) values calculated for windows of 300bp, and with a step size of 25. Gene boundaries indicated.
Figure 6: Average nucleotide diversity for the mitochondrial genomes or each taxonomic group. Error bars depict standard deviation in diversity values. Number of mitochondrial genomes included in each group indicated.
Figure 7: Average nucleotide diversity for each mtDNA gene and concatenated-gene alignments. Error bars depict standard deviation in diversity values.
Figure 8: Correlation between gene length and the number of taxonomic groups recovered as monophyletic (with clade support of BP > 50% or PP > 0.95), for PhyML, RaxML, and MrBayes.

Chapter Four
Figure 9: Toad with nasal myiasis in Suffolk, UK. Larvae were collected and preserved in alcohol prior to DNA extraction.
Figure 10: Phylogenetic relationships within Calliphoridae (in group), and representatives of Muscoidea (outgroup), based on a Bayesian analysis of the EF-1α nucleotide data set. All branches supported with a PP ≥ 0.95 are labelled.

Figure 11: Phylogenetic relationships within Calliphoridae (in group), and representatives Muscoidea (outgroup), based on a Bayesian analysis of the 28S nucleotide data set. All branches supported with a PP ≥ 0.95 are labelled.

Figure 12: Phylogenetic relationships within Calliphoridae (in group), and representatives Muscoidea (outgroup), based on a Bayesian analysis of the COX1 nucleotide data set. All branches supported with a PP ≥ 0.95 are labelled.

Figure 13: Phylogenetic relationships within Calliphoridae (in group), and representatives of Muscoidea (outgroup), based on a partitioned Bayesian analysis of the combined gene (EF-1α, 28S, COX1) data set. All branches supported with a PP ≥ 0.95 are labelled.

Chapter Five

Figure 14: Sampling locations within South and Central America.

Figure 15: Sampling locations within Cuba and Dominican Republic.

Figure 16: A rooted phylogram constructed by maximum likelihood analysis of COX1 nucleotide sequence data from a variety of populations of C. hominivorax from around the Caribbean, Central and South America.

Figure 17: An unrooted phylogram constructed by maximum likelihood analysis of the combined three-gene (COX1, EF-1α, 12S) data set sampled from a variety of populations of C. hominivorax from around the Caribbean, Central and South America.

Chapter Six

Figure 18: Polyclonal responses to L. sericata using anti-sera from each of the three rabbits (R1, R2, R3) immunised with PBS protein extracts from whole larvae L. sericata (UK population).

Figure 19: Western blot analysis with different protein concentrations of first instar larval preparations of L. sericata (UK population).

Figure 20: Western blot analysis of first instar larval preparations, with a broad range pre-stained SDS-PAGE standard (10 - 250kD), displaying protein band
(A) selected for N-terminal sequencing, (C. vicina, UK population; L. sericata, UK population; L. cuprina, Australian population; L. caesar, UK population).

Figure 21: Western blot analysis of third instar larval preparations, with a broad range pre-stained SDS-PAGE standard (10 - 250kD). C. vicina - UK population; L. sericata - UK population; L. cuprina - Australian population; L. caesar - UK population.

Figure 22: Western blot analysis of first, second and third instar larval preparations for L. sericata (UK population), with a broad range pre-stained SDS-PAGE standard (10 - 250kD).

Figure 23: Western blot analysis of first, second and third instar larval preparations for C. vicina (UK population), with a broad range pre-stained SDS-PAGE standard (10 - 250kD).

Figure 24: Western blot analysis of first instar larval of C. vicina (UK, Australian, and USA populations), with a broad range pre-stained SDS-PAGE standard (10 - 250kD).

Figure 25: Western blot analysis of first instar larval of L. sericata (UK, Australian, and USA populations), with a broad range pre-stained SDS-PAGE standard (10 - 250kD).

Figure 26: Two dimensional analysis, comprising pH gradient from 3 to 11, of first instar larval preparations from UK populations of C. vicina (A) and L. sericata (B); (I – Western immunoblot membranes processed using antiserum; II – replica gel stained with Sypro Ruby).

Figure 27: Gold labelled TEM images of cross-sections taken from first-instar C. vicina. Cross-section is taken across the larval cuticle, showing envelope, epicuticle and procuticle layers.

Figure 28: Gold labelled TEM images of larval cuticle from first-instar L. sericata larvae. The cross-sections are taken across the larval cuticle, showing envelope, epicuticle and procuticle layers.

Figure 29: Gold labelled TEM images of first instar C. vicina (A) and L. sericata (B), illustrating muscle layers found between the cuticle layers and gut cavity.

Figure 30: Gold labelled TEM images from the gut of first-instar C. vicina (A) and L. sericata (B), illustrating gut cavity and villi projections into the lumen.

Figure 31: Gold labelled TEM images of control experiments from first-instar C. vicina (A) and L. sericata (B) larvae.
List of Tables

Chapter Two
Table 1: Details of PCR primers used. 59

Chapter Three
Table 2: Details of best-fit substitution models selected by MODELGENERATOR for each gene data set, including invariable sites (+I) and rate variation among sites (+Γ). 78
Table 3: Number of described species in each taxonomic group, and total number of mitochondrial genomes available at the time of analysis. 81
Table 4: Bootstrap partition (BP) values for PhyML and RaxML phylogenetic analysis of both single gene data sets and the concatenated data set. 85
Table 5: Posterior probability (PP) values for MrBayes phylogenetic analysis of both single gene data sets and the concatenated data set. 86

Chapter Four
Table 6: Details of larval feeding behaviour and host. 118
Table 7: List of taxa included in study, location details and accession numbers. 124

Chapter Five
Table 8: Cochliomyia hominivorax specimen details. 151
Table 9: Details of primers used to amplify EF-1α and COX1. 155
Table 10: De novo sequence predictions for protein spot 1. 193
Table 11: De novo sequence predictions for protein spot 2. 194
Table 12: De novo sequence predictions for protein spot 3. 194
Table 13: BLASTP2 matches (Database – ndb95). 196
Table 14: MS-BLAST matches (Database – UniRef100). 197
Chapter One

Introduction
1 Introduction

1.1 Parasitism in Diptera (true flies)

Arthropods and terrestrial vertebrates have coexisted for over 200 million years. During this time various forms of parasitism have evolved independently within Arthropoda (Balashov, 2006). One feature that has been linked to the origin of parasitism within insects is an increased rate of mitochondrial molecular evolution (Hafner et al., 1994; Dowton and Austin, 1995; Page et al., 1998). Suggested causes for this relationship have included an increase in speciation rate (Page et al., 1998), an increase in rate of mutation (Dowton and Austin, 1995), and increased rates of gene rearrangement (Dowton, 1999; Le et al., 2000; Shao et al., 2001). Four general parasitic developmental strategies are currently recognised, including ectoparasitism (where larvae develop externally), endoparasitism (where larvae feed within the body of the host), koinobionts (parasites that allow the host to continue to develop after oviposition) and idiobionts (where the host is paralyzed or killed before oviposition) (Askew and Shaw, 1986; Gauld and Bolton, 1988; Godfray, 1993).

Within Arthropoda, parasitism has originated most frequently in Diptera, where it has evolved independently over 100 times in 21 families, and accounts for an estimated 20% of all species exhibiting parasitoid behaviour (Eggleton and Belshaw, 1992). Parasitic diptera also use a broader range of hosts than any other arthropod parasitic group, including 22 orders across five phyla (Ferrar, 1987; Eggleton and Belshaw, 1992, 1993). As might be expected, given the variety of hosts, a wide range of host-parasite relationship are found within Diptera, including parasitoidism, myiasis, micropredation and phytophagy (Stevens and Wallman, 2006).

1.2 The evolution of myiasis in Diptera

‗Myiasis‘, a term first coined by Hope (1840), has been described as „the infestation of live human and vertebrate animals with dipterous larvae, which at
least for a period, feed on the host's dead or living tissue, liquid body substances, or ingested food" (Zumpt, 1965). Zumpt (1965) proposed two ways myiasis behaviour might originate, saprophyagy (feeding on decaying organic matter) and sanguinivory (feeding on blood). Myiasis has evolved independently multiple times within Diptera, most commonly from saprophagous ancestral states (Schaefer, 1979; Ferrar, 1987). It is likely that facultative ectoparasitic behaviour evolved from saprophagic flies being occasionally attracted to dead or decaying tissue of a wounded, or immuno-compromised, living host. Through this facultative parasitic intermediate, obligate parasitism eventually developed, by which time fly larvae had become aggressive feeders on tissues of otherwise healthy hosts (Zumpt, 1965; Erzinclioglu, 1989a). However, evolutionary relationships between myiasis-causing Diptera remain ambiguous (Stevens and Wallman, 2006).

Today, arguably the most economically significant myiasis-causing flies belong to dipteran family Oestroidea, which contains three of the most notorious fly species involved in the myiasis of livestock and humans worldwide, namely Calliphoridae (blowflies), Sarcophagidae (flesh flies) and Oestridae (warble and bot flies) (Hall, 1995).

1.2.1 Calliphoridae (blowflies)

Calliphoridae (blowflies) are estimated to have originated over 105 million years ago (MYA), with over 1000 blowfly species currently recognised, comprising around 150 genera (Rognes, 1991). Distribution is worldwide, with blowflies being found in almost all localities on all continents (Otranto and Stevens, 2002). Patterns of blowfly evolution appear to indicate that many of the parasitic blowfly groups evolved after periods of geographic isolation, undergoing local divergence and speciation (Stevens et al., 2006). Furthermore, it appears that the majority of calliphorid subfamilies had diverged by the time the supercontinent Gondwana broke up, approximately 100 MYA (Stevens, 2003). This is supported by observations that even closely related taxa can exhibit very different larval feeding strategies, with saprophagic species having obligate parasitic sister taxa (Stevens, 2003; Stevens and Wallman, 2006). Fossil
evidence from this period, however, is sparse (Michelsen, 2000; Amorim and Silva, 2002), for example, four fossil pupae recovered from the Edmonton Formation, Canada, dating back to the Upper Cretaceous (65 to 105 MYA) represent the earliest recorded calliphorid fossils (McAlpine, 1970). However, while these fossil pupae represent the first pre-Tertiary evidence of muscoid flies, subsequent examinations have raised doubts over their calliphorid origin (Erzinclioglu, 1984; Wallman and Donnellan, 2001). In addition, eight fossil pupae found within bone fragments of a giant fossil bird recovered from the La Brea tar pits, California, dating back to the Pleistocene era (12,000 to 2.5 million years ago), have been assigned to Calliphoridae (Pierce, 1945).

Morphological evidence has supported eight subfamily groupings within Calliphoridae; Calliphorinae, Chrysomyinae, Helicoboscinae, Luciliinae, Melanomyinae, Polliinae, Rhiniinae and Rhinophorinae (Rognes, 1991). To date, the most comprehensive phylogenetic analyses of Calliphoridae have been based on morphological evidence, and have, on the whole, supported a monophyletic grouping (Rognes, 1991). Accumulating molecular data also appear to support this hypothesis (Stevens, 2003). Relationships within Calliphoridae, on the other hand, have undergone many revisions since the late 1940s (Stevens, 2003). Traditionally, strong phylogenetic support has been found for separate monophyletic groupings of the subfamilies Calliphorinae, Luciliinae and Chrysomyinae (Wallman et al., 2005). Molecular-clock analyses carried out by Wallman et al. (2005), using an arthropod mtDNA clock developed by Brower (1994) calibrated using divergence times based on dated geological events, have estimated that Chrysomyinae diverged from a Luciliinae-Calliphorinae approximately 21.7 MYA, with Luciliinae and Calliphorinae diverging approximately two million years later (Wallman et al., 2005). The majority of speciation within each genera is thought to have occurred between 1 and 15 MYA, with most groups displaying parasitic behaviour having originated between 5 and 15 MYA (Wallman et al., 2005; Stevens et al., 2006).

Within Calliphoridae a range of larval feeding habits exist, including saprophagy, facultative ectoparasitism (feeding on necrotic or living tissue), and
obligate parasitism (feeding exclusively on living tissue). The multiple origins of myiasis behaviour within Calliphoridae appear to have evolved in accordance with Zumpt’s (1965) hypothesis, through either sanguinivory or saprophagy. The vast majority of myiasis-causing blowflies appear to display saprophagic origins (Zumpt, 1965; Stevens, 2003), with evidence of only two sanguinivorous origins currently known, *Auchmeromyia luteola* (Congo floor maggot) and species belonging to the genus *Protocalliphora* (bird blowflies) (Stevens, 2003).

Parasitic blowfly species characteristically demonstrate low host specificity, a relatively rapid larval feeding stage (when compared to other myiasis-causing dipterans, such as Oestridae), largely cutaneous infestations, and high pathogenicity (Stevens *et al.*, 2006).

1.2.2 Sarcophagidae (flesh flies)

Sarcophagidae is thought to have originated in the early Cretaceous (Pape *et al.*, 2006), however fossil evidence is scarce and little work has been done to estimate dates of divergence (Stevens *et al.*, 2006). Approximately 2600 Sarcophagidae species are currently recognised (Pape *et al.*, 2006), with molecular evidence supporting monophyly of the group (Wells *et al.*, 2001). Traditionally three subfamilies have been recognized, Sarcophaginae, Miltogramminae, and Paramacronychiinae. Of these, Sarcophaginae appears to have diversified largely in the New World, while the majority of Miltogramminae diversity is concentrated in the Old World. In comparison, Paramacronychiinae, the smallest of the subfamilies, appears largely restricted to Neoarctic and Palaearctic regions (Pape *et al.*, 2006).

Sarcophagidae are typically ovolarviparous (eggs hatching and larvae developing within the female), and exhibit a wide range of lifestyles, including saprophagy, inhabiting plants, coprophagy (feeding on faeces), inquilinism (living commensally in the nest or burrow of another organism), and parasitoidism (Pape *et al.*, 2006). Like Calliphoridae, parasitic sarcophagids cause myiasis of relatively short duration and with a low host specificity, and high pathogenicity (Otranto and Stevens, 2002; Stevens *et al.*, 2006).
1.2.3 Oestridae (bot and warble flies)

Like Sarcophagidae, a lack of fossil evidence has made dating the origins of Oestridae difficult (Stevens et al., 2006). However, recent work by Pape (2006) proposes that Oestridae evolved from parasitism of rodents or lagomorphs. From this Pape (2006) suggested that oestrids have always been, as they are today, restricted to a limited number of host species, and that oestrid diversity can be linked to the diversification of its host species during the Cretaceous-Tertiary boundary, approximately 90 to 100 MYA (Pape, 2006).

Four subfamilies are currently recognised within Oestridae: Cuterebrinae, Gasterophilinae, Hypodermatinae, and Oestrinae (Wood, 1987; Pape, 1992). Between 18 (Zumpt, 1965), 25 (Pape, 2001) and 28 (Otranto et al., 2000) genera have been identified, comprising 151 species (Zumpt, 1965; Hall and Wall, 1995). Monophyly of Oestridae is generally supported by morphological and life history data, for example an exhaustive cladistic analysis carried out by Pape (2001), examining 118 morphological, ontological, physiological and behavioural characters, supported the monophyletic status of Oestridae within Oestroidea. Several molecular studies have also supported this monophyletic grouping of Oestridae. For example, Otranto et al. (2003) analysed nucleotide and amino acid mtDNA COX1 sequence data from 18 species of Oestridae, representing each of the four Oestridae subfamilies, providing the first molecular data set for myiasis-causing Oestridae species. In agreement with classical morphology based taxonomy, Otranto et al. (2003) found strong divergence among the four subfamilies, and supported monophyly of Oestridae. Similarly, Stevens (2003) carried out analyses using both likelihood and parsimony methods to reconstruct phylogenies from nuclear 28S rRNA and mitochondrial COX1 and COX2 sequence data. Stevens (2003) found the two markers to be largely congruent, consistently grouping oestrid taxa together with high clade support. However, earlier work by Nirmala et al. (2001), analysing partial 16S and nearly complete 18S rRNA gene sequence data by parsimony and likelihood methods, failed to recover the monophyly of Oestridae. Nirmala et al. (2001) did however successfully recover the subfamilies
Hypodermatidae, Oestrinae, and Gasterophilidae with high clade support, but with the position of Cuterebridae remaining unresolved.

Oestridae are typically obligate endoparasites of mammals, with high host specificity, and relatively low levels of pathogenicity (Stevens et al., 2006). In contrast to the largely cutaneous calliphorid myiasis, species of Oestridae can cause myiasis in nasopharyngeal tracts (e.g. *Oestrus* spp.), digestive tracts (e.g. *Gasterophilis* spp.), internal organs (e.g. *Hypoderma* spp.), and subcutaneous tissue (e.g. *Przhevalskiana* spp.) (Otranto et al., 2003). Furthermore, while calliphorid and sarcophagid myiasis typically lasts between 4-7 days (Stevens et al., 2006), oestrid infestations can last for several weeks or even months within a living host.

Within Oestridae, larval morphology can be particularly diverse. For example, larvae of the subfamilies Oestrinae and Gasterophilinae typically display large mouth hooks and well developed dorsal and or ventral spines, likely related to the nasopharyngeal region and the gastrointestinal location of their myiasis (Stevens et al., 2006). Oestridae are also characterized by the diversity of enzymes expressed by larvae to combat host immune responses (Otranto, 2001). These adaptations of larval morphology and assortment of specialised enzymes appear to be related to the larval migration through host tissue that is typical to oestrid parasitism. Consequently, these features are not found in the ectoparasitic Calliphoridae and Sarcophagidae, whose evolution does not appear to have been so closely related to that of their hosts (Stevens et al., 2006).

### 1.3 Myiasis in wild and domesticated animal populations

Cases of myiasis have been reported in a wide range of wild animals, including: wildebeest (Horak, 2005); rhebok antelope (Horak and Boomker, 1998), mice (Wilson et al., 1997), caribou (Hughes et al., 2009) and howler monkeys (Baron et al., 1996; Milton, 1996). To date little research has been carried out to investigate the impact myiasis has on these wild animal populations, with many pursuing the belief that host-parasite relationships evolve towards a benign...
state (Holmes, 1983; Gulland, 1995). However, an increasing amount of empirical evidence is suggesting that myiasis-causing flies can significantly affect the condition, fecundity, grazing behaviour, and survival of wild animal populations (Gunn and Irvine, 2003; Irvine, 2006; Hughes et al., 2009). For example, *Hypoderma tarandi* (warble fly) and *Cephenemyia trompe* (bot fly) infestations have been directly linked to changes in foraging and herding behaviour of reindeer and caribou (Folstad et al., 1991; Hagemoen and Reimers, 2002; Fauchald et al., 2007), with high larvae burdens significantly reducing fecundity of individuals (Hughes et al., 2009). Similarly, mortality rates of populations of howler monkey (*Alouatta palliate*) on Colorado Island, Panama, have been significantly correlated to parasitism by the bot fly *Alouattamyia baeri* (Milton, 1996).

While most cases of myiasis reported in wild animals have caused by members of Oestridae and Sarcophagidae, parasitism by Calliphoridae (blowflies) appear largely unreported in wild animal populations, although infestations of bird and amphibian populations are well documented. A recent study on blowfly parasitism of tree swallows (*Tachycineta bicolour*), comparing natural wetlands and wetlands reclaimed for ‘oil sands’ mining, found infestations by *Protocalliphora* species (Diptera: Calliphoridae) in up to 72% of nests at natural wetland sites, and in 100% of nests at reclaimed wetland sites (Gentes et al., 2007). The observations of this study also reported parasitic burdens being twice as high at reclaimed wetland sites, concluding that the process of oil sands mining significantly disrupts local habitat characteristics, increasing prevalence of blowfly parasitism significantly (Gentes et al., 2007). Similarly, parasitism of toads and frogs by the blowflies *Lucilia bufonivora* and *Lucilia silvarum*, resulting in significant mortality, have been reported across northern Europe and North America (Brumpt, 1934; James and Maslin, 1947; Hall, 1948; Anderson and Bennett, 1963; Bleakney, 1963; Briggs, 1975; Strijbosch, 1980; Roberts, 1998; Bolek and Coggins, 2002; Bolek and Janovy, 2004; Eaton et al., 2008).

An increasing amount of evidence is suggesting that the domestication of agricultural animals, by selection for improved produce, yield and quality, has
resulted in a significant increase in susceptibility to dipteran myiasis. For example, the woolly fleece found on modern domestic sheep are absent from more ancestral breeds, which have a stiff and hairy coat, and thin woolly undercoat. This selection for longer, thicker fleeces with thinner wool fibre diameter has drastically decreased resistance to ectoparasites and bacterial dermatitis (Stevens et al., 2006). Indeed, by far the greatest numbers of reported cases of parasitism by Calliphoridae are found within domesticated livestock. For example, up to 80% of all British sheep farms documented to be affected by myiasis caused by the blowfly Lucilia sericata, with a mortality rate of around 2% (French et al., 1992; French et al., 1995). Although mortality rates of up to 20-30% have been reported in other areas of Europe (Liebisch et al., 1983; Mashkei, 1990). The increased prevalence of blowfly myiasis in domesticated animals is also likely associated with the occurrence of high stocking densities, synchronized breeding seasons, and open wounds from shearing, branding, etc (Mullen and Durden, 2002). For example, myiasis caused by the Old and New World screwworm flies (Calliphoridae) has been directly linked to animal husbandry practises, whereby intensive farming systems, where livestock are not closely supervised, have prevented early diagnosis of infested animals (IAEA, 1998).

As may be expected, the vast majority of research concerning myiasis focuses on economically important host species (i.e. livestock), and so the extent to which other domesticated species are affected remains unclear. Although, a recent study found myiasis of domesticated rabbits, usually by the blowfly Lucilia sericata, to be a widespread problem within the UK, with over 94% of veterinary practices surveyed having treated blowfly strike in rabbits during 2005 (Bisdorff and Wall, 2006).

Finally, while the majority of reported cases of myiasis appear to occur in animals, myiasis-causing flies are also of medical importance in both developed and developing nations. For example, while wohlfahrtiosis (myiasis caused by larvae of the sarcophagid genus Wohlfahrtia) is predominantly found in sheep (Farkas, 1996), two species common to south-eastern Europe, Russia, the Middle East and North Africa, Wohlfahrtia magnifica and Wohlfahrtia vigil, are
also known to be obligate parasites which can sometimes infest humans (Sherman et al., 2000).

1.4 The economic significance of myiasis

The economic cost of myiasis for the Australian wool industry alone, the largest in the world, has been estimated at over AUS$ 250 million a year (http://www.wool.com.au, 2002), while estimated costs of $40 million have been estimated for the New Zealand sheep industry (Heath and Bishop, 1995).

Similarly, New World screwworm, Cochliomyia hominivorax (Calliphoridae), infestations are estimated to be responsible for hundreds of millions of dollars in economic loss in South American livestock industries due to livestock loss, decreased fertility and reduced milk, meat and wool yields, as well as the cost of treatment of infested livestock and insecticidal prevention methods, (Klassen and Curtis, 2005; Vargas-Teran et al., 2005). Costs of monitoring and treatment alone have been estimated at between $4.82 to $10.71 per animal (IAEA, 1998). Total annual losses due to C. hominivorax for South America and the Caribbean have been estimated at over $3600 million and $135 million, respectively (IAEA, 1998; Vargas-Teran et al., 2005).

1.5 The medical, veterinary and economic importance of Calliphoridae

1.5.1 Ecological significance

Saprophagic (detritus feeding) blowfly species are of great ecological significance within an ecosystem, performing an essential role in the decomposition of animal remains (Putman, 1983; Donovan et al., 2006). Decomposers are often regarded of the most important groups within an ecosystem, ensuring the repeated recycling of biomass (Putman, 1983).
1.5.2 Forensic entomology

Modern forensic science has seen the importance of entomological evidence becoming increasingly established worldwide. Within forensic entomology, blowflies are recognised as some of the most important and robust indicator species, commonly being among the first insects to colonise a body after death, often within hours (Smith, 1986). Immature stages of blowflies, particularly larvae, can be collected at a crime scene of suspected homicide, for example, and used to establish the minimum post-mortem interval (PMI), using the age of the insect as a form of ‘biological clock’. Additionally, toxological examination of larvae can be used to identify drugs and toxins that may present in the body at the time of death (Amendt et al., 2004).

1.5.3 ‘Maggot therapy’

Reports of the use of blowfly larvae for the treatment of non-healing wounds has been documented in Aboriginal tribes (Dunbar, 1944) and the Mayans of Central America (Weil et al., 1933). However, it was the American Civil War (1861–1865) and First World War (1914-1918) that saw the first widespread observations by surgeons of the healing effects blowfly larvae had on wounds (Sherman et al., 2000). William Baer (1872-1931), a professor of orthopaedic surgery, is largely accredited as the founder of modern ‘maggot therapy’ (Sherman et al., 2000). Baer went on to use his experiences on the battle fields of the First World War to use maggot therapy for the treatment of intractable bone infections in patients at the Baltimore Children’s Hospital (Baer, 1929). Baer (1931) was also the first to recognise the need to use sterilised larvae to prevent patients developing tetanus, etc. However, the development of mass produced antibiotics in the 1940s ultimately lead to a decline in the use of maggot therapy (Sherman et al., 2000).

More recently, the application of sterilized blowfly larvae for wound cleaning and healing is increasingly becoming recognised as an efficient mainstream medical practice, particularly in light of increasing antibiotic resistance of many bacterial strains (Bunkis et al., 1985; Church, 1996;
Sherman et al., 2000; Jones and Wall, 2008). The larvae are most commonly being used for the treatment of necrotic wounds (e.g. pressure sores, infected surgical wounds, leg ulcers, etc); encouraging wound debridement, disinfection by larval secretion (e.g. ammonia), and wound healing (Sherman et al., 2000). Larval excretory-secretory (ES) products have even been credited with the inhibiting numerous pro-inflammatory responses of activated neutrophils, which in chronic wounds would otherwise facilitate the damage of tissue (van der Plas et al., 2007).

Interestingly, the most popular blowfly species used in human maggot therapy is *Lucilia sericata* (Sherman et al., 2000), a species considered a serious pest to sheep, where infestations can cause considerable animal welfare and economic issues. However, immune responses corresponding to those reported in sheep have not been documented in humans. While this is likely due to the controlled nature of the clinically induced myiasis used in maggot therapy preventing increases in blood ammonia levels, observations have also been made that larvae used in the second and third weeks of a therapy practice are less likely to survive (Sherman et al., 2000).

In contrast, the use of maggot therapy in veterinary medicine remains rare, which in turn makes it difficult to assess success rates (Jones and Wall, 2008). However, individual successful cases have been reported on several domesticated species, including; buffalo (Iversen, 1996), a donkey (Bell, 2001), a Guernsey bull (Dicke, 1953), horses (Morrison, 2006; Sherman et al., 2007), rabbits (Kocisova et al., 2003), and even sheep (Kocisova et al., 2006).

### 1.5.4 ‘Sheep strike’

The two main causal agents of blowfly ‘sheep strike’ in cool temperate habitats (Europe and New Zealand) and sub-tropical/warm temperate habitats (Australia and South Africa), are *Lucilia sericata* and *Lucilia cuprina*, respectively (Stevens and Wall, 1997a). Today, the risk of sheep strike is more severe in Australia than any other country in the world (Phillips, 2009). This combined with the fact that the Australian wool industry is the largest in the world, have lead to *L.
*cuprina* being one of the most well studied blowfly species. Of particular focus has been investigation into host immune responses, with a view to vaccine development (Bowles *et al.*, 1987; Barrett and Trevella, 1989; Tellam *et al.*, 1994; Casu *et al.*, 1997; Tellam and Eisemann, 1998; Tellam *et al.*, 2000; Tabouret *et al.*, 2001; Tellam *et al.*, 2001; Colditz *et al.*, 2002; Tabouret *et al.*, 2003).

If left untreated, sheep strike will lead to the eventual death of the animal within as little as five days (Hall *et al.*, 1980; Broadmeadow *et al.*, 1984; Sandeman *et al.*, 1987; Guerrini, 1988). The primary cause of death, in addition to skin and tissue loss, is ammonia toxicosis, a result of the larvae excreting their nitrogenous waste as ammonia, in comparison to the majority of insects which excrete uric acid (Lennox, 1941). The ammonia is absorbed into the host’s blood circulation and lymphatic ducts, and if left untreated can progress to hyperammonaemia, alkalosis of the blood, brain damage (encephalopathy), and eventual death (Guerrini, 1988). However, observations have also been made that levels of blood ammonia decrease with successive infestations, and an increasing number of larvae are needed with each consecutive implantation to obtain steady ammonia levels, suggesting some acquired system of degrading ammonia within the host (Guerrini, 1988).

**Innate and adaptive host immune responses**

During sheep strike significant innate (non-specific) and adaptive (specific) host immune responses are observed (Fig 1). Within as little as 1-2 hours of initial contact with larvae, a rapid local inflammatory response occurs (Broadmeadow *et al.*, 1984), involving activation of natural killer cells (NK), eosinophils (E), mast cells (M), neutrophils (N), gamma delta T cells (γδ+), and the alternative complement pathway, see Fig. 1 (Otranto, 2001).
Figure 1: Schematic illustration of interactions between blowfly larvae and the host immune system, representing both innate (left) and adaptive (right) immune response pathways induced in the host by larval antigens on the wound surface (Otranto, 2001). (NK = natural killer cells; E = eosinophils; M = mast cells; N = neutrophils; γδ+ = gamma delta T cells; Th = T helper cells; IFN-γ = interferon γ; Ag = larval antigens; MHCII = major histocompatibility complex class II; APC = antigen-presenting cells; IL = interleukin; CD8+ = cytotoxic T cells; TNF-α = tumor necrosis factor α).

Approximately 48 hours after initial infestation these innate responses also trigger cellular immune responses, with NK cells activating helper T-cells through the production of interferon γ (IFN-γ). The differentiation of these helper T-cells into Type 1 (Th1) and Type 2 (Th2) sub-types is then stimulated by the release of interleukin 12 (IL-12) by antigen-presenting cells (APC) (Otranto, 2001). Type 1 T-cells go on to enhance the ‘specific’ immune responses by releasing IL-2 and IFN-γ to engage cytotoxic T cells (CD8+). Type 2 T-cells, on the other hand, release IL-4, IL-5, IL-6 and IL-10, which activate B cells and the humoral immune responses (i.e. production of immunoglobulins) approximately 96-120 hours after infection (Bowles et al., 1994; Otranto, 2001). Eventually, production of immunoglobulin E (IgE) promotes further eosinophil and mast cell recruitment at the wound site, while...
APCs release the pro-inflammatory and immunoregulatory cytokines, IL-1 and tumor necrosis factor α (TNF-α) (Otranto, 2001).

**The role of fleece rot and dermatophilosis**

Fleece rot and dermatophilosis are two key predisposing factors to sheep strike (McGuirk, 1978; O'Meara et al., 1997); both are the result of microbial activity in fleece subject to prolonged wetting. Both diseases are thought to play a role in activating an inflammatory response, facilitating wound development, and aiding the initiation of primary myiasis (Sandeman, 1996; Otranto, 2001). Resistance to fleece rot, in particular, is considered one of the best known breeding indicators for resistance to ‘body strike’ (myiasis located on the body of the animal) (Raadsma, 1987).

**1.5.5 New and Old World screwworm flies**

The New World Screwworm (NWS), *Cochliomyia hominivorax*, and Old World Screwworm (OWS), *Chrysomya bezziana*, are major parasitic pests of livestock, affecting both developed and developing countries worldwide. The world organization for animal health, Office International des Epizooties, have classified both the NWS and OWS as a List B disease, which states that they are “a transmissible disease which is considered to be of socioeconomic and/or public health importance within countries and which is significant in the international trade of animals and animal products” (IAEA, 1998).

*Cochliomyia hominivorax*, in particular, is widely considered one of the most destructive insect parasites of livestock in the Western hemisphere’s tropical and sub-tropical regions (Klassen and Curtis, 2005), with a broad trans-boundary geographic range, traditionally stretching from southern USA down through all countries of Central and South America, with the exception of Chile (Torres et al., 2004). Consequently, *C. hominivorax* has also been listed as a “priority trans-boundary animal disease for the Americas” by the Food and Agriculture Organization of the United Nations (FAO) (IAEA, 1998). Being an obligate ectoparasite, eggs of *C. hominivorax* are laid onto open wounds of a
range of living vertebrate hosts, including cattle, sheep, pigs and humans, (Klassen and Curtis, 2005). Like ‘sheep strike’, *C. hominivorax* infestations initiate significant humoral immune responses, affecting host development and growth, and ultimately leading to death of the host if left untreated. However, unlike sheep strike, *C. hominivorax* also infest humans, with mortality rates of up to 40% being reported in poorer regions of Central and South America (Vargas-Teran *et al.*, 2005).

The 1950s saw the first suggestion that a long term programme of sterile insect technique (SIT) could have the potential to eradicate wild *C. hominivorax* populations (Knipling, 1955). Subsequently, SIT programmes coordinated by the US Department of Agriculture (USDA), Food and Agriculture Organization of the United Nations (FAO), and the International Atomic Energy Agency (IAEA) were launched across the southern United States, and later down through Central America. To date, *C. hominivorax* has been eradicated from North and Central America, with a permanent biological barrier established in Panama to prevent re-introduction from endemic South American populations (Colwell *et al.*, 2004; Klassen and Curtis, 2005; Vargas-Teran *et al.*, 2005).

However, despite the successful eradication of *C. hominivorax* in North and Central America, SIT programmes started in Jamaica in 1998 have continued to fail (IAEA, 1998; Vargas-Teran *et al.*, 2005). Furthermore, molecular studies have uncovered possible genetic differentiation between South, Central and North America *C. hominivorax* populations (LaChance *et al.*, 1964; Makela and Richardson, 1978; Richardson *et al.*, 1982; Azeredo-Espin and Pavan, 1983; McInnis, 1983; Roehrdanz, 1989; Infante-Malachias and Azeredo-Espin, 1995; Infante-Malachias *et al.*, 1999; Azeredo-Espin and Lessinger, 2006). The first suggestion of genetic differentiation between populations of *C. hominivorax* came with the earliest examination of the structure of chromosome material (cytogenetics) (LaChance *et al.*, 1964). Reported observations of different chromosomal polymorphisms (Makela and Richardson, 1978; Richardson *et al.*, 1982; Azeredo-Espin and Pavan, 1983) led to speculation of the existence of *C. hominivorax* “mating types” (Makela and Richardson, 1978), or even cryptic species (Richardson *et al.*, 1982).
Preferential mating behaviour has also been reported between native females and sterile males, although whether preference was for wild male over sterile male, or between distinct populations, was unclear (McInnis, 1983). Until recently, *C. hominivorax* population genetics have found little evidence of population differentiation. However, it should be noted that much of the previously published work has focused on comparing relatively small geographic ranges, for example, comparing populations within Brazil (Infante-Malachias *et al.*, 1999), Costa Rica (Taylor and Peterson, 1994), and Central America (Krafsur and Whitten, 1993). Consequently, the impact intraspecific diversity could have for the continued, and future, success of *C. hominivorax* SIT programmes, which fundamentally rely on the rate of mating between sterile males and fertile females remaining at a higher ratio than wild type males to female, remains unclear (Mahon *et al.*, 2000).

Despite occupying a similar ecological niche to the *C. hominivorax*, albeit in different geographic ranges, very little research has been completed concerning the significance of *Chrysomya bezzianna* myiasis on livestock industries and public health (IAEA, 1998; Vargas-Teran *et al.*, 2005). Similarly, while *C. hominivorax* eradication programmes have been implemented for over 50 years, SIT has yet to be validated for the control and eradication of *Chrysomya bezzianna* (IAEA, 1998).

## 1.6 Methods of controlling blowfly myiasis

### 1.6.1 Insecticides

For the past 100 years insecticides have formed a major role in insect control, driven largely by the development of synthetic insecticides after the Second World War (Dyck *et al.*, 2005). During the 1950s, the most commonly used insecticides for the control of sheep strike were dieldrin and aldrin. Subsequent insect resistance to dieldrin led to it being replaced with the organophosphate diazinon. Despite the development of diazinon resistance in the mid-1960s, some sheep farms still use this insecticide to combat sheep strike. However, concerns over the toxic nature of diazinon decomposition, and accumulation in
the food chain, led the Australian Pesticides and Veterinary Medicines Authority (APVMA) to cancel registration of this organophosphate (Australian Pesticides and Veterinary Medicines Authority, 2006), although it is still permitted for treatment of individually struck sheep (Evans, 2009). Current sheep strike targeted insecticides provide protection for between 12-24 week periods (Bell and Sackett, 2005). However, many products result in undesirable levels of chemical residue in the fleece, resulting in the need for wool withholding periods (WHP), typically between 6 weeks and 3 months depending on chemical used, prior to shearing (Evans, 2009). Currently the only chemical with no WHP needed is the spinosyn group (e.g. Extinosad), however this reportedly provides only 3-4 weeks protection in adult sheep, and as little as 10-14 days protection in lambs (Evans, 2009).

While many different insecticides are currently used to help control sheep strike, increased insecticide resistance, along with growing concern over environmental contamination and food safety, have led to a rise in the use of integrated pest management (IPM), combining several practices, and increasingly used as alternative to insecticide use altogether. IPM was first adopted in the early 1970s, with an emphasis on a move towards more selective insecticides, whereas today biological insect control is the now the main focus of modern IPM programmes (Dyck et al., 2005). Controversy over the use of organophosphates to control sheep strike (Murray et al., 1992) has also led to the potential use of a biological insecticide to be explored (Johnson et al., 1998). To date attempts at biological insecticide control of myiasis-causing flies have included the use of entomopathogenic fungi, bacteria (e.g. Bacillus thurgiensis), and phoretic mites (Hall and Wall, 1995; Johnson et al., 1998; Mazyad and Soliman, 2006; Oliveira et al., 2006).

1.6.2 Mulesing

The first domesticated sheep to arrive in Australia, in the late 19th Century, were plain breeched breeds from the Cape of Good Hope, South Africa. In 1883, Vermont Merino rams were introduced to improve wool quality and fleece weight. However, Merino sheep also possess a pair of vertical folds along each
side of the perineal region, so establishing the trait of lose wrinkled skin on the breech in present-day Australian sheep flocks (Phillips, 2009). The vast majority of sheep strike cases occur in the breech region, where the wrinkled skin causes urine and faecal staining which increase attractiveness to flies (Watts et al., 1979; Heath and Bishop, 2006). While a return to farming plain breeched sheep has been considered, the development of the surgical procedure ‘mulesing’ in the 1930s proved so effective at reducing levels of sheep strike that Merino breeds remain the most common sheep in Australia (Phillips, 2009).

Today, a combination of breeds susceptible to sheep strike, large extensively-managed flocks, and warm climate make the risk of sheep strike more severe in Australia than any other country in the world. Consequently, mulesing remains the most common method of sheep strike control in Australia, despite now being illegal in many other countries (Phillips, 2009). The procedure of mulesing involves the removal of loose wrinkled skin around the breech of sheep, usually without application of anaesthetic, with the wound then healing to produce a large area of smooth, stretched bare skin on the breech and tail (Phillips, 2009). This procedure has been proved to significantly decrease levels of sheep strike, by reducing faecal and urine staining in the breech (Watts et al., 1979; Heath and Bishop, 2006; Lee and Fisher, 2007; Phillips, 2009). Mulesing has also been found to reduce the need for insecticide dipping, increase shearing times, and allow shearing to be timed with wool quality, rather as a means of reducing sheep strike (Heath and Bishop, 2006). However, as well as significantly reducing levels of breech strike, mulesing is also preventing effective selection for resistance to sheep strike in general, by masking sheep of both desirable and undesirable genotype (Bell and Sackett, 2005).

Following action by the animal rights organization the People for the Ethical Treatment of Animals (PETA), the representative body of the Australian wool industry, Australian Wool Innovation Ltd, announced plans to phase out the practise of mulesing by 2010 (Australian Wool Innovation Limited, 2004). However, the absence of any single method as effective as mulesing at
providing protection against sheep strike remains a substantial problem to both sheep welfare and the economy of the Australian sheep farming industry (Phillips, 2009).

### 1.6.3 Genetic approaches to improving sheep strike resistance

Despite the clear relationship between wrinkled skin and prevalence of sheep strike, selection for plain breeches is thought unlikely to be as effective at preventing blowfly myiasis as mulesing (James, 2006). Additionally, experimental crosses of naturally bare breeched sheep breeds, such as the Wiltshire Horn, with Merinos results in a reduction of fleece weight and an increase in wool fibre density, both unfavourable traits in wool production (Rathie et al., 1994). Consequently, one of the main research focal points by the Cooperative Research Centre for Sheep Industry Innovation (Sheep CRC), a partnership between Australia’s leading sheep industry organisations, for the control of sheep strike has been the improvement of genetic resistance in Merino breeds.

A long term experimental breeding programme has been undertaken since the mid 1970s at the Trangie Agricultural Research Centre (TARC), Australia, selecting sheep lineages for ‘resistance’ (R) and ‘susceptibility’ (S) to fleece rot and blowfly strike. Immunization trials using the Trangie flocks have reported considerable variation in the secretion of wound exudate between R and S flocks, with R sheep producing exudate more rapidly after primary myiasis with *Lucilia cuprina* first instar larvae, and also showing increased levels of IgG, fibrinogen and complement C3 (Sanders, 1986; O’Meara et al., 1997). However, observations that titres of serum antibody did not differ between R and S flocks, and of inconsistency between individuals in level of acquired protection after repeated induced infestations, both signify an innate quality, independent of specific antibodies, to acquired fly strike resistance (MacDiarmid et al., 1995; O’Meara et al., 1997). Increased levels of protection also did not seem to last beyond periods of repeated infestation, again implying a lack of ‘immunological memory’ in the responses involved in resistance (Sandeman, 1990). While the exact mechanisms of genetic sheep strike resistance remain
unclear, differences in inflammatory responses are suspected to be involved (O'Meara et al., 1995).

More recently resistance to *Lucilia cuprina* larval growth has been identified as a moderately heritable trait in Merino sheep (Colditz et al., 2006; Smith et al., 2008). This trait has been found to be unrelated to wool characteristics, cutaneous wheal response to immunization with larval ES products, resistance to internal parasites (nematodes, etc), or the reduced blowfly strike susceptibility associated with fleece rot resistance (Colditz et al., 2006; Smith et al., 2008). This larval growth resistance is thought to be partially regulated by anti-larval factors in serum and eosinophils. Furthermore, potential quantitative trait loci (QTL) for larval growth and survival have been identified on chromosome 11 and 18, respectively, although more extensive studies are needed to confirm these findings (Smith et al., 2008). However, while a genetic approach to sheep strike control is widely seen as the only realistic long-term solution, mulesing is currently preventing effective selection by masking sheep of both desirable and undesirable genotype (Bell and Sackett, 2005).

### 1.6.4 Grazing management practises

One of the main associated factors attracting flies to sheep in instances of sheep strike is thought to be faecal staining of the breech of a sheep. Improved grazing pastures, common throughout modern sheep farming, are known to increase instances of diarrhoea in grazing sheep, with heavy stocking levels additionally preventing accumulation of fibre content in grass. Fibre supplements have been shown to reduce diarrhoea, and hence also attractiveness to flies (Davidson et al., 2006). Additionally, pastures containing plants such as birds foot trefoil (*Lotus corniculatus*) and sulla (*Hedysarum coronarium*) have been found to reduce risk of diarrhoea, thought to be related to high tannin content. However, these types of pasture are often difficult to maintain (Waghorn et al., 1999; Leathwick and Heath, 2001).
1.6.5 Sterile Insect Technique

The concept of sterile insect technique (SIT) first came about following observations by Knipling in the 1930s of the acute sexual aggressiveness in male *C. hominivorax* (New World screwworm fly), along with observations that female *C. hominivorax* only mate once during their lifetime. These observations lead Knipling to realize that wild *C. hominivorax* populations could potentially be controlled by releasing huge quantities of sterile males. Knipling also recognized that releasing sterile males over consecutive generations could potentially lead to eradication of wild populations altogether, on condition that the wild populations were isolated enough to allow sterile:fertile mating to outnumber fertile matings (Knipling, 1955; 1979, 1985). However, it wasn’t until 1950 that the potential of X-rays to induce male sterility in flies, first reported by Muller (1928), was linked to sterile insect technique. Experiments by Bushland and Hopkins (1953) found male *C. hominivorax* to be more sensitive to the effects of radiation than females, with lower doses needed to induce sterility, and that sterile males competed equally well for females as non-sterile males.

The first field trial of *C. hominivorax* SIT was carried out on Sanibel Island, 5km off the coast of Florida. While releases of 38 sterile males per square kilometre per week over several months resulted in sterility of approximately 80% of egg masses, eradication was not achieved. This was most likely due to constant reintroduction of wild type males and fertile egg bearing females from mainland USA. A second field trial carried out on the island of Curaçao, 64km off the coast of Venezuela, where sterile males mass-reared in Florida were released at a rate of 78 males per square kilometre per week had little effect on wild *C. hominivorax* population levels. However, eradication was eventually achieved after seven weeks by increasing release rate to 150 males per square kilometre per week (Lindquist, 1955). This success lead to the creation of several SIT mass rearing centres in Florida, and in 1957 a New World screwworm SIT programme was launched across southern USA (Lindquist, 1963; Klassen and Curtis, 2005). By 1966 *C. hominivorax* had been eradicated from the USA (Klassen and Curtis, 2005). In 1972 the USA and Mexico signed a cooperative agreement to extend SIT
eradication programs into Mexico, and by the 1980s SIT programs had been extended into and beyond countries boarding Mexico to the south (Klassen and Curtis, 2005). To date, SIT has eradicated *C. hominivorax* from USA and Central America, with a weekly release of sterile males across a region known as the Darien Gap in Panama, close to the border with Colombia, creating a sterile-fly barrier to prevent the re-introduction into the Central America mainland (Colwell *et al.*, 2004; Klassen and Curtis, 2005; Vargas-Teran *et al.*, 2005). Future SIT plans are now turning their attention to persisting populations, with particular concern surrounding some Caribbean Islands which potentially hold the biggest threat to re-introduction of the *C. hominivorax* to North and Central America (Colwell *et al.*, 2004).

SIT has also been used to successfully prevent *C. hominivorax* becoming established in other regions of the world. In the late 1980s reports of *C. hominivorax* in Libya (El-Azazy, 1989; Gabaj and Beesley, 1989; Gabaj *et al.*, 1989) began to raise concern by both the Libyan Government and FAO (Food and Agriculture Organization of the United Nations). In December 1990 SIT programs were launched to prevent spread into other regions of Africa and southern Europe, using a sterile male release rate of 3.5 million per week, eventually increased to 40 million per week by May 1991. By 1992 the Libyan Government had officially declared the eradication of *C. hominivorax* (Lindquist *et al.*, 1992).

Despite SIT programmes having successfully eradicated *C. hominivorax* in USA, Central America, and Libya, the threat that intra-specific genetic differentiation poses to on-going and future SIT efforts remains the subject of much focus. A symposium on the eradication of the screwworm from the United States and Mexico stated that one of the main areas of future research should be the genetic characterization of *C. hominivorax* populations, following observations that strains used in mass-production become obsolete and must be replaced approximately every 18 months due of the inherited changes induced by prolonged captivity (Graham, 1985).
To date SIT programmes have successfully been used to eradicate several insect species of medical and veterinary significance, including the mosquito species *Culex quinquefasciatus* in Myanmar (Laven, 1967) and *Anopheles albimanus* in El Salvador (Breeland *et al*., 1974; Weidhaas, 1974), and more recently the tsetse fly *Glossina austeni* on the Island of Unguja, Zanzibar (Vreysen *et al*., 2000). However, while SIT was among the first widely used biological insect control methods, its capacity for the suppression, containment, prevention and eradication of pests is still only now becoming fully recognized (Dyck *et al*., 2005).

### 1.6.6 Genetic sexing systems

Genetic sexing systems work on a similar basis to SIT, in that they involve releasing large quantities of modified males or females, in this case genetically modified, who mate with wild type individuals to reduce the overall fitness of the population. Genetic sexing systems have several advantages over SIT, such as modification of sterilization procedures to obtain sterile males with improved competitiveness for females, and the ability to create phenotypic markers to identify fertile from sterile offspring (Saul, 1990). Genetic sexing systems have been developed for several insect species, including mosquitoes (Curtis *et al*., 1976; Seawright *et al*., 1978) and flies (McDonald, 1971; Robinson and van Heemert, 1982). One of the most promising avenues of research involves using conditional lethal genes and Y-autosome translocations, so genetically linking a particular gender to a set of deleterious mutations (Saul, 1990). However, translocation based genetic modification also has its disadvantages, including difficulty in finding appropriate conditional lethal genes, breakdown of the translocation through recombination, and effects mutations have on lowering fitness of mutant individuals, both in terms of survival to sexual maturity and mating ability (Saul, 1990).

An example of such technology is the female killing (FK) system developed for the control of *L. cuprina* in Australia (Whitten *et al*., 1977; Whitten, 1979; Foster *et al*., 1985; Foster *et al*., 1988; Foster, 1989). In this FK system modified male *L. cuprina* carry wild type alleles on a Y-linked translocation, and...
alleles for a recessive eye colour mutation on a normal set of autosomes, giving them normal eye pigmentation. Males then pass the Y-linked translocation onto their male offspring only, and the mutation onto their female offspring. When heterozygous females mate with modified males half of all female offspring will be homozygous for the mutation, making them functionally blind with white eyes, greatly reducing their chance of surviving to maturity (Whitten et al., 1977; Foster et al., 1988). Field trials of this FK system have proved to reduce wild \textit{L. cuprina} populations by both the semi-sterility caused by the translocation itself, and through the eventual increase in proportion of homozygotic individuals through sustained release (Whitten et al., 1977; Whitten, 1979; Foster et al., 1985; Foster, 1989). While genetic death rates of up to 94\% have been obtained using this FK system (Foster, 1991) the competitiveness of released males has been shown decline as frequency of released males increases (Mahon, 1995). Reanalysis of published data by Mahon (1995) has suggested that this trend is not restricted to the \textit{L. cuprina} FK system, and is also present in other mass reared insect programs. In order to eradicate target populations, this apparent frequency dependent nature of male competitiveness increases both the length of release period and the number of sterile males needed. However, Mahon (1995) also acknowledged that effective monitoring of wild populations throughout the mass release can allow appropriate adjustments in release rates to maintain levels of male competitiveness. Genetic sexing systems are also susceptible to levels of male recombination, which can rapidly separate the linkage between the mutations and the Y chromosome (Foster et al., 1980; Saul, 1984; Hooper et al., 1987; Busch-Petersen, 1989), and field trials have shown that this can result in serious genetic decline of release strains (Foster et al., 1985; Hooper et al., 1987).

While FK systems are largely considered more cost effective than SIT for long term suppression programs, with lower release rates of FK males giving higher genetic death rates in density influence populations than sterile males, to date no effective large scale FK systems for the control of \textit{L. cuprina} have been developed (Foster et al., 1988; Foster, 1991). However, recent developments of a strain of transgenic \textit{L. cuprina}, using a transposon piggyBac vector and an EGFP (Enhanced Green Fluorescent Protein) marker gene, along with a two
component FK system developed in *Drosophila melanogaster*, have the potential to improve success of future *L. cuprina* genetic sexing systems (Scott *et al.*, 2004). Under this two component system, individuals carrying both components die unless fed on a diet containing tetracycline (Scott *et al.*, 2004). The development of such a tetracycline-repressible system in *L. cuprina*, which seems to successfully control female viability, does demonstrate key progress towards the eventual development of a genetically modified strain of *L. cuprina* suitable for a male-only sterile release program (Scott *et al.*, 2004). However, significant work is still needed to provide an effective FK system for the control of sheep strike (Scott *et al.*, 2004).

### 1.6.7 Vaccine development

Currently no commercially viable vaccines are available for the control of myiasis, despite considerable resources having been assigned to their development (Elkington and Mahony, 2007). One of the largest obstacles in the development of a vaccine for sheep strike has been the apparent lack of immunity to fly strike in sheep, even after repeated infestations (Elkington and Mahony, 2007). However, while no naturally acquired immunity seems to occur in sheep, the possibility that repeated infestation can improve resistance has been suggested by the observation of resistance in some older sheep (Watts *et al.*, 1979; Sandeman *et al.*, 1985; Sandeman *et al.*, 1986; Bowles *et al.*, 1996).

While many immunization trials have attempted to increase serum levels of immunoglobulin G (IgG) directed at blowfly larval antigens, including the use of repeated whole larvae infestations, immunisation with antigens from crude larval and excretory-secretory preparations (Bowles *et al.*, 1987), immunisation with larval cuticle proteins (Barrett and Trevella, 1989), purified serine proteases (Tellam *et al.*, 1994), and extracts of the larval peritrophic membrane (Tellam and Eisemann, 1998; Colditz *et al.*, 2002). While *in vitro* feeding experiments generated promising results, with significant retardation in larval growth, *in vivo* results failed to match up, providing insufficient protection from subsequent myiasis infestations, thought to be due to much lower serum IgG levels present *in vivo* (Elkington and Mahony, 2007).
Potential target antigens for the development of a vaccine against sheep strike have also included the proteolytic digestive and excretory/secretory (ES) proteins produced by larvae (Tabouret et al., 2003). However, larval ES products (containing mostly digestive proteases) have been found to display considerable proteolytic activity, but only weak antigenicity in sheep, making them ineffective candidate vaccine antigens (Tabouret et al. 2001). In contrast, salivary gland extracts of Oestrus ovis have been found to be highly immunogenic in sheep (Innocenti et al., 1995; Tabouret et al., 2001; Tabouret et al., 2003), although no detectable enzymatic activity has been found (Tabouret et al. 2003).

One candidate vaccine antigen purified from Lucilia cuprina larvae that showed much promise during the late 1990s was peritrophin-95. This antigen forms an integral glycoprotein component of the peritrophic membrane (PM), a semi-permeable film-like envelope composed of chitin, proteoglycans and proteins, which separates ingested food and the absorptive/secretory midgut tissue. As well as assisting the digestion of host tissue, the PM also protects the absorptive/secretory midgut tissue from parasites and micro-organisms (Casu et al., 1997). Sheep vaccinated with peritrophin-95 were found to produce a strong immune response that inhibits larval growth (Casu et al., 1997; Tellam et al., 2000; Tellam et al., 2001). This larvae growth inhibition is caused by a reduction in permeability of the PM, and consequent starvation of the larvae, due to antibodies in the vaccinated host tissue binding to the peritrophin-95 on the surface of the PM causing in a layer of partially digested protein to build up (Casu et al., 1997). However, vaccination trials comparing native peritrophin-95 and recombinant forms of the protein, produced in bacteria (Escherichia coli) and baculovirus-infected insect cells, found much weaker responses from immunisation of either of the recombinant forms, compared to purified native form of the protein (Tellam et al. 2001). Investigation into the structure and function of the recombinant antibodies found the E. coli-expressed peritrophin-95 to be both un-glycosylated and folded incorrectly. By comparison, the insect cell-expressed peritrophin-95 was glycosylated and almost certainly correctly folded, but displayed differences in oligosaccharide structure from the native peritrophin-95 (Tellam et al. 2001). Additionally, even when
immunization trials used native peritrophin-95, a reduction in larval growth of over 80% was necessary to induce sufficient larval mortality (East et al. 1993; Eisemann et al. 1994; Casu et al. 1997; Tellam et al. 2001). To date, attempts to develop a functional recombinant peritrophin-95 vaccine have failed.

The extensive sheep strike immunisation trials have also revealed that protection in the immunised sheep is uncorrelated to level of antibody, suggesting that serum antibodies do not contribute significantly to providing protection against myiasis (Bowles et al., 1996). Indeed, no direct relationship has been found between level of serum antibody and scale of infestation for any myiasis-causing ectoparasites (Baron and Colwell, 1991). Notably, in sheep strike it seems that a rapid cellular response triggered within hours of first instar larvae coming into contact with the sheep’s skin was responsible for the observed protection. Probing sera, wound exudate, and skin sections from L. cuprina struck sheep, using a monoclonal antibody (MAb) raised against ovine IgE, MacDiarmid et al (1995) successfully exposed cell bound IgE. However, circulating IgE (total or L. cuprina specific) could not be detected, possibly due to the rapid antibody degradation due to enzymatic activity from larval ES products preventing the MAb recognition of IgE (MacDiarmid et al., 1995). This study found no difference in level of cell bound IgE between skin sections of struck and control sheep. The struck sheep showed no increase in IgE even after three strikes, with the exception of one individual which had twice as much IgE as the two other struck animals (MacDiarmid et al., 1995). However, other immunization trials have reported a greater amount of cell bound IgE in resistant compared to susceptible sheep (Colditz et al., 1994). Partial resistance has also been reported after induced infections at 2 weekly intervals, perhaps suggesting that more than 3 strikes are required for such blowfly ‘sensitization’, and observed levels of IgE production or cell binding to increase (MacDiarmid et al., 1995).

Overall immunisation trials have revealed considerable inconsistency in resistance between individuals of sheep, again signifying the innate quality of acquired fly strike resistance (Sandeman et al., 1985; MacDiarmid et al., 1995; O'Meara et al., 1997). Consequently, it seems that the future success of sheep
strike vaccine development requires a combined focused on blocking parasite immune evasion defences to induce rapid host cellular immune responses (more specifically hypersensitivity responses), targeting antigens involved in wound initiation, and inhibiting larval nutrition and/or growth (Elkington and Mahony, 2007).

1.6.8 Baited traps

The potential importance of trapping myiasis-causing flies in reducing sheep strike has long been recognised (Newman and Clark, 1926; Heath, 1993). For example, fly traps used in Southern Queensland, Australia, were associated with reducing sheep strike by up to 55% between 1999 and 2001 (Ward and Farrell, 2003). Over 85% of Australian sheep strike is caused by *Lucilia cuprina*, and so many traps have focused on this particular species. One example is the LuciTrap® (Miazma Pty Ltd, Queensland, Australia), a commercially available non-insecticidal trap that uses a synthetic odour-bait (Lucilure), and which claims to be selective for *L. cuprina*. An investigation into the impact of LuciTraps® in Australia found almost 60% of flies caught to be *L. cuprina* (Urech et al., 2009). LuciTraps® have additionally proven successful in capturing *L. sericata*, accounting for 20-30% of flies captured in Tasmanian trials (Horton et al., 2001). However, a study comparing the effectiveness of LuciTraps® at attracting *L. sericata* in Australia and Hungary have suggested possible behavioural differences between European and Australian *L. sericata* populations, with LuciTrap® performance in Hungary being much poorer than in Australia (Hall et al., 2003).

While baited traps have proven an effective technique for controlling some myiasis-causing flies, they are most efficient when used as part of an IPM strategy (Colwell et al., 2004). One area in which trapping seems to show particular promise is in the development of fly strike early warning systems, using the traps to monitor population dynamics of myiasis-causing flies (Wall et al., 2002; Colwell et al., 2004). Baited traps have also proven valuable in monitoring effectiveness of myiasis control programs. For example, baited traps monitoring numbers of *C. hominivorax* have been used extensively to
assess success of past and on-going SIT programs (Broce et al., 1977; Vale et al., 1988)

### 1.7 Diagnostic techniques

#### 1.7.1 Morphology

To date, the most comprehensive phylogenetic analyses of Calliphoridae have been based on morphological evidence (Rognes, 1991). However, perhaps the most widespread use of blowfly morphological character data is in species identification of forensically important species, whereby immature stages (eggs, larvae, or puparia) can be used to estimate PMI and investigate toxic substances present.

As many forensically important blowfly species exhibit almost identical morphology in immature stages of development, light microscopy (Hinton, 1981; Erzinclioglu, 1989b; Sukontason et al., 2004) and scanning electron microscopy (SEM) (Kitching, 1976; Greenberg and Szyska, 1984; Erzinclioglu, 1989b; Liu and Greenberg, 1989; Greenberg and Singh, 1995; Mendonca et al., 2008) have been extensively used in attempts to differentiate the immature stages of forensically important species. However, inconsistencies in observations seem to suggest the presence of some intra-specific morphological diversity among blowflies (Greenberg and Singh, 1995; Mendonca et al., 2008). Consequently, attempts are usually made to rear these immature stages to adults in order to aid accurate species identification (Amendt et al., 1999; Sukontason et al., 2004; Sukontason et al., 2007b). Nevertheless, applying morphological identification keys typically requires comprehensive knowledge and experience in entomology and taxonomy (Amendt et al., 1999).

#### 1.7.2 DNA-based analyses

The last decade has seen a significant increase in molecular studies on forensically important blowflies and these have proved particularly useful for the identification of immature stages, where morphological differentiation is often
ambiguous, and for use with damaged specimens (Sperling et al., 1994; Stevens and Wall, 2001; Wallman et al., 2005; Harvey et al., 2008).

At present, DNA sequence-based diagnostics, as first suggested by (Sperling et al., 1994), are the only routinely used molecular-based species identification tool (Wells and Stevens, 2008). For example polymerase chain reaction (PCR) amplification of the complete internal transcribed spacer (ITS) region of ribosomal DNA (rDNA), followed by restriction fragment length polymorphism (RFLP) analysis has been demonstrated to have the ability to identify forensically important dipteran species (Ratcliffe et al., 2003). However, single point mutations have the capability to add or remove restriction sites, changing the restriction profile, and potentially leading to mis-identification of specimen (Nelson et al., 2008).

Another method that has been advocated for taxonomic studies is random amplified polymorphic DNA (RAPD) PCR technique (Welsh and McClelland, 1990; Williams et al., 1990). However, characterization studies using this technique have demonstrated problems with the ability to replicate results (Black, 1993). For example, RAPD amplifications have been shown to be affected by a range of factors, including; relative concentrations of PCR reagents, DNA template concentration, extraction method, and DNA preservation procedure (Stevens and Wall, 1995; 1997b).

Finally, perhaps the most widely researched method for the identification of forensically important blowflies uses DNA sequencing and phylogenetic analysis (Wells et al., 2001; Wells and Sperling, 2001). This method involves analysing sequence data from an unknown specimen along with sequences from known taxa, whereby the unknown taxa pair with their closest relative in an inferred phylogeny (Brunner et al., 2002; Nelson et al., 2008). However, the accuracy and validity of such DNA based species identification has been suggested to be unreliable (Wallman et al., 2005), with some closely related clades still having little sequence variation, due to recent divergence times (Wallman et al., 2005). Targeting more variable regions or using allozymes
have been suggested as a solution to help alleviate this problem (Wallman et al., 2005).

Furthermore, current practice typically employs single gene analysis for the identification of blowfly species, the limitations in the reliability of which, for example due to horizontal gene transfer or incomplete lineage sorting at the locus in question, have been reported previously (Stevens et al., 2002; Stevens, 2003; Wells et al., 2007). Adding sequence character data has also been found to greatly improve resolving power within Insecta (Hillis et al., 1994; Russo et al., 1996). While it seems unlikely that any single gene will be capable of unambiguously identifying all blowfly species, a universal switch to multi-gene approaches, for example using expressed sequence tags (ESTs) to screen multiple independent sites across blowfly genomes, is urgently needed.

1.7.3 Rapid diagnostic tests

While research concerning human and animal disease typically focus on vaccine and drug development, the need for better diagnostics is becoming increasingly recognised (Ridley, 2006). Above all, there is a necessity for quality assured, user friendly diagnostic tests that can be used ‘in the field’ (Ridley, 2006). Such rapid tests would allow immediate diagnosis and treatment, for example helping to prevent the spread of disease to other individuals in a population.

To date, few diagnostic tools have implemented antigenic markers to distinguish morphological similar insect species (Miller, 1981; Ma et al., 1990; Stuart et al., 1994; Zeng et al., 1999). Traditionally, studies into arthropod parasites and the immune responses of their vertebrate hosts have focused on non-specific antigens, giving restricted diagnostic capacity (Otranto, 2001). However, immunological methods for the diagnosis of myiasis can play a vital role in both the early detection of infestations, and also in general population monitoring, in turn assisting the development of treatment, prevention and eradication strategies. The successful development of an enzyme-linked immunosorbent assay (ELISA) based on two conformational monoclonal
antibodies (MAbs) specific for New World screwworm (Figarola et al., 2001), lends further support to the prospect of developing successful rapid diagnostic tools for the identification of blowfly species of veterinary, medical and forensic importance.

1.8 Methods of phylogenetic inference

1.8.1 Substitution models

Even the earliest comparative analyses of DNA sequence data (e.g. Jukes and Cantor, 1969; Brown et al., 1982) acknowledged the need for substitution models to describe rates of sequence change over time in molecular phylogenetics. The rate of divergence between two sequences is only linear for a short amount of time; the more divergent two sequences are the higher the likelihood that multiple mutations occur at the same nucleotide site, leading to an underestimation of the number of nucleotide substitutions since the two sequences last shared a common ancestor (Sullivan and Joyce, 2005). The first substitution models developed (e.g. Jukes and Cantor, 1969) attempted to account for these multiple substitutions at the same site to improve accuracy of phylogenetic reconstruction.

Underestimating the amount of evolutionary change (number of substitutions) can be particularly problematic in cases where a phylogenetic tree contains long branches separated by a short internal branch, a phenomenon known as long branch attraction (Felsenstein, 1978). Long branch attraction (LBA) occurs when DNA sequences from two or more long branch lineages are assumed to be closely related, regardless of true evolutionary history, for example when rates of substitution are high the probability that these long branch lineages have independently evolved the same nucleotide base at a particular site increases. LBA is most common where rates of evolution vary significantly between taxa, or where taxa are particularly divergent (Page and Holmes, 1998). This problem can, however, be minimized by adding taxa to break up long branches, using alternative slower evolving traits, and by employing an appropriate substitution model. Model choice is therefore critical (Sullivan and Joyce, 2005).
Substitution models are also able to assign weights to each possible nucleotide substitution, with most models being reversible, meaning that they allocate the same weight to reversible substitutions (e.g. C to G is equally as likely as G to C). The Jukes-Cantor model (Jukes and Cantor, 1969) is the simplest model possible, assuming all substitutions are equally likely, while more complex models distinguish between different substitutions, for example Kimura’s two-parameters model (Kimura, 1980) assumes different rates between transitions and transversions.

It has been demonstrated that substitution models that take into account rate heterogeneity across sites provide better tree estimates (Yang, 1996; Whelan et al., 2001). Rate heterogeneity can be sub-divided into two forms: spatial heterogeneity, where substitution rate or nucleotide frequencies vary between sites, and temporal heterogeneity, where substitution rates vary in time along the branches of an evolutionary tree (Whelan, 2008). Several substitution models have been developed to assume heterogeneity. The most popular way of assuming spatial heterogeneity is to apply a gamma (Γ) distribution to substitution rate, whereas temporal heterogeneity can be accounted for by using the covarion model (Tuffley and Steel, 1998), which allows switches in substitution rates (Whelan, 2008).

Another common feature of nucleotide data is compositional bias, where frequencies of the four nucleotides based are unequal. Taxa are described as showing stationarity when they demonstrate similar patterns of compositional bias, situations where compositional bias varies between taxa are then described as deviations from stationarity (Collins et al., 1994). Examples of models that can account for composition bias, unequal rates of substitution, and among-site rate variation include proportion of invariant sites (+I) and discrete gamma-distributed rates of variable sites (+Γ) (Swofford et al., 1996; Whelan et al., 2001).
1.8.2 Distance methods

Distance methods of phylogenetic reconstruction are based on the concept of using measures of sequence dissimilarity to calculate the number of evolutionary changes that have occurred since the sequences shared a common ancestor (Page and Holmes, 1998). Pairwise calculations of these observed distances can then be used to infer the evolutionary history of those sequences by reconstructing a phylogenetic tree by placing closely related sequences under the same interior node, and with branch lengths representing the observed distances between sequences. However, distance measures can also underestimate evolutionary changes, for example the greater the divergence between two sequences the more likely it is that multiple mutations have occur at the same nucleotide site. To resolve this issue, substitution models representing different hypotheses about relative rates of mutation along the sequences can be used to correct these distance measures.

Two main objections to the use of distance methods to construct phylogenetic trees have been made: 1) that using pairwise distances to summarize sequences loses information, and 2) that tree branch lengths reconstructed from observed distances are not always easy to interpret biologically. For example, homoplasy can result in observed distances being much smaller than the true evolutionary distances (Page and Holmes, 1998).

1.8.3 Maximum parsimony

In comparison, maximum parsimony (MP) is a discrete method, in that it uses sequence data directly, or functions derived from it, rather than pairwise distance measures. Maximum parsimony methods of phylogenetic inference select the tree, or trees, with the fewest evolutionary changes (i.e. the most parsimonious tree or trees, MPT(s)). MP methods maximise the amount of evolutionary change through homology (i.e. minimises the amount of homoplasy), and assumes the implicit hypothesis of evolution itself, i.e. that evolutionary change is rare (Page and Holmes, 1998).
Nevertheless, MP has been shown to be inconsistent under some instances, for example LBA, which can cause MP to mistake homoplasy as synapomorphy (Bergsten, 2005). Maximum parsimony is particularly vulnerable to LBA as seeks the minimum number of evolutionary events (Felsenstein, 2004).

1.8.4 Maximum likelihood

Maximum likelihood (ML) methods of phylogenetic reconstruction calculate the likelihood (L) of obtaining the sequence data, given the topology of a tree. The tree that gives the sequence data the highest likelihood is selected (Page and Holmes, 1998). While ML methods are most commonly used in molecular phylogenetics (Swofford et al., 1996; Huelsenbeck and Crandell, 1997; Huelsenbeck et al., 2002; Ronquist, 2004), progress in the development of morphology based ML methods is being made (Lewis, 2001; Nylander et al., 2004; Ronquist, 2004). However, It has recently been shown that multiple maximum likelihood values may exist for a given phylogenetic tree (Zhou et al., 2006). This implies that even if the globally optimal tree is found using ML methods, the ML criterion alone may not be enough to resolve the true phylogeny for certain problems (Zhou et al., 2006).

1.8.5 Bayesian inference

Bayesian inference follows the concept of Bayes’ theorem (Barnard and Bayes, 1958), which states that the posterior probability is the probability of the hypothesis given the data, as opposed to ML which is the probability of the data given the hypothesis, and that the prior is the probability of the hypothesis before the data. The Bayesian method used in phylogenetic inference, first introduced by three independent groups in 1996 (Li, 1996; Mau, 1996; Rannala and Yang, 1996), uses Markov Chain Monte Carlo (MCMC) methods to estimate distribution of posterior probabilities and substitution parameters to make tree searches more computationally feasible (Brooks et al., 2007).
To prevent MCMC chains becoming sub-optimally trapped Metropolis-coupled MCMC, or MCMCMC, can be implemented. Under this algorithm several MCMC chains with different stationary distributions can be run in parallel, typically a two simultaneous MCMCMC runs, each with four MCMC chains, are used. Within each MCMCMC run the chains swap information at regular intervals to improve the mixing in presence of multiple local peaks in the posterior probability (Geyer, 1991; Huelsenbeck et al., 2001; Altekar et al., 2004). One of the four chains represents the ‘cold chain’, as it contains the target posterior probability, while the remaining three chains are ‘heated chains’. Heating chains makes it easier to cross between peaks in the posterior probability. After each iteration, states can be swapped between chains leading to better mixing, and preventing the cold chain becoming sub-optimally trapped.

Assuming the evolutionary model used is true and that the posterior probability distribution has been correctly sampled by the MCMC, the resulting posterior probability value represents the probability that the tree is correct given the data. The probability that a monophyletic clade is ‘true’, given the caveats of the model and data, is estimated by the proportion of trees in the MCMC sample in which the clade occurs (Brooks et al., 2007).

1.9 Choice of molecular marker

1.9.1 Mitochondrial DNA

Within insect systematics mitochondrial DNA (mtDNA) and nuclear ribosomal DNA have been the most popular molecular targets (Otranto and Stevens, 2002). Mitochondrial DNA contains regions with a range of different rates of evolution (Simon et al., 1994), making it suitable for both phylogenetic and population genetic studies, where genes with appropriate rates of evolution can be chosen for a particular temporal scale of analysis (Kocher et al., 1989; Avise, 1994). Other advantages of mtDNA sequence data include, a lack of recombination during cell division, high copy number, relative ease of isolation, availability of universal primers, the presence of both conserved and variable regions, and the expectation of mtDNA to reach reciprocal monophyly before
nuclear genes due to generally higher rates of sequence change (Avise et al., 1979; Lunt et al., 1996; Monteiro and Pierce, 2001; Funk and Omland, 2003; Dowton, 2004; Lin and Danforth, 2004). Rates of substitution rates in mtDNA are estimated to be between five and ten times that of nuclear DNA (Brown et al., 1979), making mtDNA particularly useful for inferring relationships between recently diverged species and in population genetics (Stevens and Wall, 1997b; Shao and Barker, 2006).

Currently, the vast majority of mtDNA population genetic and evolutionary studies of parasitic arthropods have used single protein coding genes, with cytochrome oxidase I (COX1), cytochrome-b (CYTB), large subunit (LSU) rRNA, and small subunit (SSU) rRNA, having received the most attention (Shao and Barker, 2006).

1.9.2 Nuclear DNA

A wide range of protein coding nuclear DNA genes have been used in insect systematics, including; elongation factor-1 alpha (EF-1α), (Cho et al., 1995; Danforth and Ji, 1998; Clark et al., 2000; Caterino et al., 2001; Cognato and Vogler, 2001; Buckley et al., 2002; Danforth, 2002), phosphoenolpyruvate carboxykinase (PEPCK) (Friedlander et al., 1996; Wiegmann et al., 2000; Sota and Vogler, 2001; Leys et al., 2002), dopa decarboxylase (DDC) (Fang et al., 1997; Tatarenkov et al., 1999; Friedlander et al., 2000), and opsin (Mardulyn and Cameron, 1999; Ascher et al., 2001; Cameron and Mardulyn, 2001; Danforth et al., 2003) to name a few.

Nuclear genes have several advantages, including; a generally low level of biased base composition, usually a slower rate of evolution than mtDNA, and the presence of regions with slower and more rapid rates of substitution (i.e. exons and introns) (Friedlander et al., 1992; Brower and DeSalle, 1994; Friedlander et al., 1994; Lin and Danforth, 2004). General observations have been made that nuclear genes have a greater resolving power than mtDNA, particularly at deeper node levels of phylogenetic inference (Reed and Sperling, 1999; Leys et al., 2000; Baker et al., 2001; Brady, 2002; Morris et al., 2002;
Danforth et al., 2003; Lin and Danforth, 2004). However, nuclear genes can also be more difficult to work with, compared to mtDNA, due to lower copy number and the not uncommon occurrence of two or more paralogous loci (Lin and Danforth, 2004).

1.9.3 Nuclear ribosomal DNA

Nuclear ribosomal DNA (rDNA) is made up of a series of tandemly repeated units, each comprising of ribosomal RNA (rRNA) genes (18S, 5.8S and 28S) separated by segments of DNA referred to as ribosomal spacer DNA. The secondary structure of rRNA molecules consists of a variety of structures (e.g. stems, hairpin loops, bulges, multi-helix junctions) created by folds in the peptide chain. These secondary structures are vital for the biological function of the RNA molecule (e.g. protein synthesis), meaning that the regions of the RNA sequence that preserve them are under distinct evolutionary constraints (Fox and Woese, 1975; Gutell et al., 1994). In eukaryotes, regions of the sequence coding for hairpin loops demonstrate a higher rate of mutation accumulation (Smit et al., 2007). The difference in rates of mutation accumulation across rDNA repeating units and across single rRNA genes (Hillis and Dixon, 1991), make rDNA sequence data an effective molecular tool for the differentiation of parasitic species, and even strains (Arnheim, 1983; Gasser, 1999). Despite the usually low rates of intraspecific variation between first and second internal transcribed spacers (ITS-1 and ITS-2), sufficient differentiation has been observed in some arthropods to distinguish between co-specific individuals, with ITS microsatellites proving valuable polymorphic markers for the study of population dynamics (Vogler and DeSalle, 1994; Onyabe and Conn, 1999; De Barro et al., 2000; Marcilla et al., 2001). Furthermore, non-coding regions such as the ITS have also been shown to exhibit much higher rates of mutation than coding mtDNA genes (Otranto and Stevens, 2002).
1.10 **Present study**

Blowflies have been an important research focal point for over 50 years, with myiasis-causing genera in particular having extensive medical and veterinary significance worldwide. The evolution of myiasis-causing dipteran groups has been linked to the development of key behavioural, physiological, and biochemical adaptations (Stevens *et al.*, 2006). By attempting to resolve how these key antigens and immunogenic proteins evolved in the blowfly, together with data from multi-gene sequencing studies, this thesis aims to explore the evolution of these parasites and the various parasitic traits which they exhibit.

1.10.1 **Aims and objectives**

The work presented in this thesis aims to address the following hypotheses;

**Chapter Three**

- The phylogenetic utility, measured by the ability to recover recognised families within Insecta, of some mitochondrial genes is better than others.
- Measuring nucleotide diversity within a gene can indicate its phylogenetic utility.

**Chapter Four**

- A congruent multi-gene phylogeny can be constructed for Calliphoridae using the genes *COX1*, *EF1-a*, and *28S*.
- The origins of different styles of parasitism within Calliphoridae reflect phylogenetic relationships within the family.

**Chapter Five**

- Populations of the New World screwworm (*Cochliomyia hominivorax*) display genetic sub-structuring in Central and South America which could affect current and future pest control programmes using sterile insect technique.
Chapter Six

- Antigenic larval protein profiles vary between different types of larval feeding behaviour (e.g. parasitic or saprophagic).

- Antigenic larval protein profiles are species-specific and present target proteins which may act as potential candidates for species-specific vaccines and diagnostic tools.
Chapter Two

Materials and Methods
2 Materials and Methods

2.1 Molecular methods

2.1.1 Sample collection and storage for DNA analysis

While the majority of specimens included in this study came from existing in-house collections at the University of Exeter, some freshly collected samples sent by colleagues, and specimens on loan from external collections were also incorporated. All specimens were stored at 4°C, either in 100% ethanol or as pinned dried specimens.

Where DNA material was available from numerous populations across the geographic range of a single species, several sequences were amplified for each gene, otherwise only a single sequence was amplified per gene for each species.

2.1.1.1 Cochliomyia hominivorax (New World screwworm)

Specimens of Cochliomyia hominivorax were obtained from locations across South American and the Caribbean (namely; Brazil, Colombia, Cuba, Dominican Republic, Ecuador, Jamaica, Peru, Venezuela, Trinidad, and Uruguay), with two or more sequences being amplified for each gene from each sample site. Furthermore, two specimens belonging to historical populations of C. hominivorax, that had been dried and pinned in 1933 and 1953, were obtained on loan from collections held at the Natural History Museum, London. See chapter 5 for further details.

2.1.2 DNA extraction

All specimens stored in 100% ethanol were re-hydrated with distilled water for a minimum of one hour prior to DNA extraction. Where adult fly specimens were available, thoracic flight muscle tissue was extracted from the ventral side of the thorax, therefore avoiding contamination from ingested protein, parasites or
eggs (Stevens, 2003). This practice also conserves nearly all taxonomically important external morphology, enabling samples to be retained as voucher specimens – unless on temporary loan from external collections. Where second or third stage larvae were used, the entire specimen was used during DNA extraction.

2.1.2.1 Salt extraction method

The majority of DNA extractions were carried out using the following salt extraction method (Aljanabi and Martinez, 1997). The tissue was homogenized in 400μl SSHB (sterile salt homogenizing buffer; sodium chloride 0.4M, Tris-HCl 10mM pH 8.0, and ethylenediaminetetraacetic acid (EDTA) 2mM pH 8.0) using a sterile plastic micropestle. To this, 40μl of 20% sodium dodecyl sulphate (SDS) and 8μl of 20mg/ml Proteinase K were added, mixed well, and incubated for 1 h at 60ºC, before being centrifuged for 1 min at 10,000g. The supernatant was decanted into clean eppendorf tubes, and 300μl of 6M sodium chloride added, before being vortexed for 30 seconds at maximum speed, and then spun down for 30 min at 10,000g. The second supernatant was again transferred to a clean tube and an equal volume (~750μl) of propan-2-ol added, mixed well, and incubated for at least one hour at -20ºC, before being centrifuged at 4ºC for 20 min at 10,000g. The resulting DNA pellet was washed with 500μl of 70% ethanol and centrifuged for 7 min at 10,000g. The supernatant was discarded, and the inside of the tube carefully dried with medical wipes, taking care not to disturb the pellet. Finally, the pellet was re-suspended in 50μl low EDTA TE (Tris EDTA) buffer, and heated to 60ºC for 1-2 hours, to promote rapid resuspension. All successfully extracted DNA samples were stored at –20ºC until analysis.

2.1.2.2 Qiagen DNeasy® Blood & Tissue Kit extraction

In some cases (e.g. historical pinned specimens on loan from external collections), only a limited amount of tissue was available, meaning that often only one attempt at DNA extraction was possible. In such instances DNA extractions were carried out using a Qiagen DNeasy® Blood & Tissue Kit
(Qiagen GmbH, Germany), according to the manufacturers protocol. Extractions were conducted in a separate ‘clean’ room to eliminate all possibility of contamination, and again all extractions stored at −20ºC until further analysis.

2.1.3 PCR amplification

DNA extractions were subject to PCR procedures to amplify regions of the nuclear gene elongation factor 1 alpha (EF-1α), the ribosomal RNA 28S, the mitochondrial cytochrome oxidase subunit 1 (COX1) or the ribosomal RNA 12S gene, see chapters 4 and 5 for further details.

Standard PCR amplification involved 1μl DNA (taken from a 1 in 10 dilution of the original DNA extraction), 2.5μl reaction buffer (160mM (NH₄)₂SO₄, 670mM Tris-HCl (pH8.8 at 25°C), 0.1% Tween-20), 1.25μl MgCl₂ (50mM), 2.5μl dNTP mix (giving final concentration of 0.2mM), 2.5μl forward and reverse external primers (5μM), 1μl of BIOTAQ™ Red DNA polymerase, and sterile H₂O to a final volume of 25μl. All amplifications were carried out in Hybaid thermal cycler machines. All PCR reactions using ‘historical’ material were carried out in a separate ‘clean’ room to avoid possible cross contamination.

2.1.3.1 Cytochrome oxidase subunit 1

Published COX1 primers (Bogdanowicz et al., 1993; Simon et al., 1994; Sperling and Hickey, 1994; Sperling et al., 1995; Lunt et al., 1996; Wells and Sperling, 1999) were used to amplify templates, with the exception of several Cochliomyia hominivorax (New World screwworm) samples where repeated attempts to amplify templates using the published primers failed. Accordingly, a new set of primers were designed in an attempt to amplify the remaining templates (Table 1).
## Materials and Methods

<table>
<thead>
<tr>
<th>Gene</th>
<th>Primer</th>
<th>Sequence (5’ to 3’)</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>COX1</td>
<td>COX1 F1**</td>
<td>ATAGACGTTGAYACWCGAGC</td>
<td>This study</td>
</tr>
<tr>
<td>COX1</td>
<td>COX1 R1**</td>
<td>ATCCATTGCACTAATCTGCC</td>
<td>This study</td>
</tr>
<tr>
<td>COX1</td>
<td>Int.F1**</td>
<td>GTATTTGCTATTATRGCCGG</td>
<td>This study</td>
</tr>
<tr>
<td>COX1</td>
<td>Int.R4**</td>
<td>AATCCTAARAAATGTGAGG</td>
<td>This study</td>
</tr>
<tr>
<td>C1-J-1751a</td>
<td></td>
<td>GGATCACCTGATATAGCATTCCC</td>
<td>Bogdanowicz et al., (1993)</td>
</tr>
<tr>
<td>C1-J-2183</td>
<td></td>
<td>CAACATTATTTTGTATTTTTTG</td>
<td>Simon et al., (1994)</td>
</tr>
<tr>
<td>C1-J-2495</td>
<td></td>
<td>CAGCTACTTTATGCTTATAG</td>
<td>Sperling et al., (1994)</td>
</tr>
<tr>
<td>C1-N-2191</td>
<td></td>
<td>CCCGCTAAAAATTTAAAACCTTC</td>
<td>Bogdanowicz et al., (1993)</td>
</tr>
<tr>
<td>C1-N-2659</td>
<td></td>
<td>GCTAATCCAGTGAATAATGG</td>
<td>Sperling &amp; Hickey (1994)</td>
</tr>
<tr>
<td>TL2-N-3013</td>
<td></td>
<td>TCCATTACATATAATCTGCAATATTAGTAG</td>
<td>Wells &amp; Sperling (1999)</td>
</tr>
<tr>
<td>UEA7</td>
<td></td>
<td>TACAGTGGAAATAGACGTTGATA</td>
<td>Lunt et al. (1996)</td>
</tr>
<tr>
<td>UEA10</td>
<td></td>
<td>TCAATGCACTAATCTGCAATTATA</td>
<td>Lunt et al. (1996)</td>
</tr>
<tr>
<td>EF-1α</td>
<td>B1**</td>
<td>CCCATYTCCGHTGGACGCGG</td>
<td>This study</td>
</tr>
<tr>
<td>C1**</td>
<td></td>
<td>GTCTCATGTCACGDACRGCG</td>
<td>This study</td>
</tr>
<tr>
<td>F.Int**</td>
<td></td>
<td>GTGCGYATCGGHACAGTACC</td>
<td>This study</td>
</tr>
<tr>
<td>R.Int**</td>
<td></td>
<td>AGTTTTCAACAGCAGCAAAGC</td>
<td>This study</td>
</tr>
<tr>
<td>EF1</td>
<td></td>
<td>ACAGCGGCGTGTTGTCCTGATGTC</td>
<td>This study *</td>
</tr>
<tr>
<td>EF2</td>
<td></td>
<td>CACATTAACATTCTGCGGATTGG</td>
<td>This study *</td>
</tr>
<tr>
<td>EF3</td>
<td></td>
<td>CCAGATTACACCAGTTTGTGA</td>
<td>This study *</td>
</tr>
<tr>
<td>EF4</td>
<td></td>
<td>CCTGGTCTACAGAGGATGGA</td>
<td>This study *</td>
</tr>
<tr>
<td>28S</td>
<td>D1.F</td>
<td>CCCCCTGAAATTTAAGCATAT</td>
<td>Friedrich &amp; Tautz (1997a)</td>
</tr>
<tr>
<td>D2.R</td>
<td></td>
<td>GTTAGACCTCTTGCTGCTCGTG</td>
<td>Hoelzel &amp; Green (1992)</td>
</tr>
<tr>
<td>D1.R</td>
<td></td>
<td>CTCTCTATTTAGCTGGCTTTCCTTC</td>
<td>Friedrich &amp; Tautz (1997a)</td>
</tr>
<tr>
<td>D2.F</td>
<td></td>
<td>GAGGGGAAGTTGAAAAGAAC</td>
<td>Hoelzel &amp; Green (1992)</td>
</tr>
<tr>
<td>D3</td>
<td></td>
<td>GCCGGGCTGTTGAACACCGG</td>
<td>Friedrich &amp; Tautz (1997b)</td>
</tr>
<tr>
<td>D7.R</td>
<td></td>
<td>CGACCTCTCCTACCTACAT</td>
<td>Friedrich &amp; Tautz (1997a)</td>
</tr>
<tr>
<td>D3-5.R</td>
<td></td>
<td>TTACACACTCTTTAGCGGA</td>
<td>Friedrich &amp; Tautz (1997b)</td>
</tr>
<tr>
<td>D35.486.R</td>
<td></td>
<td>TCAGGAAGCAGCACTACCTA</td>
<td>Friedrich &amp; Tautz (1997b)</td>
</tr>
<tr>
<td>D35.742.F</td>
<td></td>
<td>TCTCAAACTTTAAATGG</td>
<td>Friedrich &amp; Tautz (1997b)</td>
</tr>
<tr>
<td>D7.F</td>
<td></td>
<td>GACTGAGTAGGAGAAAGGT</td>
<td>Friedrich &amp; Tautz (1997a)</td>
</tr>
<tr>
<td>TI-N-24</td>
<td></td>
<td>ATTACCCCTCTCAAAGTGA</td>
<td>Simon et al., (1994)</td>
</tr>
</tbody>
</table>

Table 1: Details of PCR primers used.

* modified from Moulton, (2000) primers (3PC, 5PC, JOM, ShM)

** newly designed *C. hominivorax* specific primers.
Figure 2: Schematic representation of the genes used in this study, indicating position of primers used for PCR amplification of the gene sequence.

*COX1* reactions using the published primers applied an initial denaturation step of 94°C for 5 min; followed by 35 cycles of 95°C for 40 s, 48°C for 30 s, and 72°C for 1 min; and a final elongation step of 72°C for 10 min (Yu et al., 2005). Reactions using the new *C. hominivorax* specific *COX1* primers employed a protocol of 95°C for 5 min as the initial denaturation step; 35 cycles of 55°C for 1 min, 72°C for 1 min and 94°C for 30 s, followed by 55°C for 1 min;
and a final elongation step of 72°C for 10 min, modified from the generic PCR protocol of Hoelzel, (1992).

2.1.3.2  *Elongation factor 1- alpha*

Modified versions of existing primers (J.K. Moulton, Tennessee, *pers. Comm.*) were used to amplify EF-1α, although initial attempts proved unsuccessful for the *C. hominivorax* samples. Subsequently, a new set of EF-1α primers specific for *C. hominivorax* were designed (Table 1). PCR amplification using both the modified Moulton and *C. hominivorax* specific EF-1α primers used the adapted Hoelzel (1992) PCR protocol (Section 2.1.3.1).

2.1.3.3  *28S* ribosomal RNA

Published primers (Hoelzel and Green, 1992; Friedrich and Tautz, 1997b, 1997a) were employed to amplified the D1-D2 and D3-D7 regions of the *28S* rRNA gene, using the adapted Hoelzel (1992) PCR protocol (Section 2.1.3.1).

Comparisons between *28S* sequences of *C. hominivorax* and sister taxa *Cochliomyia macellaria* showed little variation, so the decision was made to not amplify this gene for the *C. hominivorax* samples as there would likely bit little or no phylogeographic content. However, alignments of mitochondrial rRNA *12S* sequences did reveal variation, and so gene this was amplified for the *C. hominivorax* samples instead (Chapter 5).

2.1.3.4  *12S* ribosomal RNA (*C. hominivorax* samples only)

Attempts were first made to also amplify some of the highly variable D-loop region, as it is positioned next to the *12S* gene on the *C. hominivorax* mitochondrial genome. While successful template amplification was obtained using published primers (Lessinger and Azeredo-Espin, 2000) automated sequencing was only consistently obtained for one primer (*SR-J-14233*) amplifying part of the *12S* gene (Fig. 2). However, when this region was compared between different *C. hominivorax* samples some variation was shown.
Accordingly, as this region could be easily amplified, and also because of time constraints due to this work being part of a collaboration with the University of Campinas, Brazil, all further amplification of templates focused on this region of the 12S rRNA gene only.

Reactions were carried out using a procedure of 94°C for 2 min initial denaturation step, followed by 30 cycles of 94°C for 30s, 58°C for 1 min and 60°C for 2 min, followed by a final elongation step of 60°C for 10 min (Lessinger and Azeredo-Espin, 2000).

2.1.4 Gel electrophoresis

PCR products were separated by electrophoresis gel, using medium sized (12.8 x 15cm) horizontal 1% agarose gels in 0.5x TBE (Tris-borate-EDTA) buffer, and bands visualised by ethidium bromide staining (1μl/ml). Fragment size was estimated using 100bp size standard (Promega), and appropriate bands cut out and purified using a Qiagen MinElute® PCR purification kit, according to manufacturers protocol.

2.1.5 Automated sequencing

Purified PCR products were sequenced on a gold standard ABI 3730xl DNA Analyzer platform, using the commercial sequencing facility COGENICS (formally Lark Technologies Inc.). Quality of sequence fragments were checked and edited manually, before being assembled into a single contig, using AutoAssembler 2.0 (ABI), from which a consensus sequence was exported.

2.1.6 Polyclonal antibody (PAb) production

Rabbit polyclonal antiserum was raised against PBS protein extracts from whole larvae of *Lucilia sericata*. Larval protein preparations were used as immunogen to immunise New Zealand male rabbits using standard immunization procedures (Thornton et al., 1993), at the Institute of Animal Health, Berkshire.
2.1.7 SDS PAGE gel electrophoresis

Lyophilized larval protein preparations were diluted to 0.1mg/ml in 500µl buffer with Laemmli buffer (0.5 M Tris-HCl, pH 6.8; 10% glycerol; 10% SDS; 5% β-mercaptoethanol; 0.5% bromphenol blue), homogenized with a micropestle, and boiled for 5 min in a water bath. Intra and inter-molecular disulphide bonds were prevented from reforming by the β-mercaptoethanol (Laemmli, 1970). Samples were then spun down in a Hettich MIKRO 200 microcentrifuge at 14,000 rpm for 5 min.

Proteins in each sample were separated by SDS PAGE gel electrophoresis, using a Bio-Rad Mini-PROTEAN® II dual slab cell, at 165 volts for 1 h, with wells receiving 16 µg of sample protein per well. PAGE running buffer was composed of the following (L⁻¹); Tris base 3.0g, glycine 14.4g, SDS 1.0g, made up to 1 litre with deionised water.

2.1.8 Western blotting

Proteins separated by SDS PAGE gel electrophoresis were transferred onto a polyvinylidene fluoride (PVDF) membrane using a Bio-Rad Mini Trans-Blot® electrophoretic transfer cell. Transfer buffer was prepared at (L⁻¹); Tris base 2.9g, Glycine 14.5g, Methanol 200ml, made up to 1 litre with deionised water and chilled to 4°C. A sponge pad and two filter papers, soaked in the chilled transfer buffer, and gel were laid onto the plastic gel holder cassette, black side down. The PVDF membrane, activated in 100% methanol for 2 min and then rinsed with deionised water, was placed on top of the gel, taking care to exclude all air bubbles from between the gel and membrane to ensure effective protein transfer. Two more filter papers and a second sponge pad, again all soaked in the chilled transfer buffer, were then placed on top of the membrane, and the gel holder cassette closed and placed in the electrode module in the tank filled with transfer buffer. The western blot was then run in an ice bucket at 75 volts for 2 h. Once the transfer was complete, non-specific binding of antibodies was prevented by incubating the membrane on a slow rocker at 4°C overnight with blocking buffer, consisting of phosphate-buffered saline (PBS) and 1% (w/v)
bovine serum albumin (BSA). Next the membrane was incubated with the primary antibody (rabbit anti-sera) diluted to 1:1000 in antibody dilutent buffer (PBS, 0.5% BSA), on a slow rocker for 2 h at room temperature. Unbound primary antibody was washed off using three 5 min PBS washes. The membrane was then incubated with the secondary antibody (anti-rabbit IgG), diluted to 1:5000 in antibody dilutent buffer (PBS, 0.5% BSA), for 1 h at room temperature. Unbound secondary antibody was washed off with three 5 min PBS washes, followed by a 5 min PBS-T (PBS containing Tween-20 detergent) wash. The membrane was then incubated for 5 min in alkaline phosphate substrate buffer (APSB), consisting of sodium chloride 0.584g; magnesium chloride 0.102g; Tris base 1.21g; made up to 100ml with deionised water, and pH adjusted to 9.5 with concentrated hydrochloric acid. Substrate solution, consisting of 10ml APSB, 66μl Nitroblue tertazolium, NBT, (50mg NBT in 1ml 70% dimethyl formamide), and 33μl 5-Bromo-4-chloro-3-indoylphosphate, BCIP, (25mg BCIP in 500μl 100% dimethyl formamide), was then incubated with the membrane on a slow shaker at room temperature until bands developed. Once bands were visible the membrane was rinsed with deionised water and dried between sheets of Whatman filter paper.

2.1.9 N-terminal sequencing

Protein bands of interest were separated from any impurities using SDS PAGE gel electrophoresis, using a Bio-Rad Mini-PROTEAN® II dual slab cell, at 165 volts for 1 h, before being electro transferred onto a PVDF membrane a Bio-Rad Mini Trans-Blot® electrophoretic transfer cell, and stained with Coomassie blue. PVDF blots were then sent for sequencing at the University of Bristol Proteomics Facility, where the blots were de-stained fully with 0.1% triethylamine in methanol and loaded on to an ABI procise cLc edman sequencer. Sequencing was carried out using standard cycles.
2.1.10 **Two-dimensional gel electrophoresis**

First instar larval preparations from *Lucilia sericata* and *Calliphora vicina* were sent to University of Bristol Proteomics Facility for two-dimensional (2D) gel electrophoresis.

At Bristol the larvae were resuspended in lysis buffer (7M Urea, 2M Thiourea and 4% CHAPS) using a hand-held homogeniser. Insoluble material was removed by spinning and aliquots (2 × 150µg) of the resulting supernatant were then made up to 200µl with lysis buffer containing 0.002% Bromophenol Blue, 0.5% (v/v) IPG Buffer pH3-11NL (GE Healthcare) and 1.2% (v/v) Destreak reagent (GE Healthcare). Samples were loaded onto 11 cm Immobiline DryStrip gels (pH 3–11 non-linear) by passive rehydration for a minimum of 12h. Following rehydration, the DryStrip gels were transferred to an Ettan IPGPhor 3 system (GE Healthcare) and isoelectric focusing was performed by applying 500 V for 1 hr, 1,000 V for 1 hr, and 8,000 V for 2 hr until a total of 14,000 Vhr had been achieved. Following isoelectric focusing, strips were equilibrated in SDS equilibration buffer (50 mM Tris-HCl, pH 8.8, 6 M urea, 30% (v/v) glycerol, 2% (w/v) SDS, and 0.002% (w/v) bromphenol blue) containing 1% (w/v) DTT for 15 min at room temperature followed by a further incubation in fresh SDS equilibration buffer containing 2.5% (w/v) iodoacetamide for 15 min at room temperature. Strips were then applied to 12.5% (w/v) SDS-PAGE gels and run at 5mA per gel for 1hr, 8mA per gel for an additional hour and then at 20 W/gel until completion using an Ettan DALT-6 separation unit (GE Healthcare). Gels to be used for spot picking and mass spectrometry were fixed for an hour in 50% methanol, 10% acetic acid and stained overnight using Sypro Ruby total protein stain (Invitrogen). Following de-staining in 10% methanol, 7% acetic acid, the gels were imaged using a Typhoon 9400 Variable Mode Imager (GE Healthcare). Duplicate gels were also transferred to PVDF membrane for immunodetection back at the University of Exeter (using western blotting protocol, Section 2.3.3).
Protein spots visible on both immunodetected blots and Sypro Ruby stained blots (i.e. immunoreactive proteins) were then identified for analysis by mass spectrometry (MS) back at the University of Bristol Proteomics Facility.

2.2 Numerical methods

2.2.1 Multiple sequence alignment

Sequences were aligned using either the alignment program ClustalX (Thompson et al., 1997) or the alignment editor SEAVIEW version 2.4 (Galtier et al., 1996) implementing the ClustalW algorithm (Thompson et al., 1994), with manual corrections by eye. See individual results chapters for further detail.

2.2.2 Nucleotide substitution model selection

Substitution models were selected by a series of nested hierarchical likelihood-ratio tests using in the program MODELTEST Version 3.06, (Posada and Crandall, 1998). The hierarchical likelihood ratio tests (hLRTs) assess 56 different substitution models, performing a series of nested likelihood ratio tests, comparing maximized log-likelihoods of models in a pairwise fashion (Posada and Buckley, 2004). Models are referred to as ‘nested’ when they form related versions of each other, for example the Jukes-Cantor model (Jukes and Cantor, 1969) can be considered as nested within the Kimura two-parameter model (Kimura, 1980), because if rates of transitions and transversions are equal the Kimura two-parameter model collapses to the Jukes-Cantor model (Posada and Buckley, 2004). Appropriate parameters were then applied during phylogenetic analyses.

2.2.3 PAUP*

Maximum parsimony (MP) and/or maximum likelihood (ML) criterion were implemented using PAUP* version 4.0 (Swofford, 1998). All MP phylogenetic analyses were carried out using an equal weighting scheme, with gaps being treated as missing data, tree bisection and reconnection (TBR) swapping,
random sequence addition, and 1000 replicates per search. All ML phylogenetic analyses were carried out with heuristic searching (for 100,000 rearrangements), TBR branch swapping, and random sequence addition with 10 repeats.

Node support for all trees were assessed using bootstrap analysis, using 100 replicates for all ML analyses (due to computational time limits), and 1000 replicates for all MP analyses. Additionally, Bremer decay indices (Bremer, 1988) were calculated on all MP inferred trees.

2.2.4 MrBayes

All Bayesian phylogenetic inferences were carried out using the program MrBayes 3.1 (Huelsenbeck and Ronquist, 2001), running two independent, Metropolis Coupled Markov Chain Monte Carlo (MCMC or MC$^3$) analyses starting from different random trees (nruns=2) (Ronquist and Huelsenbeck, 2003). Each MCMC run contained four chains (mcmc nchains=4), consisting of three heated chains (using the default temperature of heated chains, temp=0.2), and one cold chain, and using a sampling frequency of 10 generations, and selected models and priors applied. The convergence diagnostic (standard deviation of split frequencies) was used to assess when to stop analyses. When the two runs converge onto the stationary phase the average standard deviation of split frequencies will approach zero, indicating that the two samples of the posterior probability have become increasingly similar. Analyses were continued until the average standard deviation of split frequencies fell below the default threshold of 0.01 (stoprule=yes stopval=0.01), where sampling of the posterior distribution was deemed to be adequate (Ronquist et al., 2005). By default the convergence diagnostic is sampled every 1000 generations, and is based on the last 75% of the samples at each calculation (relburnin=yes burninfrac=0.25), for example at generation 10,000 it is calculated using the last 7500 generations. Consequently, a default burn-in of 0.25, corresponding to the first 25% of samples obtained up until convergence had been reached, was also applied to summarize substitution model parameters (sump) and trees and branch lengths (sumt). Plots of
generation versus log probability of the data (log likelihood values) produced by the sump command were then checked to ensure stationarity had been reached (i.e. plot shows no patterns in the data, resembles ‘white noise’) (Ronquist et al., 2005). Tree topology was then calculated from the remaining data after discarding burn-in samples by constructing a majority-rule consensus tree, where the posterior probability of a clade is represented by the percentage of samples that recovering that clade (Huelsenbeck and Ronquist, 2001).

All multi-locus analyses were partitioned according to individual gene data sets, unlinking model parameters and using variable rate parameters to allow all partitions to evolve under different rates.

2.2.5 RaxML

RaxML (Randomized a(x)ccelerated Maximum Likelihood) is a hill-climbing algorithm which optimizes the likelihood of an initial fast built parsimony-based tree in order to converge towards an optimum tree topology. Tree optimization involves removing all possible subtrees from the current best tree, and inserting them within neighbouring branches up to a specified distance of nodes away. If one of the topologies leads to a higher likelihood value, this new topology is adopted, and again all possible subtrees on the new best tree rearranged. After each rearrangement a global branch length optimization is also carried out on the top 20 topologies only. Tree optimization continues until topology converges to an optimum (Stamatakis et al., 2005).

2.2.6 PhyML

Like RaxML, PhyML (PHYlogenetic inferences using Maximum Likelihood) is a hill-climbing algorithm, starting with a random solution, in this case an initial tree built by distance methods, and iteratively making small changes to the both tree topology and branch lengths, to improve its likelihood until an optimum is found (Guindon and Gascuel, 2003).
2.2.7 Sliding windows analysis

Sliding window analyses were carried out using the program DnaSP version 4.50.3 (Rozas et al., 2003), to estimate the nucleotide diversity ($\pi$) across a multiple sequence alignment. During this analysis $\pi$ was calculated across a window (region of DNA), and this value assigned to the nucleotide at the midpoint of the window; the window is then moved along the alignment in a series of steps. Sites containing alignment gaps were not included in the length of the windows, ensuring all windows had the same number of nucleotides (Chapter 3).

2.3 Analytical methods

2.3.1 Mass spectrometry

Selected protein spots were cut from the gel using the Investigator ProPic Automated 2D spot picking robot and digested with trypsin using the ProGest automated digestion unit (both from Perkin Elmer Life Sciences (UK) Ltd., Beaconsfield, England). The resulting peptides were analysed by mass spectrometry using a 4700 MALDI-TOF/TOF mass spectrometer (Applied Biosystems, Foster City, CA) to give a peptide mass fingerprint and peptide sequence information, which was searched against various databases using the Mascot search program (www.matrixscience.com) to identify the protein present in the gel spot (Chapter 6).

2.3.2 Peptide de novo sequencing from MS/MS spectrum

Observed peptides were also fragmented using collision induced dissociation. De novo sequencing was performed on the resultant fragment spectra using GPS DeNovo Explorer v3.6 (Applied Biosystems). Denovo explorer was used with the following settings; trypsin was specified as the digestive enzyme, a mass tolerance of 0.2Da was used, methionine oxidation was treated as a variable modification and cysteines were assumed to be carbamido-methylated. Peptide sequence identities were ranked by the degree of fit between the actual
data and theoretical fragments of candidate peptide sequences and the peak intensity. The resulting sequence information was searched against various databases using the search programs BLASTP2 and MS-BLAST (http://blast.wustl.edu) (Chapter 6).

2.3.3 Electron microscopy gold labelling

Electron microscopy (EM) gold labelling was carried out on first instar larvae of *Lucilia sericata* and *Calliphora vicina* according to protocol by Thornton and Talbot (2006). Initial processing (embedding tissue in resin, cross section cutting and transfer to nickel grids), and use of TEM for visualization of gold were carried out by a trained electron microscopist at the Bioimaging Centre, University of Exeter.

The larvae were immersed in IEM fixative buffer, incubated at 23°C for 3 h, before being twice washed for 10 min in 0.1M phosphate buffer (pH 7.2). Samples were then dehydrated in a graded water:ethanol series (% vol/vol ethanol 30, 50, 75, 95, 100; 5 min immersion for each), before being transferred to LR White resin and incubated at 23°C for 16 h to allow infiltration of the resin. The samples were transferred to fresh resin held in gelatine capsules (filled to exclude air) and the resin polymerized by exposure to UV light for 48 h. Cross-sections approximately 70-90nm thick were cut along length of the embedded larvae using a diamond knife fitted to an ultramicrotome, and the sections transferred to Piloform-coated nickel grids. The grids were blocked by incubating in 50μl of sterile filtered PBST-BSA blocking buffer (1% w/v BSA, filtered through 0.2-μm sterile filter) for 10 min at room temperature with agitation, and washed in sterile filtered (0.2μm filter) PBST washing buffer (3 × 3 min). The grids were then incubated with polyclonal rabbit anti-sera (1:500 dilution in PBST-BSA blocking buffer) or PSBT-BSA blocking buffer (control) at 23°C for 1 h, and washed in PBST washing buffer (4 × 3 min). The grids were gold labelled by incubating in 50μl of secondary antibody-colloidal gold conjugate (1:20 dilution with PBST-BSA blocking buffer), washed PBST washing buffer (4 × 3min), and dried on Whatman filter paper at 23°C. Finally, the grids were incubated for 20 min in uranyl acetate, followed by lead citrate for
4 min, and returned to a 23ºC dessicator to dry before observation with a transmission electron microscope at 80 or 100 kV.
Chapter Three

Which mitochondrial gene (if any) is best for reconstructing insect phylogenetic relationships?
Which mitochondrial gene (if any) is best for reconstructing insect phylogenetic relationships?

In preparation for re-submission to the journal *Molecular Phylogenetics and Evolution*, with authors as follows:

L. McDonagh, L. Cibrario, J. Harrison and J. R. Stevens

School of Biosciences; University of Exeter; Prince of Wales Road, Exeter, Devon, EX4 4PS.

L. McDonagh undertook all Bayesian phylogenetic analyses, statistical analysis, mitochondrial genome data mining for sliding windows analysis, sliding windows analyses, all tables and figures included, and all drafts of the re-written manuscript.

L. Cibrario undertook original mitochondrial genome selection for phylogenetic analysis, multiple sequence alignment and substitution model selection for phylogenetic analyses, all PhyML and RaxML phylogenetic analyses, and drafts of the previously submitted manuscript.

J. Harrison checked available sequence data.

J. R. Stevens provided the original concept of the paper, guidance throughout, and commented on both original and re-written versions of the manuscript.
3.1 Introduction

While DNA sequence-based analyses have in the past employed both nuclear and mitochondrial sequence data to explore the evolutionary origins and relationships within Insecta, mitochondrial DNA (mtDNA) in particular seems to have been favoured (Hillis et al., 1996; Cameron et al., 2006; Shao and Barker, 2006). The advantages of utilizing mtDNA sequence data include; a lack of recombination during cell division, high copy number, relative ease of isolation, availability of universal primers, the presence of both conserved and variable regions, and the expectation of mtDNA to reach reciprocal monophyly before nuclear genes due to generally higher rates of evolution (Avise et al., 1979; Lunt et al., 1996; Monteiro and Pierce, 2001; Funk and Omland, 2003; Dowton, 2004; Lin and Danforth, 2004).

Insecta currently represent over half of all described organisms on earth, with over a million catalogued species (Chapman, 2006) and estimates of as high as 30 million undescribed species (Erwin, 1982). Insect mitochondrial genomes are approximately 16kb long and generally contain 13 protein-coding genes (COX1-3, CYTB, ATP6, ATP8, ND1-6 and ND4L), two ribosomal RNA (rRNA) genes, and generally 22 transfer RNA (tRNA) genes (Boore, 1999). Several different techniques have been employed to assess the phylogenetic value of mtDNA genomes for reconstructing relationships within Arthropoda. These include the analyses of gene rearrangements (Boore and Brown, 1998); number of genes (Rokas and Holland, 2000; Nardi et al., 2003); amino acid sequences (Nardi et al., 2001); purine-pyrimidine recoding of nucleotide sequences (Saitoh et al., 2006); and including all available genes, excluding the control region (Castro and Dowton, 2007).

The majority of mtDNA-based arthropod phylogenetic and population studies have frequently used only single protein coding gene analyses, with many mtDNA genes being largely disregarded and others receiving only sporadic attention (Caterino et al., 2000). However, limitations in the reliability of single-gene based analyses due, for example, to horizontal gene transfer or incomplete lineage sorting at the locus in question, are being increasingly
accepted; for examples see studies on *Drosophila* (Steinbachs *et al.*, 2001), *Protocalliphora* (Whitworth *et al.*, 2007) and Diptera (Stevens *et al.*, 2002; Stevens, 2003; Cameron *et al.*, 2007; Wells *et al.*, 2007). Cytochrome oxidase subunit 1 (*COX1*), in particular, has been among the most commonly used loci within insect系统atics (Zhang and Hewitt, 1997a), with its size and structure being among the most preserved among many species of Insecta (Lessinger *et al.*, 2000). More recently, DNA ‘barcoding’ initiatives (Hebert *et al.*, 2003a; Hebert *et al.*, 2003b) have advocated the high phylogenetic potential of a region within *COX1*. While there are obvious advantages of an international DNA barcoding programme, such as large scale standardization of sequencing efforts, debate concerning the reliability of single-gene analyses has also been raised. However, many concerns often seem to result from confusion over the use of barcoding to resolve evolutionary relationships, rather than its intended use as a tool for high through-put species assignment, often utilizing information gained from more thorough phylogenetic analyses (Moritz and Cicero, 2004).

The assessment of the phylogenetic utility of mtDNA genes has been received much attention in recent years, with for example Steinbachs *et al.* (2001) finding *ND4* to be the most effective gene at recovering known relationships within the genus *Drosophila*. Similarly, work by Cameron *et al.* (2004; 2006; 2007) has focused on using mtDNA genome sequence data to recover arthropod relationships. The authors found that phylogenetic analyses were greatly influenced by factors such as taxonomic sampling, choice of outgroup, and gene selection; the overall conclusion of this study being there were insufficient mitochondrial genome data available, at the time, to resolve basal arthropod relationships.

Here we present two approaches for assessing the utility of individual mtDNA genes to reconstruct evolutionary relationships within Insecta. The first approach utilizes a series of phylogenetic reconstructions, using a collection of 63 mtDNA genomes (55 Insecta, 7 outgroup), that had been systematically selected from all arthropod mtDNA genomes currently available. Phylogenetic trees were inferred for each individual mtDNA gene (12 protein coding and 2 rRNA), as well as a combined data set, and the ability of each to recover
monophyly of some of the major orders within Insecta tested. For the purpose
of this study nine insect orders were chosen (Archaeognatha, Coleoptera,
Diptera, Hemiptera, Hymenoptera, Lepidoptera, Orthoptera, Phthiraptera, and
Thysanura), together with three sister-lineage outgroups (Collembola,
Crustacea and Myriapoda). Additionally, three phylogenetic programs,
representing perhaps the fastest and most precise examples of Bayesian and
maximum likelihood (ML) tree building methods using models of nucleotide
substitution (Stamatakis et al., 2005), were used to analyse each data set.

The second approach aimed to quantify the level of DNA polymorphism
across each mtDNA gene, allowing unequal rates of nucleotide substitutions to
be detected, and related back to phylogenetic performance. In contrast, this
approach expanded the sample size to include sequence data from all currently
available mtDNA genomes from each of the nine Insecta orders, and three
outgroup lineages.

In summary, the work presented here seeks to build upon previous
studies (Steinbachs et al., 2001; Cameron et al., 2004; Cameron et al., 2006;
Cameron et al., 2007), representing perhaps the broadest attempt, to date, to
establish which mtDNA genes are able to convincingly retrieve currently
recognised major taxonomic orders within Insecta.
3.2 Materials and Methods

3.2.1 Phylogenetic reconstruction using selected mtDNA genomes

3.2.1.1 mtDNA genome selection

Nucleotide and amino acid sequence data for complete mtDNA genomes from 62 arthropod taxa were collected. Of these, 60 were extracted from the NCBI database, while the genomes of the blowflies (Calliphoridae), *Lucilia sericata* (AJ422212) and *Chrysomya megacephala* (AJ42604) are newly published (Stevens et al., 2008). Genomes from two Collembola species, three Myriapoda and two Crustacea genomes were also included in the analyses as outgroups. In an attempt to minimise potential phylogenetic artefacts, such as long-branch attraction, outgroups selected from Crustacea and Myriapoda were chosen with regard to the on-going debate concerning the sister-group status of these two subphyla with respect to Hexapoda (Zrzavy and Stys, 1997; Edgecombe et al., 2000). Accession numbers of genomes analysed are given in Fig. 3.

3.2.1.2 Multiple sequence alignment and substitution model selection

Preliminary alignments for the 13 mitochondrial protein-coding genes (nucleotide and amino acid) and the two rRNA genes (nucleotide only) were created using the sequence alignment editor Se-Al (Rambaut, 1996), with manual correction by eye. In an attempt to maximise the phylogenetic signal-to-noise ratio, regions containing large numbers of gaps or which could not be aligned with confidence were removed (masked) prior to phylogenetic analyses. Due to comparatively poor ATP8 alignments, a result of the short coding region of this gene, it was removed from all further phylogenetic analyses.

Two concatenated gene alignments, containing all remaining genes (14 nucleotide, protein-coding and two rRNA genes; and 12 amino acid, protein coding genes only), were created by joining individual gene alignments together, again using Se-Al. For each of the individual gene alignments substitution
models were selected and used for all subsequent phylogenetic reconstructions using the program MODELGENERATOR, version 0.83 (Keane et al., 2006).

<table>
<thead>
<tr>
<th>Gene</th>
<th>Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATP6</td>
<td>GTR + I + Γ</td>
</tr>
<tr>
<td>ATP8</td>
<td>GTR + Γ</td>
</tr>
<tr>
<td>COX1</td>
<td>GTR + I + Γ</td>
</tr>
<tr>
<td>COX2</td>
<td>GTR + I + Γ</td>
</tr>
<tr>
<td>COX3</td>
<td>GTR + I + Γ</td>
</tr>
<tr>
<td>CYTB</td>
<td>GTR + I + Γ</td>
</tr>
<tr>
<td>ND1</td>
<td>GTR + I + Γ</td>
</tr>
<tr>
<td>ND2</td>
<td>GTR + I + Γ</td>
</tr>
<tr>
<td>ND3</td>
<td>GTR + Γ</td>
</tr>
<tr>
<td>ND4</td>
<td>GTR + I + Γ</td>
</tr>
<tr>
<td>ND4L</td>
<td>GTR + Γ</td>
</tr>
<tr>
<td>ND5</td>
<td>GTR + I + Γ</td>
</tr>
<tr>
<td>ND6</td>
<td>GTR + I + Γ</td>
</tr>
<tr>
<td>SSU</td>
<td>GTR + I + Γ</td>
</tr>
<tr>
<td>LSU</td>
<td>GTR + I + Γ</td>
</tr>
</tbody>
</table>

Table 2: Details of best-fit substitution models selected by MODELGENERATOR for each gene data set, including invariable sites (+I) and rate variation among sites (+Γ).

3.2.1.3 Phylogenetic analysis

Phylogenetic trees were reconstructed for each of the individual gene alignments and the concatenated multi-gene alignment. Three different phylogenetic inference programs were employed; two ML methods, PhyML (Guindon and Gascuel, 2003) and RaxML (Stamatakis et al., 2005; Stamatakis, 2006), and the Bayesian program MrBayes 3.1 (Huelsenbeck and Ronquist, 2001). PhyML and RaxML both represent fast hill-climbing ML tree building programs, searching for the best possible tree topology by optimizing the likelihood of an initial starting tree already comprising all sequences (Stamatakis et al., 2005). By comparison, MrBayes, a program for Bayesian phylogenetic inference, uses a Markov Chain Monte Carlo (MCMC) method to exploring tree
space in a stepwise fashion and optimising parameters in an iterative manner (Ronquist et al. 2005).

All PhyML analyses implemented a heuristic ML search using an initial distance-based starting tree (BIONJ), using default settings (optimised tree topology, optimised branch lengths and rate parameters, 4 rate substitution categories), applying selected substitution models and parameters (Table 2), all free parameters were estimated by PhyML, and calculating bootstrap support using 1000 repetitions.

RaxML analyses also used a heuristic ML search, with an initial parsimony-based starting tree, selected substitution models and parameters were implemented (Table 2), all free parameters were estimated by RaxML, 4 rate substitution rate categories were used, and bootstrap support calculated using 1000 repetitions.

Bayesian analyses ran two independent Metropolis Coupled MCMC (MCMCMC) searches starting from different random trees (nruns=2), each search contained three heated chains (using the default heating temperature, temp=0.2), and one cold chain, with a sampling frequency of 10 generations, and selected models and priors applied. Analyses carried out using the combined data set was partitioned according to individual gene data sets, allocating unlinked model parameters and variable rate parameters to allow all partitions to evolve under different rates. MCMCMC runs were continued until the convergence diagnostic (standard deviation of split frequencies) fell below the default threshold (stopval=0.01), indicating sufficient convergence between the two samples of the posterior probability (Ronquist et al., 2005). The default convergence diagnostic burn-in fraction of 0.25 was used (burninfrac=0.25), consequently, a corresponding burn-in of 0.25, corresponding to the first 25% of samples obtained up until convergence had been reached, was also applied to summarize substitution model parameters (sump) and trees and branch lengths (sumt). Plots of generation versus log probability of the data (log likelihood values) produced by the sump command were also checked to ensure stationarity had been reached (i.e. plot shows no patterns in the data,
resembles ‘white noise’) (Ronquist et al., 2005). Tree topology was then calculated from the remaining data after discarding burn-in samples by constructing a majority-rule consensus tree, where the posterior probability of a clade is represented by the percentage of samples that recovering that clade (Huelsenbeck and Ronquist, 2001).

### 3.2.1.4 Quantifying the phylogenetic utility of each mtDNA gene

A lack of agreement in evolutionary relationships within Insecta has meant that a ‘true’ insect phylogeny remains elusive. Therefore, in this study the phylogenetic accuracy of each individual gene tree topology was assessed by their ability to recover monophyly of each of the 13 selected groups. Testing the monophyly of currently recognised taxonomic units in this way will not only allow comparisons between the phylogenetic utility of different mtDNA genes to be made, but will also enable the precision of current taxonomic nomenclature itself to be assessed.

Where a monophyletic grouping was recovered, clade confidence measures were recorded, as either bootstrap partitions (PhyML and RaxML) or as Bayesian posterior probabilities (MrBayes).

### 3.2.2 Quantifying nucleotide diversity using all available mtDNA genomes

#### 3.2.2.1 mtDNA genome sequences and alignment

Nucleotide sequence data from all available mtDNA genomes were collected for each of the 13 arthropod groups, giving a total of 153 mitochondrial genomes, see Table 2.

The single-gene sequence data sets (13 protein coding and two rRNA genes) were aligned separately within each taxonomic group, using the alignment editor SEAVIEW, version 2.4 (Galtier et al., 1996), implementing the CLUSTALW algorithm (Thompson et al., 1994), and alignments checked by eye.
Aligning all genes within each group independently in this manner, along with the larger taxon sample size, meant that alignments of *ATP8* sequence data were possible, and so this gene was included for this part of the study.

Likewise, it was not necessary to mask these alignments, as partitioning the data into taxonomically related groups (e.g. within an order) resulted in greater sequence similarity and fewer, if any, regions of ‘messy’ alignment.

<table>
<thead>
<tr>
<th>Insecta</th>
<th>Number of described species</th>
<th>Number of mitochondrial genomes available</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt; million $^{(1)}$</td>
<td>102</td>
<td></td>
</tr>
<tr>
<td><strong>Orders within Insecta</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coleoptera</td>
<td>350,000 $^{(2)}$</td>
<td>13</td>
</tr>
<tr>
<td>Diptera</td>
<td>150,000 $^{(3)}$</td>
<td>24</td>
</tr>
<tr>
<td>Lepidoptera</td>
<td>&gt;137,000 $^{(3)}$</td>
<td>12</td>
</tr>
<tr>
<td>Hymenoptera</td>
<td>115,000 $^{(4)}$</td>
<td>5</td>
</tr>
<tr>
<td>Hemiptera</td>
<td>80,000 $^{(5)}$</td>
<td>11</td>
</tr>
<tr>
<td>Orthoptera</td>
<td>20,000 $^{(6)}$</td>
<td>15</td>
</tr>
<tr>
<td>Phthiraptera</td>
<td>3000 $^{(7)}$</td>
<td>3</td>
</tr>
<tr>
<td>Thysanura</td>
<td>370 $^{(8)}$</td>
<td>3</td>
</tr>
<tr>
<td>Archaeognatha</td>
<td>350 $^{(9)}$</td>
<td>3</td>
</tr>
<tr>
<td><strong>Outgroups</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Crustacea</td>
<td>30,000 $^{(10)}$</td>
<td>36</td>
</tr>
<tr>
<td>Myriapoda</td>
<td>10,500 $^{(11)}$</td>
<td>8</td>
</tr>
<tr>
<td>Collembola</td>
<td>6000 $^{(12)}$</td>
<td>7</td>
</tr>
</tbody>
</table>

Table 3: Number of described species in each taxonomic group, and total number of mitochondrial genomes available at the time of analysis; (1) Chapman (2006); (2) Tree of Life; (3) The Global Lepidoptera Names Index; (4) HymAToL; (5) ‘Discover Life’, University of Georgia; (6) Myers (2001); (7) Ruppert & Barnes (1994); (8) National Center for Biotechnology Information.

### 3.2.2.2 Calculating nucleotide diversity by sliding window analysis

Nucleotide diversity ($\pi$) was calculated using a sliding window technique, whereby $\pi$ is calculated across a region of DNA, or ‘window’, with this value
being assigned to the nucleotide at the midpoint of the window; the window is then moved along the alignment in a series of steps. All sliding window analyses were carried out using the program DnaSP version 4.50.3 (Rozas et al., 2003).

Concatenated gene sequence files were first created within DnaSP for each of the 13 arthropod groups, completed by adding single-gene alignments together in the same order observed in the majority of mitochondrial genomes studied to ensure consistency in gene order, as a small number of taxa demonstrated gene rearrangement.

Nucleotide diversity was calculated within each group for each individual gene (13 protein coding, 2 rRNA), and across the whole mtDNA genome, using a window of 25 base pairs (bp) and step size of 5 bp. Additionally, a sliding window of 300 bp and step size of 25 bp was also applied to the mtDNA genome alignments, so as to take into account the larger overall length of the concatenated alignment. This larger sliding window scale also made it possible to compare the average level of mtDNA genome nucleotide diversity between different taxonomic groups to be made.

In all analyses sites containing alignment gaps were not included in the length of the windows, ensuring all windows have the same number of nucleotides.
Chapter 3

3.3 Results

3.3.1 Phylogenetic Reconstruction

3.3.1.1 Nucleotide versus amino acid sequence data

Slowly evolving characters are often favoured over faster evolving characters, particularly when inferring deep level phylogenetic relationships, and this has led to the preferential use of amino acid data over nucleotide data. Amino acids have also been suggested to have an advantage over nucleotide data due to their higher number of character states, with 20 amino acid states (based on the universal genetic code) and only four nucleotide states, which could make convergence less likely in amino acid data. However, amino acids represent composite characters, as they essentially involve combining three characters (nucleotides) together into a single unit (Simmons and Freudenstein, 2002). Composite coding results in two main problems; firstly there is a loss of hierarchic information when unordered states are used, and secondly, since the genetic code is degenerate, amino acids are likely to be subject to homoplasy, for example when different codons that assign the same amino acid have evolved independently in different lineages (Simmons and Freudenstein, 2002; Simmons et al., 2002a, 2002b; Cameron et al., 2004). Furthermore, while several different mitochondrial genetic codes are currently recognised, the assumptions made by the choice of genetic code may be broken by small deviations from the expected code by individual taxa, leading to inaccuracies translated amino acid sequence (Cameron et al., 2004).

In the study presented here, similar topology was recovered from phylogenetic reconstructions using both nucleotide and amino acid sequence data, however, the nucleotide sequence data was found to recover much higher levels of clade support. Using nucleotide sequence data additionally allowed the inclusion of the two rRNA genes, which, not being protein coding, cannot be translated into amino acid sequence data. Consequently, taking into account both the arguments given above and the greater clade support given by the nucleotide
data in this study, all subsequent discussions concentrate exclusively on the nucleotide sequence phylogenies.

3.3.1.2 Measuring clade support

Where monophyletic groupings were recovered, clade support was calculated by either bootstrap partitions (BP) or Bayesian posterior probabilities (PP), see Table 3. However, it should be kept in mind that these are not equivalent measures of clade support (Alfaro et al., 2003). Extensive investigation into comparisons between these two measures has suggested that PP are a more precise estimator of phylogenetic accuracy, with a PP of 0.7 corresponding more closely to an accuracy of 70%, whereas a BP of 70% would, by comparison, correspond to an accuracy of around 95% (Hillis and Bull, 1993; Wilcox et al., 2002). For example; a threshold of over 75% BP may used for high confidence in groupings, whereas a PP of over 0.95 would be required for an equivalent acceptable support of monophyly. Consequently, all values below 50% BP or 90% PP have been removed from this study as insufficient evidence of successful monophyly retrieval.
<table>
<thead>
<tr>
<th>PhyML</th>
<th>COX1</th>
<th>COX2</th>
<th>COX3</th>
<th>ND1</th>
<th>ND2</th>
<th>ND3</th>
<th>ND4</th>
<th>ND4L</th>
<th>ND5</th>
<th>ND6</th>
<th>CYTB</th>
<th>ATP6</th>
<th>LSU</th>
<th>SSU</th>
<th>Concatenated</th>
</tr>
</thead>
<tbody>
<tr>
<td>Insecta</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Diptera</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>97</td>
<td>-</td>
<td>99</td>
<td>-</td>
<td>99</td>
<td>-</td>
<td>81</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>100</td>
</tr>
<tr>
<td>Lepidoptera</td>
<td>100</td>
<td>90</td>
<td>100</td>
<td>97</td>
<td>99</td>
<td>97</td>
<td>100</td>
<td>97</td>
<td>100</td>
<td>93</td>
<td>100</td>
<td>75</td>
<td>100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hemiptera</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Hymenoptera</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Coleoptera</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>98</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>87</td>
</tr>
<tr>
<td>Pthiraptera</td>
<td>100</td>
<td>-</td>
<td>98</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>78</td>
<td>-</td>
<td>76</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>100</td>
</tr>
<tr>
<td>Orthoptera</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Thysanura</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Archaeognatha</td>
<td>99</td>
<td>77</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>79</td>
<td>-</td>
<td>-</td>
<td>77</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>100</td>
</tr>
<tr>
<td>Myriapoda</td>
<td>98</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Collembola</td>
<td>100</td>
<td>95</td>
<td>-</td>
<td>-</td>
<td>95</td>
<td>-</td>
<td>100</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>100</td>
</tr>
<tr>
<td>Crustacea</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>RaxML</th>
<th>COX1</th>
<th>COX2</th>
<th>COX3</th>
<th>ND1</th>
<th>ND2</th>
<th>ND3</th>
<th>ND4</th>
<th>ND4L</th>
<th>ND5</th>
<th>ND6</th>
<th>CYTB</th>
<th>ATP6</th>
<th>LSU</th>
<th>SSU</th>
<th>Concatenated</th>
</tr>
</thead>
<tbody>
<tr>
<td>Insecta</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Diptera</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>95</td>
<td>-</td>
<td>92</td>
<td>-</td>
<td>79</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>100</td>
</tr>
<tr>
<td>Lepidoptera</td>
<td>100</td>
<td>-</td>
<td>100</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>99</td>
<td>-</td>
<td>100</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Hemiptera</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Hymenoptera</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Coleoptera</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>91</td>
<td>-</td>
<td>95</td>
<td>-</td>
<td>-</td>
<td>100</td>
</tr>
<tr>
<td>Pthiraptera</td>
<td>100</td>
<td>-</td>
<td>98</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>83</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>99</td>
</tr>
<tr>
<td>Orthoptera</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Thysanura</td>
<td>83</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>78</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Archaeognatha</td>
<td>99</td>
<td>85</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>79</td>
<td>-</td>
<td>83</td>
<td>-</td>
<td>83</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>99</td>
</tr>
<tr>
<td>Myriapoda</td>
<td>99</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Collembola</td>
<td>99</td>
<td>97</td>
<td>100</td>
<td>94</td>
<td>100</td>
<td>94</td>
<td>100</td>
<td>77</td>
<td>82</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>100</td>
</tr>
<tr>
<td>Crustacea</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>75</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 4: Bootstrap partition (BP) values for PhyML and RaxML phylogenetic analysis of both single gene data sets and the concatenated data set (containing all 14 gene data sets), only cases where arthropod groupings were recovered with a BP of over 75% are shown.
<table>
<thead>
<tr>
<th>MrBayes</th>
<th>COX1</th>
<th>COX2</th>
<th>COX3</th>
<th>ND1</th>
<th>ND2</th>
<th>ND3</th>
<th>ND4</th>
<th>ND4L</th>
<th>ND5</th>
<th>ND6</th>
<th>CYTB</th>
<th>ATP6</th>
<th>LSU</th>
<th>SSU</th>
<th>Concatenated</th>
</tr>
</thead>
<tbody>
<tr>
<td>Insecta</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1</td>
</tr>
<tr>
<td>Diptera</td>
<td>-</td>
<td>1</td>
<td>-</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td>1</td>
<td>-</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1</td>
</tr>
<tr>
<td>Lepidoptera</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td>1</td>
<td>-</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>1</td>
</tr>
<tr>
<td>Hemiptera</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Hymenoptera</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Coleoptera</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>-</td>
<td>1</td>
<td>-</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.98</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Pthiraptera</td>
<td>1</td>
<td>-</td>
<td>1</td>
<td>-</td>
<td>0.96</td>
<td>-</td>
<td>1</td>
<td>-</td>
<td>1</td>
<td>0.99</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>1</td>
</tr>
<tr>
<td>Orthoptera</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Thysanura</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Archaeognatha</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.98</td>
<td>-</td>
<td>1</td>
<td>0.96</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Myriapoda</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1</td>
</tr>
<tr>
<td>Collembola</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>-</td>
<td>0.96</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Crustacea</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>0.98</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.96</td>
<td>-</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 5: Posterior probability (PP) values for MrBayes phylogenetic analysis of both single gene data sets and the concatenated data set (containing all 14 gene data sets), only cases where arthropod groupings were recovered with a PP of over 0.95 are shown.
3.3.1.3 Single-gene analyses

Single-gene phylogenies produced by both of the fast ML methods (PhyML and RaxML) and the Bayesian analyses (MrBayes) were found to agree on support of monophyly for several insect orders. For example, all three methods recovered high levels of clade support (i.e. BP over 75%; PP over 0.95) for the monophyly of Archaeognatha, Diptera, Lepidoptera, and Phthiraptera, by three or more genes. Coleoptera was also recovered with high support for at least three genes by MrBayes and RaxML, but only one gene (ND5) by PhyML. However, three of the insect orders (Hemiptera, Hymenoptera, and Orthoptera) along with the class Insecta itself, were found to be paraphyletic in all single-gene and concatenated gene analyses.

Of the outgroup lineages, Collembola was found to be monophyletic by multiple genes for all three phylogenetic methods. Conversely, Crustacea was retrieved by just ND2 (RaxML and MrBayes) and COX1 (MrBayes only), and Myriapoda recovered by only COX1 (all three methods).

While no gene appears universally best suited for recovering monophyly, the MrBayes analysis of COX1 did successfully recover as many groups as the concatenated gene analysis (Fig. 4). Of the remaining genes; ND2, ND4 and ND5 also appeared to perform well, with ND4 performing equally as well as COX1 under PhyML and RaxML, and ND2 recovering the same number of groups under RaxML.
Figure 3: Phylogenetic relationships within Insecta (in group), and representatives of Myriapoda, Collembolla, and Crustacea (outgroups), based on a partitioned Bayesian analysis of the concatenated mtDNA nucleotide data set, using 14 partitions (12 protein-coding genes, and two rRNA genes). All branches supported with a PP ≥ 0.95 are labelled.
Figure 4: Number of taxonomic groups recovered as monophyletic (with clade support of BP > 75% or PP > 0.95) for all single-gene and concatenated phylogenetic reconstructions using PhyML, RaxML and MrBayes.
3.3.1.4 Concatenated-gene analyses

As might be expected, all trees inferred from the concatenated gene data set retrieved almost all groups as monophyletic, and with very high support. All three phylogenetic methods recovered Archaeognatha, Coleoptera, Diptera, Lepidoptera, Phthiraptera, and Collembola, suggesting high phylogenetic corroboration between individual gene data sets for these groups. Additionally, Crustacea and Myriapoda were recovered by MrBayes only, each with a PP of 1.0. However, none of the three methods managed to recover monophyly of Hemiptera, Hymenoptera, Orthoptera, Thysanura, or Insecta from the concatenated data set, although there was some support for Thysanura in several single-gene trees.

3.3.1.5 Comparing phylogenetic methods

While clade support measures within each single-gene analysis generally agreed between the three different phylogenetic methods used, some exceptions were also apparent. One notable example being that of Lepidoptera, which appears well supported by almost all single-gene analyses using PhyML and MrBayes, but is only recovered by three genes using RaxML, albeit always with very high BP support. Interestingly, MrBayes, a method reported to have an increased sensitivity to phylogenetic signal (Alfaro et al., 2003), does seem to perform better than the other two methods for the majority of the data sets tested here, recovering more monophyletic groupings than either of the other two methods in over half of the single-gene trees, as well as the concatenated tree (Fig. 4). RaxML appeared perform the poorest, recovering the same number of groups as PhyML for six genes, but fewer than PhyML for another six genes, only outperforming PhyML in the ND2 and ATP6 gene trees.

A recent study comparing a wide range of commonly used phylogenetic programs similarly found that MrBayes outperformed all other programs, both in terms of speed and tree quality (Williams and Moret, 2003). However, neither PhyML nor RaxML were tested in this study, and subsequent claims have been made by the creators of both programs that they are each able to outperform
MrBayes (Guindon and Gascuel, 2003; Stamatakis et al., 2005; Stamatakis, 2006).

However, while the data presented here found that MrBayes outperformed both PhyML and RaxML in terms of detecting phylogenetic content in sequence data, all Bayesian analyses were considerably slower to complete.

### 3.3.2 Nucleotide diversity

#### 3.3.2.1 Single-gene levels of nucleotide diversity

Sliding widow analyses for each of the mtDNA genes (13 protein coding, including ATP8; and 2 rRNA) revealed that nucleotide diversity not only varied between different genes, but also in some examples, considerably within the same gene (Fig. 5 and 7). The highest average nucleotide diversity was found in ND1 and ND2, while COX1 and the small subunit rRNA (SSU rRNA) displayed the lowest. Furthermore, regions of extreme high and low diversity were also observed within several of the genes, for example ND5 and COX1 (Fig. 5), and this was in turn reflected in larger standard deviations of $\pi$ (Fig. 7).
Figure 5: Sliding window analyses of the mtDNA genome alignments for Hemiptera, Lepidoptera, Diptera, and Insecta. Nucleotide diversity ($\pi$) values calculated for windows of 300bp, and with a step size of 25. Gene boundaries indicated.
Figure 6: Average nucleotide diversity for the mitochondrial genomes or each taxonomic group. Error bars depict standard deviation in diversity values. Number of mitochondrial genomes included in each group indicated.
Figure 7: Average nucleotide diversity for each mtDNA gene and concatenated-gene alignments. Error bars depict standard deviation in diversity values.
Interestingly, the section of COX1 relating approximately to the region designated for DNA barcoding displays lower nucleotide diversity than other areas within the gene. However, when tested no significant correlation was found between the number of groups recovered by each gene with high support (BP > 75%; PP > 0.95), and nucleotide diversity (Pearson correlation analysis; PhyML $r = -0.35$, df = 13, $P > 0.05$; RaxML $r = -0.15$, df = 13, $P > 0.05$; MrBayes $r = -0.15$, df = 13, $P > 0.05$).

3.3.2.2 Levels of nucleotide diversity across the mtDNA genome

Of the 13 arthropod groups studied, Lepidoptera displayed the lowest overall level of nucleotide diversity across the genome, and also displayed the smallest standard deviation of $\pi$ (Fig. 5 and 6). In comparison, a group of similar sample size, Hemiptera (n = 11), showed an average $\pi$ over twice that of Lepidoptera - the second highest among the groups studied - and also displayed one of the largest standard deviations. Insecta, the largest single grouping examined, could perhaps be expected to also show a high level of nucleotide diversity due to the inclusion of a large number of taxa (102) from several different orders. However, although the nucleotide diversity of Insecta is relatively high (with average of $\pi$ 0.31) compared to the other groupings assessed, Insecta does not seem very different from the much smaller sized outgroups, two of which (Crustacea and Myriapoda) being sub-phyla, may be expected to contain more distantly related taxa and therefore show a greater deal of DNA polymorphism. Hemiptera and Hymenoptera also both show relatively high average nucleotide diversity across the mtDNA genome, along with high standard deviations. When relating these relationships back to the concatenated data set phylogenetic reconstructions (Fig. 3) both Hemiptera and Hymenoptera also exhibit relatively long branches, despite efforts to select taxa to break up as many long branches as current taxonomic sampling effort would allow.

However, despite examples such as that of Lepidoptera (which exhibit a low average $\pi$ and apparent high phylogenetic signal) no significant correlation was found between nucleotide diversity of a taxonomic group and phylogenetic support, as measured by the number of genes recovering that group as
monophyletic (Pearson correlation analysis; PhyML $r = -0.50$, df = 13, $P > 0.05$; RaxML $r = -0.10$, df = 13, $P > 0.05$; MrBayes $r = -0.30$, df = 13, $P > 0.05$).

Lastly, nucleotide diversity across the genome was found to display a significant positive correlation with the standard deviation in $\pi$ (Pearson correlation analysis, $r = 0.97$, df = 11, $P < 0.05$). This correlation implies that groups with higher mean nucleotide diversity between taxa across the concatenated genes also exhibit more regions of extreme high and low $\pi$ values. One possible cause of this could be due to inherently different rates of evolution within genes, and perhaps indicating the need for improved taxonomic sampling balance to prevent long branch attraction.

### 3.3.3 The importance of gene length

While level of nucleotide diversity does not seem to reflect phylogenetic signal, a factor that does appear to be a strong indicator of the utility of a gene to recovering monophyletic groupings within Insecta is gene length. The four of the longest genes, (COX1, ND5, ND4 and LSU) performed particularly well (with the exception of the Bayesian analysis of ND5) retrieving considerably more groups, and producing higher clade support for those groups. Similarly, ATP8, ND4L, and ND3 are among the shortest genes in the mtDNA genome, and appear universally poor in the phylogenetic reconstructions. When compared, a significant positive correlation was found between gene length (as given by alignments used in sliding window analyses) and the number of taxonomic groups recovered as monophyletic, with clade support over 75% BP and 0.95 PP (Pearson correlation analysis; PhyML $r = 0.7$, df = 12, $P < 0.05$; RaxML $r = 0.73$, df = 12, $P < 0.05$; MrBayes $r = 0.81$, df = 12, $P < 0.05$; Fig. 8).

Several mtDNA based studies have specifically addressed the issue of gene length in determining phylogenetic utility. Among these, Cummings (1995), and Zardoya and Meyer (1996) identified gene length as a significant factor in ensuring the recovery a known phylogeny, suggesting that longer genes performed better. Although Steinbachs et al. (2001), reported only a weak correlation between the ability of a particular gene to recover a known genealogy and gene length in *Drosophila*. 
Figure 8: Correlation between gene length and the number of taxonomic groups recovered as monophyletic (with clade support of BP > 50% or PP > 0.95), for PhyML, RaxML, and MrBayes. (see text for Pearson correlation values)
3.4 Discussion

3.4.1 Should mtDNA markers be used to resolve deep level relationships?

With molecular clock estimates dating the origin of Insecta at around 420-430 MYA (Gaunt and Miles, 2002), we first address the argument that mitochondrial sequence data in general evolve too quickly to be of use resolving deep nodes (Shao and Barker, 2006; Whitfield and Kjer, 2008), with claims that mitochondrial genes are only suitable for resolving relationships younger than 5-10 MYA (Lin and Danforth, 2004).

However, all comparisons of resolving power between nuclear and mtDNA sequence data to date have focused on only a few mitochondrial genes, and have not yet focused on phylogenetic utility at a mitogenomic scale. So while some argue that mtDNA is unsuitable for higher level phylogenetics, we have shown here that the mitochondrial genome is capable of recovering many major insect orders, and that given suitable markers, phylogenetic signal could even be retrieved by multi-gene analyses using just a few key genes.

3.4.1.1 Detecting phylogenetic ‘signal’

In this study we have attempted to measure the phylogenetic signal within mitochondrial genes by testing their ability to recover monophyly of recognised taxonomic groupings. However, the failure of any phylogenetic reconstructions (single-gene or concatenated) to retrieve monophyly of three orders within Insecta (Hemiptera, Hymenoptera and Orthoptera) may on first inspection raise concerns over the accuracy of their currently accepted, largely morphology based, taxonomy. For example, while phylogenetic reconstructions included only two Orthopteran and three Hymenopteran genomes from those available, perhaps suggesting ineffective taxon sample size, all 11 currently available Hemipteran genomes were included, and yet the group was still recovered as paraphyletic in all analyses. Nevertheless, 6 out of the 9 orders were successfully recovered, and so their recovery and clade support can be judged
as good measures of strength of phylogenetic signal within a gene. Furthermore, a lack of monophyly should not be accepted as proof of paraphyly, given the limited number of mitochondrial genomes currently available. Indeed, the current state of insect systematics has been criticised for a fundamental lack of co-ordination in choice of molecular marker (Caterino et al., 2000). By focusing on whole mitochondrial genomes, therefore immediately being limited to those available at the time, unavoidable taxonomic sampling bias within Insecta is to be expected.

All phylogenetic reconstructions were completed using three different tree building methods. Comparing different phylogenetic programs in this way has enabled the sensitivity of phylogenetic signal within individual genes to be examined. As expected, phylogenetic signal was the greatest within the concatenated-gene data set, where almost identical relationships are recovered regardless of optimality criterion used (PhyML, RaxML or MrBayes). However, single-genes, particularly those that performed less well, did demonstrate some inconsistencies between the different tree building methods used. RaxML and PhyML are both hill-climbing maximum likelihood methods, optimizing the likelihood of an initial fast built tree in order to converge towards an optimum tree topology. However, in this study RaxML appeared to perform the weaker than PhyML in terms of recovering monophyly of taxonomically recognised families. One of the main differences between PhyML and RaxML is the way the initial tree is built, with RaxML using a parsimony-based tree and PhyML and using a distance-based tree. While it is unclear from this study why PhyML appeared to recover more monophyletic groups than RaxML, these findings do highlight how sensitive molecular data can be to the use of different methods of phylogenetic inference, and in the case of PhyML and RaxML even very similar methods. Overall, MrBayes appeared the most sensitive method to picking up phylogenetic signal, agreeing with opinions that Bayesian PP support more correct monophyletic groupings than the fast ML methods (PhyML and RaxML), and also requiring fewer character data to recover high clade support (Alfaro et al 2003).
3.4.2 Which mitochondrial gene (if any) is best?

As might be expected, phylogenetic signal was maximised when all available mitochondrial genes were analysed together in the concatenated-gene data set. These findings agree with the recent conclusions of Fenn et al. (2008) who, when investigating the phylogenetic utility of mtDNA genomes in recovering relationships within Orthoptera, found that monophyly was most successfully recovered when all data were analysed simultaneously using nucleotide sequence data, and regardless of optimality criterion. Furthermore, this study also directly compared the protein coding and rRNA sequence data, finding that although the rRNA genes did contribute signal when included in the multi-gene analysis, the majority of phylogenetic content seemed to come from the protein coding regions of the mitochondrial genome (Fenn et al., 2008).

While the rapid advances in DNA sequencing techniques and computational power may one day lead to realistic phylogenetic inference on a mitogenomic scale, our findings suggest that concentrating on a few key mtDNA genes might actually be sufficient. For example, while it is apparent that single-gene phylogenies varied greatly in their ability to recover monophyly, four mtDNA genes in particular (ND2, ND4, ND5 and COX1) did appear to perform well, recovering between six and eight evolutionary relationships in agreement with established taxonomy. In particular, we have confirmed that COX1 is indeed a comparatively good phylogenetic marker, being one of the largest genes in the mitochondrial genome, thus potentially delivering a high degree of phylogenetic information (Lunt et al., 1996).

Nevertheless, the limitations in the reliability of single-gene analyses should not be forgotten. For example; while the Bayesian inference of COX1 overwhelmingly provides the best results from all single-gene analyses, it still fails to recover the monophyly of Diptera, a group well supported by multiple other single gene analyses. Indeed, introgressive hybridization associated with infections of the endosymbiotic bacteria Wolbachia, are estimated to affect between 15 and 75% of all insects (Werren et al., 1995; West et al., 1998; Jeyaprakash and Hoy, 2000; Werren and Windsor, 2000), and have been found
to interrupt monophyly of the diptera genus *Protocalliphora* within *COX1* sequence data (Whitworth *et al.*, 2007).

Our findings agree with many of the concerns raised by Meier *et al.* (2006) and Cameron *et al.* (2006), suggesting that single-gene based mtDNA phylogenies do not appear well suited to the task of a providing a universal taxonomic tool for the study of insects. Indeed, sequence length appears to be a strong indicator of ability to recover monophyletic groupings within Insecta, confirming that the addition of sequence character data greatly improves the degree of variation being sampled, and so provides greater resolving power (Hillis *et al.*, 1994; Russo *et al.*, 1996). Accordingly, we recommend that reconstructions of insect phylogenetic relationships only be made from multi-gene analyses, using as many mitochondrial genes as possible, particularly focusing on the four key genes *COX1, ND2, ND4* and *ND5*.

However, undeniably, current phylogenetic marker selection is often made on the availability of universal primers, rather than in consideration of suitability to hypotheses being tested. Therefore, the fundamental need for international large scale taxonomic sampling for other mitochondrial genes, akin to that already directed at *COX1*, is required if robust multi-gene analyses are to become universal.

### 3.4.2.1 Detecting ‘noise’ in the data

One of the most surprising finding of this study has been the apparent lack of relationship between estimates of DNA polymorphism, as quantified by measures of nucleotide diversity, and phylogenetic signal. While the rate of DNA polymorphism appears to exceed the capacity of such a technique to provide clues to phylogenetic utility, there does seem to be some evidence that it can still yield important information about the data set itself. For example; while the very low levels of recombination encountered by mitochondrial genomes mean that individual genes can be assumed to have same genealogical history (Hurst and Jiggins, 2005), we have shown that different mitochondrial genes, and regions within genes, exhibit different rates of DNA
polymorphism. High levels of nucleotide diversity may be detecting unequal rates of sequence evolution among lineages, or differences in rate of change among sites within genes, both of which can be accounted for when applying appropriate substitution models to the data during phylogenetic inference. The apparent lack of correlation of phylogenetic signal with phylogenetic signal simply could be due to the employment of suitable substitution models. However, clarification of this would require more exhaustive investigation than provided here, carrying out phylogenetic reconstructions under a series of different substitution models.
Chapter Four

Molecular systematics of blowflies (Diptera: Calliphoridae):
A multi-gene approach.
4.1 Introduction

4.1.1 Calliphoridae (Diptera: Oestroidea)

Calliphoridae (blowflies) are estimated to have originated over 100 million years ago (MYA) (Stevens, 2003). To date, the most comprehensive phylogenetic analyses of Calliphoridae have been based on morphological evidence, which have generally supported a monophyletic grouping (Rognes, 1991), with a few exceptions (Rognes, 1997; Pape and Arnaud, 2001). Accumulating molecular data also appear to support monophyly of Calliphoridae (Nirmala et al., 2001; Otranto and Stevens, 2002; Stevens, 2003). Relationships within Calliphoridae, on the other hand, continue to undergo revision (Stevens, 2003). Morphological evidence typically supports eight subfamily groupings within Calliphoridae; Calliphorinae, Chrysomyinae, Helicoboscinae, Luciliinae, Melanomyinae, Polliinae, Rhiniinae and Rhinophorinae (Rognes, 1991). Although, strong phylogenetic support has only been found for the monophyly of Calliphorinae, Luciliinae and Chrysomyinae (Wallman et al., 2005). Of these, Chrysomyinae has proved the most difficult to resolve. For example, phylogenies based on 28S ribosomal RNA (rRNA) and the mitochondrial genes cytochrome oxidase subunits 1 and 2 (COX1 and COX2) have shown considerable incongruence due to different rates of substitution between the nuclear and mitochondrial genes, ultimately resulting in a failure to recover monophyly (Stevens, 2003).

Over 1000 species are currently recognised within Calliphoridae, comprising around 150 genera, (Rognes, 1991). It has been estimated that most blowfly sub-families had diverged by the time the supercontinent Gondwana broke up, approximately 100 MYA (Stevens, 2003). The earliest documented blowfly fossil is thought to belong to the extinct taxon Cretiformia fowleri, which existed in Canada in the Upper Cretaceous, 65 to 105 MYA (McAlpine, 1970). Molecular clock analysis using fossil dating has suggested that Calliphorinae, Chrysomyinae and Luciliinae shared a common ancestor as recently as 25 MYA (Wallman et al., 2005). However, examinations of the Cretiformia fowleri fossil have raised doubts over its calliphorid origin (Erzinclioglu, 1984; Wallman and Donnellan, 2001).
While morphological evidence has suggested a Calliphorinae-Chrysomyinae sister-grouping (Rognes, 1997), in contrast, molecular data instead supports a Calliphorinae-Luciliinae grouping, with respect to Chrysomyinae (Stevens, 2003; Wallman, 2005). In the absence of reliable fossil record evidence for Calliphoridae, molecular clock analyses carried out by Wallman et al. (2005) implemented an arthropod mtDNA clock developed by Brower (1994), which was calibrated using divergence times of taxa based on dated geological events inferred from biogeographical and paleoclimatological evidence (Brower, 1994). These calliphorid molecular clock studies have estimated that Chrysomyinae diverged from a Luciliinae-Calliphorinae grouping an estimated 21.7 MYA, with Luciliinae and Calliphorinae diverging approximately two million years later (Wallman et al., 2005). The majority of speciation within each genera is thought to have occurred between 1 and 15 MYA, with most groups displaying parasitic behaviour having originated between 5 and 15 MYA (Wallman et al., 2005; Stevens et al., 2006).

### 4.1.1.1 Auchmeromyiinae

**Auchmeromyia**

The five species that make up the genus *Auchmeromyia* represent the only known group of sanguivorous larvae within Calliphoridae to feed on mammals, sometimes including humans (Goddard, 2007). *A. luteola* (the Congo floor maggot), for example, is a well documented occasional parasite of humans in sub-Saharan Africa (Dutton et al., 1904), where eggs are laid in dwellings with sandy floors, allowing larvae to burrow, and emerge at night for a blood meal (Noireau, 1992).

**Cordylobia**

Within the genus *Cordylobia*, the most commonly reported species is *Cordylobia anthropophaga* (the tumbu fly), which is the prime causal agent of human myiasis in Africa (Tamir et al., 2003). *Cordylobia* also contains two less well documented species, *Co. ruandae* and *Co. rodhaini*, which cause myiasis
in various mammals, particularly rodents, and occasionally humans (Tamir et al., 2003). During cases of human infestation, eggs are usually laid on damp clothes, the eggs then hatch in response to warmth from the human body, and larvae burrow into the skin and causing furuncular myiasis (Tamir et al., 2003; Adeyinka and Mbanaso, 2004). Despite cases of human myiasis due to Co. anthropophaga being widely reported (Edungbola, 1982; Ogunniyi et al., 1994; Ugwu and Nwadiaro, 1999), increasingly from visitors returning from sub-Saharan Africa (Robert and Yelton, 2002; Fusco et al., 2005; Goksu et al., 2007), only a single phylogenetic study, by Stevens (2003), has included Co. anthropophaga.

4.1.1.2 Bengaliinae

Bengalia

A total of 41 species of Bengalia are currently recognised, based on morphological characters (Rognes, 2005). Adult Bengalia flies (Robineau-Desvoidy, 1830) are traditionally found in Afrotropical and Oriental regions, and are known to be aggressive feeders on termites, and ant pupae (Senoir White et al., 1940; Pont, 1980). However, little is known of Bengalia breeding habits and larval feeding behaviour (Rognes, 2005), although larvae of the closely related Verticia fasciventris (Bengaliinae) have been found developing in the heads of termite soldier (Sze et al., 2008).

The position of Bengalia with respect to Calliphoridae also remains ambiguous, with both a sub-family sister lineage to Auchmeromyiinae (Rognes, 2005), and a separate sister-family to Calliphoridae itself (Lehrer, 2003), having been proposed.

4.1.1.3 Calliphorinae

Calliphora
The majority of *Calliphora* species are found in Holarctic and Australasian regions (Zumpt, 1965), with many species appearing to be more cold adapted than other calliphorids (Faucherre *et al.*, 1999). For example, the distribution of *C. vicina* seems to be limited to areas where temperatures do not rise above 30°C for prolonged periods (Donovan *et al.*, 2006). While most Calliphora are oviparous (egg laying), observations have been made in New Zealand of *C. stygia* exhibiting oviparity during cool weather, and viviparity (larvae developing in body of the female) during hot weather months (Zumpt, 1965).

The majority of the species belonging to *Calliphora* are thought to be earthworm parasites (Wallman and Adams, 1997), although saprophagy and parasitism of vertebrates are also common. This genus is also morphologically very similar, making species differentiation sometimes difficult, particularly in egg and larval stages (Wallman and Donnellan, 2001).

One *Calliphora* species that has been particularly well studied is *C. vicina*, an important indicator species in forensically entomology in Europe and North America, and an occasional agent of myiasis in animals, and occasionally humans (Delhaes *et al.*, 2001). Possible intra-specific variation within *C. vicina* has been implied by observations of significant differences in diapause response to photoperiod and temperature between populations from England and Finland (McWatters and Saunders, 1996, 1998; Donovan *et al.*, 2006). However, phylogenetic analyses have failed to find evidence of genetic sub-structuring between specimens from different geographical regions (Stevens and Wall, 2001).

**Onesia**

Approximately 61 species belonging to the genus *Onesia* have currently been described, being largely found in Australasian/Oceanian and Palaeartic regions (Xue *et al.*, 2009). While many *Onesia* species are parasitoids of earthworms (Pape *et al.*, 2008), they are also often found feeding on carrion (Wallman and Donnellan, 2001). While *Onesia* has traditionally been granted genus status, several phylogenetic studies have found evidence classifying
Onesia as a sub-group within Calliphora (Wallman and Adams, 1997; Wallman and Donnellan, 2001).

4.1.1.4 Chrysomyinae

Chrysomya

Chrysomya include both saprophagic and secondary facultative ectoparasitic species (Baumgartner and Greenberg, 1984). In India, Ch. megacephala has been identified as one of the main blowfly species involved in the infestation of drying fish, a major economic issue in Asia, Africa and the Pacific (Wall et al., 2001).

Within Chrysomya, the taxonomic status of C. albiceps and C. rufifacies, in particular, remain ambiguous, having been described as synonyms (Ullerich, 1963; Zumpt, 1965; Kurahashi, 1989) and separate species (Holdaway, 1933; Tantawi and Greenberg, 1993). Traditionally, both C. albiceps and C. rufifacies have Old World ranges of distribution, with C. albiceps being predominantly found from northwest India down to southern Africa, C. rufifacies in India and on several islands in the Pacific, and both species being present in southern Pakistan (Senoir White et al., 1940) and Iran (Parchami-Araghi, 1995). More recently, both species have been introduced to South America (Baumgartner and Greenberg, 1984). These taxa have been found to successfully hybridize under laboratory conditions (Ullerich, 1963).

4.1.1.5 Luciliinae

Lucilia

Lucilia Robineau-Desvoidy is a relatively small blowfly genus, comprising around 27 morphological similar species, also known as ‘greenbottles’ (Aubertin, 1933; Stevens and Wall, 1996a). The majority of Lucilia are saprophagic, with a small number having become facultative ectoparasites causing cutaneous
myiasis on a number of mammals, both wild and domesticated (Hall and Wall, 1995).

**Lucilia bufonivora**

*Within Lucilia, L. bufonivora, referred to as L. elongata in North America (Zumpt 1965), is a specialist parasite of Anura (frogs and toads). Reports have been made of the high host specificity displayed by L. bufonivora, preferentially infesting the common, or European, toad Bufo bufo (Brumpt, 1934; Strijbosch, 1980).*

*Lucilia bufonivora* is thought to have diverged from its sister taxa, *L. silvarum*, very recently (Stevens and Wall, 1996a), with Townsend (1935) even grouping both species under the separate sub-genus ‘*BufoLucilia*’. While *L. silvarum* is largely considered a generalised saprophage, several reports of causing myiasis in anurans have been documented in North America (James and Maslin, 1947; Hall, 1948; Anderson and Bennett, 1963; Bleakney, 1963; Briggs, 1975; Roberts, 1998; Bolek and Coggins, 2002; Bolek and Janovy, 2004; Eaton et al., 2008). However, many of these reports are of second and third instar larvae, so whether the species identified is responsible for causing the primary myiasis, or are secondary parasites, is unclear (Bolek and Coggins 2002). Furthermore, early reports of anuran myiasis often neglected to distinguish between *L. bufonivora* and *L. silvarum* (Stradler, 1930; Hall, 1948).

Field studies of amphibians in the German region of North Rhine-Westphalia revealed myiasis infestation rates of up to 70%, causing significant mortality (Weddling and Kordg, 2008). Nevertheless, *L. bufonivora* remain poorly studied (Neumann and Meyer, 2008).

**Lucilia sericata and Lucilia cuprina**

The morphologically similar species *L. sericata* and *L. cuprina* are estimated to have diverged between 0.35 and 0.39 MYA (Stevens et al., 2002; Wallman et al., 2005). Today, *L. sericata* and *L. cuprina* represent two of the most
economically important myiasis-causing blowflies worldwide, being the main causal agents of blowfly ‘sheep strike’ in cool temperate (Europe and New Zealand) and sub-tropical/warm temperate regions (Australia and South Africa), respectively (Stevens and Wall, 1997a).

Within *L. sericata* populations, some intra-specific variation in oviposition behaviour has been reported, for example, UK populations demonstrating a greater probability of laying eggs when exposed to sheep wool, than Australian or Danish populations (Crombie, 1944; Cragg, 1950; Cragg and Cole, 1956). However, the majority of molecular analyses have failed to find clearly distinct geographical *L. sericata* strains (Sperling *et al.*, 1994; van der Leij, 1995; Stevens and Wall, 1997b). In comparison, two morphologically distinct *L. cuprina* sub-species, *L. cuprina cuprina* Wiedmann, primarily found in Neotropical, Oriental, and southern Neoarctic regions, and *L. cuprina dorsalis* Robineau-Desvoidy, found in East and sub-Saharan Afrotropical, and Australasian areas, have been identified (Waterhouse and Paramonov, 1950; Norris, 1990).

*L. sericata* and *L. cuprina* are sympatric in many areas around the world (Zumpt, 1965; Spradberry, 1991). While they can be distinguished morphologically, for example by integument colour of their abdomens and front femurs (Ulliyett, 1945), these morphological characters are often ambiguous, leading to much speculation of hybridization in areas where both species occur (Ulliyett, 1945). Furthermore, both *L. sericata* and *L. cuprina* have demonstrated low frequency hybridization, under laboratory conditions, with offspring displaying either intermediate or *L. cuprina* morphology (Mackerras, 1933; Ulliyett, 1945; Waterhouse and Paramonov, 1950).

Random amplification of polymorphic DNA (RAPD) analysis has identified a distinctive sub-grouping of Hawaiian *L. cuprina*, which appeared to be more closely related to *L. sericata* than other *L. cuprina* (Stevens and Wall, 1996b). On closer inspection, these Hawaiian flies appeared to be morphologically identifiable as *L. cuprina cuprina*, while other *L. cuprina* tested (including specimens from Africa, Europe, and Australasia) appeared to be *L.
L. cuprina dorsalis (Stevens and Wall, 1997b). Additionally, a sub-group of L. cuprina from Townsville, Australia, were identified which grouped in with the Hawaiian subset. While these Townsville specimens also appeared to be morphologically L. cuprina cuprina, molecular data identified them as an intermediate between the predominant-type L. cuprina dorsalis and the Hawaiian type, L. cuprina cuprina (Stevens and Wall, 1996b). Similarly, phylogenetic analysis using sequence data from COX1-COX2 and 28S, from a range of L. sericata and L. cuprina populations, found considerable incongruence between the mitochondrial and nuclear rRNA data when including Hawaiian samples (Stevens et al., 2002). When examined, Hawaiian L. cuprina specimens were found to contain L. cuprina-type 28S sequence data and L. sericata-type COX1 and COX2 sequence data (Stevens et al., 2002). As a result, it seems that extant Hawaiian L. cuprina populations are likely to have a L. cuprina x L. sericata hybrid origin (Stevens and Wall, 1996b; Stevens et al., 2002).

Recent analysis of South African L. sericata and L. cuprina populations, using 28S and COX1 sequence data, have also revealed several specimens with conflicting haplotypes for each gene (Tourle et al., 2009). Interestingly, of the seven L. cuprina specimens demonstrating L. sericata COX1 haplotypes, only one displayed an intermediate morphology, with the rest displaying L. cuprina type morphology. Molecular studies using mtDNA have also found Hawaiian haplotypes grouping with specimens from Taiwan and Thailand (Harvey et al., 2008). While comparisons of Australian populations have revealed paraphyletic groupings that merge L. sericata and L. cuprina (Wallman et al., 2005; Nelson et al., 2007). This apparent hybridization between wild populations of L. sericata and L. cuprina is particularly important when considering the difference in style of parasitism exhibited by these species. For example, L. sericata is often used in medical applications (maggot therapy) as the larvae tend to feed on dead tissue, whereas L. cuprina are more aggressive feeders on living tissue.
Dyscritomyia

The genus *Dyscritomyia* (Grimshaw 1901), which contains 35 recognised species, is endemic to the Hawaiian Islands (Hardy, 1981), being predominantly found above 600m (2000ft) (Wells *et al.*, 2002). *Dyscritomyia* differs from the majority of other calliphorids in that is it viviparous, with eggs being retained inside the adult female, and individual larvae being produced once they have reached second or third instar stage (Pollock, 1974). Still relatively little is known about larval feeding behaviour of *Dyscritomyia*. While reports of saprophagic feeding on molluscs have also been documented (Terry, 1912; Perkins, 1913; Sweezey, 1914), larval mouth hook structures suggest parasitic feeding (Pollock, 1974).

*Dyscritomyia* is currently placed within the sub-family Luciliinae, with phylogenetic analysis using *COX1-COX2* sequence data having suggested a monophyletic sister lineage to *Lucilia* (Wells *et al.*, 2002). Understanding the evolution and adaptive radiation within *Dyscritomyia* could potentially provide new insight into the biogeography of the Hawaiian archipelago as a whole (Wells *et al.*, 2002). However, to date, very little molecular research has focused on *Dyscritomyia*.

Hemipyrellia

The eight species that make up *Hemipyrellia* are morphologically very similar to the genus *Lucilia*, also sharing similar habitats (Zumpt, 1956; Fan, 1997). Several Hemipyrellia species are considered to be of forensic importance (Chen *et al.*, 2004; Lee *et al.*, 2004; Sukontason *et al.*, 2007a; Sukontason *et al.*, 2008), with some also being occasional agents of myiasis in animals (Zumpt and Ledger, 1967; Roy and Dasgupta, 1971).

Traditionally, *Hemipyrellia* have been included within the genus *Lucilia* (Zumpt, 1956). Molecular evidence has provided conflicting accounts of the position of *Hemipyrellia* within Calliphoridae. For example, Chen *et al.* (2004) grouped *H. ligurriens* together with several *Lucilia* species when comparing
COX1 sequence data. However, work by Stevens and Wall (1997a), using 12S rRNA data, and Wallman et al. (2005), using COX1, COX2, and ND4-ND4L sequence data, and found several Hemipyrellia species (H. fernandica, and H. fergusoni and H. ligurriens, respectively) to be distinct from Lucilia species, supporting a separate status of genus for Hemipyrellia.

4.1.2 The evolution of myiasis in Calliphoridae

The defining characteristic of blowflies is the necessity to lay eggs on a proteinaceous substrate, often including the tissues of living animals; larvae then develop by feeding on the protein-rich tissue of the host, a process known as myiasis. Within Calliphoridae, a range of larval feeding habits exist, including; coprophagy, saprophagy, sanguinivory and ectoparasitism, both facultative and obligate (Table 4). However, understanding of the evolution of these different behaviours is often restricted by a lack of widely agreed theories of evolutionary relationship and classification within Calliphoridae (Stevens, 2003).

Phylogenetic studies have found support for Zumpt’s (1965) hypothesis of saprophagous or sanguinivorous origins of parasitism within Calliphoridae (Stevens et al., 2006). For example, Auchmeromyia luteola (the Congo floor maggot) and various species Protocalliphora (bird blowflies) are thought to represent sanguinivorous origins to parasitism (Stevens, 2003). However, most examples of parasitic behaviour within Calliphoridae appear to be saprophagic in origin (Stevens, 2003; Stevens and Wallman, 2006). Work by Stevens (2003), comparing 28S sequence data from a range of blowfly species, suggested at least five independent origins of obligate parasitism within Calliphoridae. Additionally, three potential origins of myiasis-causing behaviour have been identified within the genus Lucilia alone, corresponding to L. sericata, L. cuprina, and L. caesar groupings (Stevens and Wall, 1996a, 1997a).
4.1.3 Choice of molecular marker

Within insect systematics the preferred molecular targets have been mtDNA and nuclear ribosomal DNA (Otranto and Stevens, 2002), largely due to the relative ease with which these markers can be amplified (Baker et al., 2001). When analysed in combination, nuclear genes have generally been found to outperform mitochondrial genes at resolving deeper level nodes, displaying lower levels of homoplasy and higher clade support (Reed and Sperling, 1999; Leys et al., 2000; Baker et al., 2001; Brady, 2002; Morris et al., 2002; Danforth et al., 2003; Lin and Danforth, 2004). However, substitution rates in mitochondrial genes are estimated to be between five and ten times that of nuclear DNA (Brown et al., 1979), making mtDNA useful for inferring relationships between recently diverged species and for population level analysis (Stevens and Wall, 1997b; Shao and Barker, 2006). The vast majority of evolutionary studies of parasitic arthropods have used single protein-coding gene analysis (Shao and Barker, 2006). However, it is becoming increasingly common within insect systematics to combine mitochondrial genes and nuclear genes, providing two or more unlinked sets of data that have evolved under different constraints (Lin and Danforth, 2004). Accordingly, this study uses three gene data sets, representing a mitochondrial protein-coding gene, a nuclear protein-coding gene, and a nuclear non-coding gene.

Cytochrome oxidase 1

Cytochrome oxidase 1 (COX1) has been one of the most commonly used loci within insect systematics (Zhang and Hewitt, 1997b), and has been used extensively within blowfly systematics (Gleeson and Sarre, 1997; Otranto and Puccini, 2000; Stevens et al., 2002; Wells et al., 2002; Harvey et al., 2003; Stevens, 2003; Wells et al., 2004; Otranto et al., 2005a; Wells et al., 2007; Wells and Williams, 2007; Harvey et al., 2008; Tourle et al., 2009). As a mitochondrial gene, COX1 sequence data have several advantages including; of a lack of recombination during cell division, high copy number, relative ease of isolation, availability of universal primers, and the presence of both conserved and variable regions. Mitochondrial genes are also expected to
reach reciprocal monophyly before nuclear genes, due to generally higher rates of sequence change (Avise et al., 1979; Lunt et al., 1996; Monteiro and Pierce, 2001; Funk and Omland, 2003; Dowton, 2004; Lin and Danforth, 2004), making \textit{COX1} particularly useful for inferring relationships between recently diverged species and in population genetics (Stevens and Wall, 1997b; Shao and Barker, 2006).

\textit{Elongation factor-1\alpha}

Despite the fact that far more phylogenetic analyses have used mtDNA or nuclear rDNA sequence data (Caterino et al., 2000), nuclear genes may offer several advantages over mitochondrial genes, for example by having a generally low level of biased base composition (Friedlander et al., 1992; Brower and DeSalle, 1994; Friedlander et al., 1994; Lin and Danforth, 2004). While a wide range of protein coding nuclear DNA genes have been used in insect systematics, by far the most popular gene has been \textit{elongation factor-1 alpha} (\textit{EF-1\alpha}) (Cho et al., 1995; Danforth and Ji, 1998; Clark et al., 2000; Caterino et al., 2001; Cognato and Vogler, 2001; Buckley et al., 2002; Danforth, 2002; Ghiselli et al., 2007).

General observations have been made that nuclear genes tend to have a greater resolving power than mitochondrial genes, particularly at deeper node levels of phylogenetic inference (Reed and Sperling, 1999; Leys et al., 2000; Baker et al., 2001; Brady, 2002; Morris et al., 2002; Danforth et al., 2003; Lin and Danforth, 2004). However, within Insecta \textit{EF-1\alpha} has been used very successfully for studies among species groups and genera (Cho et al., 1995; Belshaw and Quicke, 1997; Mitchell et al., 1997; Reed and Sperling, 1999). For example, including \textit{EF-1\alpha} sequence data into analyses of the genus \textit{Papilio} was found to improve resolution of relationships among species, also supporting congruent relationships with the mitochondrial genes \textit{COX1} and \textit{COX2} (Reed and Sperling, 1999).
**28S rRNA**

The secondary structure of ribosomal RNA (rRNA) molecules consists of a variety of structures created by folds in the polypeptide chain (e.g. stems, hairpin loops, bulges, multi-helix junctions), which is ultimately determined by sequence of nucleotides in the rRNA gene. These secondary structures are vital for the biological function of the RNA molecule (e.g. protein synthesis), meaning that the regions of the RNA sequence that preserve them are under distinct evolutionary constraints (Fox and Woese, 1975; Gutell et al., 1994). For example, in eukaryotes, regions of the sequence coding for hairpin loops demonstrate a higher rate of mutation accumulation (Smit et al., 2007). The difference in rates of mutation accumulation across rRNA genes make them effective molecular tools for the differentiation of species, and even strains (Arnheim, 1983; Gasser, 1999). A popular rRNA gene within insect systematics is the nuclear 28S large subunit rRNA, which displays both conserved and highly variable regions (D expansions) suitable for resolving relationships at a range of different hierarchical levels, and even having the ability to distinguish between closely related species (Otranto et al., 2005b).

### 4.1.4 Aims

The defining characteristic of blowflies is their need for larvae to develop on a proteinaceous substrate, often a living vertebrate host. As well as including some vital insect indicator species used in forensic entomology, Calliphoridae also represent some of the most economically significant myiasis-causing flies in the world, notorious for their parasitism of livestock. While phylogenetic studies of Calliphoridae have traditionally focused on morphological evidence, molecular data is becoming increasingly favoured. However, despite being the focus of research for over 50 years, key taxonomic relationships within this family remain ambiguous.

The work presented here reconstructs a multi-locus phylogeny, using two protein coding genes (mitochondrial and nuclear) and one nuclear rRNA gene, in an attempt to resolve evolutionary relationships within Calliphoridae. Using
sequence data from three independent genes in this way, will not only allow the different evolutionary rates of each gene to improve phylogenetic resolution at different levels of divergence, but also enable any inconsistencies in evolutionary history between the gene trees and species tree to be explored. Taxonomic sampling has particularly focused on populations of *L. sericata* and *L. cuprina*, in an attempt to reveal the extent hybridisation is occurring between these important blowfly species.

Finally, patterns of blowfly evolution suggest that parasitism is a trait that has evolved numerous times within Calliphoridae, with even closely related taxa exhibiting very different larval feeding strategies. Accordingly, the multi-loci phylogenetic framework will also be used in an attempt to elucidate the evolution of these key parasitic behaviours, providing an ideal model for understanding the evolution of metazoan parasites.
<table>
<thead>
<tr>
<th>Sub-family</th>
<th>Species</th>
<th>Larval feeding behaviour</th>
<th>Host</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>Auchmeromyiinae</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>Auchmeromyia luteola</em></td>
<td>Obligate (Sanguinivorous)</td>
<td>Animals, Humans</td>
<td>6, 21</td>
</tr>
<tr>
<td></td>
<td><em>Cordylobia anthropophaga</em></td>
<td>Obligate</td>
<td>Animals, Humans</td>
<td>6, 21</td>
</tr>
<tr>
<td></td>
<td><strong>Bengaliinae</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>Bengalia depressa</em></td>
<td>Obligate</td>
<td>Termites and ant pupae</td>
<td>3, 8</td>
</tr>
<tr>
<td></td>
<td><strong>Calliphorinae</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>Calliphora dubia</em></td>
<td>Facultative (secondary)</td>
<td>Carrion</td>
<td>16, 21</td>
</tr>
<tr>
<td></td>
<td><em>Calliphora quadramaculata</em></td>
<td>Facultative (secondary)</td>
<td>Carrion, Sheep</td>
<td>16, 21</td>
</tr>
<tr>
<td></td>
<td><em>Calliphora stygia</em></td>
<td>Facultative (secondary)</td>
<td>Carrion, Sheep</td>
<td>16, 21</td>
</tr>
<tr>
<td></td>
<td><em>Calliphora vicina</em></td>
<td>Facultative (secondary)</td>
<td>Carrion, Humans, Dogs, Monkeys, Sheep</td>
<td>16, 21</td>
</tr>
<tr>
<td></td>
<td><em>Calliphora vomitoria</em></td>
<td>Facultative (secondary)</td>
<td>Carrion, Sheep</td>
<td>16, 21</td>
</tr>
<tr>
<td></td>
<td><em>Cynomya cadaverina</em></td>
<td>Saprophagic</td>
<td>Vertebrate</td>
<td>4, 16</td>
</tr>
<tr>
<td></td>
<td><em>Cynomya mortuorum</em></td>
<td>Facultative (secondary)</td>
<td>Vertebrate</td>
<td>4, 16, 21</td>
</tr>
<tr>
<td></td>
<td><em>Onesia tibialis</em></td>
<td>Facultative (primary)</td>
<td>Earthworms</td>
<td>12, 16, 21</td>
</tr>
<tr>
<td></td>
<td><strong>Chrysomyinae</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>Chrysomya albiceps</em></td>
<td>Facultative (secondary)</td>
<td>Sheep</td>
<td>6, 13, 21</td>
</tr>
<tr>
<td></td>
<td><em>Chrysomya bezziana</em></td>
<td>Facultative (primary)</td>
<td>Carrion, Humans, Ruminants, Horses</td>
<td>16, 21</td>
</tr>
<tr>
<td></td>
<td><em>Chrysomya chloropyga</em></td>
<td>Facultative (secondary)</td>
<td>Carrion, Ruminants</td>
<td>16, 21</td>
</tr>
<tr>
<td></td>
<td><em>Chrysomya megacephala</em></td>
<td>Facultative (secondary)</td>
<td>Carrion, Ruminants, Donkeys</td>
<td>6, 13, 16, 21</td>
</tr>
<tr>
<td></td>
<td><em>Chrysomya rufifacies</em></td>
<td>Facultative (secondary)</td>
<td>Carrion, Sheep, Cattle</td>
<td>6, 13, 16, 21</td>
</tr>
<tr>
<td></td>
<td><em>Cochliomyia hominivorax</em></td>
<td>Obligate</td>
<td>Humans, Animals</td>
<td>13, 16, 21</td>
</tr>
<tr>
<td></td>
<td><em>Cochliomyia macellaria</em></td>
<td>Facultative (secondary)</td>
<td>Humans, Animals</td>
<td>13, 16, 21</td>
</tr>
<tr>
<td></td>
<td><em>Compsomyia fulvicrura</em></td>
<td>Saprophagic</td>
<td>Carrion</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td><em>Phormia regina</em></td>
<td>Facultative (?)</td>
<td>Carrion, Sheep</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td><em>Protocalliphora azurea</em></td>
<td>Obligate</td>
<td>Birds</td>
<td>4, 6, 21</td>
</tr>
<tr>
<td></td>
<td><em>Protocalliphora sialia</em></td>
<td>Facultative (primary)</td>
<td>Carrion, Vertebrates</td>
<td>16, 21</td>
</tr>
<tr>
<td></td>
<td><em>Protophormia terraenovae</em></td>
<td>Facultative (secondary)</td>
<td>Carrion, Sheep</td>
<td>4, 6, 16, 21</td>
</tr>
</tbody>
</table>

Table 6: Details of larval feeding behaviour and host (? unclear whether spp. capable of initiating primary myiasis, * outgroup taxa)

<table>
<thead>
<tr>
<th>Sub-family</th>
<th>Species</th>
<th>Larval feeding behaviour</th>
<th>Host</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Helicoboscinae</td>
<td><em>Eurychaeta palpalis</em></td>
<td>Saprophagic</td>
<td>Slugs and Snails</td>
<td>4</td>
</tr>
<tr>
<td>Luciliinae</td>
<td><em>Dyscritomyia luciloiides</em></td>
<td>Facultative (?)</td>
<td>Carion, Vertebrate and/or Invertebrate</td>
<td>2, 7, 9</td>
</tr>
<tr>
<td></td>
<td><em>Dyscritomyia robusta</em></td>
<td>Facultative (?)</td>
<td>Carion, Vertebrate and/or Invertebrate</td>
<td>2, 7, 9</td>
</tr>
<tr>
<td></td>
<td><em>Hemipyrellia fergusoni</em></td>
<td>Saprophagic</td>
<td>Carion</td>
<td>11, 18, 20, 24</td>
</tr>
<tr>
<td></td>
<td><em>Hemipyrellia fernandica</em></td>
<td>Saprophagic</td>
<td>Carion</td>
<td>11, 18, 20, 24</td>
</tr>
<tr>
<td></td>
<td><em>Lucilia amputacea</em></td>
<td>Facultative (secondary)</td>
<td>Frogs</td>
<td>4, 6, 13, 21</td>
</tr>
<tr>
<td></td>
<td><em>Lucilia butonivora</em></td>
<td>Obligate</td>
<td>Frogs and/or toads</td>
<td>4, 6, 13</td>
</tr>
<tr>
<td></td>
<td><em>Lucilia caesar</em></td>
<td>Facultative (secondary)</td>
<td>Carion, Humans, Sheep</td>
<td>6, 13, 16, 21</td>
</tr>
<tr>
<td></td>
<td><em>Lucilia cluvia</em></td>
<td>Saprophagic</td>
<td>Carion</td>
<td>26</td>
</tr>
<tr>
<td></td>
<td><em>Lucilia cuprina</em></td>
<td>Facultative (primary)</td>
<td>Carion, Humans, Dogs, Sheep, Goats, Cattle</td>
<td>6, 13, 16, 21</td>
</tr>
<tr>
<td></td>
<td><em>Lucilia illustris</em></td>
<td>Facultative (secondary)</td>
<td>Carion, Sheep</td>
<td>6, 13, 16, 21</td>
</tr>
<tr>
<td></td>
<td><em>Lucilia mexicana</em></td>
<td>Saprophagic</td>
<td>Carion</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td><em>Lucilia papuensis</em></td>
<td>Saprophagic</td>
<td>Carion</td>
<td>19, 15</td>
</tr>
<tr>
<td></td>
<td><em>Lucilia porphyrina</em></td>
<td>Facultative (primary)</td>
<td>Vertebrate</td>
<td>6, 13, 21</td>
</tr>
<tr>
<td></td>
<td><em>Lucilia richardsii</em></td>
<td>Facultative (?)</td>
<td>Vertebrate</td>
<td>6, 13, 21</td>
</tr>
<tr>
<td></td>
<td><em>Lucilia sericata</em></td>
<td>Facultative (primary)</td>
<td>Carion, Humans, Dogs, Sheep, Goats, Cattle</td>
<td>6, 13, 16, 21</td>
</tr>
<tr>
<td></td>
<td><em>Lucilia silvarum</em></td>
<td>Facultative (?)</td>
<td>Frogs and/or toads</td>
<td>4, 23</td>
</tr>
<tr>
<td></td>
<td><em>Lucilia thatuna</em></td>
<td>Facultative (?)</td>
<td>Vertebrate</td>
<td>4, 23</td>
</tr>
<tr>
<td>Polleniinae</td>
<td><em>Pollenia rudis</em></td>
<td>Facultative (primary)</td>
<td>Earthworms</td>
<td>4, 21</td>
</tr>
<tr>
<td>Muscoidea *</td>
<td><em>Mesembrina meridiana</em></td>
<td>Facultative (?)</td>
<td>Carion, Vertebrates</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td><em>Musca domestica</em></td>
<td>Facultative (secondary)</td>
<td>Coprophagous, Carion, Vertebrates</td>
<td>6, 17, 25</td>
</tr>
<tr>
<td></td>
<td><em>Stomoxys calcitrans</em></td>
<td>Saprophagic</td>
<td>Coprophagous, Carion</td>
<td>1, 14</td>
</tr>
</tbody>
</table>

Table 4: Details of larval feeding behaviour and host (? unclear whether spp. capable of initiating primary myiasis, * outgroup taxa)

4.2 Methods and Materials

4.2.1 Sample collection

The specimens used for DNA sequencing came from in-house collections at the University of Exeter, freshly collected samples sent by colleagues, and specimens on loan from external collections (Table 5). All specimens were stored at 4°C, either in 100% ethanol, or as dried pinned specimens.

4.2.1.1 Lucilia bufonivora specimens

Two larval specimens collected from myiasis of a toad in the UK, and believed to be *Lucilia bufonivora*, were included in the study. These larvae were collected from the infected nasal passages of a live toad (Fig. 9) and were provided to us as dead, alcohol preserved early-stage larvae. Due to their condition and stage of development these larvae were not identified morphologically.

![Figure 9: Toad with nasal myiasis in Suffolk, UK. Larvae were collected and preserved in alcohol prior to DNA extraction (photograph taken by Michael Porter).](image)

4.2.1.2 Lucilia cuprina hybrid specimen

DNA extracted from the legs of a suspected *L. cuprina* hybrid specimen was characterized. The legs were from a specimen included the study by Tourle *et al.* (2009), and have already been confirmed as being from a *L. sericata x L. cuprina* hybrid. Tourle *et al.* (2009) amplified COX1 and 28S sequence data for this
specimen, the work presented here aims to confirm the findings of this study, and also add further data from EF-1α.

### 4.2.2 DNA extraction

All DNA extractions were carried out using either a salt extraction method (Aljanabi and Martinez, 1997), or Qiagen DNeasy® Blood & Tissue Kit (Qiagen GmbH, Germany). See Chapter 2 for further details.

### 4.2.3 PCR amplification

DNA extractions were subject to PCR procedures to amplify regions of the nuclear gene *elongation factor-1 alpha* (EF-1α), the ribosomal DNA 28S, and the mitochondrial *cytochrome oxidase I* (COX1).

#### 4.2.3.1 Cytochrome oxidase 1

Published COX1 primers (Bogdanowicz *et al.*, 1993; Simon *et al.*, 1994; Sperling and Hickey, 1994; Sperling *et al.*, 1995; Lunt *et al.*, 1996; Wells and Sperling, 1999) were used to amplify templates. See Chapter 2 for primer and PCR amplification details.

#### 4.2.3.2 Elongation factor-1α

Modified versions of existing primers (McDonagh *et al.*, 2009a) were used to amplify EF-1α. See Chapter 2 for primer and PCR amplification details.

#### 4.2.3.3 28S

Published primers (Hoelzel and Green, 1992; Friedrich and Tautz, 1997b, 1997a) were employed to amplify the D1-D7 expansion regions and related core elements of the large subunit 28S rRNA. See Chapter 2 for primer and PCR amplification details.
4.2.4 PCR product purification and sequencing

PCR products were separated by electrophoresis gel, and appropriate sized bands cut out and purified. Purified PCR products were the sequenced using the commercial sequencing facility COGENICS (formally Lark Technologies Inc.). See Chapter 2 for further details.

4.2.5 Sequence proofreading

Sequence fragments were checked for quality and edited manually before being assembled into a single consensus sequence, using AutoAssembler 2.0 (Applied Biosystems, Inc.). Any ambiguities in the consensus sequence were resolved, or standard IUPAC/IUB codes (Leonard, 2003) used where the nucleotide could not be read from the chromatogram or if polymorphisms in amplicon sequencing were suspected. Edited consensus sequences were then export into a text file for alignment.

4.2.6 Multiple sequence alignment

Multiple sequence alignment was carried out using the alignment editor SEAVIEW version 2.4 (Galtier et al., 1996) implementing the ClustalW algorithm (Thompson et al., 1994), with manual corrections by eye.

4.2.7 Nucleotide substitution model selection

Appropriate nucleotide substitution model parameters were selected by a series of nested hierarchical likelihood-ratio tests using the program MODELTEST Version 3.06, (Posada and Crandall, 1998).

4.2.8 Bayesian phylogenetic inference

All phylogenetic reconstructions were carried out using the program MrBayes 3.1 (Huelsenbeck and Ronquist, 2001). All phylogenetic analyses implemented two independent Metropolis Coupled MCMC (MCMCMC) searches starting from...
different random trees (nruns=2), each search contained three heated chains (using the default heating temperature, temp=0.2), and one cold chain, with a sampling frequency of 10 generations, and selected models and priors applied. The multi-gene phylogeny was reconstructed using a partitioned data set, with unlinked model parameters, and variable rate parameters to allow each gene data set (partition) to evolve under different rates.

Analyses were continued until the convergence diagnostic (standard deviation of split frequencies) fell below the default threshold (stopval=0.01), indicating sufficient convergence between the two samples of the posterior probability (Ronquist et al., 2005). The default convergence diagnostic burn-in fraction of 0.25 was used (burninfrac=0.25), consequently, a corresponding burn-in of 0.25, corresponding to the first 25% of samples obtained up until convergence had been reached, was also applied to summarize substitution model parameters (sump) and trees and branch lengths (sumt). Plots of generation versus log probability of the data (log likelihood values) produced by the sump command were also checked to ensure stationarity had been reached (i.e. plot shows no patterns in the data, resembles ‘white noise’) (Ronquist et al., 2005). Tree topology was then calculated from the remaining data after discarding burn-in samples by constructing a majority-rule consensus tree, and the probability that a monophyletic clade is ‘true’, given the caveats of the model and data, estimated by the proportion of trees in the MCMC sample in which the clade occurs (Brooks et al., 2007).

### 4.2.9 Partition homogeneity test

The level of phylogenetic congruence between single gene data sets was tested using the partition homogeneity test (parsimony-based ILD test) in PAUP* (Swofford, 1998). The test was implemented under parsimony, with random taxa addition, no swapping, and 1000 replicates.
<table>
<thead>
<tr>
<th>Sub-family</th>
<th>Species</th>
<th>Location</th>
<th>EF-1α</th>
<th>COX1</th>
<th>28S</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Auchmeromyiinae</td>
<td>Auchmeromyia luteola</td>
<td>Kenya</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Cordylobia anthropophaga</td>
<td>Cameroon</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Bengaliinae</td>
<td>Bengalia depressa</td>
<td>Kenya</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Calliphorinae</td>
<td>Calliphora dubia</td>
<td>Australia</td>
<td>+</td>
<td></td>
<td>+</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>Calliphora quadramaculata</td>
<td>New Zealand</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Calliphora stygia</td>
<td>Australia</td>
<td>+</td>
<td></td>
<td>+</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>Calliphora vicina</td>
<td>Australia</td>
<td>+</td>
<td></td>
<td>+</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td></td>
<td>UK</td>
<td>+</td>
<td></td>
<td>+</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Calliphora vomitoria</td>
<td>UK</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>USA</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Cynomyinae</td>
<td>Cynomya cadaverina</td>
<td>Canada</td>
<td>+</td>
<td></td>
<td>+</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Cynomya mortuorum</td>
<td>UK</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Onesia tibialis</td>
<td>Australia</td>
<td>+</td>
<td></td>
<td>+</td>
<td>11</td>
</tr>
<tr>
<td>Chrysomyinae</td>
<td>Chrysomya albiceps</td>
<td>Kenya</td>
<td>+</td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Chrysomya bezziana</td>
<td>Indonesia</td>
<td>+</td>
<td></td>
<td>+</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Chrysomya chloropyga</td>
<td>Tanzania</td>
<td>+</td>
<td></td>
<td>+</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Chrysomya megacephala</td>
<td>India</td>
<td>+</td>
<td></td>
<td>+</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Chrysomya rufifacies</td>
<td>Australia</td>
<td>+</td>
<td></td>
<td>+</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>Cochliomyia hominivorax</td>
<td>Brazil</td>
<td>+</td>
<td></td>
<td>+</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>Cochliomyia macellaria</td>
<td>Brazil</td>
<td>+</td>
<td></td>
<td></td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Compsomyiops fulvicrura</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>Phormia regina</td>
<td>Czech Republic</td>
<td>+</td>
<td></td>
<td></td>
<td>3, 7</td>
</tr>
</tbody>
</table>

Table 7: List of taxa included in study, location details and accession numbers (+ this study).

<table>
<thead>
<tr>
<th>Sub-family</th>
<th>Species</th>
<th>Location</th>
<th>EF-1α</th>
<th>COX1</th>
<th>28S</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chrysomyinae</td>
<td>Protocalliphora azurea</td>
<td>Belgium</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Protocalliphora sialis</td>
<td>USA</td>
<td>+</td>
<td>AF295559</td>
<td>+</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Protophormia terraeovae</td>
<td>UK/DD/Czech Republic</td>
<td>+</td>
<td>AF295553</td>
<td>+</td>
<td>3</td>
</tr>
<tr>
<td>Luciliinae</td>
<td>Dyscritomyia lucilioides</td>
<td>Hawaii, USA</td>
<td>+</td>
<td>AY074903</td>
<td>+</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>Dyscritomyia robusta</td>
<td>Hawaii, USA</td>
<td>+</td>
<td>AY074899</td>
<td>+</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>Hemipyrella fergusoni</td>
<td>Tanzania</td>
<td>+</td>
<td></td>
<td>+</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>Hemipyrella fernandica</td>
<td>Tanzania</td>
<td>+</td>
<td></td>
<td>+</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Lucilia ampullacea</td>
<td>UK/DD/Korea</td>
<td>+</td>
<td>EU925394</td>
<td>+</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>Lucilia bufonivora</td>
<td>UK</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>UK</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Lucilia caesar</td>
<td>UK</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Lucilia clivia</td>
<td>USA</td>
<td>+</td>
<td>DQ453490</td>
<td>+</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>Lucilia cuprina</td>
<td>Hawaii, USA</td>
<td>+</td>
<td>AJ417704</td>
<td>+</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Kenya</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>New Zealand</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Perth, Australia</td>
<td>+</td>
<td>AJ417707</td>
<td>+</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td></td>
<td>South Africa</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Townsville, Australia</td>
<td>+</td>
<td>DQ453492</td>
<td>+</td>
<td>13</td>
</tr>
<tr>
<td>Lucilia cuprina hybrid</td>
<td>South Africa</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lucilia illudis</td>
<td>UK</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lucilia mexicana</td>
<td>USA</td>
<td>+</td>
<td>DQ453492</td>
<td>+</td>
<td>13</td>
<td></td>
</tr>
</tbody>
</table>

Table 5: List of taxa included in study, location details and accession numbers (+ this study).

<table>
<thead>
<tr>
<th>Sub-family</th>
<th>Species</th>
<th>Location</th>
<th>EF-1α</th>
<th>COX1</th>
<th>28S</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Luciliinae</td>
<td>Lucilia papuensis</td>
<td>Australia</td>
<td>+</td>
<td>AY842609</td>
<td>+</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>Lucilia porphyrina</td>
<td>Australia</td>
<td>+</td>
<td>AY842610</td>
<td>+</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>Lucilia richardii</td>
<td>UK</td>
<td>+</td>
<td>AJ417714</td>
<td>+</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Lucilia sericata</td>
<td>New Zealand, Perth, Australia</td>
<td>+</td>
<td>AJ417713</td>
<td>+</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>UK, USA, Zimbabwe</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Lucilia silvarum</td>
<td>UK</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Lucilia thatuna</td>
<td>USA</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Polleniinae</td>
<td>Pollenia rudis</td>
<td>UK</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Helicoboscinae</td>
<td>Eurychaeta palpalis</td>
<td>-</td>
<td>FJ025672</td>
<td>FJ025612</td>
<td>FJ025512</td>
<td>15</td>
</tr>
<tr>
<td>Muscoidea *</td>
<td>Mesembrina meridiana</td>
<td>UK</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Stomoxys calcitrans</td>
<td>-</td>
<td>AF503149</td>
<td>AB479528</td>
<td>AJ551427</td>
<td>5, 17, 10</td>
</tr>
</tbody>
</table>

Table 2: List of taxa included in study, location details and accession numbers (+ this study).

4.3 **Results**

4.3.1 **Substitution models**

The best-fit model for all three single-gene data sets was the General Time Reversible model (GTR) with proportion of invariable sites (+I) and gamma distributed rate variation among sites (+Γ). However, during the combined-gene analysis a partitioned data set was still implemented, with model parameters unlinked and variable rate parameters, allowing each gene to evolve under different rates.

4.3.2 **Congruence between gene trees**

The level of phylogenetic congruence between single gene data sets was tested using the partition homogeneity test (parsimony-based ILD test), implemented under parsimony, with random taxa addition, no swapping, and 1000 replicates. The ILD test revealed a significant difference \( P < 0.01, n = 1 \) in topology between the three genealogies (Fig. 10-12, pages 123-125).

4.3.3 **Auchmeromyiinae**

All gene trees (Fig. 10-12), and the multi-gene tree (Fig. 13), confirm the subfamily status of Auchmeromyiinae, represented in this study by *Auchmeromyia luteola* and *Cordylobia anthropophaga*. While COX1 (Fig. 12) failed to resolve the position of Auchmeromyiinae within a Bengaliinae-Calliphoridae-Helicoboscinae-Luciliinae grouping, EF-1α (Fig. 10) suggested an Auchmeromyiinae-Bengaliinae sister lineage. In contrast the 28S gene phylogeny (Fig. 11) failed to recover any robust resolution for the position of Auchmeromyiinae within Calliphoridae. Nevertheless, the multi-gene phylogeny recovered an Auchmeromyiinae-Bengaliinae lineage with high clade support, although the relationship between this lineage and Chrysominae is unresolved.
4.3.4 Bengaliinae

The position of *Bengalia depressa* in all three gene trees supports the separate sub-family status of Bengaliinae; being recovered with high clade support as a sister lineage to Auchmeromyiinae by the *EF-1α* gene tree (Fig. 10) and the overall multi-gene tree (Fig. 13).

4.3.5 Calliphoridae

Calliphoridae is recovered as monophyletic by both *EF-1α* (Fig. 10) and *28S* (Fig. 11) gene trees, and the overall multi-gene tree. However, Calliphoridae is found to be paraphyletic in the *COX1* tree (Fig. 12), with the Helicoboscinae taxon, *Eurychaeta palpalis*, being recovered within a clade of Calliphoridae taxa. All gene trees support a Calliphorinae-Luciliinae sister lineage, with respect to Chrysomyinae.

Within Calliphorinae, the genus *Calliphora* were recovered as paraphyletic, with *Calliphora vicina* and *Calliphora vomitoria* grouping with the two *Cynomya* taxa, *Cy. cadaverina* and *Cy. mortuorum*, in all three gene trees. Additionally, *Onesia tibialis* grouped with *Calliphora dubia* in the *EF-1α* (Fig. 10) and *28S* (Fig. 11) trees, and with *Eurychaeta palpalis*, albeit in a clade with *C. dubia*, in the *COX1* tree (Fig. 12). The multi-gene tree (Fig. 13) recovered *O. tibialis* with *C. dubia* with high support, to the exclusion of *E. palpalis*. The *EF-1α* and *COX1* gene trees recovered the UK and Australian *Calliphora vicina* taxa together with high support. The *28S* gene tree failed to resolve a monophyletic grouping for these two taxa, placing them within an unresolved clade containing the *C. vomitoria*, *Cy. cadaverina*, and *Cy. mortuorum*. In contrast, the UK and USA specimens of *C. vomitoria* were only recovered as monophyletic by the *EF-1α* tree, being paraphyletic in the *COX1* tree, and forming a polytomy within the *28S* tree. In contrast, the multi-gene tree (Fig. 13) recovered high clade support for a *C. vicina – Cynomya* sister group to *C. vomitoria*.
4.3.6 Chrysomyinae

COX1 (Fig. 12) was the only gene tree to retrieve Chrysomyinae as a monophyletic clade, although with a posterior probability (PP) of just 75%. All Chrysomyinae taxa were also grouped together within the EF-1α tree (Fig. 10), this time with a PP value of 99%, however this clade also included Pollenia rudis (Polleniinae). In contrast, Chrysomyinae remained largely unresolved in the 28S tree (Fig. 11). Combining the three data sets was found to recover Chrysomyinae as a monophyletic clade, but with no significant clade support (Fig. 13).

While only EF-1α retrieved the genus Chrysomya as a monophyletic clade, the multi-gene tree recovered monophyly with PP of 100%. All gene trees grouped Chrysomya albiceps and Chrysomya ruffifacies together, although high PP values were only recovered in the EF-1α (Fig. 10) and COX1 (Fig. 12) trees. While Chrysomya bezzianna and Chrysomya megacephala are grouped together in the 28S (Fig. 11) and EF-1α trees, with high support for this clade given by EF-1α, the multi-gene phylogeny fails to recover this relationship, instead placing Chrysomya bezzianna as a sister lineage to Chrysomya megacephala and Chrysomya chloropyga. All gene trees grouped the two Protocalliphora taxa together with PP values of 100%. However, none of the gene trees, or the multi-gene tree (Fig. 13), recovered monophyly for the Chrysomyinae tribe Phormini, represented in this study by Phormia regina, Protophormia terraenovae, Protocalliphora azurea and Protocalliphora sialia. Finally, Cochliomyia hominivorax, Cochliomyia macellaria, and Compsomyiops fulvicrura consistently grouped together, with two of the gene trees (EF-1α, and COX1) and the multi-gene tree displaying high PP grouping the two Cochliomyia taxa as a sister group to Compsomyiops fulvicrura.

4.3.7 Luciliinae

The sub-family Luciliinae was recovered as monophyletic by all three genes, receiving PP values of 100 in the EF-1α (Fig. 10) and COX1 (Fig. 12) trees, and the overall multi-gene tree (Fig. 13). Within Luciliinae, the two Dyscritomyia...
taxa, *D. lucilioides* and *D. robusta*, were grouped together with high support by all genes. COX1 placed *Dyscritomyia* in a clade with the two *Hemipyrellia fernandica*, *Hemipyrellia fergusoni*, *L. cluvia*, and *L. mexicana*, separate from the remaining *Lucilia* taxa. Within this clade each of the three genera represented (*Dyscritomyia*, *Hemipyrellia*, *Lucilia*) recovered PP values of 100. While EF-1α also recovered these *Hemipyrellia* and *Lucilia* (*L. cluvia* and *L. mexicana*) clades as separate sister lineages, *Dyscritomyia* was recovered within the main *Lucilia* clade. In contrast, 28S (Fig. 11) recovered *Dyscritomyia* as a separate sister lineage to all other Luciliinae taxa, grouping *Hemipyrellia* with *L. caesar* and *L. porphyrina*, with respect to all other *Lucilia* taxa. The multi-gene phylogeny (Fig. 13) recovered both *Dyscritomyia* and *Hemipyrellia* as distinct clades, separate from all *Lucilia* taxa, but unresolved in their relationship to *Lucilia* or each other.

The two specimens believed to be *L. bufonivora* were consistently recovered together with high support, being placed in a clade with the UK and USA *L. silvarum* taxa by EF-1α (Fig. 10) and 28S (Fig. 11), and grouped with the USA *L. silvarum* and *L. richardsii* by COX1 (Fig. 12). COX1 was the only gene not to recover the two *L. silvarum* taxa together. The position of *L. richardsii* within *Lucilia* remains unclear, being grouped with the USA *L. silvarum* by COX1, as a sister lineage to the *L. sericata* clade by EF-1α, and within the *L. sericata* clade by 28S. The multi-gene tree (Fig. 13) recovered the *L. bufonivora* and *L. silvarum* taxa together with high support, but without the two *L. silvarum* forming a monophyletic clade with respect to *L. bufonivora*. *L. richardsii* was recovered in the multi-gene tree as a sister lineage to the *L. sericata* and *L. cuprina* clade, albeit without significant support. While the position of *L. papuensis* within *Lucilia* differed in each gene tree, combining the three genes recovered high support grouping *L. papuensis* with *L. ampullacea*, *L. cluvia* and *L. mexicana*.

Within *Lucilia*, EF-1α (Fig. 10) recovers the *L. sericata* and *L. cuprina* taxa as two separate clades, with PP values over 0.95. In contrast, COX1 (Fig. 12) grouped the Hawaiian and South Africa hybrid *L. cuprina* taxa together, forming a sister lineage to the *L. sericata* clade. The 28S gene tree (Fig. 11),
however, while grouping the two hybrid taxa within other *L. cuprina* taxa, failed to recover either *L. cuprina* or *L. sericata* monophyletically. Analysing the combined-gene data set (Fig. 13) resulted in support of a sub-grouping of the South African hybrid and the Hawaiian (suspected hybrid) taxa within a monophyletic *L. cuprina* clade, again suggesting mixed hybrid signals between genes.

### 4.3.8 Larval feeding habits within Calliphoridae

The Muscoidea outgroup taxa support the theory of a saprophagous ancestor to Calliphoridae. The blowfly species used in this study include 8 saprophagic taxa, representing the sub-families Luciliinae (*L. papuensis, H. fergusoni, H. fernandica*), Calliphorinae (*Cy. cadaverina*), Chrysomyinae (*Co. fulvicrura*), and Helicoboscinae (*E. palpalis*). Four of the seven blowfly sub-families represented in this study also include obligate parasitic species, namely; Auchmeromyiinae (*A. luteola* and *Co. anthropophaga*), Bengaliinae (*B. depressa*), Chrysomyinae (*C. hominivorax* and *P. azurea*), and Luciliinae (*L. bufonivora*). The remaining taxa are all facultative parasites, with larvae being capable of feeding on dead or living host tissue. The facultative parasitic taxa were also sub-division into those capable of initiating myiasis in an otherwise healthy host (primary facultative parasites), and those capable or only infesting existing myiasis (secondary facultative parasites). However, a lack of information, or conflicting reports, regarding whether a species is capable of initiating myiasis, or not, have prevented several taxa from being classified in this way. Taxa that specialize on parasitizing non-mammalian hosts are also indicated on the multi-gene tree (Fig. 13), revealing several independent origins of similar parasitic behaviour.
Figure 10: Phylogenetic relationships within Calliphoridae (in group), and representatives of Muscoidea (outgroup), based on a Bayesian analysis of the EF-1α nucleotide data set. All branches supported with a PP ≥ 0.95 are labelled. This topology only recovers support for the sub-families Calliphorinae and Lucilinae as monophyletic sister clades, and for Auchmeromyiinae as a separate clade, although failing to find support to resolve its position within Calliphoridae.

(Green - facultative parasites, • - primary agents of myiasis, o - secondary agents of myiasis; Blue - obligate parasites, ■ - sanguinivorous origin; Red - saprophagic species).
Figure 11: Phylogenetic relationships within Calliphoridae (in group), and representatives Muscoidea (outgroup), based on a Bayesian analysis of the 28S nucleotide data set. All branches supported with a PP ≥ 0.95 are labelled. This topology fails to recover support for any of the sub-families within Calliphoridae.

(Green - facultative parasites, • - primary agents of myiasis, o - secondary agents of myiasis; Blue - obligate parasites, ■ – sanguinivorous origin; Red - saprophagic species).
Figure 12: Phylogenetic relationships within Calliphoridae (in group), and representatives Muscoidea (outgroup), based on a Bayesian analysis of the COX1 nucleotide data set. All branches supported with a PP ≥ 0.95 are labelled. This topology only recovers support for the sub-families Auchmeromylinae and Luciliinae.

(Green - facultative parasites, • - primary agents of myiasis, o - secondary agents of myiasis; Blue - obligate parasites, ■ – sanguinivorous origin; Red - saprophagic species).
Figure 13: Phylogenetic relationships within Calliphoridae (in group), and representatives of Muscoidea (outgroup), based on a partitioned Bayesian analysis of the combined gene (EF-1α, 28S, COX1) data set. All branches supported with a PP ≥ 0.95 are labelled. The ILD test has indicated significant phylogenetic conflict between the three genes, consequently this topology should not be used to represent accurate phylogenetic relationships, and instead be used to reflect areas of conflict between genes.

(Green - facultative parasites, • - primary agents of myiasis, o - secondary agents of myiasis; Blue - obligate parasites, ■ – sanguinivorous; Red - saprophagic species).
4.4 Discussion

4.4.1 Taxonomic review of Calliphoridae

Multiple-loci phylogenies not only permit the use of genes that have evolved at different rates, so improving phylogenetic signal at different levels of divergence, but also allow the identification of experimental errors in species identification, and sequencing (Monteiro and Pierce, 2001). For example, despite *Onesia* traditionally being classified as a separate genus within Calliphorinae, this study consistently recovered *O. tibialis* within the genus *Calliphora*. These findings agree with the conclusions of other authors, who have reported *Onesia* to be a sub-genus within *Calliphora* (Wallman and Adams, 1997; Wallman and Donnellan, 2001; Wallman et al., 2005). Whether this incorrect genus classification is true for other, or even all, *Onesia* taxa is unclear. Only a small number of the 61 species belonging to *Onesia* have been included in molecular phylogenetic studies, and if the position of this group within Calliphoridae is to be resolved sampling of *Onesia* needs to greatly improve.

This study also attempted to resolve the position of the endemic Hawaiian group *Dyscritomyia* within Calliphoridae. While all three genes recovered *Dyscritomyia* within Luciliinae, position of the genus within this sub-family differed between gene trees. In agreement with Wells et al., (2002), *Dyscritomyia* were recovered as a separate sister lineage to *Lucilia* by 28S rRNA. COX1 placed *Dyscritomyia* away from the main clade of *Lucilia* species, grouping it with *L. mexicana*, *L. cluvia*, and the two *Hemipyrellia* taxa. In contrast, EF-1α recovered *Dyscritomyia* within the main *Lucilia* clade. However, combining all three gene data sets recovered high clade support for a separate *Dyscritomyia* lineage, distinct from all other Luciliinae species.

The close relationship between the morphologically similar *Hemipyrellia* and *Lucilia* has also been confirmed in this study. While individual gene trees supported the convention of placing *Hemipyrellia* as a sub-genera within *Lucilia*,
combining the three gene data sets provides enough phylogenetic signal to differentiate *Hemipyrellia* as a separate sister genus to *Lucilia* with PP of 100%.

The phylogenetic analyses presented in this chapter have also allowed the identification of two larvae collected from an infested toad in the UK as *L. bufonivora*. While these specimens have not been identified morphologically, the consistent division of these taxa away from *L. silvarum*, the only other *Lucilia* species known to infest toads, suggests that they are very likely *L. bufonivora*. The work presented here represents the first molecular phylogenetic study to include *L. bufonivora*. As well as confirming *L. bufonivora* and *L. silvarum* as sister taxa, this study also suggests a close relationship with *L. sericata*, *L. richardsii*, and *L. cuprina*.

### 4.4.2 Hybridisation and introgression

Hybridization between animal species is usually considered a rare occurrence, largely triggered by changes in geographical range, or habitat alteration due to environmental disturbance (Mallet, 2005). As a result, rates of hybridization and introgression are considered negligible. Even if hybridization is rare between two species, the few viable hybrids that do occur can potentially have significant evolutionary consequences, by allowing the transfer of alleles between species (Mallet, 2005).

The most likely species to hybridize are those that are recently diverged and closely related. For example hybridization within *Heliconius* (Lepidoptera) has been reported as most common between species with mtDNA sequence divergence of less than 2% (Mallet et al., 1998). While hybridizations between *L. sericata* and *L. cuprina* have been demonstrated to produce fertile offspring under laboratory conditions (Ullyett, 1945), no official reports of wild hybrid populations have been made (Stevens and Wall, 1996b; Stevens et al., 2002). Numerous morphological (Holloway 1991b, 1991a; Stevens and Wall, 1996a) and genetic (Stevens and Wall, 1996b, 1997b, 2001; Stevens et al., 2002; Stevens, 2003; Wallman et al., 2005; Harvey et al., 2008) studies have focused on trying to separate these ambiguous species, however only a small number
have alluded to putative hybrid populations (Stevens and Wall, 1996b, 1997b; Nelson et al., 2007; Tourle et al., 2009).

Hybridization between morphologically similar species can often be very cryptic (Mallet, 2005). By comparing molecular data from three different genes, here two *L. cuprina* specimens, from Hawaii and South Africa, have been found to contain introgressed mtDNA COX1 haplotypes from *L. sericata*. The widespread implications of hybridization in regions where *L. sericata* and *L. cuprina* occur in sympatry, become apparent when considering that both species are used in forensic entomology, to calculate post mortem interval (PMI). For example differences in life cycle and development times between these two species, even at the same temperature (O’Flynn 1983; Grassberger and Reiter, 2001), could lead to inaccurate PMI estimates when hybrid specimens are misidentified (Stevens et al., 2002; Tourle et al., 2009). In terms of adult phenotype, *L. cuprina* seems to be dominant over *L. sericata*, as indicated by the findings of Ullyett (1945) and Stevens et al. (2002). However, blowflies are most commonly used in forensic entomology in their immature stages (eggs, larvae, pupae), where morphological distinction between the *L. sericata* and *L. cuprina* is most ambiguous.

The use of sterilized *L. sericata* larvae in the treatment of non-healing wounds is becoming increasingly recognised as an efficient mainstream medical practice, particularly in light of increasing antibiotic resistance of many bacterial strains (Bunkis et al., 1985; Church, 1996; Sherman et al., 2000; Jones and Wall, 2008). While the larvae of *L. sericata* preferentially feed on dead tissue, *L. cuprina* are more aggressive feeders on living tissue, making them highly unsuitable for such a medical application. Hybrids between the two species have been reported to display intermediate or *L. cuprina* morphology (Mackerras, 1933; Ullyett, 1945; Waterhouse and Paramonov, 1950); however, backcrossing hybrids are often very difficult to distinguish morphologically from parent species, in consequence, rates of backcrossing can easily be underestimated (Mallet, 2005). In light of recognition of hybridization between *L. sericata* and *L. cuprina* in South Africa, Tourle et al. (2009) recommended that
strains used for such medical purposes be routinely genetically checked to ensure specimens are not hybrids.

*L. sericata* and *L. cuprina* also represent two key causal agents of myiasis in livestock, particularly affecting sheep. The level of sheep myiasis in Australia, home to the world’s largest wool industry, is currently controlled by the use of mulesing, a procedure involving the removal of skin from the breech. In 2004 the representative body of the Australian wool industry, Australian Wool Innovation Ltd, announced plans to phase out the practise of mulesing by 2010 (Australian Wool Innovation Limited, 2004). No single method as effective as mulesing at controlling myiasis exists, and as a result an increase in the need for insecticides is expected. One possible implication of hybridisation between *L. sericata* and *L. cuprina* is the potential introgression of insecticide resistance (Stevens *et al.*, 2002; Tourle *et al.*, 2009) as has been documented between species of black fly (Diptera: Simuliidae) (Boakye and Meredith, 1993). Recent advances towards the development of a female killing (FK) system for the control of *L. cuprina* in Australia (Scott *et al.*, 2004) could also come under threat if hybridization proves more common than expected.

### 4.4.3 Evolution of parasitism within Calliphoridae

Patterns of blowfly evolution appear to indicate that many of the parasitic blowfly groups evolved after periods of geographic isolation, undergoing local divergence and speciation (Stevens *et al.*, 2006). This is supported by the observation that even closely related taxa can exhibit very different larval feeding strategies (Stevens, 2003; Stevens and Wallman, 2006).

The majority of the taxa included in this study are facultative parasites, with obligate parasitism appearing to have developed independently several times within Calliphoridae. In the case of Auchmeromyiinae, obligate parasitism seems to have evolved twice, once by saprophagy (*Co. anthropophaga*), and once by sanguinivory (*A. luteola*). The grouping of *Compsomyiops fulvicrura*, with *Cochliomyia macellaria* and *Cochliomyia hominivorax*, is particularly interesting, as it represent a transition from saprophagy, to facultative and
obligate parasitism, between three closely related taxa. *L. bufonivora* and *L. silvarum* also represent two recently diverged sister taxa (Stevens and Wall, 1996a) that display different forms of parasitism, with *L. bufonivora* being an obligate parasite, and *L. silvarum* a suspected facultative parasite.

While two of the three saprophagic *Lucilia* species represented in this study – *L. cluvia* and *L. mexicana* – were consistently grouped together with high support by all gene data sets (Fig. 10-12), incongruence between gene data sets meant that the position of the only other saprophagic *Lucilia* species – *L. papuensis* – remained ambiguous.

Relatively little is known about the larval feeding habits of the Hawaiian genus *Dyscritomyia*. The combined-gene data set used in this study supports separate genus lineages for both *Dyscritomyia* and the saprophagic *Hemipyrellia*. However single-gene analyses differ in support of which genus is more closely related to *Lucilia*. The 28S gene tree implies that *Dyscritomyia* diverged before *Hemipyrellia*, suggesting independent origins of parasitism in *Dyscritomyia* and *Lucilia*. *COX1* also appears to support this theory, by recovering *Dyscritomyia* in a separate clade along with the saprophagic *L. cluvia*, *L. mexicana* and *Hemipyrellia*. In contrast, *EF-1α* recovers *Dyscritomyia* within *Lucilia*, albeit grouped with the saprophagic *L. papuensis*. Combining the three data sets results in a lack of resolution for the placement of *Dyscritomyia* and *Hemipyrellia*, although strong support grouping each of these genera away from *Lucilia* is found.

Classifying the larval feeding status of a blowflies can be problematic, with reports of saprophagy or secondary facultative parasitism often changing according to location within a species geographic range (Stevens and Wallman, 2006). Similarly, many of the facultative parasites included in this study could not be reliably classified as being responsible either for primary or secondary myiasis. Zumpt (1965) proposed that myiasis originates from saprophagous (feeding on decaying organic matter) or sanguinivorous (feeding on blood) larval feeding behaviour, and that facultative ectoparasitic behaviour evolves from saprophagic flies being occasionally attracted to dead or decaying tissue.
of a wounded, or immuno-compromised, living host. Through this facultative parasitic intermediate, obligate parasitism eventually develops, by which time fly larvae had become aggressive feeders on tissues of otherwise healthy hosts (Zumpt, 1965; Erzinclioglu, 1989a). If this theory is correct for Calliphoridae, as has been suggested by Stevens (2003), it seems likely that primary facultative behaviour – having the ability to initiate parasitism on an otherwise healthy animal – developed from secondary facultative behaviour – opportunistic parasitism of an existing myiasis, representing an evolutionary step towards obligate parasitism. If correct, using the taxa in this study that could be subdivided into primary or secondary causal agents of myiasis, it seems that primary facultative parasitism may have evolved multiple times within Calliphoridae (Fig 10-12).

Blowfly parasitism of vertebrates appears to be largely focused towards domesticated species, such as livestock animals. It seems very likely that the parasitism of sheep by species of Lucilia arose in direct response to the domestication of sheep and the opportunities it has created, for example increased perennial skin wrinkles providing attractive environments for larval development (Stevens et al., 2006). While most of the parasitic blowflies represented in this study have mammalian host species, commonly domesticated animals such as cattle and sheep, several taxa also reveal more specialist parasitism. Up to eight independent switches from a mammalian host are demonstrated, B. depressa (suspected parasite of termites), O. tibialis (earth worm parasite), P. rudis (earthworm parasite), L. ampullacea (frog parasite), L. bufonivora and L. silvarum (toad and frog parasites), P. azurea (bird parasite), and E. palpalis (parasite of slugs and snails). With the exception of L. bufonivora and L. silvarum all of these species are unrelated, representing separate developments of specialist parasitism.

4.4.4 Future work

While the findings presented here confirm hybrid populations in Hawaii and South Africa, more in depth taxonomic sampling, particularly from regions where these species are sympatric, is necessary to determine the true scale of
hybridisation and introgression, and to allow hypotheses of a single or multiple ancient introgressions of mtDNA haplotypes between the two species to be tested.
Chapter Five

Phylogenetic analysis of the New World screwworm fly, *Cochliomyia hominivorax*, suggests genetic isolation of some Caribbean island populations following colonization from South America.
Phylogenetic analysis of the New World screwworm fly, *Cochliomyia hominivorax*, suggests genetic isolation of some Caribbean island populations following colonisation from South America

Submitted to the journal *Medical and Veterinary Entomology*, accepted 15th October 2008, with authors as follows:

L. McDonagh¹, R. García² and J.R. Stevens¹

¹ Molecular Systematics Group, Hatherly Laboratories, School of Biosciences; University of Exeter; Prince of Wales Road, Exeter, Devon, EX4 4PS.

² IAEA Consultant, Tuxtla, Mexico

L. McDonagh undertook all DNA extractions, PCR amplification, sequence alignment, phylogenetic analyses, and all drafts of the manuscript.

R. García provided all Cuban specimens.

J.R. Stevens provided guidance throughout, and commented on earlier versions of the manuscript.

**N.B:** *This work was carried out as part of the ‘FAO/IAEA Coordinated Research Project on Enabling Technologies for the Expansion of Screwworm SIT Programmes’.*
5.1 Introduction

*Cochliomyia hominivorax*, the New World screwworm fly (NWS), is widely considered the most destructive insect parasite of livestock in the Western hemisphere's tropical and sub-tropical regions (Klassen and Curtis, 2005). With such a broad trans-boundary geographic range, historically stretching from North America down through South America, with the exception of Chile (Torres et al., 2004), *C. hominivorax* is responsible for hundreds of millions of dollars in economic loss in the livestock industry due to animal loss, decreased fertility, reduced milk, meat and wool yields, and the cost of treatment and prevention methods (Klassen and Curtis, 2005; Vargas-Teran et al., 2005).

Since its first use in Florida in 1958, sterile insect technique (SIT) has successfully been used to eradicate *C. hominivorax* from all of mainland North America and most of Central America (Klassen and Curtis, 2005). Today, *C. hominivorax* has been eradicated as far south as Panama, where an ongoing weekly release of 50 million sterile males close to the border with Colombia has created a sterile fly barrier in Darien Province to prevent the reintroduction of NWS fly from South America (Klassen and Curtis, 2005; Vargas-Teran et al., 2005). In the Caribbean, *Cochliomyia hominivorax* has also been eradicated from the Virgin Islands and Puerto Rico. However, no such programmes have been attempted on Cuba or Hispaniola, leaving reservoirs for the potential reintroduction of *C. hominivorax* into areas where the fly has been eliminated (Klassen and Curtis, 2005).

Despite the overwhelming success of the NWS fly SIT programme in the USA and Central America, significant differences in the efficacy of the technique have been apparent. For example, while initial field trials conducted in Florida resulted in almost 100% egg sterility, further field trials using the same southern USA strain of irradiated flies produced very different results on an island off the coast of Venezuela, with the release of sterile male flies resulting in observed egg sterility of just 15% (Klassen and Curtis, 2005). Most recently, a SIT programme in Jamaica in 1999 – 2005, using flies originating from Mexico, also failed (A. Robinson, International Atomic Energy Agency, pers. comm, 2007),
and the programme is currently suspended. Reasons for the failure of SIT-based control programmes in some locations are unknown, but it has been hypothesized that failure may be related to the mating incompatibility between sterile and wild flies, or to the existence of sexually incompatible cryptic species (Richardson, 1982; Richardson et al., 1982). However, the existence of cryptic species within *C. hominivorax* remains controversial (LaChance et al., 1982; Taylor and Peterson, 1994).

Recent molecular studies have provided evidence of some genetic differentiation between South, Central and North American *C. hominivorax* populations (Roehrdanz, 1989; Infante-Malachias and Azeredo-Espin, 1995; Infante-Malachias et al., 1999). Furthermore, observations of genetic homogeneity among North American samples and diversity among South American populations, have lead to suggestions of a South American origin, with a possible founder effect giving rise to the North American and Caribbean populations (Infante-Malachias et al., 1999; Azeredo-Espin and Lessinger, 2006). Considerable genetic diversity (> 2%) has also been reported between Jamaican and mainland populations of *C. hominivorax* (Roehrdanz, 1989). While the contribution this genetic differentiation may have had towards the failure of current SIT programmes in Jamaica remains unclear (Dyck et al., 2005), differences between populations at a genetic level are likely to reflect the level of genetic flow and degree of isolation. Consequently assessing the level of genetic divergence between populations of *C. hominivorax* could be a useful tool in the identification of target populations for future SIT programmes. For example, the more isolated a population is, the higher the sustainability of establishing a screwworm-free area, as the chance of reintroduction from other extant populations is lower (Robinson et al., 2009).

Comparative PCR-RFLP mtDNA studies between *C. hominivorax* and its sister taxa *C. macellaria*, have also highlighted the degree of variability seen within *C. hominivorax*. For example, while 14 different *C. hominivorax* haplotypes were identified from samples taken from around the Caribbean and South America, only a single haplotype was identified in *C. macellaria* (Taylor et al., 1996). These *C. hominivorax* haplotypes could also be subdivided into
three distinct lineages, North and Central America (including Costa Rica and Cuba), South America, and Jamaica (Taylor et al., 1996). Taylor et al. (1996) also used the mtDNA data to predict divergence times; dating the divergence of North and South American haplotypes at approximately 0.55 to 1.65 MYA, just after the continental movement and the meeting of the two land masses around 1.5 to 2.5 MYA, (Taylor et al., 1996). While total geographic conformity was not supported by these haplotypes, some support for partial phylogeographic structuring was identified, suggesting possible multiple origins of C. hominivorax populations throughout the Caribbean (Taylor et al., 1996).

**Molecular markers**

The protein-coding gene COX1 and the rRNA gene 12S represent two of the most widely used mitochondrial genes within insect systematics (Caterino et al., 2000). Mitochondrial genes can offer several advantages during phylogenetic studies including, of a lack of recombination during cell division, high copy number, relative ease of isolation, availability of universal primers, and the presence of both conserved and variable regions. Mitochondrial genes are also expected to reach reciprocal monophyly before nuclear genes, due to generally higher rates of sequence change (Avise et al., 1979; Lunt et al., 1996; Monteiro and Pierce, 2001; Funk and Omland, 2003; Dowton, 2004; Lin and Danforth, 2004), making them particularly suitable for use in this study where suspected genetic differentiation between different populations of the same species is being investigated. COX1, in particular, has previously been successfully used to confirm genetic differentiation within the oriental fruit fly Bactrocera dorsalis (Shi et al., 2005), the small tortoiseshell butterfly Aglais urticae (Vandewoestijne et al., 2004), and the Korean firefly Pyrocoelia rufa (Lee et al., 2003).

Despite the fact that far more phylogenetic analyses have used mtDNA or nuclear rDNA sequence data (Caterino et al., 2000), nuclear genes can offer several advantages over mitochondrial genes, for example by having a generally low level of biased base composition (Friedlander et al., 1992; Brower and DeSalle, 1994; Friedlander et al., 1994; Lin and Danforth, 2004). While a wide range of protein coding nuclear DNA genes have been used in insect
systematics, by far the most popular gene has been *elongation factor-1 alpha* (*EF-1α*) (Cho *et al.*, 1995; Danforth and Ji, 1998; Clark *et al.*, 2000; Caterino *et al.*, 2001; Cognato and Vogler, 2001; Buckley *et al.*, 2002; Danforth, 2002; Ghiselli *et al.*, 2007). Furthermore, while general observations have been made that nuclear genes tend to have a greater resolving power at deeper node levels of phylogenetic inference (Reed and Sperling, 1999; Leys *et al.*, 2000; Baker *et al.*, 2001; Brady, 2002; Morris *et al.*, 2002; Danforth *et al.*, 2003; Lin and Danforth, 2004), *EF-1α* has successfully resolved relationships among species groups and genera (Cho *et al.*, 1995; Belshaw and Quicke, 1997; Mitchell *et al.*, 1997; Reed and Sperling, 1999). For example, within the genus *Papilio* including *EF-1α* sequence data into analyses was found to improve resolution of relationships among species, supporting congruent relationships with the mitochondrial genes *COX1* and *COX2* (Reed and Sperling, 1999). More recently, *EF-1α* has also successfully been used alongside the mitochondrial genes *COX1* and *ITS* to investigate genetic differentiation between Eastern Atlantic Mediterranean, and Black Sea populations of the barnacles *Chthamalus stellatus*, *Chthamalus montagui*, and *Euraphia depressa* (Shemesh *et al.*, 2009). Among the three genes used in the study by Shemesh *et al.* (2009), *COX1* sequences were found to be most variable, and *ITS* sequences the least variable. Interestingly, while *EF-1α* sequences were less variable that *COX1*, they appeared the most informative in terms of population structuring, with the observed genetic divergences also corresponding to major geological transformations during the end of the Messinian and Pleiocene periods (Shemesh *et al.*, 2009). Shemesh *et al.* (2009) suggest these findings indicate different haplotype sorting for mitochondrial and nuclear genes, and recommend that both nuclear and mitochondrial markers should be sequenced, even if nuclear markers appear less variable than mitochondrial ones (Shemesh *et al.*, 2009).

In an attempt to utilise both nuclear and mitochondrial markers, the work presented in this chapter assesses genetic variation within and between populations of *C. hominivorax* using DNA sequence data from *EF-1α* (nuclear protein-coding), *COX1* (mitochondrial protein-coding) and *12S* (mitochondrial rRNA). Specimens from a range of populations (including South America, the
Caribbean, and North America (historic specimens) were sampled in an attempt to assess genetic divergence in *C. hominivorax*. 
5.2 Materials and methods

5.2.1 Samples and DNA extraction

Details of all specimens analysed in this study are given in Table 6. Briefly, specimens of *C. hominivorax* were obtained from various locations across South America and the Caribbean, including Brazil, Colombia, Cuba, the Dominican Republic, Ecuador, Jamaica, Peru, Venezuela, Trinidad and Uruguay (Fig. 14). Two potentially key islands, Cuba and the Dominican Republic, were sampled more intensively (Fig. 15). Two Calliphoridae outgroup taxa from the closely related sub-family Luciliinae (Stevens, 2003) were used.

All specimens were stored in 100% ethanol at 4°C, except the Uruguayan samples, which arrived as DNA extracts direct from the University of Campinas, Brazil, and were re-hydrated with distilled water for a minimum of 1 h prior to polymerase chain reaction (PCR) amplification. Where adult flies specimens were available, thoracic flight muscle tissues were extracted, avoiding contamination from ingested protein, parasites or eggs (Stevens, 2003), and the remainder of the adult fly kept as a voucher specimen. The entire specimen of second and third instar larvae was used. DNA extraction was carried out using a salt extraction method (Aljanabi and Martinez, 1997), see Chapter 2 for protocol. Successfully extracted DNA samples were stored at −20 °C until analysis.
<table>
<thead>
<tr>
<th>Sample site location</th>
<th>Code</th>
<th>Capture date</th>
<th>Life stage</th>
<th>COI</th>
<th>12S</th>
<th>EF1α</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cuba – East</td>
<td>A</td>
<td>Oct-05</td>
<td>L3</td>
<td>FM86778</td>
<td>FM866410</td>
<td>FM867795</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>FM86779</td>
<td>FM866409</td>
<td>FM867796</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>Oct-05</td>
<td>L3</td>
<td>FM86776</td>
<td>FM867706</td>
<td>FM867797</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>FM86777</td>
<td>FM866411</td>
<td>FM867798</td>
</tr>
<tr>
<td>Santiago de Cuba</td>
<td></td>
<td></td>
<td></td>
<td>FM86779</td>
<td>FM866409</td>
<td>FM867796</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>Nov-05</td>
<td>L3</td>
<td>FM86769</td>
<td>FM86714</td>
<td>FM867805</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>FM86768</td>
<td>FM86713</td>
<td>FM867806</td>
</tr>
<tr>
<td>Pinar del Río</td>
<td></td>
<td></td>
<td></td>
<td>FM86770</td>
<td>FM866408</td>
<td>FM867794</td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>Oct-05</td>
<td>L3</td>
<td>FM86780</td>
<td>FM866407</td>
<td>FM867793</td>
</tr>
<tr>
<td>Ciego de Ávila</td>
<td></td>
<td></td>
<td></td>
<td>FM86771</td>
<td>FM867707</td>
<td>FM867800</td>
</tr>
<tr>
<td>Cuba - Isla de la Juventud</td>
<td>E</td>
<td>Nov-05</td>
<td>L3</td>
<td>FM86774</td>
<td>FM867708</td>
<td>FM867799</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>FM86775</td>
<td>FM867707</td>
<td>FM867800</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>Nov-05</td>
<td>L2-L3</td>
<td>FM86771</td>
<td>FM867712</td>
<td>FM867803</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>FM86770</td>
<td>FM867711</td>
<td>FM867804</td>
</tr>
<tr>
<td></td>
<td>G</td>
<td>Nov-05</td>
<td>L2-L3</td>
<td>FM86773</td>
<td>FM867709</td>
<td>FM867802</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>FM86772</td>
<td>FM867710</td>
<td>FM867801</td>
</tr>
<tr>
<td>Ciro Redondo</td>
<td>H</td>
<td>Oct-05</td>
<td>A</td>
<td>FM86761</td>
<td>FM86721</td>
<td>FM867811</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>FM86760</td>
<td>FM86722</td>
<td>FM867812</td>
</tr>
<tr>
<td></td>
<td>I</td>
<td>Oct-05</td>
<td>L3</td>
<td>FM86762</td>
<td>FM86720</td>
<td>FM867814</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>FM86763</td>
<td>FM86719</td>
<td>FM867813</td>
</tr>
<tr>
<td>Duarte</td>
<td>J</td>
<td>Oct-05</td>
<td>L2-L3</td>
<td>FM86759</td>
<td>FM86724</td>
<td>FM867818</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>FM86758</td>
<td>FM86723</td>
<td>FM867817</td>
</tr>
<tr>
<td>Neiba</td>
<td>K</td>
<td>Oct-05</td>
<td>L3</td>
<td>FM86756</td>
<td>FM86726</td>
<td>FM867815</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>FM86757</td>
<td>FM86725</td>
<td>FM867816</td>
</tr>
<tr>
<td>Polo</td>
<td>L</td>
<td>Oct-05</td>
<td>L2-L3</td>
<td>FM86764</td>
<td>FM86718</td>
<td>FM867809</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>FM86765</td>
<td>FM86717</td>
<td>FM867810</td>
</tr>
<tr>
<td>Santo Domingo</td>
<td>M</td>
<td>Sep-05</td>
<td>A</td>
<td>FM86767</td>
<td>FM86716</td>
<td>FM867807</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>FM86766</td>
<td>FM86715</td>
<td>FM867808</td>
</tr>
</tbody>
</table>

Table 8: *Cochliomyia hominivorax* specimen details.
<table>
<thead>
<tr>
<th>Sample site location</th>
<th>Code</th>
<th>Capture date</th>
<th>Life stage</th>
<th>Accession number</th>
<th>COI</th>
<th>12S</th>
<th>EF1α</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brazil</td>
<td>Minas Gerais</td>
<td>Nov-00</td>
<td>A</td>
<td>FM867752</td>
<td>FM866402</td>
<td>FM867788</td>
<td></td>
</tr>
<tr>
<td>Sao Paulo</td>
<td></td>
<td></td>
<td></td>
<td>FM867753</td>
<td>FM867731</td>
<td>FM867823</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>FM867746</td>
<td>FM867736</td>
<td>FM867829</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>FM867747</td>
<td>FM867737</td>
<td>FM867828</td>
<td></td>
</tr>
<tr>
<td>Jamaica</td>
<td></td>
<td>Jan-05</td>
<td>A</td>
<td>FM867743</td>
<td>FM867740</td>
<td>FM867832</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>FM867742</td>
<td>FM867741</td>
<td>FM867833</td>
<td></td>
</tr>
<tr>
<td>Trinidad</td>
<td></td>
<td>Aug-05</td>
<td>L3</td>
<td>FM867746</td>
<td>FM867736</td>
<td>FM867829</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>FM867747</td>
<td>FM867737</td>
<td>FM867828</td>
<td></td>
</tr>
<tr>
<td>Venezuela/Colombia border</td>
<td></td>
<td>May-05</td>
<td>L3</td>
<td>FM867743</td>
<td>FM867740</td>
<td>FM867832</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>FM867742</td>
<td>FM867741</td>
<td>FM867833</td>
<td></td>
</tr>
<tr>
<td>Uruguay</td>
<td></td>
<td></td>
<td>DNA extract</td>
<td>FM867744</td>
<td>FM867739</td>
<td>FM867830</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>FM867745</td>
<td>FM867738</td>
<td>FM867831</td>
<td></td>
</tr>
<tr>
<td>Chigorodo</td>
<td>Choco</td>
<td>Mar/Apr 07</td>
<td>L2-3</td>
<td>FM867783</td>
<td>FM866404</td>
<td>FM867790</td>
<td></td>
</tr>
<tr>
<td>Colombia</td>
<td>Turbo</td>
<td></td>
<td></td>
<td>FM867782</td>
<td>FM866405</td>
<td>FM867791</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Napo</td>
<td></td>
<td></td>
<td>FM867787</td>
<td>FM866406</td>
<td>FM867792</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Esmeraldas</td>
<td></td>
<td></td>
<td>FM867786</td>
<td>FM867729</td>
<td>FM867821</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Guayas</td>
<td></td>
<td></td>
<td>FM867755</td>
<td>FM867727</td>
<td>FM867819</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>FM867754</td>
<td>FM867728</td>
<td>FM867820</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>FM867749</td>
<td>FM867734</td>
<td>FM867826</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>FM867748</td>
<td>FM867735</td>
<td>FM867827</td>
<td></td>
</tr>
<tr>
<td>Peru</td>
<td>Chiclayo</td>
<td>May-07</td>
<td>L2-3</td>
<td>FM867749</td>
<td>FM867734</td>
<td>FM867826</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Tambopata</td>
<td></td>
<td></td>
<td>FM867748</td>
<td>FM867735</td>
<td>FM867827</td>
<td></td>
</tr>
<tr>
<td>USA</td>
<td>Texas</td>
<td>1953</td>
<td>A</td>
<td>FM867751</td>
<td>FM867732</td>
<td>FM867825</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1933</td>
<td></td>
<td>FM867750</td>
<td>FM867733</td>
<td>FM867824</td>
<td></td>
</tr>
</tbody>
</table>

Table 6: *Cochliomyia hominivorax* specimen details.
Figure 14: Sampling locations within South and Central America.

5.2.1.1 ‘Historical’ North American samples and DNA extraction

Two specimens belonging to historical Texan populations of *C. hominivorax*, on loan from the collection held at the Natural History Museum, London, were also sampled. DNA extractions from these specimens, which had been dried and pinned in 1933 and 1953, were carried out using a DNeasy® Blood & Tissue Kit (Qiagen Ltd., Crawley, U.K.). DNA extraction was performed by removing thoracic flight muscle from the ventral side of the thorax, leaving as much of the taxonomically important morphology intact as possible. To ensure that DNA obtained was that of the historic specimen and not a result of contamination from contemporary *C. hominivorax* material, all molecular work for the historical samples was carried out in a separate ‘clean’ room, using a separate set of pipettes, etc. Extracted DNA was stored at −20 °C until analysis.

5.2.2 DNA sequences

Each DNA extraction was subjected to PCR amplification of an approximately 680-bp region of the mitochondrial *COX1*, an approximately 780-bp region of *EF-1α* and a 600–700bp region of the mitochondrial 12S gene. Using published primers (Otranto *et al.*, 2005a), *COX1* sequence data were obtained from approximately half of the Cuban and Dominican Republic samples. Accordingly, a new set of *COX1* primers were designed (two forward, two reverse, Table 7) in an attempt to amplify the remaining templates. Initial attempts to utilize more generic arthropod *EF-1α* primers (provided by J.K. Moulton, pers. comm, 2004) proved unsuccessful and a new set of primers specific for *C. hominivorax* were designed. All 12S sequences were amplified using published primers (Lessinger and Azeredo-Espin, 2000), although successful PCR product sequencing was consistently only obtained for the reverse primer (*SR-J-14233*). At least two sequences were amplified for each gene from each of the locations (Table 6). Details of all published primers are given in Chapter 2.
Table 9: Details of primers used to amplify EF-1α and COX1.

See Chapter 2 for details of PCR amplification protocol. All PCR reactions using historical North American material were carried out in a separate, clean room to avoid possible contamination.

The resulting PCR products were separated by gel electrophoresis, see Chapter 2 for details, with appropriate bands cut out and purified using a MinElute® PCR purification kit (Qiagen UK), according to the manufacturer’s protocol. Purified PCR products were sequenced in both directions on a gold standard ABI 3730xl DNA Analyzer platform, using the commercial sequencing facility COGENICS (formally Lark Technologies Inc.).

Sequence fragments were checked for quality and edited manually before being assembled into a single consensus sequence, using AutoAssembler 2.0 (Applied Biosystems, Inc.). Any ambiguities in the consensus sequence were then resolved, or standard IUPAC/IUB codes (Leonard, 2003) used where the nucleotide could not be read from the chromatogram, or if polymorphisms in amplicon sequencing were suspected. Edited consensus sequences were then exported into a text file for alignment.
5.2.3 Sequence alignment and phylogenetic analysis

All multiple sequence alignments were carried out using ClustalX (Thompson et al., 1997), with final adjustments performed by eye. Phylogenetic analyses were carried out using PAUP* 4.0, (Swofford, 1998), with both maximum parsimony (MP) and maximum likelihood (ML) criteria being used to infer trees for individual gene datasets and a combined dataset.

Parsimony tree inference was employed using an equal weighting scheme, gaps treated as missing data, tree bisection and reconnection (TBR) swapping, and random sequence addition, with 1000 replicates per search. The parsimony-based partition homogeneity test (incongruence length difference [ILD] test) (Farris et al., 1995) was used to evaluate the phylogenetic congruence between the genes and to verify whether the three datasets were combinable.

For ML analyses, parameters produced from a series of nested hierarchical likelihood-ratio tests using MODELTEST Version 3.06, (Posada and Crandall, 1998), with heuristic searching (for 100,000 rearrangements), TBR branch swapping and random sequence addition with 10 repeats, were used to implement the most appropriate model of evolution for each of the three genes and the combined dataset separately. Node support for all tree topologies was assessed using bootstrap analysis, using 100 replicates for ML analyses (restricted for computational time limits) and 1000 replicates for all MP analyses, with additional Bremer decay indices (Bremer, 1988) calculated on all parsimony trees.

When differences between taxa are very small, a reduction in the accuracy of algorithmic-based distance methods can result in incorrect tree topology. By comparison, methods based on optimality criteria, such as parsimony and ML, provide the advantage of choosing one tree over another based on a mathematical ranking system, whereas distance methods cluster taxa according to the conventions of the particular algorithm being used, although even then the best tree may not be found (Swofford and Sullivan,
Parsimony has also been shown to be valuable under circumstances where rates of evolution are low (Swofford and Sullivan, 2003), such as between gene sequences from populations of the same species. Accordingly, data were analysed by both parsimony and ML methods, with detailed analysis of inter-gene phylogenetic congruence and levels of relative support for clades performed using parsimony-based methods.
5.3 Results

5.3.1 Phylogenetic analysis: EF-1α

The EF-1α sequence dataset consisted of 743 aligned nucleotide sites, of which 50 were found to be parsimony informative. Parsimony analysis of the EF-1α data found a large number of most-parsimonious trees (MPTs) (L = 101). A strict consensus tree resulted in a loss of all resolution within the Caribbean and South American NWS taxa, but grouped together both USA samples, albeit with no node support.

The best-fit ML model for the EF-1α dataset was found to be a Hasegawa-Kishino-Yano (HKY) (Hasegawa et al., 1985), including invariable sites (+I), and rate variation among sites (+Γ) (base frequencies = [A 0.2339; C 0.3010; G 0.2259]; number of substitution types = 2; T ratio = 1.6875; rate = Γ; α = 0.8377; proportion invariant sites = 0.6982). Two ML trees were found for the EF-1α dataset and node support was evaluated using bootstrap values. Strong support for monophyly of the NWS samples was found, although intra-specific bootstrap support was very low.

5.3.2 Phylogenetic analysis: 12S

The 12S sequence data contained 1239 characters, of which just 53 were parsimony informative sites. A strict consensus tree of MPTs (L = 409) resulted in a lack of meaningful resolution within the NWS taxa.

The best-fit ML model for the 12S dataset was found to be Felsenstein 81 (F81) (Felsenstein, 1981), including rate variation among sites (+Γ) (base frequencies = [A 0.4230; C 0.1164; G 0.0689]; number of substitution types = 1; rate = Γ; α = 0.6480).

The ML analysis found a single tree which displayed a lack of resolution for the South American and Caribbean samples, but grouped the two USA samples together, with a strong bootstrap value of 99%.
5.3.3 Phylogenetic analysis: COX1

The COX1 sequence data, although the smallest dataset contained 78 parsimony informative sites. Moreover, by contrast with the EF-1α and 12S datasets, a strict consensus of all MPTs (L = 274) did conserve some resolution within the screwworm clade (Fig. 3), with a distinct 10-taxon ‘Cuban clade’ (with the exception of one Brazilian sample) preserved, although with relatively low bootstrap support.

The best-fit ML model for the COX1 dataset was found to be a transitional model (TIM), including invariable sites (+I), and rate variation among sites (+Γ) (base frequencies = [A 0.3083; C 0.1625; G 0.1511]; number of substitution types = 6 [generalized time-reversible]; r matrix = [A – C 1.0000; A – G 3.3693; A – T 2.1386; C – G 2.1386; C – T 7.5244]; rate = Γ; α = 0.7496; proportion invariant sites = 0.4459).

Two ML trees were found, again with the noticeable Cuban clade in both topologies. Mapping bootstrap values on to one of the ML trees revealed strong support for the Cuban clade (Fig. 16).

5.3.4 Combined dataset

The parsimony-based ILD test, performed on a partitioned combined three-gene (COX1, 12S and EF-1α) dataset, revealed a significant difference (P < 0.01, n = 1) in topology between the genealogies. This suggests the data partitions have significantly different signals under a 95% significance level, and as such are not compatible. However, in an attempt to identify clades which are congruent between datasets, a combined dataset analysis was performed.

Maximum parsimony analysis of the combined data resulted in MPTs with a length of 924, with a strict consensus preserving an even larger 12-taxon Cuban clade, including the Minas Gerais taxon from Brazil (not shown); this Cuban grouping received high Bremer support in the combined dataset. Throughout the combined tree, regions of apparent incongruence between the
genealogies (as implied by the ILD test) are identified by negative Bremer support values, indicating where the data partition did not support a particular node. From this it is possible to see that the Cuban clade is not one of the areas of conflict between genealogies, although some of the topology within this clade does receive negative Bremer values. The large numbers of MPTs found in all datasets (combined and single-gene) reflect the largely uniform nature and lack of resolving power of these sequence data for *C. hominivorax*.

The best-fit ML model was found to be a generalized time reversible (GTR) (Tavaré, 1986) including invariable sites (+I), and rate variation among sites (+Γ) (base frequencies = [A 0.3418; C 0.1735; G 0.1361]; number of substitution types = 6; r matrix = [A – C 1.9887; A – G 2.7059; A – T 2.7872; C – G 1.0890; C – T 4.6986]; rate = Γ; α = 0.6723; proportion invariant sites = 0.5283).

Conversely, node support on the single ML tree, found during the likelihood analysis (Fig. 17) resulted in a loss of support for the original 10-taxon Cuba grouping. However, grouping of the samples from Texas away from the remaining South American and Caribbean samples is very strongly supported (Fig. 17), suggesting a clear North-South genetic divide.
Figure 16: A rooted phylogram constructed by maximum likelihood analysis of COX1 nucleotide sequence data from a variety of populations of *C. hominivorax* from around the Caribbean, Central and South America. Two *Lucilia* species are used as outgroup taxa. Bootstrap values are shown at each node receiving >50% bootstrap support. This topology reveals moderate support for a clade of Cuban samples (also containing a single Brazilian sample), suggesting possible genetic sub-structuring within *C. hominivorax*. While other Cuban samples are found elsewhere in the tree, their position is not well supported.
Figure 17: An unrooted phylogram constructed by maximum likelihood analysis of the combined three-gene (COX1, EF-1α, 12S) data set sampled from a variety of populations of *C. hominivorax* from around the Caribbean, Central and South America. Two *Lucilia* species are used as outgroup taxa. Bootstrap values are shown at each node receiving >50% bootstrap support. This topology reveals strong support separating the two Texan (North American) samples from all other populations, suggesting a strong North-South genetic divide within *C. hominivorax*. 
5.4 Discussion

5.4.1 Regional genetic variability of NWS fly populations

A small group of Cuban samples can be seen in both COX1 (Fig. 16) and multi-gene (Fig. 17) parsimony-based genealogies, although node support for this grouping is only found in the COX1 tree. While this clade may suggest that some specimens from geographically diverse populations appear to show a greater affinity for one another than they do for flies from other locations, this small Cuban clade was the only major grouping within the South American and Caribbean samples to be supported by any of the gene trees. However, one notable anomaly within this Cuban clade is the presence of a Brazilian sample (Minas Gerais). Whereas the remaining C. hominivorax taxa appear indiscriminately placed within the tree, this Minas Gerais sample appears strongly associated with Cuban screwworm populations. This may be the result of one of a number of factors, including: historical lineage sorting; transportation of fly-infested livestock, stochastic convergent evolution (in view of the low number of informative characters identified), phylogenetic error (where inaccurate relationships have been recovered by poorly resolved trees), or laboratory error.

However, by far the most striking relationship found in this study was that of the strongly supported separation of the two Texan samples from the South American and Caribbean samples, suggesting a possible north/south divide. Previously, a nucleotide divergence of > 2% was observed between Jamaican and Mexican samples (Roehrdanz, 1989) and it has been suggested that such a result supports a north/south differentiation, with Jamaican populations originally colonized from South America (Azeredo-Espin and Lessinger, 2006).

A previous mtDNA-based study by Taylor et al. (1996), which incorporated the findings of Roehrdanz (1989), compared C. hominivorax samples from around the Caribbean, Central and South America, and showed haplotypes to subdivide into three lineages. The lineages identified by Taylor et al. (1996) were ‘North and Central America’ (which contained Costa Rican and
Cuban samples, the only significant clade in their cladogram, with bootstrap support of 90%), ‘South America’, and ‘Jamaica’ (with the Jamaican and one of the Trinidad samples sharing a comparatively different haplotype of their own), although total geographic conformity was not supported and only some support for partial phylogeographic structuring was present. Although samples from the Caribbean in the study by Taylor et al. (1996) were limited, the authors noted that the single Cuban sample appeared to be more closely related to Central American populations, whereas the two Dominican Republic samples grouped with those from South America, suggesting the possibility of multiple origins of *C. hominivorax* throughout the Caribbean.

Several phylogenetic studies have reported evidence of distinct Cuban populations of *C. hominivorax*, which appear genetically divergent from other populations in the Caribbean (Taylor et al., 1996; Lyra et al., 2009; Torres and Azeredo-Espin, 2009). Significantly, while the results presented in this chapter provide some evidence supporting this distinct Cuban clade, albeit poorly supported, they do not link this Cuban group with a North American lineage, as previously hypothesized by Taylor et al. (1996).

### 5.4.2 Mitochondrial vs. nuclear genealogy

Population studies using mtDNA have commonly used sequence data from a single gene, with *COX1* one of the most commonly used (Liu et al., 2006). It is now widely recognized, however, that the use of single-gene phylogenies (especially mtDNA gene phylogenies) to represent overall species evolution can frequently be misleading (Stevens and Wall, 1996b; Stevens et al., 2002; Wells et al., 2007) and multi-gene phylogenies are now recognized as essential. Where intra-specific relationships based on single-gene sequences are sometimes poorly resolved (Shao and Barker, 2007), multiple concatenated nucleotide sequences may contain sufficient genetic differentiation to resolve such relationships. Similarly, combined molecular and morphological studies have identified the possible existence of distinct geographical ‘races’ in the Old World screwworm fly, *Chrysomya bezziana* (Hall et al., 2001).
Chapter 5  

Phylogenetic analysis of Cochliomyia hominivorax

The findings of this study have shown that the range of different molecular markers used (nuclear protein-coding, mitochondrial protein-coding, and mitochondrial rRNA) demonstrate very different powers of resolution. As may be expected from the slowest evolving marker, the nuclear gene EF-1α failed to support any genetic sub-structuring within C. hominivorax. This contradicts the findings of Shemesh et al. (2009), who reported that EF-1α sequence data was able to support genetic differentiation within several species of barnacle. However, these different conclusions are most likely due to differences in dates of divergence between populations, with C. hominivorax populations diverging being much more recently, compared to the populations of barnacles. This theory may be supported by the fact that one of the mitochondrial markers (12S) also failed to recover population sub-structuring within C. hominivorax. The COX1 sequence data, however, do appear to support some geographic differentiation within C. hominivorax, although with only moderate node support. While the phylogenetic signal in the COX1 data may be explained as the result of the uni-parental inheritance and non-recombining nature of the mitochondrial gene, the lack of similar signal in the 12S data suggests that some regions of the mtDNA are experiencing faster rates of evolution than others. Further mtDNA based phylogenetic analyses of C. hominivorax will provide an interesting insight into whether such intra-specific relationships are universal throughout the C. hominivorax mitochondrial genome.

5.4.3 General findings of the FAO/IAEA research project

The work presented in this chapter forms part of the ‘FAO/IAEA Coordinated Research Project on Enabling Technologies for the Expansion of Screwworm SIT Programmes’. This collaborative project focused on the screwworm species, Cochliomyia hominivorax (the New World screwworm), Chrysomya bezziana (the Old World screwworm), and Wohlfahrtia magnifica. Three other studies within the project focused on the population genetics of C. hominivorax, using microsatellite data (Griffiths et al., 2009; Torres and Azeredo-Espin, 2009) and mitochondrial DNA markers (Lyra et al., 2009).
Griffiths *et al.* (2009) used four microsatellite loci to study populations of *C. hominivorax* from Brazil, Venezuela, Jamaica, and Trinidad. The findings of this study suggested single inter-breeding populations on the two Caribbean islands, Trinidad and Jamaica, but no distinct population sub-structuring on the South American mainland populations. Microsatellite data from other insect species have revealed non-endemic island populations to have up to two fold lower level of genetic diversity than mainland populations (Bonizzoni *et al.*, 2000; Michel *et al.*, 2005), due to small founder numbers, lower populations size after foundation, and restricted gene flow with other populations (Torres and Azeredo-Espin, 2009). However, Torres *et al.* (2009), also using microsatellite data, reported high levels of genetic diversity in Caribbean island populations, including Cuba, Dominican Republic, Jamaica, and Trinidad and Tobago. Possible explanations for this high genetic variation included multiple successive colonisations of Caribbean islands from mainland populations, large founder populations and no severe bottleneck, or a rapid increase in population size after colonisation. Additionally, Torres *et al.* (2009) also challenged the general view that *C. hominivorax* were first introduced to the Caribbean islands with domesticated animals, and have since remained genetically isolated. Instead suggesting that the populations may have been on the islands longer than previously thought, which could also account for the high levels of genetic diversity. In contrast, using mtDNA PCR-RFLP data, Lyra *et al.* (2009) found distinct population structure and low variability among Caribbean island populations, proposing that each island represents an independent evolutionary unit, connected by a restricted gene flow. Mainland South America populations were found to have, low, but significant, population structure and high variability, supporting a South American origin of *C. hominivorax*. Lyra *et al.* (2009) suggested that this conflict between nuclear (Torres and Azeredo-Espin, 2009) and mtDNA diversity could be due to either, small founder female populations followed by rapid increase in population size recovering nuclear diversity but not mitochondrial diversity, or multiple colonisations with the same mtDNA haplotype.
Summary

Both nuclear (Griffiths et al., 2009; Torres and Azeredo-Espin, 2009) and mtDNA (Lyra et al., 2009) have suggested that Caribbean islands represent independent evolutionary units, connected by limited gene flow. In agreement with the findings presented in this chapter, both Lyra et al. (2009) and Torres et al. (2009) reported Cuban populations to be the most distinct from the other islands. For example, Lyra et al. (2009) identified one particular mtDNA haplotype dominant in Cuban populations that seems very different to others from South America and other Caribbean islands. Further studies are being carried out to investigate whether this haplotype is unique to Cuba (Lyra et al., 2009). Lyra et al. (2009) suggests that this differentiation of Cuban populations from other islands indicates a North American origin, with other islands having a South American origin. However, the work presented in this chapter, being unique to the project by including North America samples, does not support association between Cuba and North America.

Of the Caribbean islands, only the Virgin Islands and Puerto Rico have successfully eradicated C. hominivorax to date, although the Jamaican government initiated a SIT programme in 1999. However, no such programmes have yet been undertaken on Cuba or Hispaniola, with the result that these islands harbour reservoirs of C. hominivorax, this could potentially lead to the reintroduction of the fly to regions across the Caribbean and Central America that are currently screwworm-free (Klassen and Curtis, 2005). Population genetic data studies have suggested that Cuba is distinct from all other C. hominivorax populations; however whether this genetic divergence reflects differential mating types that affect the success of SIT is unclear. While population genetic data can help identify levels of gene flow between populations, it does not directly indicate whether the populations are compatible reproductively.

While the findings of the FAO/IAEA research project suggest that there is likely considerable population genetic sub-structuring within C. hominivorax, in order to directly assess how this may affect on-going and future SIT programmes extensive mating compatibility studies are needed. Indeed
preliminary data from mating studies comparing wild Jamaican and Cuban flies with the current release strain have observed evidence of asymmetric mating, with wild males refusing to mate with released females (R. Garcia, unpublished data, quoted in Vreysen et al., [2006]).
Chapter Six

An investigation into the diversity of antigenic larval proteins, concentrating on four key blowfly species, *Calliphora vicina*, *Lucilia caesar*, *Lucilia cuprina* and *Lucilia sericata*. 
A sub-section of the following chapter was submitted to the journal *Forensic Science International: Genetics*, and accepted 9th February 2009.

(Appendix A).

**Development of an antigen-based rapid diagnostic test for the identification of blowfly (Calliphoridae) species of forensic significance.**

L. McDonagh ¹, C. Thornton ¹, J.F. Wallman ², J.R. Stevens ¹

¹ Hatherly Laboratories, School of Biosciences, University of Exeter, Prince of Wales Road, Exeter, EX4 4PS, UK

² Institute for Conservation Biology, School of Biological Sciences, University of Wollongong, New South Wales 2522, Australia

L. McDonagh undertook all sample preparation, gel electrophoresis, western blotting, EM work, and drafted and led on writing the manuscript.

C. Thornton provided guidance throughout, and commented on drafts of the manuscript.

J.F. Wallman reared, provided and prepared all Australia specimens.

J.R. Stevens provided guidance throughout, commented on drafts and helped write the manuscript.
Chapter 6  
An investigation into the diversity of antigenic larval proteins

6.1 Introduction

6.1.1 The evolution of myiasis within Calliphoridae

Due to the parasitic nature of myiasis-causing larvae, it is likely that their evolution has been significantly affected by the immune response of the host, with different parasitic strategies having developed a multitude of defences against both specific (B-cell and T-cell) and non-specific (natural killer cell) immune responses (Casu et al., 1994; Stevens and Wallman, 2006). Understanding the myiasis induced host immune responses themselves, and the ways in which larval strategies have evolved to cope with these defences, will in turn facilitate the development of diagnostic tools for use in early detection of myiasis in livestock, along with vaccine development (Boulard et al., 1996; Otranto, 2001; Stevens et al., 2006).

Lucilia

*Lucilia sericata* and *Lucilia cuprina* represent two primary facultative ectoparasites of the greatest medical and veterinary significance worldwide, being the prime cause of sheep ‘fly strike’ in cool temperate habitats (Europe and New Zealand) and sub-tropical/warm temperate habitats (Australia and South Africa) respectively (Stevens and Wall, 1997a). Due to its impact on some of the largest wool industries in the world, *Lucilia cuprina* in particular has been the focus of much research into improvement of fly strike resistance and vaccine development (East and Eisemann, 1993; Casu et al., 1997; O'Meara et al., 1997; Tellam and Bowles, 1997; Tellam and Eisemann, 1998; Reed et al., 1999). The third *Lucilia* species, *Lucilia caesar*, is considered largely to be a northerly Palaearctic species (Brinkmann, 1976), and is a secondary facultative ectoparasite of sheep (Stevens and Wall, 1997a).
Calliphora

In the UK, Calliphora vicina, a synanthropic species largely considered to be Holarctic in distribution (Rognes, 1991; Hwang and Turner, 2005; Whitworth, 2006), is one of the most common forensically important blowflies (Ames et al., 2006).

6.1.2 The cycle of myiasis in sheep

Hypersensitivity in sheep

First instar calliphorid larvae are able to induce rapid local inflammatory responses in sheep within 1-2 hours of contact (Broadmeadow et al., 1984). These rapid local inflammatory responses, along with shredding of epidermal cells, larval proteolytic enzyme secretion, and oral spine action, work to initiate wound development on the skin of infested sheep (Sandeman et al., 1987; Sandeman, 1990). While inflammatory responses appear to form a fundamental part of wound development, it remains unclear whether such responses impede or actually facilitate larval growth.

Second and third instar larval stages use a combination of large mouth hooks and secreted proteases to break down host tissue. In contrast, the much smaller, and more vulnerable, first instar stages are more dependent on their excretory-secretory (ES) products to damage tissue and initiate lesions. However, larval ES products have been shown to be only weakly antigenic in sheep (Tabouret et al., 2001), suggesting that merely contact with the larval cuticle is sufficient to induce the hypersensitivity response observed.

Cell-mediated immune responses

Approximately 48 hours after initiation of myiasis by L. cuprina, significant cellular immune responses can be observed, with increased levels of granulocytes (neutrophils and eosinophils) being found primarily on the wound
Chapter 6  An investigation into the diversity of antigenic larval proteins

surface, and increased concentration of T-cells (CD4⁺, CD8⁺ and γδ) on the surface of the dermis, (Bowles et al., 1992; Elhay et al., 1994).

**Antibody-mediated (humoral) immune responses**

Approximately 96-120 hours after infection an increase in B-cell production is observed, with responses in secondary myiasis being greater than those in cases of primary myiasis (Bowles et al., 1994).

**6.1.3 Immunomodulation of host responses**

In addition to degrading antibodies, larval excretory-secretory (ES) products of *L. cuprina* have also been observed to be involved in host immunomodulation, significantly reducing antibody production and weakening T-cell responsiveness (Kerlin and East, 1992). Larval immunosuppression has also been attributed to excretion of ammonia, which can cause a reduction in level of mature neutrophils, eosinophils and lymphocytes, as well as down-regulating humoral responses (i.e. inhibiting serum globulin production) (Guerrini, 1988, 1997; Colditz et al., 2002). The role immunomodulation plays in the suppression of naturally acquired immunity remains unclear, and indeed the systems behind such immune evasion mechanisms in *L. cuprina* have yet to be studied in detail (Elkington and Mahony, 2007).

By comparison, parasite modulation of host immune response is particularly well established in Oestridae (Stevens and Wallman, 2006). For example, inflammatory responses in myiasis induced by *Hypoderma* (Hypodermosis) are suppressed by action of enzymes secreted by the larvae (Otranto, 2001). Furthermore, within *L. cuprina* trypsin-like larval ES enzymes have been discovered showing a significant amino acid sequence similarity (42-45% homology) with the *Hypoderma* trypsin proteases hypodermin A (*HA*) and hypodermin B (*HB*), both of which are involved in larval defence against host immune responses (specific and non-specific) (Casu et al., 1994). More specifically, *HA* and *HB* stimulate breakage of α and β chains of complement C3, which in turn inhibits the inflammatory response in primary myiasis (Boulard,
1989). HA also inhibits lymphocyte production (i.e. preventing specific cellular responses, and essentially hindering ‘immune memory’ systems) (Chaubadie and Boulard, 1992), and degrades bovine IgG, inhibiting its biological function by cleaving molecules into antigen-binding (Fab) and antibody Fc domains (Pruett, 1993). However, while no inflammatory responses are seen in cases of Hypoderma-induced myiasis, acute inflammatory responses are observed in calliphorid myiasis.

6.1.4 Levels of resistance to sheep strike

While no naturally acquired total immunity against fly strike seems to occur, the potential that repeated infestation can improve resistance has been implied by observations of greater resistance in some older sheep (Watts et al., 1979; Sandeman et al., 1985; Sanders, 1986; Sandeman, 1990; Bowles et al., 1996; Elkington and Mahony, 2007). Indeed, immunisation trials carried out by Bowles et al., (1996) investigating several different L. cuprina larval antigens, including one linked with the surface proteins of the first instar larvae, observed particularly high induced levels of protection. Notably, it seemed that a rapid cellular-mediated response – triggered within hours of first instar larvae coming into contact with the sheep’s skin – seemed responsible for the observed protection. The speed of this cellular immune response suggested ‘hypersensitivity’ to the first instar larvae. More specifically, links have been made with the stimulation of type I (IgE-mediated, anaphylactic or atopic response) and type III (Arthus-type, immune complex-mediated) hypersensitivity responses to increased levels of resistance in sheep (Sanders, 1986; Bowles et al., 1987). In particular, type III Arthus-type responses are found to be notably greater in sheep with an induced resistance due to repeated infestation.

In order to explore immune responses to fly strike further, specially bred sheep lines for ‘resistance’ (R) and ‘susceptibility’ (S) to fleece rot and blowfly strike have been reared since the mid 1970’s at the Trangie Agricultural Research Centre (TARC), Australia. Subsequent immunization trials using the Trangie flocks have reported considerable variation in the secretion of wound exudate between R and S flocks, with R sheep producing exudate more rapidly
after the primary contact with the *L. cuprina* first instar larvae, and also showing increased levels of IgG, fibrinogen and complement C3 (Sanders, 1986; O'Meara *et al.*, 1997). However, observations that titres of serum antibody did not differ between R and S flocks, and of inconsistency between individuals in level acquired protection after repeated induced infestations, both suggest that the innate immune response is key to the acquisition of fly strike resistance (MacDiarmid *et al.*, 1995; O'Meara *et al.*, 1997). Furthermore, any increased level of protection detected did not seem to last beyond the period of repeated infestation, again suggesting a lack of ‘immunological memory’ in the responses involved in resistance (Sandeman, 1990).

### 6.1.5 Methods for control of myiasis

Current methods for managing myiasis in sheep include the use of sheep dips (i.e. insecticides), animal husbandry practices (e.g. mulesing, tail docking, etc), bait trapping, and biological control agents. However, the approaching 2010 phase out of mulesing in Australia (Australian Wool Innovation Limited, 2004) is expected to be followed by a significant increase in level of fly strike in sheep. The economic significance of this becomes apparent when the current cost of myiasis to the Australian wool industry, the largest in the world, of over AUS$ 250 million (http://www.wool.com.au, 2002) is considered.

### 6.1.6 Immunisation trials and vaccine development

Currently no commercially viable vaccines exist for the control of myiasis, despite considerable resources having been assigned to their development (Elkington and Mahony, 2007). One of the largest obstacles in the development of a vaccine has been the apparent lack of naturally acquired immunity to fly strike in sheep, even after repeated infestations (Elkington and Mahony, 2007). It has been argued that this apparent lack of acquired immunity is due to the relatively short duration of calliphorid myiasis infestations, allowing insufficient time for exposure of larval antigens to the host’s immune system to build an effective response (Tellam and Bowles, 1997). However, myiasis typically consists of successive and overlapping infestations, prolonging the overall
exposure time to larval antigens (Elkington and Mahony, 2007). Furthermore, significant immune responses are observed in sheep (O’Donnell et al., 1980; Bowles et al., 1992) and yet protective immunity from subsequent infestations does not occur (Elkington and Mahony, 2007).

Many immunisation attempts have been made to try and increase serum levels of immunoglobulin G (IgG) directed at blowfly larval antigens, including the use of repeated whole larvae infestations and immunisation with antigens from crude larval and excretory-secretory preparations (Bowles et al., 1987), immunisation with larval cuticle proteins (Barrett and Trevella, 1989), purified serine proteases (Tellam et al., 1994), and extracts of the larval peritrophic membrane (Tellam and Eisemann, 1998; Colditz et al., 2002). While in vitro feeding experiments generated promising results, with significant retardation in larval growth, in vivo results failed to match up, providing insufficient protection from subsequent myiasis infestations, thought to be due to much lower serum IgG levels present in vivo (Elkington and Mahony, 2007).

On closer investigation in vitro disruption of larval growth was found to be caused by blockage of the peritrophic membrane, preventing nutritional intake to the midgut epithelial cells and eventual starvation (Tellam et al., 1994). This could also explain why such vaccines failed to work successfully in vivo, as the larval systems of extracellular digestion in the lumen and subsequent absorption of nutrients across the peritrophic membrane, also act to protect the digestive epithelial cells, protecting them from contact with the ingested antibodies and preventing intracellular damage (Barrett and Trevella, 1989). Perhaps more significantly, larvae are also able to degrade the IgG antibodies themselves using secreted tryptic and chymotryptic enzymes, with a large proportion of wound exudate at the site of myiasis being found to consist of degraded IgG (Sandeman et al., 1995). The extent of protection against host immune responses afforded by the ability to break down IgG antibodies during natural occurrences of myiasis is not fully understood (Elkington and Mahony, 2007). For example, while up to 60% of IgG present in wound exudate is thought to be degraded within 6 hours of initial infection, resulting degraded Fab (antigen binding domains) and heavy chain variable regions of antibodies are
still able to damage the peritrophic membrane (Sandeman et al., 1995). Furthermore, while first instar larvae are capable of creating a wound within 8 hours of hatching (Sandeman et al., 1987; Young et al., 1996), serum IgG levels do not peak until 72 hours after the onset of myiasis (Barrett and Trevella, 1989; Bowles et al., 1994). By this time the larvae have developed into third instar and as such are much less susceptible to action by antibody than the smaller, more vulnerable, first instar stages (Eisemann et al., 1990; East et al., 1992; Johnson and Moutsopoulos, 1992).

It seems that more successful vaccine development programmes should share a combined focus on blocking parasite immune evasion defences, inducing rapid host cellular immune responses (more specifically hypersensitivity responses), targeting antigens involved in wound initiation, and inhibiting larval nutrition and/or growth (Elkington and Mahony, 2007).

6.1.7 DNA-based species identification of blowfly species

6.1.7.1 Phylogenetic approach

At present, DNA sequence-based analysis is the only routinely used molecular-based species identification tool for Calliphoridae (Wells and Stevens, 2008). Several loci have been advocated for species identification through phylogenetic inference, whereby the position of unknown specimens within a phylogenetic tree reveals the species taxonomy (Stevens and Wall, 2001; Wallman and Donnellan, 2001; Wells and Sperling, 2001). The most commonly used markers include: regions of the mitochondrial cytochrome oxidase subunits I (COX1) and II (COX2) (Malgorn and Coquoz, 1999; Wells and Sperling, 2001; Ratcliffe et al., 2003; Saigusa et al., 2005; Wells and Williams, 2007; Ying et al., 2007; Nelson et al., 2008; Song et al., 2008), and the nuclear internal transcribed spacers (ITS) (Malgorn and Coquoz, 1999; Harvey et al., 2008). Mitochondrial DNA (mtDNA) in particular is favoured due to its high copy number; ease of isolation; and conserved sequence across taxa with supposed high mutation regions, making discrimination between species, and even subspecies, possible (Funk and Omland, 2003; Harvey et al., 2008).
Chapter 6  An investigation into the diversity of antigenic larval proteins

Limitations

Current practise typically employs single gene phylogenies for the identification of blowfly species. However, limitations in the reliability of single locus phylogenies have been reported (Stevens et al., 2002; Stevens, 2003; Wells et al., 2007). Such phylogenies only infer evolutionary relationships for the particular gene used, and as such may not represent the true species phylogeny, for example, due to horizontal gene transfer or incomplete lineage sorting at the locus in question. Despite mtDNA being expected to reach reciprocal monophyly before nuclear genes (Funk and Omland, 2003), phylogenetic inference from mtDNA genes has been widely reported to show paraphyly within Calliphoridae, as well as other organisms (Wells et al., 2007; Wells and Stevens, 2008). In particular, blowfly phylogenies based on mitochondrial and nuclear genes have exposed conflicting evolutionary relationships (Stevens and Wall, 1996b; Stevens et al., 2002; Wells et al., 2007; Wells and Stevens, 2008), with, for example, a COX1 genealogy rejecting reciprocal monophyly for L. cuprina when compared with a 28S rRNA phylogeny (Stevens et al., 2002; Wells et al., 2007; Harvey et al., 2008; Wells and Stevens, 2008). Such findings imply incomplete lineage sorting for some recently diverged blowflies, and so may prove problematic when relying on phylogenetic methods of diagnosis which themselves assume reciprocal monophyly.

6.1.7.2 Multi-gene phylogenetic approach

While some loci are more routinely used than others, no single gene is approved for species identification and, in view of the problems outlined, it appears unlikely that any single gene will be capable of unambiguously identifying all calliphorid species. Consequently, a universal switch to multi-gene approaches, for example using ESTs to screen multiple independent sites across blowfly genomes, is urgently needed.

Limitations
Chapter 6

An investigation into the diversity of antigenic larval proteins

The accuracy of phylogenetic inference, particularly when used for species determination of unidentified specimens, is also highly dependent on the experience of the analyst (Wells and Stevens, 2008). An important factor commonly misunderstood is the influence of taxonomic sampling, as both the number and choice of taxa represented can have a significant impact on the phylogenetic relationships inferred. In cases where many closely related species are compared, insufficient sampling may lead to intra- and interspecific ranges of genetic variation overlapping, affecting the ability to determine the identity of unknown samples (Moritz and Cicero, 2004; Harvey et al., 2008). A second and potentially large source of error is the use of uncritical BLAST searching of databases that rely heavily on individual investigators being responsible for the accuracy and taxonomy of submitted sequence data. At present, the importance of such sources of error remain largely unknown. However, as several recently documented cases illustrate, even some commonly-cited sequence records appear to have been mislabelled, leaving the reliability of reference sequence data in dispute (Wells and Stevens, 2008). Consequently, though it is generally accepted that such DNA-based identification of forensically important blowflies, for example, is expected to increase in popularity in the near future. Emphasis is now also being placed on the need to investigate alternative approaches for the identification of blowfly species (Wells and Stevens, 2008).

6.1.7.3 An antigenic approach

Recent work by Roelfstra et al., (2009), focusing on immunogenic profiling of *Gasterophilus intestinalis* (Oestridae) crude larval extracts, has revealed antigenic differentiation between different instar stages of larval development. Roelfstra et al., (2009) found second instar *G. intestinalis* larvae to be more immunogenic than third instar, and attributed this to higher levels of enzyme production in the more migrational second instar stage of *G. intestinalis*. Using one dimensional electrophoresis analysis, a total of fourteen immune-reactive second instar larvae protein bands were identified and analysed by MS. Of these, significant matches were found for three proteins from 4 of the 14 bands. These proteins were identified as actin, gylceraldehyde 3-phosphate...
dehydrogenase (GAPDH) and haemoglobin. Similarly, thirteen bands were chosen from blots of third instar larvae, with 10 proteins being significantly matched; representing five larval serum proteins (α or β chains), arylphorin, three hemoglobins and a murein lipoprotein. Of these, larval serum protein and arylphorin, homologous to *Calliphora vicina* and *Drosophila melanogaster*, respectively, are thought to be fundamental in development, being reabsorbed by fat body cells prior to pupation (Roelfstra et al., 2009). Arylphorin has also been shown to cause a considerable increase growth *in vivo* in several insect species (Hakim et al., 2005).

Roelfstra et al., (2009) additionally found antigenic differentiation between different types of host (naturally infested horses vs. immunised mice) in second instar stages. Two dimensional analyses revealed almost twice as many antigenic proteins on second instar blots probed with horse serum than those probed with the mouse serum. From these two dimensional blots 26 protein spots were chosen for MS analysis, of which just 7 proteins were matched, including paramyosin, serum albumin, tubulin, enolase, tropomyosin, GAPDH, and haemoglobin. The most interesting protein match was enolase, thought to be a key surface protein in several pathogens for attacking host tissue (Bernal et al., 2004). By comparison, two dimensional analyses of third instar larvae revealed similar antigenic protein profiles between horse and mouse serum. A total of 39 third instar protein spots were selected, of which 19 were successfully characterized by MS, representing 8 different proteins (filamin, heat shock protein, serum albumin precursor, phosphoenolpyruvate carboxykinase (PEPCK), enolase, fumarase, beta-actin, and haemoglobin). Finally, Roelfstra et al (2009) imposed three criteria to be met before a protein was considered successfully identified, these included; that the probability based MOWSE score from both MS and MS/MS was significant, matched peptide masses were abundant in spectra, and that the estimated molecular weights (MW) of the significant matches corresponded with the experimental observed values. In total, this study successfully identified 15 antigenic proteins (Roelfstra et al., 2009).
To date, few diagnostic tools have used antigenic protein markers to
distinguish morphological similar species of insects (Miller, 1981; Ma et al.,
1990; Stuart et al., 1994; Zeng et al., 1999). However, the successful
development of enzyme-linked immunosorbent assays (ELISA) for the detection
of Cochliomyia hominivorax (Figarola et al., 2001) and Oestrus ovis (Yilma,
1992; Deconninck et al., 1995; Goddard et al., 1999), lend further support to the
potential of rapid diagnostic tools for the identification of calliphorid species.

6.1.8 Aims

Due to the range of different styles of parasitism found within Calliphoridae
(saprophagic, obligate and facultative), blowfly speciation is likely to have been
accompanied by the evolution of different larval proteins relating to their
differing life history strategies (Stevens, 2003; Stevens et al., 2006). The work
presented here aims to investigate these similarities and dissimilarities in larval
antigens in four blowfly species of medical and veterinary importance, namely;
Calliphora vicina, Lucilia sericata, Lucilia cuprina, and Lucilia caesar. This in
turn will allow a comparison to be made between the antigenic profiles of two
different larval feeding behaviours exhibited by these species (e.g. the
saprophagic C. vicina, and facultative Lucilia species). Finally, studying several
members of the genus Lucilia, who share the same form of parasitism (i.e.
facultative ectoparasitism), will also permit the discriminatory power of such
antigen markers between individual species to be explored, and in doing so
potentially reveal candidates for the development of diagnostic tools.
6.2 Methods and Materials

6.2.1 Taxonomic sampling

Four blowfly species of forensic and economic significance were chosen for evaluation: Calliphora vicina, Lucilia sericata, Lucilia cuprina and Lucilia caesar. C. vicina, a saprophagic species (i.e. the larvae feed exclusively on dead or decaying matter), is a key indicator species in forensic entomology, used extensively in the UK, and elsewhere, to estimate post-mortem interval (PMI). The three Lucilia species (L. sericata, L. cuprina, L. caesar) are all facultative parasites (i.e. able to feed on either necrotic or living tissue), and are the primary taxa responsible for fly strike in sheep.

6.2.2 Sample collection

In-house colonies at the University of Exeter were used for the collection of UK populations of Calliphora vicina and Lucilia sericata, first, second, and third instar. Lyophilized samples (first and third instar only) from Australian populations of Calliphora vicina, Lucilia cuprina, and Lucilia sericata were provided by James Wallman and colleagues (University of Wollongong, NSW, Australia), and samples (first, second, and third instar) from USA populations of Calliphora vicina and Lucilia sericata provided by Jeff Wells and colleagues (West Virginia University, USA). All fresh samples were lyophilized prior to analysis.

6.2.3 Polyclonal antibody (PAb) production

Rabbit polyclonal antiserum was raised against PBS protein extracts from whole larvae (first, second and third instar) of L. sericata. See Chapter 2 for further details.
6.2.4 Gel electrophoresis and Western blotting

Larval protein extracts for each of the four blowfly species were separated using SDS-PAGE gel electrophoresis, electrophoretically transferred to polyvinylidene difluoride membranes (PVDF), and probed with the immunized rabbit antiserum to expose the antigenicity of the proteins present. See Chapter 2 for details.

Where first instar larvae were used, 5mg samples were collected, homogenised using a plastic micropestle, and diluted to 0.1mg/ml in Laemmli buffer. Where second or third instar samples were used, a single whole larva was homogenised, and diluted to 0.1mg/ml in Laemmli buffer.

6.2.5 N-terminal sequencing

Protein bands of interest were separated from any impurities using SDS-PAGE gel electrophoresis, electro transferred onto a PVDF membrane, and N-terminal sequencing carried out at the University of Bristol Proteomics Facility. See Chapter 2 for further details.

6.2.6 Two-dimensional gel electrophoresis

Crude first instar larval preparations from *L. sericata* and *Calliphora vicina* were separated by two-dimensional (2D) gel electrophoresis, and diagnostic protein spots selected for analysis by mass spectrometry (MS). See Chapter 2 for further details.

6.2.7 Mass spectrometry (MS)

Protein spots were analysed by MALDI-TOF and MS/MS on a 4700 MALDI-TOF/TOF tandem mass spectrometer (Applied Biosystems), with five MSMS (sequencing) runs being carried out for each spot. Combined peptide mass fingerprint and MS/MS queries were carried out with MASCOT on the MSDB database (i.e. scores > 74 being significant at p < 0.05 for the eukaryote
database; scores > 60 being significant at p < 0.05 for the *Drosophila* database). See Chapter 2 for further details.

### 6.2.7.1 Peptide *de novo* sequencing from MS/MS spectrum

Observed peptides were also fragmentted using collision induced dissociation, and *de novo* sequencing performed on the resultant fragment spectra. Peptide sequence identities were then ranked by the degree of fit between the actual data and theoretical fragments of candidate peptide sequences and the peak intensity. The resulting sequence information was then used to perform sequence similarity searches with BLASTP2 with the non-redundant database (nrdb95), and MS-BLAST (Gish, 1996-2006 http://blast.wustl.edu; (Shevchenko *et al.*, 2001) and using the platform ButterflyBase (Papanicolaou *et al.*, 2008) on three different databases (Arthropod, *Cochliomyia hominivorax*, and UniRef100). See Chapter 2 for further details.

### 6.2.8 Electron microscopy (EM) gold labelling

Electron microscopy (EM) gold labelling was carried out on first instar larvae of *L. sericata* and *Calliphora vicina* according to protocol by Thornton and Talbot, (2006). See Chapter 2 for details.
6.3 Results

6.3.1 Sample collection

Initially problems were encountered producing sufficient first instar larvae samples from existing in house *L. sericata* and *Calliphora vicina* UK colonies. However, fresh stocking of colonies improved egg laying, and hatching success rates.

6.3.2 Rabbit polyclonal antiserum

Anti-sera from all three rabbits immunised with PBS protein extracts of *L. sericata* whole larvae (first, second and third instar) reacted strongly (Fig. 18), and antiserum from rabbit 1 (R1) was selected for use in all subsequent immunoblotting experiments.

![Figure 18: Polyclonal responses to *L. sericata* using anti-sera from each of the three rabbits (R1, R2, R3) immunised with PBS protein extracts from whole larvae *L. sericata* (UK population).](image-url)
6.3.3 Gel electrophoresis and Western blotting

A range of larval protein concentrations were assessed (Fig. 19), with a concentration of between 0.1mg/ml and 0.05mg/ml found to be the most suitable for displaying antigen banding profiles clearly.

![Figure 19: Western blot analysis with different protein concentrations of first instar larval preparations of L. sericata (UK population). (A = 8mg/ml; B = 1mg/ml; C = 0.1mg/ml; D = 0.01mg/ml)](image)

Strong polyclonal cross-reactivity was demonstrated for each of the blowfly taxa, with unique protein banding profiles being found for all four species tested (Fig. 20-25). Comparing immunoblot profiles not only revealed variation in banding pattern between the different species (Fig. 20 and 21), but also between the different instar stages (Fig. 22 and 23). Overall, the greatest amount of inter-specific antigenic diversity was found between first instar larvae (Fig. 20), with third instar larvae displaying very similar profiles in each of the four species (Fig. 21).

While several protein bands appeared to be shared among all four taxa, taken as a whole, first instar protein banding profiles observed appear species specific. Several candidate protein markers with diagnostic potential can clearly be seen (Fig. 20). However, repeated attempts to sequence one of these markers (Fig. 20; band A) by N-terminal peptide sequencing failed, possibly due to the band being predominantly carbohydrate or a very heavily glycosylated glycoprotein, which can result in blank sequencing cycles, reduced peaks or altered retention times.
Figure 20: Western blot analysis of first instar larval preparations, with a broad range pre-stained SDS-PAGE standard (10 - 250kD), displaying protein band (A) selected for N-terminal sequencing, (C. vicina, UK population; L. sericata, UK population; L. cuprina, Australian population; L. caesar, UK population).

Figure 21: Western blot analysis of third instar larval preparations, with a broad range pre-stained SDS-PAGE standard (10 - 250kD). C. vicina - UK population; L. sericata - UK population; L. cuprina - Australian population; L. caesar - UK population.
Figure 22: Western blot analysis of first, second and third instar larval preparations for \textit{L. sericata} (UK population), with a broad range pre-stained SDS-PAGE standard (10 - 250kD).

Figure 23: Western blot analysis of first, second and third instar larval preparations for \textit{C. vicina} (UK population), with a broad range pre-stained SDS-PAGE standard (10 - 250kD).
6.3.3.1 Biogeographic diversity

First instar larvae were compared between three populations (UK, Australia and USA) or *L. sericata* and *C. vicina*, as this is the stage of larval development that showed the greatest differentiation in between species, and therefore is most likely to show any intra-specific differentiation; as opposed to later instar stages which seem to be more similar.

![Western blot analysis of first instar larval of C. vicina (UK, Australian, and USA populations), with a broad range pre-stained SDS-PAGE standard (10 - 250kD).]

Figure 24: Western blot analysis of first instar larval of *C. vicina* (UK, Australian, and USA populations), with a broad range pre-stained SDS-PAGE standard (10 - 250kD).

![Western blot analysis of first instar larval of L. sericata (UK, Australian, and USA populations), with a broad range pre-stained SDS-PAGE standard (10 - 250kD).]

Figure 25: Western blot analysis of first instar larval of *L. sericata* (UK, Australian, and USA populations), with a broad range pre-stained SDS-PAGE standard (10 - 250kD).
While no antigenic differentiation was found between UK, Australian, and USA populations of *L. sericata* (Fig. 25), some differentiation, most noticeably a single protein band present in the UK sample and absent in the Australian and USA samples was found within *C. vicina* (band B, Fig. 24). While wider sampling of these populations would be needed to confirm this possible antigenic differentiation, the limited availability of Australian and USA samples meant that this was not possible in the time span of this project. Furthermore, comparisons with previous western blots of first instar from UK populations of *C. vicina* (e.g. Fig. 23) do not seem to display band B, suggesting either some abnormality in the sample, or highlighting the lack of sensitivity in the method used.

### 6.3.4 Two dimensional gel electrophoresis

Two-dimensional (2D) gel electrophoresis separates proteins first by isoelectric point and then by molecular mass. The major advantage of this technique is the ability to distinguish between different isoforms of a protein with similar molecular weights, for example a protein that has been phosphorylated. Three immunoreactive protein spots visible on both the western blots and Sypro Ruby stained blots (Fig. 26, spots 1-3) were chosen for analysis by mass spectrometry (MS). These spot were chosen as they were easily identifiable, and can clearly be seen to be present in *L. sericata* and absent in *C. vicina*. 
Figure 26: Two dimensional analysis, comprising pH gradient from 3 to 11, of first instar larval preparations from UK populations of *C. vicina* (A) and *L. sericata* (B); (I – Western immunoblot membranes processed using antiserum; II – replica gel stained with Sypro Ruby).
6.3.5 Mass Spectrometry

6.3.5.1 MALDI-TOF/TOF

MSMS (sequencing) runs (five carried out for each spot) detected 7 MS/MS spectra in total, but no significant protein matches were found for using either the eukaryote or the *Drosophila* database. This was unexpected since spots (for example, protein spot 1) provided particularly good spectra.

6.3.5.2 De novo peptide sequencing

*De novo* sequencing interpretation of the tandem mass spectra (MS/MS) predicted numerous possible peptide sequences (Tables 8-10). BLAST searching the *de novo* predicted sequences allows all high-scoring pairs (regions of high local sequence similarity) between an individual predicted peptide and a protein sequence from the database to be detected. While MASCOT searches expose and list proteins with peptides that match exactly to the uploaded MS/MS spectra, BLAST searches sort hits by their total scores, which are calculated for each protein in the database by adding up scores of individual high scoring pairs (HSPs) that are higher than a specified threshold, and so are therefore sensitive to the total number of HSPs. However, this also means that a HSP with a low score can still be included if, for that particular protein, the score for all HSPs is higher. Consequently, BLAST searches can compensate for the limited accuracy of *de novo* sequence prediction by scoring hits even if no identical peptides have been found.
### Table 10: De novo sequence predictions for protein spot 1

<table>
<thead>
<tr>
<th>Size</th>
<th>Rank</th>
<th>Sequence</th>
<th>Score</th>
<th>Modifications*</th>
</tr>
</thead>
<tbody>
<tr>
<td>1202Da</td>
<td>1</td>
<td>C*EDDHDAALR</td>
<td>74.8102</td>
<td>C [1];</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>EC*DDHDAALR</td>
<td>74.8102</td>
<td>C [2];</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>C*ETEHDAAALR</td>
<td>73.5261</td>
<td>C [1];</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>EC*TEHDAALR</td>
<td>73.5261</td>
<td>C [2];</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>C*EDDHWAALR</td>
<td>55.2643</td>
<td>C [1];</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>EC*DDHWALR</td>
<td>55.2643</td>
<td>C [2];</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>C*ETEHWALR</td>
<td>51.6609</td>
<td>C [1];</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>EC*TEHWALR</td>
<td>51.6609</td>
<td>C [2];</td>
</tr>
<tr>
<td>1007Da</td>
<td>1</td>
<td>GSSTWRVSK</td>
<td>29.6629</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>SGSTWRVSK</td>
<td>29.4669</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>GSSTWRASR</td>
<td>29.2755</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>GSSTWRGTR</td>
<td>29.2755</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>GSSTWRWK</td>
<td>29.2755</td>
<td></td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>SGSTWRWK</td>
<td>29.0795</td>
<td></td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>SGSTWRASR</td>
<td>28.88</td>
<td></td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>SGSTWRGTR</td>
<td>28.88</td>
<td></td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>NCDRWGTGR</td>
<td>25.8535</td>
<td></td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>NCDRWASAR</td>
<td>25.7656</td>
<td></td>
</tr>
<tr>
<td>2339Da</td>
<td>1</td>
<td>C<em>C</em>VLCKM*NSLNDVATHNWR</td>
<td>67.635</td>
<td>C [1,2]; M [7];</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>C<em>C</em>VKESM*NSLNDVATHNWR</td>
<td>67.437</td>
<td>C [1,2]; M [7];</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>C<em>C</em>VKESM*SNLNDVATHNWR</td>
<td>67.4343</td>
<td>C [1,2]; M [7];</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>C<em>C</em>VLCKM*SNLNDVATHNWR</td>
<td>67.3167</td>
<td>C [1,2]; M [7];</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>C<em>C</em>VKESFSNLNDVATHNWR</td>
<td>65.0279</td>
<td>C [1,2];</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>C<em>C</em>VKESFSNLNDVATHNWR</td>
<td>65.0254</td>
<td>C [1,2];</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>C<em>C</em>VLCKFNSLNDVATHNWR</td>
<td>64.4143</td>
<td>C [1,2];</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>C<em>C</em>VYKAVDLNSM*HPGNWR</td>
<td>64.2083</td>
<td>C [1,2]; M [13];</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>C<em>C</em>VLCKFSNLNDVATHNWR</td>
<td>64.1111</td>
<td>C [1,2];</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>C<em>C</em>VYKAVDNLNSM*HPGNWR</td>
<td>64.1094</td>
<td>C [1,2]; M [13];</td>
</tr>
</tbody>
</table>

(C) Carboxymethyl, (M) Oxidation.
### Table 11: *De novo* sequence predictions for protein spot 2

<table>
<thead>
<tr>
<th>Size</th>
<th>Rank</th>
<th>Sequence</th>
<th>Score</th>
<th>Modifications*</th>
</tr>
</thead>
<tbody>
<tr>
<td>1011.5Da</td>
<td>1</td>
<td>LKTRHVEE</td>
<td>7.4454</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>ADRRHPME</td>
<td>7.2388</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>KPGPLKC*PN</td>
<td>7.1239</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>LKTRHPME</td>
<td>6.9803</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>MAASHGGGAPR</td>
<td>6.5503</td>
<td></td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>ACPARHVEE</td>
<td>6.4959</td>
<td></td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>AAVTRHVEE</td>
<td>6.4207</td>
<td></td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>ADRRHVEE</td>
<td>6.4075</td>
<td></td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>AAVTRHPME</td>
<td>6.358</td>
<td></td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>PKGPLKC*PN</td>
<td>6.3533</td>
<td>C [7];</td>
</tr>
<tr>
<td>2211Da</td>
<td>1</td>
<td>KC<em>NFLKEAVSRDVM</em>NNNR</td>
<td>52.1924</td>
<td>C [2]; M [14];</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>KC<em>NM</em>LKEAVSRDVFNNNR</td>
<td>51.6528</td>
<td>C [2]; M [4];</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>KC<em>NM</em>LKEAVSRDVM*NNNR</td>
<td>51.417</td>
<td>C [2]; M [4,14];</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>KC<em>M</em>NLKSLVSRRDFLDNR</td>
<td>51.294</td>
<td>C [2]; M [3];</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>KC<em>M</em>NLKSLVSRRDVM*LDNR</td>
<td>51.294</td>
<td>C [2]; M [3,14];</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>KC<em>FNKLEAVSRDVM</em>LDNR</td>
<td>51.0918</td>
<td>C [2]; M [14];</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>KC<em>M</em>NLKEAVSRDVM*LDNR</td>
<td>51.0918</td>
<td>C [2]; M [3,14];</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>KC<em>M</em>NLKEAVSRDVM*LDNR</td>
<td>51.0918</td>
<td>C [2]; M [3];</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>KC*NFLKEAVSRDVFNNNR</td>
<td>49.1931</td>
<td>C [2];</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>KC*FNKLEAVSRDVFLDNR</td>
<td>48.6589</td>
<td>C [2];</td>
</tr>
</tbody>
</table>

### Table 12: *De novo* sequence predictions for protein spot 3

<table>
<thead>
<tr>
<th>Size</th>
<th>Rank</th>
<th>Sequence</th>
<th>Score</th>
<th>Modifications*</th>
</tr>
</thead>
<tbody>
<tr>
<td>1011Da</td>
<td>1</td>
<td>HVTASNVKR</td>
<td>11.9658</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>HLSAAAGTKR</td>
<td>10.8442</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>VHTASNVKR</td>
<td>10.6328</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>LKVHVPYR</td>
<td>10.2745</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>HVTTAAGTKR</td>
<td>10.0811</td>
<td></td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>HDRAAGTKR</td>
<td>10.0318</td>
<td></td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>LHSAAAGTKR</td>
<td>9.8165</td>
<td></td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>HLSASNVKR</td>
<td>9.7376</td>
<td></td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>KLHVHPYR</td>
<td>9.6744</td>
<td></td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>HLSAAPPVYR</td>
<td>9.0774</td>
<td></td>
</tr>
<tr>
<td>1204Da</td>
<td>1</td>
<td>VAPTVPGMNR</td>
<td>15.21</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>AHKSTESTM*R</td>
<td>14.2236</td>
<td>M [9];</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>VAPTVPGKSW</td>
<td>14.1621</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>GLPVTVPGMNR</td>
<td>13.9156</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>AHKSTESTFR</td>
<td>13.5463</td>
<td></td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>LGPVTVPGMNR</td>
<td>13.3933</td>
<td></td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>AVPVTVPGMNR</td>
<td>13.3933</td>
<td></td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>TAAPC*SLLTNR</td>
<td>13.1168</td>
<td>C [5];</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>GDAPC*ALETNR</td>
<td>13.0331</td>
<td>C [5];</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>GLPVTVPGKSW</td>
<td>12.9854</td>
<td></td>
</tr>
</tbody>
</table>
Sequence similarity searches by BLASTP2 using the non-redundant database (nrdb95) matched a number of positive hits for protein spots 1 and 2, but only borderline matches for protein spot 3 (Table 11). In an attempt to explore other protein databases, searches were also carried out using the ButterflyBase platform (MS-BLAST). While these searches failed to find matches using either the Arthropod or Cochliomyia hominivorax (containing over 18k ESTs) databases, positive matches were found for proteins spots 1 and 2 using the UniRef100 database (Table 12). Nevertheless, under both BLASTP2 and MS-BLAST predicted peptide matches represented a very wide range of organisms (fungi, bacteria, protozoa, plants). With the possible exception of Apis mellifera (European honey bee), it seems very doubtful that any of the matches are in fact homologous to larval L. sericata proteins, particularly as these are proteins absent in C. vicina.

While rearing conditions, (e.g. nutrition, egg laying substrate, etc) were identical between L. sericata and C. vicina colonies, larvae from neither colony were reared from sterilized eggs or fed on sterile media. Consequently, one possibility could be that the protein spots sampled were not larval proteins at all (e.g. bacteria). Additionally, if this is the case, it seems that despite identical rearing conditions, these antigens demonstrate the possible association of another organism with one blowfly species (L. sericata) and not from another (C. vicina). On the other hand, if this were the case significant MASCOT matches with the MS/MS spectra might also be expected.

It is more probable that a combination of inaccuracies in the de novo interpretation of the MS/MS spectra, and a lack of homologous proteins in databases would account for the BLAST matches found. This would indicate a problem with accurately matching L. sericata specific peptides to any proteins already in the databases (i.e. a lack of blowfly protein sequence information or anything related closely enough). Additionally, being saprophagic, C. vicina may be expressing the ancestral larval antigenic protein profile, with the apparent protein spots identified here being exclusively found in facultative ectoparasitic calliphorids, or even L. sericata itself.
<table>
<thead>
<tr>
<th>Spot 1</th>
<th>Positive hits</th>
<th>HSP score</th>
<th>Total score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Putative protein kinase</td>
<td><em>Plasmodium falciparum</em></td>
<td>59</td>
<td>260</td>
</tr>
<tr>
<td>phi Mu50B-like protein</td>
<td><em>Staphylococcus aureus</em> phage phi 11</td>
<td>65</td>
<td>65</td>
</tr>
<tr>
<td><strong>Borderline hits</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Putative AP2/EREBP transcription factor</td>
<td><em>Arabidopsis thaliana</em></td>
<td>55</td>
<td>101</td>
</tr>
<tr>
<td>Peptidyl-tRNA hydrolase</td>
<td><em>Wolbachia</em> endosymbiont strain TRS of <em>Brugia malayi</em></td>
<td>61</td>
<td>61</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Spot 2</th>
<th>Positive hits</th>
<th>HSP score</th>
<th>Total score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unknown</td>
<td>Environmental sequence</td>
<td>66</td>
<td>66</td>
</tr>
<tr>
<td>Conserved hypothetical protein</td>
<td><em>Pseudomonas aeruginosa</em></td>
<td>66</td>
<td>66</td>
</tr>
<tr>
<td>Putative periplasmic protein</td>
<td><em>Wolinella succinogenes</em></td>
<td>64</td>
<td>64</td>
</tr>
<tr>
<td><strong>Borderline hits</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unknown</td>
<td>Environmental sequence</td>
<td>63</td>
<td>63</td>
</tr>
<tr>
<td>Unknown</td>
<td><em>Candida glabrata</em> strain</td>
<td>63</td>
<td>63</td>
</tr>
<tr>
<td>Karyogamy protein regulatory protein</td>
<td><em>Saccharomyces cerevisiae</em></td>
<td>61</td>
<td>61</td>
</tr>
<tr>
<td>Putative THO complex 3</td>
<td><em>Oryza sativa</em></td>
<td>61</td>
<td>61</td>
</tr>
<tr>
<td>Conserved hypothetical protein</td>
<td><em>Pseudomonas putida</em></td>
<td>60</td>
<td>60</td>
</tr>
<tr>
<td>Conserved hypothetical protein</td>
<td><em>Pseudomonas syringae</em></td>
<td>60</td>
<td>60</td>
</tr>
</tbody>
</table>

| Spot 3 | Positive hits - none | | |
| **Borderline hits** | | | |
| Putative exported protein | *Bordetella bronchiseptica* strain RB50 | 61 | 61 |
| Putative exported protein | *Bordetella pertussis* strain Tohama I | 61 | 61 |

Table 13: BLASTP2 matches, showing the score of the HSP, highest-scoring pairs (local hit), and total score (global hit)

(Database - nrdb95; 2078555 sequences; 597106688 residues)
<table>
<thead>
<tr>
<th>Spot 1</th>
<th>Positive hits</th>
<th>HSP score</th>
<th>Total score</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Putative uncharacterized protein</td>
<td>Plasmodium falciparum</td>
<td>49</td>
</tr>
<tr>
<td></td>
<td>Putative uncharacterized protein</td>
<td>Plasmodium falciparum</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td>Predicted protein</td>
<td>Apis mellifera</td>
<td>76</td>
</tr>
<tr>
<td></td>
<td>Pc16g12050 protein</td>
<td>Penicillium chrysogenum</td>
<td>73</td>
</tr>
<tr>
<td></td>
<td><strong>Borderline hits</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Putative uncharacterized protein</td>
<td>Coccidioides immitis</td>
<td>67</td>
</tr>
<tr>
<td>Spot 2</td>
<td>Positive hits</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Cyclic nucleotide-binding protein</td>
<td>Desulfatibacillum alkenivorans</td>
<td>66</td>
</tr>
<tr>
<td></td>
<td>Transposase</td>
<td>Syntrophomonas wolfei</td>
<td>76</td>
</tr>
<tr>
<td></td>
<td><strong>Borderline hits</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Tubulin-tyrosine ligase family protein</td>
<td>Tetrahymena thermophila</td>
<td>70</td>
</tr>
<tr>
<td></td>
<td>Lipocalin family protein</td>
<td>Chlorobium phaeobacteroides</td>
<td>69</td>
</tr>
<tr>
<td></td>
<td>Chromosome 18 scaffold</td>
<td>Vitis vinifera</td>
<td>68</td>
</tr>
<tr>
<td></td>
<td>Apolipoprotein N-acyltransferase</td>
<td>Thioalkalivibrio sp</td>
<td>68</td>
</tr>
<tr>
<td></td>
<td>Putative uncharacterized protein</td>
<td>Vitis vinifera</td>
<td>68</td>
</tr>
<tr>
<td></td>
<td>Putative uncharacterized protein</td>
<td>Clostridium scindens</td>
<td>68</td>
</tr>
<tr>
<td></td>
<td>Putative uncharacterized protein (Fragment)</td>
<td>Vitis vinifera</td>
<td>67</td>
</tr>
<tr>
<td></td>
<td>Chromosome undetermined SCAF15013</td>
<td>Tetraodon nigroviridis</td>
<td>67</td>
</tr>
</tbody>
</table>

Table 14: MS-BLAST matches showing the score of the HSP, highest-scoring pairs (local hit), and total score (global hit)

(Database - UniRef100; 7,534,635 sequences; 2,695,776,793 residues)
6.3.6 Localization of antibody binding by electron microscopy (EM) gold labelling

The cross-sections of first instar larvae examined by immunogold labelling and transmission electron microscopy (TEM) using serum raised to *L. sericata* larval proteins, indicates strong gold localisation on cuticle layers of *L. sericata* and *C. vicina* (Figs 27 and 28). The cuticle has three distinct layers: the outermost envelope (env), the middle protein-rich epicuticle (epi) and the chitin-rich lamellar procuticle (pro). Gold particles appeared in greatest concentration in the epicuticle layer, becoming less concentrated in the procuticle, and even sparser in deeper lying tissue (Figs 27-29). Gold localization was also observed, albeit to a lesser extent, along membranes lining the gut (Fig 30).

It seems the majority of first instar larval proteins that provoke an antigenic response form part of the cuticle layers, with much of the gold being localised on the epicuticle rather than the cuticle envelope itself.

Immunogold labelling was also carried out on cross-sections that had not been probed with the serum raised to *L. sericata* larval proteins. These ‘control’ cross-sections revealed almost no gold particle binding (Fig 31), confirming that the gold was only binding to the particles causing an antigenic response (Fig. 27 – 30).
Figure 27: Gold labelled TEM images of cross-sections taken from first-instar *C. vicina*. Cross-sections illustrate the larval cuticle, and show the envelope, epicuticle and procuticle layers. Gold particles are seen binding to all antigenic proteins within the cross-section, showing an apparent accumulation of antigenic molecules within the epicuticle layer.
Figure 28: Gold labelled TEM images of larval cuticle from first-instar *L. sericata* larvae. The cross-sections are taken across the larval cuticle, showing envelope, epicuticle and procuticle layers. Gold particles are seen binding to all antigenic proteins within the cross-section, showing an apparent accumulation of antigenic molecules within the epicuticle layer.
Figure 29: Gold labelled TEM images of first instar *C. vicina* (A) and *L. sericata* (B), illustrating muscle layers found between the cuticle layers and gut cavity. Gold particles binding to antigenic proteins reveal no general patterns of localisation.
Figure 30: Gold labelled TEM images from the gut of first-instar *C. vicina* (A) and *L. sericata* (B), illustrating gut cavity and villi projections into the lumen. Gold labelling reveals an apparent localisation of antigenic proteins along the lining of the gut, with few antigenic molecules found within the gut lumen itself.
Figure 31: Gold labelled TEM images of control experiments from first-instar *C. vicina* (A) and *L. sericata* (B) larvae. These control cross-sections had not been probed with the serum raised to *L. sericata* larval proteins (rabbit anti-sera); the lack of gold particle binding demonstrates that gold is only binding to particles causing an antigenic response in Fig. 27-30, which had been probed with the rabbit anti-sera.
6.4 Discussion

While morphologically very distinct from the three *Lucilia* species at the adult fly stage, the larvae of *C. vicina*, particularly first instar stages, exhibit very similar gross morphology to *Lucilia*. Of the three morphologically similar *Lucilia* species, *L. sericata* and *L. cuprina* in particular, are also notoriously difficult to discriminate – being almost identical, even as adult flies. Consequently, species identification at early stages of an infestation, either myiasis or saprophagy, can prove difficult and time consuming. The primary objective of this work was to investigate whether reproducible antigenic differences between first instar larvae were potentially capable of discriminating between the four calliphorid species, therefore exploring the hypothesis that physiological differences may exist where morphological diversity is limited. In addition to distinct differences in larval antigens being observed between the two genera (as might be expected with the different styles of parasitism represented by each), antigenic protein profiles were also found to be species specific within *Lucilia*. In addition, antigenic protein profiles proved to also be instar stage specific, with the greatest amount of differentiation between the four species being found in first instar larvae.

Interestingly, while genetic differentiation has been reported between Australian and UK populations of *L. sericata* (Stevens et al., 2002), no antigenic differences were observed between these populations, or indeed with a population from the USA. In comparison some differentiation was detected between UK, Australian, or USA populations of *C. vicina*, with one band in particular seeming to be present in the UK sample, and absent from the Australian and USA samples. However, the protein band in question could not be found on other western blots of first instar UK *C. vicina* samples, and the limited supply of Australian and USA samples meant that repeating the blot comparing each of the populations directly was not possible at this time.

In contrast to the immunogenic profiling of *G. intestinalis* (Oestridae) larval extracts by Roelfstra et al., (2009), the results presented here have been fundamentally limited by an apparent lack of similar antigens in current protein
databases. This is perhaps unsurprising considering that the antigens identified here represent a relatively small number of species specific proteins. By comparison, Roelfstra et al., (2009) focused on antigenic differences between second and third instar stages of a single species, and so the majority of the antigens identified may be expected include ubiquitous developmental proteins found in many other related organisms.

An additional problem that may be affecting attempts at protein characterization could be caused by glycosylation of antigens, which can affect N-terminal sequencing. Electron microscopy (EM) gold labelling of L. sericata and C. vicina first instar revealed the majority of larval antigens to form part of the cuticle layers, being most concentrated in the epicuticle. Antigens that occur in particularly proteolytic environments (e.g. parasite gut or digestive system, host/parasite boundary, etc.) are commonly found to be glycosylated, with the glycosylation itself possibly forming a large part of the native protein (Willadsen, 2006). Eisemann (1994) highlighted the importance of such oligosaccharide defences in blowfly larvae by feeding L. cuprina larvae on a serum-free medium containing lectins of varying specificity. In these trials a 2µM concentration of wheat germ lectin was found to result in a 50% larval growth inhibition, with 100% mortality being achieved using a concentration of 25µM, due to the lectins binding to glycosylated proteins linked to the peritrophic membrane. Studies investigating the effect of the peritrophic membrane antigen PM95 on limiting larval growth, found strong immunogenic responses for both the oligosaccharide and polypeptide proportions of the antigens (Tellam et al., 2001). As of yet the significance of glycosylation in protective antigenicity, and the role of this in preventing the development of efficient vaccines remains unclear and requires further investigation (Willadsen, 2006). However, it seems possible that the majority of the antigens observed in blowfly larvae would be heavily glycosylated as a result. It is therefore also possible that the antigenic diversity observed between different blowfly species, and in particular between closely related species, is actually reflecting oligosaccharide differences, rather than the polypeptides themselves.
To date the majority of studies investigating blowfly larval antigens have focused almost exclusively on identifying proteins for use in the development of a vaccine for the prevention of myiasis. For example, Tellam and Eisemann, (1998) homogenized and sequentially extracted whole first instar *L. cuprina* larvae, using a series of buffers (of increasing solubilising power), and testing the antigenic power (using immunization trials with sheep) of each protein fraction sequentially, but focusing specifically on those proteins that effected larval growth. One protein fraction in particular was chosen, as it contained proteins that reduced larval growth by 84% *in vitro*. Immunofluorescence and immunogold localisation found these proteins to be primarily located on the peritrophic membrane, and to a lesser extent on the larval cuticle, basement membranes, and microvilli of digestive epithelial cells.

Vast resources have been dedicated to the investigation into blowfly larval antigens, almost exclusively focusing on the identification of vaccine candidates. An opportunity that has, thus far, been rarely considered is the diagnostic potential of such antigens. Here, we have begun to explore this novel approach, and have proved that larval antigens show great potential in the development of successful diagnostic tools for the identification of calliphorid species. One example of the way the diagnostic potential of blowfly larval antigens could be utilised is through the development of a lateral flow diagnostic tool. Lateral flow devices (LFD) are a format of semi-quantitative immunoassay for detecting the presence (or absence) of a target antigen (i.e. species-specific antigenic protein) in a complex sample of proteins (e.g. whole blood, serum, urine, etc). The most publicised example of the use of this technology is home pregnancy test first introduced by Unipath in 1988. Such technology is now used extensively for the rapid ‘on-site’ diagnosis of a wide range of human diseases including visceral leishmaniasis (Sundar *et al*., 1998), HIV/AIDS (Iweala, 2004), malaria (Moody, 2002), SARS (Kammila *et al*., 2008) and invasive aspergillosis (Thornton, 2008).

Lateral flow devices work as follows: a species-specific monoclonal antibody (MAb) is immobilised to a defined capture zone on a porous nitrocellulose membrane, while the same MAb conjugated to colloidal gold
particles serves as the detection reagent. Samples of solubilised antigens are added to a release pad containing the antibody–gold conjugate. The antibody–gold conjugate binds to the target antigen, passes along the porous membrane by capillary action, and binds to the MAb immobilised in the capture zone. Once an antigen extract is prepared and applied to the LFD, the test result is recorded within 10-15 min. Bound antigen-antibody-gold complex is seen as a red line with an intensity that is proportional to the antigen concentration. Anti-mouse immunoglobulin immobilised to the membrane in a separate zone acts as an internal control. In the absence of the target antigen, no complex is formed in the zone containing the solid-phase antibody, and a single control line is seen. In the presence of the target antigen, two lines are clearly visible. One of the key advantages of this technology is the simplicity of the test, typically requiring little or no sample or reagent preparation.

A potential application of such a diagnostic tool would be for use within forensic entomology. Modern forensic science has seen the importance of entomological evidence becoming well established worldwide, particularly when facing crime scenes involving a suspected homicide or manslaughter. Within forensic entomology, blowflies are recognised as some of the most important and robust indicator species, as they are usually among the first insects to colonise a body after death, often within hours (Smith, 1986). Typically, the immature stages of blowflies, particularly larvae, are collected at a suspected crime scene and are used to establish the minimum post-mortem interval (PMI). The immature stages, particularly early stage larvae, of several forensically important blowfly genera, e.g. Lucilia and Calliphora, are notoriously difficult to identify on the basis of morphology; the lack of species-specific larval morphological characters can make taxonomic keys very difficult to use, particularly for non-specialist entomologists (Malgorn and Coquoz, 1999). Accurate species determination is essential for PMI calculations, as growth and development rates can be highly species-specific. Currently, forensic entomologists are increasingly favouring DNA based methods of species identification (Malgorn and Coquoz, 1999; Vincent et al., 2000; Stevens and Wall, 2001; Wallman and Adams, 2001; Wallman and Donnellan, 2001; Wells and Sperling, 2001; Chen et al., 2004; Saigusa et al., 2005; Harvey et al., 2008),
despite the well documented problems encountered when using these techniques (see Section 6.1.7).

The findings presented here demonstrate that the identification of species specific larval antigens, for the development of diagnostic tools, is currently being prevented by a lack of homologous proteins in existing databases. Consequently, the decision was made not to select any further protein bands or spots for analysis at present, as it is highly likely that they too represent proteins currently unrepresented in databases.

A blowfly genome project is currently being carried out in collaboration between the ARC Special Research Centre for Environmental Stress and Adaptation Research (CESAR), in Melbourne, and Massey University in New Zealand. To date, a bacterial artificial chromosome (BAC) library has been produced containing 55,000 clones, with an average insert of 100 kb. Additionally, sequencing of around 5000 clones from a cDNA library produced from *L. cuprina* embryos has identified 1853 genes, with sequence similarity to proteins in the database GenBank being found for 78% of these expressed sequence tags (ESTs). A high resolution genetic map of *L. cuprina* will eventually be produced through the hybridization of cDNA clones to BAC filters (http://cesar.org.au, accessed 26/07/09). However, this work is being funded through the Australian Wool Innovation Ltd, and as such is not publically available.

### 6.4.1 Future work

A grant proposal for the construction of a cDNA library suitable for 454 sequencing of *L. sericata* is currently being planned in collaboration with colleagues based at the Tremough Campus of the University of Exeter. This will then in turn provide an appropriate genomic framework with which to identify the unknown proteins described here.
Chapter Seven

General discussion
7.1 Introduction

Insects first evolved around 600 million years ago, an estimated 300 million years before warm-blooded vertebrates (Wall, 2007). During the time insects and vertebrates have coexisted, various forms of parasitism have independently evolved (Balashov, 2006). Parasitism has originated most frequently within Diptera, where it has evolved over 100 times in 21 families, and accounts for an estimated 20% of all known species exhibiting parasitoid behaviour (Eggleton and Belshaw, 1992). The dipteran family Calliphoridae contains some of the most economically significant myiasis-causing flies in the world, notorious for their parasitism of livestock.

It has been estimated that the sub-families within Calliphoridae had diverged by the time the supercontinent Gondwana broke up, approximately 100 MYA (Stevens, 2003). However, the majority of within genera speciation is thought to have occurred as recently as 1 to 15 MYA, although most groups displaying parasitic behaviour are thought to have originated between 5 and 15 MYA (Wallman et al., 2005; Stevens et al., 2006). Many of the parasitic sub-families and genera are also associated with distinct geographic regions, suggesting multiple origins of parasitism after periods of geographic isolation (Stevens et al., 2006).

The evolution of myiasis itself has been linked to the development of key behavioural, physiological, and biochemical adaptations (Stevens et al., 2006). By attempting to resolve how this key parasitic behaviour evolved in the blowfly, this thesis aims to explore the evolution of host-parasite immunology, providing a model for understanding the evolution of metazoan parasites in general.

7.2 Phylogenetic signal within the mitochondrial genome

Mitochondrial genes have been widely used within insect systematics. However, the majority of phylogenetic studies have used data from only a single gene, commonly COX1, and few have systematically assessed which gene(s) in the mitochondrial genome are best suited for studying particular insect orders.
Chapter three of this thesis presents a comprehensive analysis of 62 hexapod mitochondrial genomes, including 55 from Insecta. Three different tree building programs were used, using both fast maximum likelihood and Bayesian inference methods, to assess the ability of single-gene and concatenated-gene phylogenies to recover currently recognised insect orders as monophyletic groupings.

The greatest amount of phylogenetic signal was recovered when all protein coding and rRNA genes were analysed together, regardless of optimality criterion used (PhyML, RaxML or MrBayes). Of the single-gene analyses COX1 out-performed all other genes, with the Bayesian inference even performing equally as well as the concatenated analysis. Noticeable differences were also observed between the three different phylogenetic programs used, PhyML, RaxML, and MrBayes. These differences were most pronounced in data sets with poorer phylogenetic signal (i.e. single-gene analyses). Despite reports that RaxML presents a more exhaustive search of tree space problems than PhyML (Stamatakis et al., 2005), in general RaxML appeared to perform the worst of the three methods. Instead, the findings of this chapter agree with claims by Alfaro et al. (2003), that Bayesian topology searches support more correct monophyletic groupings than fast maximum likelihood, and also requiring fewer character data to recover high clade support.

Additionally claims have been made that mitochondrial genes are only suitable for resolving relationships younger than 5-10 MYA (Lin and Danforth, 2004). However chapter three demonstrates that mtDNA is capable of recovering many major insect orders, and that given suitable markers, sufficient phylogenetic signal can be retrieved by multi-gene analyses using just a few key genes.

Single-gene analyses varied significantly in their ability to recover monophyly of major insect orders. However, four genes in particular, ND2, ND4, ND5, and COX1, did appear to perform better than the others, recovering most evolutionary relationships in agreement with established taxonomy. Being one of the largest genes in the mitochondrial genome, thus potentially delivering a
high degree of phylogenetic information, COX1 was found to perform the best out of these four genes, confirming the long held impression that COX1 is indeed a remarkably good phylogenetic marker. Consequently this work is particularly significant in the context of current DNA barcoding efforts focusing on COX1. Large scale standardized DNA barcoding schemes have been proposed to improve both the identification of unknown specimens, and recognition of new species (Hebert et al., 2003a; Stoeckle, 2003). Nevertheless, while COX1 appears to be the best single phylogenetic marker within the mitochondrial genome, single-gene analyses remain deeply flawed.

While mitochondrial genes are essentially linked, non-independent markers, the work presented in this chapter also confirms sequence length as an effective indicator of ability to recover monophyletic groupings within Insecta. Accordingly, strong recommendations are made that reconstructions of insect evolutionary relationships only be made from multiple-gene analyses. Frequently phylogenetic markers are chosen based on the availability of universal primers, rather than in consideration of their suitability to the hypotheses being tested. In recognition of criticisms that insect systematics has demonstrated a fundamental lack of co-ordination in the choice of molecular markers used (Caterino et al., 2000), large scale standardization of taxonomic sampling of other mitochondrial genes, akin to that already directed at COX1, are should also be strongly encouraged if multi-gene analyses are to become universally applied.

Finally, this chapter explored the relationship between the rate of DNA polymorphism, calculated by level of nucleotide diversity using a sliding window technique, and phylogenetic signal. While rates of DNA polymorphism in mitochondrial genes were found to exceed the capacity to provide clues to phylogenetic utility, some evidence was found that high levels of nucleotide diversity could instead be reflecting unequal rates of sequence evolution among lineages, or differences in rate of change among sites within genes, both of which can be accounted for when applying appropriate substitution models.
7.3 Molecular systematics of Calliphoridae

The defining characteristic of blowflies is the need for larval stages to feed on a proteinaceous substrate, commonly tissue of a vertebrate host. When larvae feed on tissue of a living host this process is known as myiasis. A range of larval feeding habits have developed within Calliphoridae, including: coprophagy, saprophagy, sanguinivory, ectoparasitism, both facultative and obligate. Understanding of the origin and evolution of these different behaviours is often restricted by a lack of widely agreed theories of evolutionary relationship and taxonomic classification within Calliphoridae (Stevens, 2003).

Insect systematics have traditionally favoured mtDNA and nuclear ribosomal DNA molecular markers (Otranto and Stevens, 2002), largely due to the relative ease with which these markers can be amplified (Baker et al., 2001). The vast majority of phylogenetic studies relating to parasitic arthropods have used single protein-coding genes (Shao and Barker, 2006). When nuclear and mitochondrial genes have been analysed together, nuclear genes have generally been found to outperform mitochondrial genes at resolving deeper level relationships (Reed and Sperling, 1999; Leys et al., 2000; Baker et al., 2001; Brady, 2002; Morris et al., 2002; Danforth et al., 2003; Lin and Danforth, 2004). However, substitution rates in mitochondrial genes are estimated to be between five and ten times that of nuclear DNA (Brown et al., 1979), making mtDNA particularly useful for inferring relationships between recently diverged species and in population-level analyses (Stevens and Wall, 1997b; Shao and Barker, 2006). It is now becoming increasingly common within insect systematics to combine mitochondrial and nuclear genes, providing two unlinked sets of data that have evolved under essentially different constraints (Lin and Danforth, 2004). Focusing on molecular markers with different rates of substitution not only improves phylogenetic resolution at different levels of divergence, but also allows comparisons between evolutionary histories of gene trees and species trees to be explored.

Previous phylogenetic studies of Calliphoridae have reported incongruence between mitochondrial and nuclear rRNA genealogies (Stevens,
Chapter 7

General discussion

In an attempt to improve resolution, the work presented in Chapter 4 combines sequence data from mitochondrial (COX1) and nuclear (EF-1α) protein-coding genes, and a nuclear rRNA (28S).

As well as attempting to resolve key ambiguous relationships within Calliphoridae, an increased sampling effort was also placed on populations of *L. sericata* and *L. cuprina*, in an attempt to assess the extent to which hybridisation is occurring between these two species. Despite *L. sericata* and *L. cuprina* demonstrating low frequency hybridisation under laboratory conditions (Ulyett, 1945), only a small number of studies have reported wild hybrid populations (Stevens and Wall, 1996b, 1997b; Nelson *et al.*, 2007; Tourle *et al.*, 2009). The work presented in Chapter 4 identifies two independent *L. cuprina* samples containing introgressed mtDNA COX1 haplotypes from *L. sericata*. These two species are morphologically very similar and, while hybrids have demonstrated intermediate or *L. cuprina* morphology (Mackerras, 1933; Ulyett, 1945; Waterhouse and Paramonov, 1950; Stevens *et al.*, 2002), backcrossing hybrids may be very difficult to distinguish from parent species, meaning that rates of backcrossing can easily be underestimated (Mallet, 2005). Accurate species identification of these two species is particularly important considering their use in medical and forensic applications. Hybridisation could also significantly impact current and future methods for the control of these myiasis-causing flies, for example, by the transfer of insecticide resistance genes.

Patterns of blowfly evolution suggest multiple independent origins of parasitic behaviour within Calliphoridae, with even closely related taxa exhibiting different larval feeding strategies. Within the sub-family Auchmeromyiinae, parasitism seems to have uniquely developed independently by both saprophagy and sanguinivory routes.

While the majority of the parasitic taxa represented in this study have mammalian host species, a small number represent more specialist parasites. Convergent evolution of such specialist parasitic behaviour is also suggested by several pairs of unrelated taxa parasitizing similar host species (e.g. the
earthworm parasites, *Onesia tibialis* and *Pollenia rudis*). Of the blowflies parasitizing mammals, the genus *Lucilia* in particular seems to have readily exploited the opportunity to parasitize domesticated animals such as sheep and cattle. Rates of blowfly myiasis in healthy wild animal populations are largely unreported, suggesting that infestations are much more common in domesticated animals, such as livestock. While this may also be due to high stocking densities of livestock providing more opportunities for myiasis-causing flies, the selective breeding of animals to improve milk, meat, and wool yields has also increased the susceptibility of these animals to parasitism and disease.

In summary, the evolution of myiasis within Calliphoridae not only represents convergent evolution of specialist parasitism between taxa in the same family, but also highlights the role recent animal domestication practices may have played in the adaptive radiation of larval feeding behaviour in some genera.

### 7.4 Intra-specific diversity of *Cochliomyia hominivorax*

*Cochliomyia hominivorax*, the New World screwworm fly, is widely considered the most destructive insect parasite of livestock in tropical and sub-tropical regions of the Western hemisphere, and is responsible for the loss of hundreds of millions of dollars in the livestock industry (Klassen and Curtis, 2005; Vargas-Teran *et al.*, 2005). Since the 1950s, *C. hominivorax* populations in North and Central America have been the target of virtually continuous eradication attempts by sterile insect technique (SIT). While successful eradication programmes have also been implemented on some Caribbean islands, the remaining endemic populations may acts as reservoirs for the potential reintroduction of *C. hominivorax* into areas where the fly has already been eliminated. However, in some areas, such as Jamaica, SIT-based control programmes have failed. Reasons for the failure of SIT-based programmes in some locations are unknown. It has been hypothesized that failure of SIT may be related to mating incompatibility between sterile and wild flies, or the existence of sexually incompatible cryptic species.
Molecular evidence has supported some genetic differentiation between populations of *C. hominivorax* (Roehrdanz, 1989; Infante-Malachias and Azeredo-Espin, 1995; Taylor *et al.*, 1996; Infante-Malachias *et al.*, 1999). Higher levels of diversity among South American populations has also suggested a South American origin to the species, with founder effects giving rise to populations in Central and North America, and throughout parts of the Caribbean (Infante-Malachias *et al.*, 1999; Azeredo-Espin and Lessinger, 2006). While both nuclear and mtDNA markers have been used to investigate intra-specific diversity within *C. hominivorax*, molecular studies have focused heavily on analysis of mitochondrial data (Taylor *et al.*, 1996; Azeredo-Espin and Lessinger, 2006).

While the phylogenetic study presented in Chapter 4 used the nuclear rRNA gene 28S, preliminary comparisons between published 28S sequences of *C. hominivorax* and its sister taxa *C. macellaria* showed little variation. Consequently, this gene was replaced with the mitochondrial rRNA 12S, which displayed considerably more variation between the two Cochliomyia taxa, and so would be more likely to reveal variation between populations of *C. hominivorax*. Accordingly, using two protein-coding genes (nuclear EF-1α, and mitochondrial COX1) and one rRNA gene (mitochondrial 12S), the work presented in Chapter 5 investigates intra-specific phylogenetic relationships and associated biogeographic patterns between *C. hominivorax* populations. This study particularly focused on locations representing populations involved in, or earmarked for, forthcoming SIT programmes. In addition, the study uniquely included two historic specimens from North America, collected in Texas in 1933 and 1953, prior to initiation of the SIT-based eradication programme.

Phylogenetic analysis of these data indicated sub-structuring of fly populations on several of the larger Caribbean islands, suggesting a period of isolation and/or founder effects following colonization from South America. Under a parsimony optimality criterion, both COX1 and the combined three-gene data set recovered a distinct Cuban clade, although only COX1 recovers node support for this group. Other studies have also identified Cuban populations as distinct from other Caribbean islands (Taylor *et al.*, 1996; Lyra *et al.*...
al., 2009; Torres and Azeredo-Espin, 2009). Significantly, our findings do not support a North American origin for Cuban flies, as has previously been suggested. Instead, a distinct and well supported separation is observed between the two North American taxa and the rest of the samples from South America and the Caribbean. While the study by Taylor et al. (1996) included only a single Cuban sample, the authors noted that this taxon appeared more closely related to Central American populations, suggesting multiple origins for Caribbean C. hominivorax populations. In contrast, the work presented in Chapter 5 implied that while the Cuban samples do seem to form a distinct clade, this clade tends to group with the South American and other Caribbean populations, relative to the two North American taxa. While this may suggest a South American origin to Cuban populations, followed by periods of isolation and/or founder effects, the inclusion of Central American samples would be necessary to confirm this hypothesis.

Three different molecular markers were used in this study, two mitochondrial and one nuclear. When comparing the three genealogies, the mitochondrial COX1 sequence data reveal a greater degree of geographic differentiation within C. hominivorax than either the mitochondrial 12S or the nuclear EF-1α data. While the phylogenetic signal in COX1 may be explained as the result of the uni-parental inheritance and non-recombining nature of the mitochondrial gene, the lack of similar signal in 12S suggests that some regions of the mtDNA are experiencing much faster rates of evolution than others.

7.5 Antigenic diversity within Calliphoridae

Due to the parasitic nature of myiasis-causing flies, it is likely that their evolution has been significantly affected by host immune responses. Understanding these immune responses, and the ways in which larval strategies have evolved to cope with these defences will, in turn, facilitate the development of both diagnostic tools and vaccines to monitor and control blowfly populations. The work presented in Chapter 6 attempts to characterize antigenic proteins present in four blowfly species of medical and veterinary importance. In particular, this
work aimed to identify antigens specific to particular types of parasitic behaviour and, potentially, to isolate species-specific markers.

The four species investigated in this chapter included three *Lucilia* species (*L. sericata*, *L. cuprina*, and *L. caesar*), and a single *Calliphora* species (*C. vicina*). Globally, *L. sericata*, *L. cuprina* and, to a lesser extent, *L. caesar*, represent the key causal agents of myiasis in sheep (‘sheep strike’), costing the Australian wool industry alone, an estimated AUS$ 250 million a year. Of these three *Lucilia* species, *L. sericata* tends to be the primary cause of sheep strike in cool temperate habitats such as Europe or New Zealand, while *L. cuprina* is more significant in sub-tropical/warm temperate habitats such as Australia and South Africa. In contrast, *L. caesar* is considered a secondary facultative parasite of sheep, restricted largely to northerly Palaearctic regions. Differences in parasitic behaviour between the species are perhaps most apparent when they occur in sympatry. For example, both *L. sericata* and *L. cuprina* are found in Australia and South Africa; however, in these regions *L. sericata* is considered an occasional parasite of sheep, while *L. cuprina* represents an aggressive agent of myiasis. This study also included *C. vicina*, one of the most common indicator species used in forensic entomology, and an occasional cause of myiasis in animals. Including these four species in the study not only permitted an assessment of different forms of parasitic behaviour to be made, but also allowed comparisons between two different genera, and also between closely related species within the same genus.

Comparisons between first, second, and third instar larval stages revealed the greatest antigenic diversity between species to be at the first instar stage. This finding supports observations by Broadmeadow *et al.* (1984) that first instar larvae are capable of initiating significant local inflammatory responses in sheep within 1-2 hour of contact. In contrast to second and third instar stages, first instar larvae use excretory-secretory (ES) products to damage tissue and initiate wound development. However, Tabouret *et al.* (2001) demonstrated that the larval ES products of other closely related myiasis-causing flies are only weakly antigenic in sheep. These findings combined with the rapid onset of immune responses in sheep strike suggest
that merely contact with the larval cuticle may be sufficient to induce hypersensitivity responses.

While the two blowfly genera represented in this study are morphologically distinct as adult flies, immature stages of development display very similar gross morphology. While DNA sequence-based analysis is currently the only routinely used molecular species identification tool for Calliphoridae, the findings presented in Chapter 6 explore the potential for identifying species-specific antigenic markers. Consequently, the objective of this part of the research was to investigate whether reproducible antigenic differences between first instar larvae were capable of discriminating between the four calliphorid species, and exploration of the hypothesis that physiological differences may exist where morphological diversity is limited.

The vast majority of studies investigating larval antigens in myiasis-causing flies have focused almost exclusively on identifying vaccine candidates for the prevention of fly strike. Despite huge resources having been dedicated to antigen research, no commercially viable vaccines exist for the control of blowfly myiasis, largely due to an apparent lack of naturally acquired immunity (Elkington and Mahony, 2007). The potential of such antigens in the development of diagnostic techniques, however, has been largely neglected. Blowfly diagnostic tools could potentially have wide-reaching applications, including use in monitoring fly populations as part of livestock myiasis control programmes, and in forensic entomology.

While the work presented in Chapter 6 exposed distinct antigenic profiles for each of the four blowfly species, isolating and identifying individual antigens proved much more problematic. In contrast to work by Roelfstra et al (2009), who successfully identified 15 antigenic proteins from the myiasis-causing fly Gasterophilus intestinalis (Oestridae), this study has been fundamentally limited by an apparent lack of homologous proteins in current databases. Unfortunately, while a blowfly genome project is currently being completed, this work is being funded through the Australian Wool Innovation Ltd, and as such is not publicly available.
7.6 Future work

The work explored in this thesis has uncovered several directions for further work. Phylogenetic analysis of Calliphoridae has supported recent reports of putative hybrid populations between the two key myiasis-causing blowflies in sheep, *L. sericata* and *L. cuprina*. In addition to the problem hybridisation may cause in species identification (by either morphological or molecular methods), potential introgression of insecticide resistance genes, for example, could significantly impact current and future methods of controlling myiasis in livestock. The work presented here exposes the need for a global study investigating the extent of hybridisation and introgression between *L. sericata* and *L. cuprina*.

The work presented in this thesis also explored differences in patterns of antigenic diversity associated with different styles of parasitism, and in doing so attempted to identify potential species-specific antigen markers. However, antigen characterization was severely hampered by a lack of homologous proteins in current molecular databases. However, the production of cDNA libraries have rapidly advanced vaccine development for several ectoparasites (e.g. the scabies mite, *Sarcoptes scabiei*) by the detection of vaccine candidates through immunoscreening and identification of ESTs (Walton et al., 2004); for example, by isolating targets with homology to proteins from other species known to be involved in growth and survival (Nisbet and Huntley, 2006). Consequently, construction of a cDNA library suitable for 454 sequencing of *L. sericata* is currently being planned, and will provide an appropriate genomic framework with which to identify the unknown proteins described here. Using a more targeted suppressive-substrate hybridization (SSH) construction of a cDNA library could also allow the restriction of amplification to only those proteins that differ between the transcriptomes of the different species (Nisbet and Huntley, 2006).

This thesis has demonstrated that larval antigens have great potential in the development of successful diagnostic tools for the identification of calliphorid species. Once antigen markers have been successfully characterized, species-specific murine monoclonal antibodies (MAbs) can be
produced, using the diagnostic proteins identified with the rabbit PAb as immunogens. Raising these MAbs will then allow the development of rapid diagnostic tools for the identification of blowfly species, for example, exploiting lateral flow technology (McDonagh et al., 2009b). Rapid diagnostic tools utilising this technology have already been developed for a wide range of human diseases including visceral leishmaniasis (Sundar et al., 1998), HIV/AIDS (Iweala, 2004), malaria (Moody, 2002), SARS (Kammla et al., 2008) and invasive aspergillosis (Thornton, 2008).
References


Australian Pesticides and Veterinary Medicines Authority (2006). The reconsideration of approvals of the active constituent diazinon, registrations of products containing diazinon and approval of their associated labels (Part 2): Preliminary review of findings, Canberra, Australia.


Bell, K. and Sackett, D. (2005). The economic and research implications of managing Merino sheep without mulesing. Sheep Updates, Department of Agriculture and Food Western Australia.


progressive necrotizing bacterial infection. American Journal of Tropical Medicine and Hygiene 64(3-4): 159-161.


IAEA (1998). Thematic plan for sterile insect technique for Old and New World screwworm, Vienna, Austria


Leys, R., Cooper, S. J. B. and Schwarz, M. P. (2002). Molecular phylogeny and historical biogeography of the large carpenter bees, genus *Xylocopa*
(Hymenoptera: Apidae). *Biological Journal of the Linnean Society* **77**: 249-266.


Mazyad, S. A. M. and Soliman, M. (2006). Biological and ecological studies on the myiasis producing larvae of *Megaselia scalaris* with special


Vreysen, M. J. B., Barclay, H. J. and Hendrichs, J. (2006). Modelling of preferential mating in area-wide control programmes that integrate the
release of strains of sterile males only or both sexes. *Annals of the Entomological Society of America* **99** 607 - 616.


Appendix A
Phylogenetic analysis of New World screwworm fly, <i>Cochliomyia hominivorax</i>, suggests genetic isolation of some Caribbean island populations following colonization from South America

L. Mc DONAGH¹, R. GARCÍA² and J. R. STEVENS¹

¹Hatherly Laboratories, School of Biosciences, University of Exeter, Exeter, U.K. and ²International Atomic Energy Agency, Tuxtla, Mexico

Abstract. Larval infestations of the New World screwworm (NWS) fly, <i>Cochliomyia hominivorax</i>, cause considerable economic losses through the direct mortality and reduced production of livestock. Since the 1950s, NWS populations in North and Central America have been the target of virtually continuous eradication attempts by sterile insect technique (SIT). Nevertheless, in some areas, such as Jamaica, SIT-based control programmes have failed. Reasons for the failure of SIT-based programmes in some locations are unknown, but it is hypothesized that failure may be related to the mating incompatibility between sterile and wild flies or to the existence of sexually incompatible cryptic species. Accordingly, the current research investigates intraspecific phylogenetic relationships and associated biogeographic patterns between NWS populations from the Caribbean and South America, which represent those populations involved in, or earmarked for, forthcoming SIT programmes. Uniquely, this study also includes analyses of two North American samples, collected in Texas in 1933 and 1953 prior to initiation of the SIT-based eradication programme. The study utilizes three nucleotide datasets: elongation factor-1α (nuclear); cytochrome oxidase subunit 1 (mitochondrial), and 12S rRNA (mitochondrial). Phylogenetic analysis of these data, representing populations from across the Caribbean, South America and Texas, indicates sub-structuring of fly populations on several of the larger Caribbean islands, suggesting a period of isolation and/or founder effects following colonization from South America; significantly, our findings do not support a North American origin for Cuban flies. The importance of these findings in the light of proposed SIT programmes in the region is discussed.

Key words. <i>Cochliomyia hominivorax</i>, 12S rRNA, Calliphoridae, cytochrome oxidase subunit 1, elongation factor-1α, evolution, New World screwworm fly, sterile insect technique.

Introduction

Since its first use in Florida in 1958, the sterile insect technique (SIT) has successfully been used to eradicate the New World screwworm (NWS) fly, <i>Cochliomyia hominivorax</i> (Coquerel), from all of mainland North America and most of Central America (Klassen & Curtis, 2005). Today, <i>C. hominivorax</i> has been eradicated as far south as Panama, where an ongoing weekly release of 50 million sterile males close to the border with Colombia has created a sterile fly barrier in Darien Province to prevent the reintroduction of NWS fly from South America (Klassen & Curtis, 2005; Vargas-Terán et al., 2005). <i>Cochliomyia hominivorax</i>
has also been eradicated from the Virgin Islands and Puerto Rico in the Caribbean, and the Jamaican government initiated a SIT programme in 1999. No such programmes have been attempted on Cuba or Hispaniola, leaving, as a consequence, reservoirs for the potential reintroduction of Cochliomyia hominivorax into areas where the fly has been eliminated (Klassen & Curtis, 2005).

However, despite the overwhelming success of the NWS fly SIT programme in the U.S.A. and Central America, since its outset significant differences in the efficacy of the technique have been apparent, with variations in induced egg sterility ranging from 100% to 15%. For example, whereas the first field trials conducted in Florida resulted in almost 100% egg sterility, further field trials using the same southern U.S.A. strain of irradiated flies produced very different results: the release of sterile male flies on an island off the coast of Venezuela resulted in observed egg sterility of just 15% (Klassen & Curtis, 2005). Most recently, a SIT programme in Jamaica in 1999–2005, using flies originating from Mexico, also failed (A. Robinson, International Atomic Energy Agency, personal communication, 2007) and the programme is currently suspended. Reasons for the failure of SIT-based control programmes in some locations are unknown, but it is hypothesized that failure may be related to the mating incompatibility between sterile and wild flies, or to the existence of sexually incompatible cryptic species, as proposed by Richardson et al. (1982a, 1982b). (N.B. The existence of cryptic species within C. hominivorax remains controversial. See, e.g.: LaChance et al. [1982]; Taylor & Peterson [1994].) To investigate these hypotheses, an assessment was undertaken, in collaboration with the International Atomic Energy Agency (IAEA), Vienna, of genetic variation within and between populations of C. hominivorax using DNA sequence data from nuclear and mitochondrial (mtDNA) genes to type fly specimens and populations from across South America, the Caribbean and from two, now extinct, southern U.S.A. populations.

Although both nuclear and mitochondrial DNA sequence data have been used to explore population genetic variability in C. hominivorax, molecular studies have focused heavily on analysis of mitochondrial markers (e.g. Taylor et al. 1996; Azeredo-Espin & Lessinger, 2006). In insects and arthropods in general, the majority of mtDNA-based evolutionary and population studies have frequently used only single protein coding genes from a limited range of those potentially available (Shao & Barker, 2007), with many mtDNA genes being largely disregarded and others receiving only sporadic attention (Caterino et al., 2000). Nevertheless, phylogenetic studies of insects based on mtDNA have been far more frequent than those using nuclear DNA (Lessinger et al., 2000), with cytochrome oxidase subunit I (COI) being among the most commonly used (Zhang & Hewitt, 1997). In addition to practical and operational considerations (which certainly account for part of its continued popularity), mtDNA also appears particularly useful for inferring relationships between closely related and recently diverged species, a result of the apparently more rapid lineage arrangement, compared with nuclear DNA (Wells et al., 2007).

However, in previous investigations of the mitochondrial genome of C. hominivorax, studies have found COI to be among the most conserved gene sequences among several dipteran species (Lessinger et al., 2000). Accordingly, in addition to COI, this study also utilized sequence data from a second mitochondrial gene, 12S, and a nuclear gene, elongation factor-1α (EF-1α). All three genes have been widely used as phylogenetic markers in insects (Caterino et al., 2000); EF-1α is almost the only single-copy nuclear gene to have been used as a molecular marker in such studies (Caterino et al., 2000).

As outlined above, genetic differences between populations and the possible existence of reduced mating success between divergent populations and/or the existence of sexually incompatible cryptic species (as proposed by Richardson et al., 1982a, 1982b) may be critical factors in determining the ultimate success of screwworm eradication programmes.

**Materials and methods**

**Samples and DNA extraction**

Details of all specimens analysed in this study are given in Table 1. Briefly, specimens of C. hominivorax were obtained from various locations across South America and the Caribbean, including Brazil, Colombia, Cuba, the Dominican Republic, Ecuador, Jamaica, Peru, Venezuela, Trinidad and Uruguay (Fig. 1). Two potentially key islands, Cuba and the Dominican Republic, were sampled more intensively (Fig. 2). Two Calliphoridae outgroup taxa from the closely related subfamily Luciínæ (Stevens, 2003) were used.

All specimens were stored in 100% ethanol at 4 °C, except the Uruguayan samples, which arrived as DNA extracts direct from the University of Campinas, Brazil, and were re-hydrated with distilled water for a minimum of 1 h prior to polymerase chain reaction (PCR) amplification. Where adult flies specimens were available, thoracic flight muscle tissues were extracted, avoiding contamination from ingested protein, parasites or eggs (Stevens, 2003), and the remainder of the adult fly kept as a voucher specimen. The entire specimen of second- and third-stage larvae was used. DNA extraction was carried out using a salt extraction method (Aljanabi & Martínez, 1997). Successfully extracted DNA samples were stored at −20 °C until analysis.

*Historical* North American samples and DNA extraction

Two specimens belonging to historical Texan populations of C. hominivorax, on loan from the collection held at the Natural History Museum, London, were also sampled. DNA extractions from these specimens, which had been dried and pinned in 1933 and 1953, were carried out using a DNeasy® Blood & Tissue Kit (Qiagen Ltd., Crawley, U.K.). DNA extraction was performed by removing thoracic flight muscle from the ventral side of the thorax, leaving as much of the taxonomically important morphology intact as possible.

To ensure that DNA obtained was that of the historic specimen and not a result of contamination from contemporary C. hominivorax material, all molecular work for the historical samples was carried out in a separate ‘clean’ room, using a...
Table 1. Cochliomyia hominivorax specimen details.

<table>
<thead>
<tr>
<th>Sample site location</th>
<th>Code</th>
<th>Capture date</th>
<th>Life stage</th>
<th>Accession number COI</th>
<th>Accession number 12S</th>
<th>Accession number EF-1α</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cuba, east</td>
<td>Las Tunas</td>
<td>A</td>
<td>Oct 2005</td>
<td>L3</td>
<td>FM867778</td>
<td>FM866410</td>
</tr>
<tr>
<td></td>
<td>Santiago de Cuba</td>
<td>B</td>
<td>Oct 2005</td>
<td>L3</td>
<td>FM867779</td>
<td>FM866409</td>
</tr>
<tr>
<td>Cuba, west</td>
<td>Pinar del Río</td>
<td>C</td>
<td>Nov 2005</td>
<td>L3</td>
<td>FM867780</td>
<td>FM867706</td>
</tr>
<tr>
<td>Cuba, centre</td>
<td>Ciego de Ávila</td>
<td>D</td>
<td>Oct 2005</td>
<td>L3</td>
<td>FM867769</td>
<td>FM867714</td>
</tr>
<tr>
<td>Cuba, Isla de la Juventud</td>
<td>Ciro Redondo</td>
<td>E</td>
<td>Nov 2005</td>
<td>L3</td>
<td>FM867780</td>
<td>FM866408</td>
</tr>
<tr>
<td></td>
<td>Micro</td>
<td>F</td>
<td>Nov 2005</td>
<td>L2, L3</td>
<td>FM867771</td>
<td>FM867712</td>
</tr>
<tr>
<td></td>
<td>Melvis</td>
<td>G</td>
<td>Nov 2005</td>
<td>L2, L3</td>
<td>FM867773</td>
<td>FM867711</td>
</tr>
<tr>
<td>Dominican Republic, northwest</td>
<td>Monte Cristi</td>
<td>H</td>
<td>Oct 2005</td>
<td>A</td>
<td>FM867760</td>
<td>FM867722</td>
</tr>
<tr>
<td></td>
<td>Duarte</td>
<td>I</td>
<td>Oct 2005</td>
<td>L3</td>
<td>FM867762</td>
<td>FM867720</td>
</tr>
<tr>
<td>Dominican Republic, south</td>
<td>Neiba</td>
<td>J</td>
<td>Oct 2005</td>
<td>L2, L3</td>
<td>FM867759</td>
<td>FM867724</td>
</tr>
<tr>
<td></td>
<td>Polo</td>
<td>K</td>
<td>Oct 2005</td>
<td>L3</td>
<td>FM867758</td>
<td>FM867723</td>
</tr>
<tr>
<td>Dominican Republic, centre</td>
<td>Santo Domingo</td>
<td>L</td>
<td>Oct 2005</td>
<td>L2, L3</td>
<td>FM867764</td>
<td>FM867718</td>
</tr>
<tr>
<td></td>
<td>Nigua</td>
<td>M</td>
<td>Sep 2005</td>
<td>A</td>
<td>FM867765</td>
<td>FM867717</td>
</tr>
<tr>
<td>Brazil</td>
<td>Minas Gerais</td>
<td>–</td>
<td>Nov 2000</td>
<td>A</td>
<td>FM867785</td>
<td>FM866402</td>
</tr>
<tr>
<td>São Paulo</td>
<td>–</td>
<td>Jan 2005</td>
<td>A</td>
<td>FM867752</td>
<td>FM867731</td>
<td>FM867823</td>
</tr>
<tr>
<td>Trinidad</td>
<td>–</td>
<td>Aug 2005</td>
<td>L3</td>
<td>FM867753</td>
<td>FM867730</td>
<td>FM867822</td>
</tr>
<tr>
<td>Venezuela/Colombia border</td>
<td>–</td>
<td>May 2005</td>
<td>L3</td>
<td>FM867747</td>
<td>FM867737</td>
<td>FM867828</td>
</tr>
<tr>
<td>Uruguay</td>
<td>–</td>
<td>–</td>
<td>DNA extract</td>
<td>FM867744</td>
<td>FM867739</td>
<td>FM867830</td>
</tr>
<tr>
<td>Colombia</td>
<td>Chigorodo</td>
<td>–</td>
<td>Mar/Apr 2007</td>
<td>L2, L3</td>
<td>FM867745</td>
<td>FM866404</td>
</tr>
<tr>
<td></td>
<td>Choco</td>
<td>–</td>
<td>–</td>
<td>FM867783</td>
<td>FM866404</td>
<td>FM867790</td>
</tr>
<tr>
<td></td>
<td>Turbo</td>
<td>–</td>
<td>–</td>
<td>FM867782</td>
<td>FM866405</td>
<td>FM867791</td>
</tr>
<tr>
<td>Ecuador</td>
<td>Napo</td>
<td>–</td>
<td>Jun 2007</td>
<td>L2, L3</td>
<td>FM867787</td>
<td>FM866406</td>
</tr>
<tr>
<td>Esmeraldas</td>
<td>Guayas</td>
<td>–</td>
<td>–</td>
<td>FM867786</td>
<td>FM867729</td>
<td>FM867821</td>
</tr>
<tr>
<td>Chilcayo</td>
<td>Guayas</td>
<td>–</td>
<td>–</td>
<td>FM867755</td>
<td>FM867727</td>
<td>FM867819</td>
</tr>
<tr>
<td>Tambopata</td>
<td>Chilcayo</td>
<td>–</td>
<td>May 2007</td>
<td>L2, L3</td>
<td>FM867754</td>
<td>FM867728</td>
</tr>
<tr>
<td>U.S.A.</td>
<td>Texas</td>
<td>–</td>
<td>1953</td>
<td>A</td>
<td>FM867751</td>
<td>FM867732</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1933</td>
<td>–</td>
<td>FM867750</td>
<td>FM867733</td>
<td>FM867824</td>
</tr>
</tbody>
</table>

L2, first-stage larva; L3, second-stage larva; A, adult.

Extracted DNA was then stored at −20 °C until analysis.

DNA sequences

Each DNA extraction was subjected to PCR amplification of an approximately 680-bp region of the mitochondrial COI, an approximately 780-bp region of EF-1α and a 600–700-bp region of the mitochondrial 12S gene.

Using published primers (Otranto et al., 2005), COI sequence data were obtained from approximately half of the Cuban and Dominican Republic samples. Accordingly, a new set of COI primers were designed (two forward, two reverse; Table 2) in an attempt to amplify the remaining templates. Initial attempts to utilize more generic arthropod EF-1α primers (provided by
J. K. Moulton, personal communication, 2004) proved unsuccessful and a new set of primers specific for *C. hominivorax* were designed. All 12S sequences were amplified using published primers (Lessinger & Azeredo-Espin, 2000), although successful PCR product sequencing was consistently only obtained for the reverse primer (SR-J-14233). At least two sequences were amplified for each gene from each of the locations (Table 1).

Standard PCR amplification involved 1 μL template DNA, 2.5 μL reaction buffer (160 mM [NH₄]₂SO₄, 670 mM Tris-HCl [pH 8.8 at 25 °C], 0.1% Tween-20), 1.25 μL MgCl₂ (50 mM), 0.5 μL forward and reverse external primers (10 mM), 1 μL of BIOTAQ™ Red DNA polymerase, and sterile H₂O to a final volume of 25 μL. All amplifications were carried out in Hybaid thermal cycler machines (Thermo Hybaid, Ashford, U.K.). COI reactions using the primers of Otranto *et al.* (2005) included an initial denaturation step of 94 °C for 5 min, followed by 35 cycles of 95 °C for 40 s, 48 °C for 30 s and 72 °C for 1 min, and a final elongation step of 72 °C for 10 min (Yu *et al.*, 2005). Reactions using the new screwworm-specific COI and EF-1α primers used modifications of the same generic PCR protocol (Hoelzel & Green, 1992). The new COI primer reactions employed a protocol whereby an initial denaturation step of 95 °C for 5 min was followed by 35 cycles of 55 °C for 1 min, 72 °C for 1 min, 94 °C for 30 s and 55 °C for 1 min, and a final elongation step of 72 °C for 10 min, modified from generic PCR practice. The EF-1α reactions used an initial denaturation step of 95 °C for 5 min, followed by 35 cycles of 58 °C for 1 min, 72 °C for 1 min, 94 °C for 30 s and 55 °C for 1 min, with a final elongation step of 72 °C for 10 min. All 12S reactions were carried out using an initial denaturation step of 94 °C for 2 min, followed by 30 cycles of 94 °C for 30 s, 58 °C for 1 min and 60 °C for 2 min, with a final elongation step of 60 °C for 10 min (Lessinger & Azeredo-Espin, 2000). All PCR reactions using ‘historical’ North American material were carried out in a separate, clean room to avoid possible contamination.

The resulting PCR products were separated by gel electrophoresis, with appropriate bands cut out and purified using a MinElute® PCR purification kit (Qiagen UK), according to the manufacturer’s protocol. Purified PCR products were sequenced on an ABI 3730 automated sequencer (Applied Biosystems Inc., Carlsbad, CA, U.S.A.), using a commercial sequencing facility (Lark Technologies, Inc., Saffron, Walden, UK).

**Sequence alignment and phylogenetic analysis**

Sequence fragments were checked for quality and edited manually before being assembled into a single consensus
sequence, using AutoAssembler 2.0 (Applied Biosystems, Inc.), from which the consensus sequence was exported. Sequences were then aligned using ClustalX (Thompson et al., 1997), with final adjustments performed by eye. Phylogenetic analyses were carried out using PAUP, (Swofford, 1998), with both maximum parsimony (MP) and maximum likelihood (ML) criteria being used to infer trees for individual gene datasets and a combined dataset.

 Parsimony tree inference was employed using an equal weighting scheme, gaps treated as missing data, tree bisection and reconnection (TBR) swapping, and random sequence addition, with 1000 replicates per search. The parsimony-based partition homogeneity test (incongruence length difference [ILD] test) (Farris et al., 1995) was used to evaluate the phylogenetic congruence between the genes and to verify whether the three datasets were combinable.

 For likelihood analyses, parameters produced from a series of nested hierarchical likelihood-ratio tests using MODELTEST Version 3.06, (Posada & Crandall, 1998), with heuristic searching (for 100000 rearrangements), TBR branch swapping and random sequence addition with 10 repeats, were used to implement the most appropriate model of evolution for each of the three genes and the combined dataset separately. Node support for all tree topologies was assessed using bootstrap analysis, using 100 replicates for ML analyses (restricted for computational time limits) and 1000 replicates for all MP analyses, with additional Bremer decay indices (Bremer, 1988) calculated on all parsimony trees.

 When differences between taxa are very small, a reduction in the accuracy of algorithmic-based distance methods can result in incorrect tree topology. By comparison, methods based on optimality criteria, such as parsimony and ML, provide the advantage of choosing one tree over another based on a mathematical ranking system, whereas algorithmic methods cluster taxa according to the conventions of the particular algorithm being used, although even then the best tree may not be found (Swofford & Sullivan, 2003). Parsimony has also been shown to be valuable under circumstances where rates of evolution are low (Swofford & Sullivan, 2003), such as between gene sequences from populations of the same species. Accordingly, data were analysed by both parsimony and ML methods, with detailed analysis of inter-gene phylogenetic congruence and levels of relative support for clades performed using parsimony-based methods.

### Table 2. Primer details.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Forward primer</th>
<th>Sequence 5′–3′</th>
<th>Reverse primer</th>
<th>Sequence 5′–3′</th>
</tr>
</thead>
<tbody>
<tr>
<td>EF-1α</td>
<td>External amplification</td>
<td>B1</td>
<td>CCCATYTCCGGHTGGCAGCGG</td>
<td>C1</td>
</tr>
<tr>
<td></td>
<td>Internal amplification</td>
<td>F. Int</td>
<td>GTGGYATCGGHACATGACC</td>
<td>Int. R1</td>
</tr>
<tr>
<td>COI</td>
<td>External amplification*</td>
<td>COI F1</td>
<td>ATAGACGTTGAYACCWGAGC</td>
<td>COI R1</td>
</tr>
<tr>
<td></td>
<td>Internal amplification†</td>
<td>COI Int.F1</td>
<td>GTATTTCGCTATTATGCGG</td>
<td>COI Int.R4</td>
</tr>
</tbody>
</table>

*Position on Cochliomyia hominivorax mtDNA genome (AF260826) 3492–4171 bp.  
†Position on Cochliomyia hominivorax mtDNA genome (AF260826) 3759–3895 bp.

### Results

#### Phylogenetic analysis: EF-1α

The EF-1α sequence dataset consisted of 743 aligned nucleotide sites, of which 50 were found to be parsimony informative. Parsimony analysis of the EF-1α data found a large number of most-parsimonious trees (MPTs) (L = 101). A strict consensus tree resulted in a loss of all resolution within the Caribbean and South American NWS taxa, but grouped together both U.S.A. samples, albeit with no node support.

The best-fit ML model for the EF-1α dataset was found to be a Hasegawa–Kishino–Yano (HKY) (Hasegawa et al., 1985), including invariable sites (+I), and rate variation among sites (+G) (base frequencies = [A 0.2339; C 0.3010; G 0.2259]; number of substitution types = 2; T ratio = 1.6875; rate = gamma [rate variation]; shape = 0.8377; proportion invariant sites = 0.6982).

Two ML trees were found for the EF-1α dataset (not shown) and node support was evaluated using bootstrap values. Strong support for monophyly of the NWS samples was found, although intraspecific bootstrap support was very low.

#### Phylogenetic analysis: 12S

The 12S sequence data contained 1239 characters, of which just 53 were parsimony informative sites. A strict consensus tree of MPTs (L = 409) resulted in a lack of meaningful resolution within the NWS taxa.

The best-fit ML model for the 12S dataset was found to be Felsenstein 81 (F81) (Felsenstein, 1981), including rate variation among sites (+G) (base frequencies = [A 0.4230; C 0.1164; G 0.4689]; number of substitution types = 1; rate = gamma; shape = 0.6480; proportion invariant sites = 0).

The ML analysis found a single tree which displayed a lack of resolution for the South American and Caribbean samples, but grouped the two U.S.A. samples together, with a strong bootstrap value of 99%.

#### Phylogenetic analysis: COI

The COI sequence data, although the smallest dataset, contained 78 parsimony informative sites. Moreover, by contrast with the EF-1α and 12S datasets, a strict consensus of all MPTs (L = 274) did conserve some resolution within the screwworm...
clade (Fig. 3), with a distinct 10-taxon ‘Cuban clade’ (with the exception of one Brazilian sample) preserved, although with relatively low bootstrap support.

The best-fit ML model for the COI dataset was found to be a transitional model (TIM), including invariable sites (+I), and rate variation among sites (+G) (base frequencies = [A 0.3083; C 0.1625; G 0.1511]; number of substitution types = 6 [generalized time-reversible]; r matrix = [A–C 1.0000; A–G 3.3693; A–T 2.1386; C–G 2.1386; C–T 7.5244]; rate = gamma; shape = 0.7496; proportion invariant sites = 0.4459).

Two ML trees were found, again with the noticeable Cuban clade in both topologies. Mapping bootstrap values on to one of the ML trees revealed strong support for the Cuban clade (Fig. 3).

**Combined dataset**

The parsimony-based ILD test, performed on a partitioned combined three-gene (COI, 12S and EF-1α) dataset, revealed a significant difference ($P < 0.01, n = 1$) in topology between the genealogies. This suggests the data partitions have significantly different signals under a 95% significance level, and as such are not compatible. However, in an attempt to identify clades which are congruent between datasets, a combined dataset analysis was performed.

Maximum parsimony analysis of the combined data resulted in MPTs with a length of 924, with a strict consensus preserving an even larger 12-taxon Cuban clade, including the Minas Gerais taxon from Brazil (not shown); this Cuban grouping received high Bremer support in the combined dataset. Throughout the combined tree, regions of apparent incongruence between the genealogies (as implied by the ILD test) are identified by negative Bremer support values, indicating where the data partition did not support a particular node. From this it is possible to see that the Cuban clade is not one of the areas of conflict between genealogies, although some of the topology within this clade does receive negative Bremer values. The large number of MPTs found in all datasets (combined and single-gene) reflect the largely uniform nature and lack of resolving power of these sequence data for *C. hominivorax*.

**Fig. 3.** Rooted phylogram constructed by maximum likelihood analysis of COI sequence data for *Cochliomyia hominivorax*. Bootstrap values are shown at each node with > 50% bootstrap support. Two *Lucilia* species were used as the outgroup.
The best-fit ML model was found to be a generalized time-reversible (GTR) (Tavaré, 1986) including invariable sites (+I), and rate variation among sites (+G) (base frequencies = [A 0.3418; C 0.1735; G 0.1361]; number of substitution types = 6; r matrix = [A–C 1.9887; A–G 2.7059; A–T 2.7872; C–G 1.0890; C–T 4.6986]; rate = gamma; shape = 0.6723; proportion invariant sites = 0.5283).

Conversely, node support on the single ML tree, found during the likelihood analysis, (Fig. 4), resulted in a loss of support for the original 10-taxon Cuba grouping. However, grouping of the samples from Texas away from the remaining South American and Caribbean samples is very strongly supported (Fig. 4), suggesting a clear North-South genetic divide.

Discussion

Regional genetic variability of NWS fly populations

Both COI and multi-gene parsimony-based genealogies group populations from Cuba together. Indeed, the Cuban clade was the only major grouping within the South American and Caribbean samples to be supported. The majority of the Dominican Republic specimens were generally also found to group together in the ML trees (not shown), albeit without node support, signifying that specimens from geographically diverse populations appear to show a greater affinity for one another than they do for flies from other locations. Significantly, however, Dominican Republic populations appeared to be more closely related to flies from other areas of South America than they did to populations from Cuba.

One notable anomaly within the Cuban clade is the presence of a Brazilian sample (Minas Gerais). Whereas the remaining non-Cuban or non-Dominican Republic taxa appear indiscriminately placed within the tree, the Minas Gerais sample appears strongly associated with Cuban screwworm populations. This may be the result of one of a number of factors, including: historical lineage sorting; transportation of fly-infested livestock, or, in view of the low number of informative characters identified, stochastic convergent evolution. In addition, the possibility of a laboratory error cannot be ruled out.

Fig. 4. Unrooted phylogram constructed by maximum likelihood analysis of the combined three-gene (COI, EF-1α, rRNA 12S) sequence dataset for Cochliomyia hominivorax. Bootstrap values are shown at nodes with > 50% bootstrap support.
Possibly the most striking relationship found in this study was that of the strongly supported separation of the two Texan samples from the South American and Caribbean samples, suggesting a possible north/south divide. Previously, a nucleotide divergence of $>2\%$ was observed between Jamaican and Mexican samples (Roehrdanz, 1989) and it has been suggested that such a result supports a north/south differentiation, with Jamaican populations originally colonized from South America (Azeredo-Espin & Lessinger, 2006).

A previous mtDNA-based study by Taylor et al. (1996), which incorporated the findings of Roehrdanz (1989), compared *C. hominivorax* samples from around the Caribbean, Central and South America, and showed haplotypes to subdivide into three lineages. The lineages identified by Taylor et al. (1996) were ‘North and Central America’ (which contained Costa Rican and Cuban samples, the only significant clade in their cladogram, with bootstrap support of 90%), ‘South America’, and ‘Jamaica’ (with both the Jamaican and one of the Trinidad samples sharing a comparatively different haplotype of their own), although total geographic conformity was not supported and only some support for partial phylogeographic structuring was present. Although samples from the Caribbean in Taylor et al. ’s (1996) study were limited, the authors noted that the single Cuban sample appeared to be more closely related to Central American populations, whereas the two Dominican Republic samples grouped with those from South America, suggesting the possibility of multiple origins of *C. hominivorax* throughout the Caribbean.

Significantly, although the results of our study provide support for a separate Cuban clade, they do not link this Cuban group with a North American lineage, as previously hypothesized by Taylor et al. (1996). Furthermore, although the distinctiveness of the majority of Cuban samples may be indicative of a sustained period of isolation and/or founder effects, their placement relatively close to other Caribbean and South American samples and well apart from the Texan specimens (Fig.4) suggests that, on the contrary, Cuban *C. hominivorax* may have originated from a southern lineage, which colonized the Caribbean islands from South America.

**Mitochondrial vs. nuclear genealogy**

Population studies using mtDNA have commonly used sequence data from a single gene, with COI one of the most commonly used (Liu et al., 2006). It is now widely recognized, however, that the use of single-gene phylogenies (especially mtDNA gene phylogenies) to represent overall species evolution can frequently be misleading (e.g. Stevens & Wall, 1996; Stevens et al., 2002; Wells et al., 2007) and multi-gene phylogenies are now recognized as essential. Where intraspecific relationships based on single-gene sequences are sometimes poorly resolved (Shao & Barker, 2007), multiple concatenated nucleotide sequences may contain sufficient genetic differentiation to resolve such relationships. Similarly, combined molecular and morphological studies have identified the possible existence of distinct geographical ‘races’ in the Old World screwworm fly, *Chrysomya bezziana*, (Hall et al., 2001).

In the current study, the use of a multi-gene analysis was found to both increase the size of the Cuban clade, previously reflected in only one of the three single-gene phylogenies (COI), and to increase support for the grouping of Cuban samples. This suggests that although the phylogenetic signal within the COI dataset provides most of the discriminatory power to identify a Cuban clade, the two other genes also offer some, albeit weak, support for such a grouping. As noted, however, a congruence test (ILD) performed on the partitioned three-gene dataset (COI, 12S and EF-1α) indicated significant incongruence between the overall genealogies. This potentially lends further support to the affinity between the Cuban samples because, even with significant conflict between genealogies, a Cuban grouping is still preserved.

When comparing the three genealogies, it seems the mitochondrial COI sequence data reveal a greater degree of geographic differentiation within *C. hominivorax* than either the mitochondrial 12S or the nuclear EF-1α data. However, although the phylogenetic signal in the COI data may be explained as the result of the uniparental inheritance and non-recombining nature of the mitochondrial gene, the lack of similar signal in the 12S data suggests that some regions of the mtDNA are experiencing faster rates of evolution than others. Further mtDNA-based phylogenetic analyses of *C. hominivorax* will provide an interesting insight into whether such intraspecific relationships are universal throughout the *C. hominivorax* mitochondrial genome. One suggested explanation for such a discrepancy between phylogeographic patterns between the nuclear and mitochondrial molecular markers refers to the increased movement of livestock between and within North, Central and South America over the past 400–500 years (Taylor et al., 1996). For example, since the transportation of livestock began to increase significantly from the 16th century onwards, populations of *C. hominivorax* that were formerly isolated or that experienced only limited gene flow have had considerably more opportunity to mix (Taylor et al., 1996; Azeredo-Espin & Lessinger, 2006).

Of the Caribbean islands, only the Virgin Islands and Puerto Rico have successfully eradicated *C. hominivorax* to date, although the Jamaican government initiated a SIT programme in 1999. However, no such programmes have yet been undertaken on Cuba or Hispaniola, with the result that these islands harbour reservoirs of *C. hominivorax*; this could potentially lead to the reintroduction of the fly to regions across the Caribbean and Central America that are currently screwworm-free (Klassen & Curtis, 2005).

**Acknowledgements**

The authors thank Dr Martin Hall (Natural History Museum, London) for making available historical *Cochliomyia hominivorax* specimens from Texas, U.S.A., Dr Lucie Evans (University of Exeter) for technical assistance, Dr Kevin Moulton (University of Tennessee, Knoxville, TN) for providing information on unpublished EF-1α primer sequences from *Musca domestica*, and Dr Alan Robinson (International Atomic Energy Agency [IAEA]) for help and support throughout. We are grateful to the IAEA (co-ordinated research project ‘Enabling Technologies
for the Expansion of SIT for Old and New World Screwworm Fly’ and the Natural Environment Research Council, U.K. (grant no. NER/M/S/2003/00081) for financial support.

Conflicts of interest

All authors declare no conflicts of interests.

References


Accepted 15 October 2008
Development of an antigen-based rapid diagnostic test for the identification of blowfly (Calliphoridae) species of forensic significance

Laura McDonagh, Chris Thornton, James F. Wallman, Jamie R. Stevens

ABSTRACT

In this study we examine the limitations of currently used sequence-based approaches to blowfly (Calliphoridae) identification and evaluate the utility of an immunological approach to discriminate between blowfly species of forensic importance. By investigating antigenic similarity and dissimilarity between the first instar larval stages of four forensically important blowfly species, we have been able to identify immunoreactive proteins of potential use in the development of species-specific immunodiagnostic tests. Here we outline our protein-based approach to species determination, and describe how it may be adapted to develop rapid diagnostic assays for the ‘on-site’ identification of blowfly species.

1. Introduction

Modern forensic science has seen the importance of entomological evidence becoming well established worldwide, particularly when facing crime scenes involving a suspected homicide or manslaughter. One of the defining characteristics of the dipteran group Calliphoridae (blowflies) is the necessity to lay eggs on a proteinaceous substrate, usually the living or necrotic tissue of an animal host; larvae then develop by feeding on this protein-rich matter. Within forensic entomology, blowflies are recognised as one of the most important and robust indicator species, as they are usually among the first insects to colonise a body after death, often within hours [1]. Typically, the immature stages of blowflies, particularly larvae, are collected at a suspected crime scene and are used to establish the minimum post-mortem interval (PMI), using larval age as a form of ‘biological clock’. However, the immature stages, particularly early stage larvae, of several forensically important blowfly genera, e.g. Lucilia and Calliphora, are notoriously difficult to identify on the basis of morphology; the lack of species-specific larval morphological characters can make taxonomic keys very difficult to use, particularly for non-specialist forensic entomologists [2]. Accurate species determination is essential to PMI calculations, as growth and development rates can be highly species-specific; consequently, forensic entomologists are increasingly favouring molecular methods of identification [2–10].

At present, DNA sequence-based analysis, as first suggested by [11], is the only routinely used molecular-based species identification tool [12]. Several loci have been advocated for species identification through phylogenetic inference, whereby the position of unknown specimens within a phylogenetic tree reveals the species taxonomy [7,9,10]. The most commonly used markers include: regions of the mitochondrial cytochrome oxidase subunits I (COI) and II (COII) [2,6,10,13–17] and the nuclear internal transcribed spacers (ITS) [2,18]. Mitochondrial DNA (mtDNA), in particular, is favoured due to its high copy number, ease of isolation, and conserved sequence across taxa with supposed high mutation regions making discrimination between species, and even sub-species, possible [18,19].

Current practice typically employs single gene phylogenies for the identification of forensically important blowflies. However, limitations in the reliability of single locus phylogenies have been reported [20–22]. Such phylogenies only infer evolutionary relationships for the particular gene used, and as such may not represent the true species phylogeny, for example, due to horizontal gene transfer or incomplete lineage sorting at the locus in question. Despite mitochondrial DNA being expected to reach reciprocal monophyly before nuclear genes [19], phylogenetic inference from mtDNA genes has been widely reported to show paraphyly within Calliphoridae, as well as other organisms [20,22]. In particular, blowfly phylogenies based on mitochondrial and nuclear genes have exposed conflicting evolutionary relationships [12,22], with, for example, a COI genealogy rejecting reciprocal monophyly for Lucilia cuprina when compared with a 28S rRNA phylogeny [12,18,22]. Such findings imply incomplete lineage sorting for some recently diverged forensically important
blowflies, and so may prove problematic when relying on phylogenetic methods of diagnosis which themselves assume reciprocal monophyly. See Wells and Stevens [12] for an up to date review of this subject.

While some loci are more routinely used than others, no single gene is approved for use in forensic entomology and, in view of the problems outlined, it appears unlikely that any single gene will be capable of unambiguously identifying all calliphorid species of forensic importance. Consequently, due to the legal implications of forensic evidence, recognition is needed that if the use of molecular methods of species identification is to be continued, a switch to multi-gene approaches, for example using ESTs to screen multiple independent sites across blowfly genomes, is urgently needed. To date, few studies have used multi-gene phylogenies to determine the identity of forensically important flies [22].

The accuracy of phylogenetic inference, particularly when used for species determination of unidentified specimens, is also highly dependent on the experience of the analyst [12]. An important factor commonly misunderstood is the influence of taxonomic sampling, as both the number and choice of taxa represented can have a significant impact on the phylogenetic relationships inferred. In cases where many closely related species are compared, insufficient sampling may lead to intra- and interspecific ranges of genetic variation overlapping, so affecting the ability to determine the identity of unknown samples [18,23]. A second and potentially immense source of error is through the use of uncritical BLAST searching of databases that rely heavily on individual investigators being responsible for the accuracy and taxonomy of submitted sequence data. At present, the importance of such sources of error remain largely unknown, however, as several recently documented cases illustrate, even some commonly cited forensically important sequence records appear to have been mislabelled, leaving the reliability of reference sequence data in dispute [12]. Consequently, though it is generally accepted that such DNA-based identification of forensically important blowflies is expected to increase in popularity in the near future, emphasis is now also being placed on the need to investigate alternative approaches [12]. In this paper we present our preliminary findings on the development of such an alternative: an antigen-based rapid diagnostic test for the identification of blowfly species of forensic significance.

Due to the presence of several different styles of parasitism within Calliphoridae (saprophage, obligate and facultative), blowfly speciation is likely to have been accompanied by the within Calliphoridae (saprophage, obligate and facultative), blowfly species of forensic significance. An antigen-based rapid diagnostic test for the identification of such DNA-based identification of forensically important blowflies is expected to increase in popularity in the near future, emphasis is now also being placed on the need to investigate alternative approaches [12]. In this paper we present our preliminary findings on the development of such an alternative: an antigen-based rapid diagnostic test for the identification of blowfly species of forensic significance.

2. Results and discussion

2.1. Western blotting and protein characterisation

First instar larval proteins were resolved by SDS-PAGE and then transferred to PVDF membranes. Membranes were then probed with rabbit antiserum raised against L. sericata larval extracts, which reacted with a series of bands for each of the four blowfly
species tested, Fig. 1. While many of the protein bands were common among all four taxa, a number also appeared to be unique to the each of four taxa.

While initially investigating antigenic similarity and dissimilarity between first instar larval stages of four forensically important blowfly species, our findings have now revealed the potential species discriminatory power of such a technique. Several species-specific diagnostic proteins markers showing potential for use in species determination have been located, and we are currently in the process of characterising these proteins. For example, a particularly striking protein band (Fig. 2; band II, 35.8 kDa) appearing to be specific to L. sericata has been located and is currently undergoing characterisation.

A systematic proteomic approach is being used to purify and identify diagnostic immunodominant bands. N-terminal peptide sequencing and MALDI-MS of diagnostic protein markers will be used to perform searches for statistically similar protein sequences in core databases.

Two-dimensional SDS-PAGE gel electrophoresis, which achieves resolution in two dimensions separating proteins first by pH and then by molecular mass, is also planned for each of the four taxa. The major advantage of this technique is the ability to distinguish between different isoforms of a protein with similar molecular weights, for example a protein that has been phosphorylated (by addition of a negatively charged group), so helping to identify further potential diagnostic markers.

3.2. Species-specific monoclonal antibodies

The Western blotting experiments carried out in this study utilised rabbit PAb. Our efforts are now focusing towards the production of species-specific murine monoclonal antibodies (MAbs) using the diagnostic proteins identified with the rabbit PAb as immunogens. Raising species-specific MAbs will then allow the development of a rapid diagnostic tool exploiting lateral flow technology, an innovative application to the identification of forensically important blowflies.

3.3. Lateral flow technology

Lateral flow diagnostic tests are a format of semi-quantitative immunoassay for detecting the presence (or absence) of a target antigen (the identified species-specific antigenic protein) in a complex sample of proteins (e.g. whole blood, serum, urine). The most publicised example of the use of this technology is home pregnancy test first introduced by Unipath in 1988. Such technology is now used extensively for the rapid ‘on-site’ diagnosis of a wide range of human diseases including visceral leishmaniasis [26], HIV/AIDS [27], malaria [28], SARS [29] and invasive aspergillosis [30].

Lateral flow devices work as follows. A species-specific MAb is immobilised to a defined capture zone on a porous nitrocellulose membrane, while the same MAb conjugated to colloidal gold particles serves as the detection reagent. Samples of solubilised antigens are added to a release pad containing the antibody–gold conjugate. The antibody–gold conjugate binds to the target antigen, passes along the porous membrane by capillary action, and binds to the MAb immobilised in the capture zone. Once an antigen extract is prepared and applied to the LFD, the test result is recorded within 10–15 min. Bound antigen–antibody–gold complex is seen as a red line with an intensity that is proportional to the antigen concentration. Anti-mouse immunoglobulin immobilised to the membrane in a separate zone acts as an internal control. In the absence of the target antigen, no complex is formed in the zone containing the solid-phase antibody, and a single control line is seen. In the presence of the target antigen, two lines are clearly visible. One of the key advantages of this technology is the simplicity of the test, typically requiring little or no sample or reagent preparation.

While we believe existing single loci DNA-based methods of species determination to be problematic, we recognise that an extensive global study would be required to provide identification kits for all forensically important blowfly species. We anticipate that our research could develop ‘on-site’ rapid diagnostic kits for several of the more common blowflies. By reducing the work load of blowfly samples needing to be sent for sequencing analysis, such test kits could prove an innovative supporting role to existing DNA sequence-based methods, allowing forensic entomologists to focus on the more challenging cases of blowfly identification.

Acknowledgements

Thanks to Dr. Kate Heesom and Dr. Will Mawby (University of Bristol, UK) for protein characterisation. Thanks also to Professor Richard Wall (University of Bristol, UK) for larval preparations of L. caesar. This work was funded by the University of Exeter. This work was presented as an invited keynote paper in the Forensic Entomology section of the ‘DNA in Forensics’ meeting held in Ancona, Italy, 27th–30th May, 2008.
References