

2012

Beyond barcoding: A mitochondrial genomics approach to molecular phylogenetics and diagnostics of blowflies (Diptera: Calliphoridae)

Leigh Nelson

University of Wollongong, lan51@uow.edu.au

Christine L. Lambkin

Queensland Museum

Philip Batterham

University Of Melbourne

James F. Wallman

University of Wollongong, jwallman@uow.edu.au

Mark P. Dowton

University of Wollongong, mdowton@uow.edu.au

See next page for additional authors

Follow this and additional works at: <https://ro.uow.edu.au/scipapers>



Part of the [Life Sciences Commons](#), [Physical Sciences and Mathematics Commons](#), and the [Social and Behavioral Sciences Commons](#)

Recommended Citation

Nelson, Leigh; Lambkin, Christine L.; Batterham, Philip; Wallman, James F.; Dowton, Mark P.; Whiting, Michael F.; Yeates, David K.; and Cameron, Stephen L.: Beyond barcoding: A mitochondrial genomics approach to molecular phylogenetics and diagnostics of blowflies (Diptera: Calliphoridae) 2012, 131-142. <https://ro.uow.edu.au/scipapers/4707>

Beyond barcoding: A mitochondrial genomics approach to molecular phylogenetics and diagnostics of blowflies (Diptera: Calliphoridae)

Abstract

Members of the Calliphoridae (blowflies) are significant for medical and veterinary management, due to the ability of some species to consume living flesh as larvae, and for forensic investigations due to the ability of others to develop in corpses. Due to the difficulty of accurately identifying larval blowflies to species there is a need for DNA-based diagnostics for this family, however the widely used DNA-barcoding marker, *cox1*, has been shown to fail for several groups within this family. Additionally, many phylogenetic relationships within the Calliphoridae are still unresolved, particularly deeper level relationships. Sequencing whole mt genomes has been demonstrated both as an effective method for identifying the most informative diagnostic markers and for resolving phylogenetic relationships. Twenty-seven complete, or nearly so, mt genomes were sequenced representing 13 species, seven genera and four calliphorid subfamilies and a member of the related family Tachinidae. PCR and sequencing primers developed for sequencing one calliphorid species could be reused to sequence related species within the same superfamily with success rates ranging from 61% to 100%, demonstrating the speed and efficiency with which an mt genome dataset can be assembled. Comparison of molecular divergences for each of the 13 protein-coding genes and 2 ribosomal RNA genes, at a range of taxonomic scales identified novel targets for developing as diagnostic markers which were 117-200% more variable than the markers which have been used previously in calliphorids. Phylogenetic analysis of whole mt genome sequences resulted in much stronger support for family and subfamily-level relationships. The Calliphoridae are polyphyletic, with the Polleninae more closely related to the Tachinidae, and the Sarcophagidae are the sister group of the remaining calliphorids. Within the Calliphoridae, there was strong support for the monophyly of the Chrysomyinae and Luciliinae and for the sister-grouping of Luciliinae with Calliphorinae. Relationships within *Chrysomya* were not well resolved. Whole mt genome data, supported the previously demonstrated paraphyly of *Lucilia cuprina* with respect to *L. sericata* and allowed us to conclude that it is due to hybrid introgression prior to the last common ancestor of modern *sericata* populations, rather than due to recent hybridisation, nuclear pseudogenes or incomplete lineage sorting.

Keywords

diptera, calliphoridae, mitochondrial, genomics, approach, molecular, phylogenetics, diagnostics, beyond, blowflies, barcoding, CMMB

Disciplines

Life Sciences | Physical Sciences and Mathematics | Social and Behavioral Sciences

Publication Details

Nelson, L., Lambkin, C. L., Batterham, P., Wallman, J. F., Dowton, M. P., Whiting, M. F., Yeates, D. K. & Cameron, S. L. (2012). Beyond barcoding: A mitochondrial genomics approach to molecular phylogenetics and diagnostics of blowflies (Diptera: Calliphoridae). *Gene*, 511 (2), 131-142.

Authors

Leigh Nelson, Christine L. Lambkin, Philip Batterham, James F. Wallman, Mark P. Dowton, Michael F. Whiting, David K. Yeates, and Stephen L. Cameron

Beyond barcoding: A mitochondrial genomics approach to molecular phylogenetics and diagnostics of blowflies (Diptera: Calliphoridae)

Submitted to *Gene* for publication as a research article

Leigh A. Nelson¹, Christine L. Lambkin², Philip Batterham³, James F. Wallman⁴, Mark Dowton⁵, Michael F. Whiting⁶, David K. Yeates¹ & Stephen L. Cameron^{7*}

¹ Australian National Insect Collection, CSIRO Ecosystem Sciences, Canberra, ACT, 2601, Australia.

² Biodiversity Program, Queensland Museum, South Brisbane, QLD, 4101, Australia.

³ Department of Genetics, and Bio21 Institute, The University of Melbourne, Parkville, 3010, Australia.

⁴ Institute for Conservation Biology & Environmental Management, School of Biological Sciences, University of Wollongong, Wollongong, NSW, 2522, Australia.

⁵ Centre for Biomedical Sciences, School of Biological Sciences, University of Wollongong, Wollongong, NSW, 2522, Australia.

⁶ Department of Biology, Brigham Young University, Provo, UT, 84602, United States of America.

⁷ Earth, Environment & Biological Sciences School, Science & Engineering Faculty, Queensland University of Technology, GPO Box 2434, Brisbane, QLD 4001, AUSTRALIA.

* *Corresponding author:* Stephen Cameron (sl.cameron@qut.edu.au); phone: +61 7 3138 2869; fax: +61 7 3138 2330.

Abbreviations: PCG: protein coding genes; *atp6*, *atp8*: ATP synthase subunits 6 and 8 genes; *cob*: cytochrome oxidase *b* gene; *cox1-cox3*: cytochrome oxidase *c* subunit 1-3 genes; *nad1-6*, *nad4L*: NADH dehydrogenase subunits 1-6 and 4L; rRNA: ribosomal RNA; *rns*, *rnl*: mitochondrial small and large ribosomal RNA subunits; tRNA: transfer RNA; *trnX*: transfer RNA genes with isotype given in standard code abbreviation; EF-1 α : elongation factor 1 alpha; 28S: nuclear large ribosomal RNA subunit; ITS:

internal transcribed space of the nuclear rRNA genes; mt: mitochondrial; bp: base pair; PCR: polymerase chain reaction; RFLP-PCR: restriction fragment length polymorphism – PCR; RAPD: random amplification of polymorphic DNA; SNPs: single nucleotide polymorphisms; sec: seconds; min: minutes; ML: maximum likelihood; BA: Bayesian analyses; NGS: next-generation sequencing; N_E : effective population size.

Abstract

Members of the Calliphoridae (blowflies) are significant for medical and veterinary management, due to the ability of some species to consume living flesh as larvae, and for forensic investigations due to the ability of others to develop in corpses. Due to the difficulty of accurately identifying larval blowflies to species there is a need for DNA-based diagnostics for this family, however the widely used DNA-barcoding marker, *cox1*, has been shown to fail for several groups within this family. Additionally, many phylogenetic relationships within the Calliphoridae are still unresolved, particularly deeper level relationships. Sequencing whole mt genomes has been demonstrated both as an effective method for identifying the most informative diagnostic markers and for resolving phylogenetic relationships.

Twenty-seven complete, or nearly so, mt genomes were sequenced representing 13 species, seven genera and four calliphorid subfamilies and a member of the related family Tachinidae. PCR and sequencing primers developed for sequencing one calliphorid species could be reused to sequence related species within the same superfamily with success rates ranging from 61-100%, demonstrating the speed and efficiency with which an mt genome dataset can be assembled. Comparison of molecular divergences for each of the 13 protein-coding genes and 2 ribosomal RNA genes, at a range of taxonomic scales identified novel targets for developing as diagnostic markers which were 117-200% more variable than the markers which have been used previously in calliphorids. Phylogenetic analysis of whole mt genome sequences resulted in much stronger support for family and subfamily-level relationships. The Calliphoridae are polyphyletic, with the Polleninae more closely related to the Tachinidae, and the Sarcophagidae are the sister group of the remaining calliphorids. Within the Calliphoridae, there was strong support for the monophyly of the Chrysomyinae and Luciliinae and for the sister-grouping of Luciliinae with Calliphorinae. Relationships within *Chrysomya* were not well resolved. Whole mt genome data, supported the previously demonstrated paraphyly of *Lucilia cuprina* with respect to *L. sericata* and allowed us to conclude that it is due to hybrid introgression prior to the last common ancestor of modern *sericata* populations, rather than due to recent hybridisation, nuclear pseudogenes or incomplete lineage sorting.

Keywords: mtDNA, DNA diagnostics; forensic entomology; Calliphoridae; phylogenetics.

Introduction

Globally, blowflies of the family Calliphoridae (Diptera: Oestroidea) have great veterinary, medical and forensic importance due to life cycles that involve development in carcasses or in live animals (myiasis) (Stevens & Wallman, 2006). While calliphorids share the myiasis-causing habit along with close relatives the flesh flies (Sarcophagidae) and botflies (Oestridae), they exhibit the widest life history range of any of the oestroid families, including not only saprophages, but also coprophages, sanguivores, ectoparasites and parasitoids (Stevens, 2003). The most economically significant blowflies are those that are ectoparasites of livestock; different species are either primary (able to pierce the skin and initiate infestations) or secondary (species that colonise infested animals) strike flies and either obligate or facultative parasites. For example, sheep blowflies or green bottles (*Lucilia* spp.) are the most important parasites affecting sheep. *Lucilia* are primary strike flies capable of initiating fly strikes on moist skin but most effectively in the presence of a moist protein source such as faeces, urine-soaked wool, fleece rot or skin abrasions. *Lucilia cuprina* is responsible for 90% of strikes in Australia and costs the industry AU\$280 million in annual stock losses, degraded wool quality, management labour and pesticide applications (Knights et al. 2008). *Lucilia* species may also be carrion-breeders, along with other calliphorid genera; such flies are therefore forensically important for the estimation of the postmortem interval (PMI) (Amendt et al., 2004; Catts & Goff, 1992; Greenberg 1991). Members of the genus *Chrysomya* in particular represent a large proportion of the insects recovered from human corpses analysed in cases of suspicious death or murder (Levot, 2003).

Identification of calliphorids in either veterinary or forensic settings is important, however morphological identifications are often hampered by the need to rear specimens to adulthood before they can be keyed (e.g. Wallman, 2001). Calliphorid faunas are composed of large numbers of species, often with highly localized populations, exhibiting differing degrees of dispersal abilities and host use, and so the need for diagnostic tools which complement morphologically based identifications is acute. Additionally, the ability to discriminate between fly populations within a species is valuable to forensic entomologists in determining if bodies have been moved or interfered with and to veterinary management programs in

preventing the spread of pesticide-resistance genes into susceptible populations (Wells & Stevens, 2008). Genetic markers, including sequence profiles, restriction site maps and single-nucleotide polymorphisms (SNPs), are therefore the logical choice in attempting to design identification tools that are flexible enough to be useful at multiple levels (Wells & Stevens, 2008). Accordingly, molecular diagnostic protocols for calliphorids have been developed based on both RFLP-PCR (Sperling et al., 1994; Malgorn & Coquoz 1999; Ratcliffe et al., 2003; Nelson et al., 2008) and DNA sequences (Wells & Stevens, 2008). Of those genes that have been applied across a range of species and so can be compared for their utility as species and population level markers, the majority are mitochondrial in origin (Table 1). The cytochrome oxidase *c* subunits I and II (*cox1*, *cox2*) are the most studied, followed by nicotinamide dehydrogenase subunits 4 and 4L (*nadh4*, *nadh4L*) and large ribosomal subunit (*lrRNA*); only recently have nuclear protein coding genes been sequenced (CAD by Kutty et al., 2010; EF-1 α by McDonagh & Stevens, 2011).

For the last decade, molecular diagnostics research in general has been dominated by the DNA barcoding paradigm, the proposal that, for metazoans, the 600 bp 5' end of the *cox1* gene is able to accurately identify known species and can "flag" undescribed species that are significantly molecularly diverged from known species (Hebert et al., 2003). While only one study has explicitly tested DNA-barcoding in calliphorids (Nelson et al., 2007), many studies have used either the 5' end or the entire *cox1* gene in combination with other genes and so are useful for assessing barcode efficacy (see Table 1). Collectively, these studies have shown that *cox1* barcodes can identify many calliphorid species, however, they have also demonstrated instances where barcoding fails due to insufficient variability between species e.g. Australian *Calliphora* spp. (Wallman & Donnellan, 2001), species paraphyly e.g. *Chrysomya chloropyga* and *Ch. putoria* (Wells et al., 2004), *Wolbachia* mediated introgression e.g. *Protocalliphora* (Whitworth et al., 2007) and possible hybrid introgression e.g. *Lucilia cuprina* and *L. sericata* (Wells et al., 2002; Stevens et al., 2002; Wells et al., 2007; Toule et al. 2009; McDonagh & Stevens, 2011). These studies have also demonstrated very low rates of intraspecific variability, suggesting that a *cox1*-barcode approach may have limited utility in determining the population or geographic origin of calliphorid

specimens (e.g. Nelson et al., 2007). Additionally, relationships at taxonomic ranks higher than the species are often not well supported, which leads to problems in the interpretation of potentially new, cryptic species (c.f. Lessard et al., 2009). The frailty of phylogenetic hypotheses based on *cox1* alone is demonstrated by recent multi-gene analyses of calliphorid evolution where nodal support is much lower for backbone nodes than for relationships below the subfamily level (Kutty et al., 2010). The development of additional genetic markers for use in calliphoid diagnostics which can be applied to specific veterinary or forensic questions, or to improve our understanding of the evolution of the Calliphoridae is thus needed, rather than the sequencing of ever more specimens for the same standard genes.

One approach to improving the range of mitochondrial genes that can be applied to calliphorids is to sequence entire mitochondrial (mt) genomes from additional species. Comparisons of variation within each gene region will enable workers to more rapidly target potentially informative genes, and the availability of genomic sequences will aid in the design of primers for wide scale sequencing programs. This approach has previously been applied on a small scale to determine markers for interspecific variability in termites (Cameron & Whiting, 2007) and to determine how representative the barcode region is of the entire genome in eutherian mammals (Luo et al. 2011). At present whole mt genomes of four calliphorid species are available, representing two subfamilies (Lessinger et al., 2000; Junquera et al., 2004; Stevens et al. 2008); the present study adds 27 additional mt genomes representing four calliphorid subfamilies, six genera, 12 species, multiple populations from four calliphorid species, plus a representative of a closely related family, the Tachinidae. In analyses of these new mt genomes we sought to address three questions:

1. Can mt genomes be economically sequenced for use in the design of molecular diagnostics studies?
2. What is the pattern of variability for each gene at the population, species, genus and subfamily levels and how does this compare to the arbitrarily-chosen mitochondrial genes used in previous calliphorid systematics and diagnostics?

3. Does a whole mt genome phylogeny of the Calliphoridae help to resolve relationships within this important family?

Materials and Methods

Specimen Collection and DNA Extraction

Flies were sourced from the field and established laboratory cultures (Table 2). Field caught flies were identified by CLL & JFW using the Australian National Insect Collection (CSIRO Ecosystem Sciences) and the Diptera collection of the School of Biological Sciences, University of Wollongong as references. Whole-genomic DNA extraction was undertaken using the DNeasy Tissue Kit (QIAGEN, Valencia, CA). Whole specimens were used for DNA extractions with a ventral mid-line incision made to allow buffers to digest thoracic muscle tissue while leaving the exoskeleton intact for vouchering. For one very large species, *Rutilia georlingiana*, thoracic flight muscle was removed after a lateral incision with fine forceps before DNA extraction using the same kit. For *L. cuprina* samples DI190 and DI213, five individual flies were sequenced from each collection locality (codes DI190.1-5 and DI213.1-5) to assess variability within lab and field populations, respectively.

PCR Amplification and Sequencing

Initially, the mt genome of a single *L. cuprina* specimen (sample #DI190.1) was amplified in six overlapping fragments via long PCRs using a combination of published near-universal insect mt primers (ex Simon et al., 1994; Skerratt et al., 2002; Bybee et al., 2004; Cameron et al., 2007) plus novel primers designed for this specimen (supplementary table S1). Each amplicon was sequenced via primer walking with novel primers designed from the sequence data for DI190.1. Following completion of the DI190.1 mt genome, we attempted to use the same combination of PCR and sequencing primers to amplify and sequence all remaining calliphorid species. Where sequence variation between species resulted in either

PCR or sequencing failure, primers were redesigned for use in the species in which the failure occurred (see Table 3 for success/failure results). A full list of the primer combinations used for long PCRs for each sample, replacement primers used in specific specimens and the sequences of all primers used in this study are provided in supplementary table S1.

Long PCRs were performed, using Elongase (Invitrogen, Carlsbad, CA) with the following cycling conditions: 92°C for 2 min; 40 cycles of 92°C for 30 s, 50°C for 30 s, and 60°C for 12 min; and a final extension step of 60°C for 20 min. Sequencing was performed using ABI BigDye ver3 dye terminator sequencing technology and run on ABI 3770 or ABI 3740 capillary sequencer. Sequencing PCR conditions were 28 cycles of 94°C/10sec, 50°C/5 sec, 60°C/4min. Heteroplasmic regions in several genomes were resolved by cloning using the Topo-TA cloning chemistry (Invitrogen).

Sequence Annotation and Alignment

Editing and assembly of chromatograms was performed using Sequencher ver. 4 and 5 (Gene Codes Corp., Ann Arbor, Michigan, USA). Transfer RNA inference and secondary structure prediction was conducted with tRNAscan-SE, using invertebrate mt codon predictors (Lowe & Eddy, 1997). Reading frames between tRNAs were found in Sequencher and protein-coding genes (PCGs) identified using translated BLAST searches (blastx) (Altschul et al. 1997), as implemented by the NCBI website (<http://www.ncbi.nlm.nih.gov/>). Annotations of the ribosomal RNA genes were done by eye with reference to previously published insect mt rRNA gene secondary structures (c.f. Cameron & Whiting 2008). Additionally, the alignment process (see below) resulted in revised annotations of the *Hypoderma* and *Exorista* mt genomes. The *nad3* and *nad6* genes in *Hypoderma* are each longer than the consensus alignment of the genes for calyptrates by 3 bp (or 1 amino acid residue) at the 5' end. In *Exorista*, *trnSI* as annotated on Genbank is missing 1 bp from both the 5' and 3' ends.

To the 27 mt genomes newly sequenced here, sequences from GenBank of the full/nearly complete mt genomes of four additional calliphorids, and five outgroup species from the families Muscidae, Oestridae, Tachinidae and Sarcophagidae were added for analysis (genomes originally published in Lessinger et al., 2000; Junqueira et al., 2004; Stevens et al., 2008; Oliviera et al., 2008; Weigl et al., 2010; Nelson et al., 2012). This taxon selection includes all the available mt genomes of the superfamily Oestroidea (which contains the calliphorids) and the muscoid grade (sister-group/paraphyletic clade containing Oestroidea) (Kutty et al., 2010). Each of the 37 mt genes were aligned separately using MUSCLE (Edgar, 2004), as implemented in MEGA version 5.03 (Tamura et al., 2011). For PCGs, nucleotide sequences were aligned by reference to translated amino acid codons and back-translated for analysis as DNA. Individually aligned gene datasets were concatenated in MacClade (Maddison & Maddison 2003) for phylogenetic analysis.

Divergence and Phylogenetic analysis

Genetic variability for each protein coding and rRNA gene was assessed by pair-wise sequence divergences under the Kimura 2-parameter (K2P) substitution model implemented in MEGA. K2P was chosen in order to allow ready comparison with previous DNA barcoding analyses which almost universally use this model despite criticisms of its biological reality (Collins et al. 2012). In addition regions of the mt genome used previously in calliphorid diagnostics and systematics were analysed for their variability across the sequenced mt genomes and compared to the 15 mt genes. The regions analysed were *cox1-5'* (= *cox1*-barcode region), *cox1-3'*, and *nad4-nad4l* (see Table 1).

Phylogenetic trees were constructed using maximum likelihood (ML) and Bayesian (BA) analyses. Replicate analyses were performed for each of two datasets, one including all codon positions for the PCGs (ALL dataset) and one which removed third codon positions (NO-3RDS) as these have been shown to be particularly susceptible to rate heterogeneity (Dowton et al., 2009b; Song et al., 2010). In both ML and BA analyses, datasets were partitioned by gene with tRNAs concatenated into a single partition (16

partitions in total). Our previous analyses of dipteran mitochondrial phylogenomics (Cameron et al. 2007) have shown that partitioning by gene gives similar results to more complicated partitioning schemes such as by codon for each gene. The muscid outgroup *Haematobia irritans* was used as the root in all analyses (following Kutty et al., 2010; Wiegmann et al., 2011).

Analyses were performed with the RaxML Black-Box webserver (<http://phylobench.vital-it.ch/raxml-bb/index.php>; Stamatakis et al., 2008) for likelihood and MrBayes ver 3.2 (Ronquist et al., 2012) for Bayesian analysis. Datasets for RAxML were partitioned and analysed with a Gamma model of rate heterogeneity. All Bayesian analyses were run with default priors and unlinked partitions, with appropriate models of molecular evolution selected for each partition using AIC as implemented in ModelTest (Posada & Crandall 1998). Each dataset was analysed using two independent runs, each of four chains (three hot and one cold chain), for three million generations with sampling every 1000 generations. Parameter convergence was achieved by all analyses within three million generations as determined using Tracer ver. 1.4 (Rambaut & Drummond 2009) and the standard deviation of split frequencies for the independent runs was below 0.01. Completed Bayesian analyses were examined for asymptotic behaviour of each parameter and of total tree likelihood; trees collected prior to this asymptotic point were treated as burn-in and discarded (generally the first 30-60,000 generations). Partition model details and Bayesian run files are available for each analysis from SLC upon request.

Results & Discussion

Genome Organization and Structure

A total of 27 mt genomes, representing 13 species, were sequenced in this study and have been lodged in GenBank (accession numbers XXXXXX-XXXXXX). These 27 mt genomes include representatives of 11 species, three genera, two calliphorid and one tachinid subfamilies previously unsequenced. Four of the sequenced species, *Ch. megagephala*, *Ch. rufifacies*, *L. cuprina* and *L. sericata*, include multiple

specimens, allowing genomic variability within these species to be assessed (see below). Complete genomes were amplified for all specimens, with the exception of *Pollenia rudis*. For this species the control region could not be amplified, resulting in the failure to sequence *trnI* and *trnQ*, and only partial sequences were obtained for *trnM* and *rrnS*. For several additional specimens, only partial control region sequences were obtained due to difficulties associated with sequencing this region (Table S2). It is estimated that 200-800 bp is missing from each of these genomes between the poly-T and poly-A stretches at the *rrnS* and *trnI* ends of the control region, respectively (Lessinger et al., 2004). In common with other studies analysing whole mt genomes, attempts to align equivalent portions of the control region of all specimens are difficult due to high levels of variation (Zhang & Hewitt, 1997). Because of the high rates of evolution experienced by this region and the difficulties encountered, control region sequences were not included in further analyses. The full length genome sequences used in analyses ranged from 14,340 – 14,830 bp in length.

All of the species sequenced in this study possessed the 37 genes found in a typical Metazoan mt genome (13 PCGs, 2 rRNA genes, 22 tRNA genes), and in the same insect ancestral arrangement found in most other dipterans (Beckenbach, 2011). In addition, all *Chrysomya* species sequenced in this study also possessed duplicate *trnI* and partial *trnQ* genes in the control region close to the *rrnS* gene, confirming and extending the original finding of this duplication by Lessinger et al. (2004). This duplication has been identified in all six *Chrysomya* species examined to date and is likely a synapomorphic feature of the genus. In each species both copies of *trnI* had the identical sequences suggesting that it is maintained by concerted evolution as there are sequence differences between *trnI* genes from different *Chrysomya* species. This tRNA duplication does not occur in the other Chrysomyinae genera examined, *Cochliomyia* or *Protophormia*. tRNAScan-SE did detect a second *trnI* copy in *Protophormia* within the control region, however this second copy had very low sequence identity to the copy in the ancestral position, including a different anticodon sequence, unlike the sequence identical copies found in *Chrysomya*. This suggests that the duplicated *trnI* found in *Protophormia* is not homologous to the duplicated *trnI* copies found in *Chrysomya*. Previous analyses have uncovered convergence between tRNA gene

rearrangements, undermining their value as putative molecular synapomorphies (e.g. Dowton et al., 2009a; Cameron et al., 2011; Kilpert et al., 2012). However, the consistent finding of the duplicated tRNAs across the diversity of *Chrysomya* species (see below), suggests that convergence is likely not at issue here and that this duplication may be taxonomically diagnostic for members of this genus. Adult *Chrysomya* are relatively easy to diagnose with morphology however larvae are more challenging as all larval instars are undescribed for many species. This marker could thus also have practical value for diagnosing the presence of *Chrysomya* spp. in mixed larval collections.

With the exception of *nad1* and *cox1*, the PCGs of all species had standard ATN start codons (Table S2). For *nad1*, the non-canonical codon TTA (encoding Leu) was observed for four species (*Ch. rufifacies*, *Ch. albiceps*, *P. rudis* and *H. ligurriens*). For all specimens, the *cox1* start codon was identified as TCG (encoding Ser), which has been widely reported in other Diptera (c.f. Cameron et al., 2007). The four genes in which the 3' end abuts another PCG (*atp8*, *atp6*, *atp8* and *nad4L*) all have complete stop codons in all species, as do 5 of the 9 tRNA flanked genes. Incomplete stop codons (T, poly-adenylated to TAA post-transcriptionally) were found for *cox1*, *cox2*, *nad5* and *nad4* in all calliphorid species. A summary of the mt genes of each species sequenced is given in Table S2. Nucleotide base composition was similar for the mt sequences of all species with a strong AT bias (mean = 76.85%), typical of dipteran mitochondrial DNA and in the middle of the range reported for holometabolous insects (Negrisolo et al., 2011).

Mt Genome Resequencing

The first calliphorid mt genome sequenced in this study, *L. cuprina* specimen DI190.1, was amplified and sequenced using seven insect near-universal primers plus 42 primers designed through primer walking of this specimen (see Table S1). The utility of “recycling” these 49 primers to sequence other calliphorid flies ranged from 100% for conspecifics and the sister species *L. sericata*, to 69% for the most distantly related calliphorid, *Pollenia*. In general, the success rate of primer reuse was proportional to the

molecular distance of the target species from *L. cuprina* (Figure 2). Because the control region of the *Pollenia* specimen failed to amplify, the unamplified portion would likely have required the redesign of additional primers due to the high variability in this region (see Table 3). The mt genome of *Rutilia*, a member of the Tachinidae which are closely related to the calliphorids, was sequenced by recycling 61% of the primers from *L. cuprina*. Additionally, for several species, redesigned primers were successful in sequencing other closely related species. For example, 41/49 (84%) of the *L. cuprina* primers were successful for *Ch. megacephala*, however the eight primers that were redesigned for this later species also worked for *Ch. saffrana* which was sequenced with only a single specific primer. Similarly, *Ch. rufifacies* required 13 redesigned primers (75% recycling), which were also successfully used on *Ch. albipes* (one additional primer) and *Protophormia* (no additional primers). In all instances where multiple specimens were sequenced for a given calliphorid species, the same combination of primers was successful for all specimens. The redesigned primers are located at different points in the mt genome; however a third (15/42) of those are located in the control region which makes up approximately 6.5% of the genome, thus highlighting the high sequence variability in this region. Depending on the use for which the mt genome sequence is being collected, e.g. diagnostic design or phylogenetics as applied below, sequencing the control region may not be necessary and the whole coding region can be sequenced with a higher primer recycling rate than is reported here.

These rates of primer recycling are comparable to those for resequencing whole mt genomes of *Reticulitermes* termites, which ranged from 100% for conspecifics to 87% for a closely related congeneric species (Cameron & Whiting, 2007). The present study attempts resequencing over much broader taxonomic ranges and it is therefore not surprising that the primer success rates drop for members of different subfamilies (or families in the case of *Rutilia*) from the species for which the initial genome was sequenced. Yamauchi et al. (2004; 2005) applied a somewhat different approach, designing degenerate primers for pancrustacean arthropods (79 novel primers in total) and then trialling a large number of overlapping, short PCRs to amplify and sequence two very distantly related insect species, the cockroach, *Periplaneta fuliginosa*, and the dragonfly, *Orthetrum triangulare*. For each species roughly half of the

short PCR primer pairs yielded products which could be sequenced (25 and 26 of 48 pairs respectively), however the region from *rrnS-nad2* failed to amplify for *Orthetrum*, such that approximately 850bp of the coding region was not sequenced (partial *nad2* and *rrnS* sequences, *trnI*, *trnQ* and *trnM* entirely missing) (Yamauchi et al., 2004). The present study achieved a higher success rate, a minimum of 61% primer recycling for *Rutilia* vs. 52% for *Periplaneta*, but was taxonomically much more tightly focused, with all but one species belonging to a single family. Attempting to sequence species from a different insect order to the one for which a set of primers was designed would likely be unsuccessful.

Molecular Divergence

Distances were calculated for individual protein-coding and rRNA genes in order to examine the pattern of molecular variability across the different genes within the mt genome and at different taxonomic levels, i.e. within the Calliphoridae¹, and within major subfamilies, genera and species (Tables 4 & 5). In addition, the variability of those regions of the mt genome used in previous diagnostic and phylogenetic studies of calliphorids was compared at the same taxonomic levels. For each of the taxonomic levels considered there was a high degree of correspondence between the three measures of variability calculated, percentage of informative sites, maximum percentage divergence and mean percentage divergence, i.e. genes with high variability in one measure typically had variability across all measures. The most variable genes differed depending on the taxonomic scale: *nad6* for within Calliphoridae, Chrysomyinae and *Chrysomya*, and *cob* for within *Lucilia* and the *L. sericata/cuprina* lineage. Variability at the fine scales was somewhat poorly predicted by variability at the next highest taxonomic level, e.g. within the *Ch. megacephalalsaffranae* clade *nad4* had the most variability and *rrnS* the highest percentage of informative sites, whereas *nad6* was most variable within *Chrysomya* as a whole. It was also difficult to extrapolate from the variability within one group to another. None of the genes identified as most variable for the *Ch. megacephalalsaffranae* clade by any of the three measures were also found to be most variable for the *L. cuprina/sericata* clade. Additionally, the maximum and mean % divergences differed greatly between the two species in each pair, with *Ch. megacephalalsaffranae* being less than

¹ Excluding *Pollenia* as it did not fall within a monophyletic Calliphoridae. See Phylogenetic Analysis section below.

half as variable as *L. cuprina/sericata* (1.515 and 0.806% vs. 3.920 and 2.103%, respectively), despite wide sampling for each species pair (Australia and Asia vs. Australia, North America and Europe). This pattern, whereby the most variable gene is difficult to predict based on comparisons of related taxonomic groups, has previously been noted for termites (Cameron & Whiting, 2007) and moths (Cameron & Whiting, 2008) and is likely a general pattern of mt genome evolution with implications for the design of studies aimed at developing or characterising molecular diagnostic markers.

Tables 4 and 5 facilitate the identification of the most variable genes in the calliphorid mt genome at different taxonomic levels (entries in bold). In general, the genes previously used for diagnostics and phylogenetic analyses of calliphorids were less variable than those chosen after comparison of all genes in the mt genome. The partial *cox2* region used by Wallman et al. (2005) had the highest percentage of informative sites for comparisons within *L. sericata*, however this was the only comparison and the only variability measure where an arbitrarily chosen gene region outperformed the gene chosen from the whole mt genome. For the majority of taxonomic comparisons, the 3' end of *cox1*, first used by Wallman & Donnellan (2001) but widely used in calliphorid diagnostics since then, was the most variable. *cox1*-3' was between 72-85% as variable as the most variable mt gene, whereas the other regions, *cox1*-5' (=CO1 barcode), *cox2*-3' and *nad4l-nad4* were approximately half as variable. This is consistent with previous analyses which have found that the arbitrarily chosen gene regions widely used in insect systematics are much less variable than those which would be chosen following examination of the entire mt genome (Cameron & Whiting, 2007; 2008) or even portions of the mt genome (Roe & Sperling, 2007).

Phylogenetic Analysis

A phylogenetic analysis of the 30 calliphorid mt genomes (plus six calyptrate outgroups) was conducted to examine evolution within this family. Replicate analyses were performed with maximum likelihood (ML) vs Bayesian analysis (BA), for each of two datasets, all positions included (ALL) vs third codon positions excluded (NO3RD). There was a high degree of topological congruence across the four

analyses with only the positions of *Protophormia terraenovae* and *Chrysomya saffrana* differing between analyses (see below). Nodal support was consistently higher in BA than ML analyses as has been noted before (Cummings et al., 2003) however it was also consistently higher for analyses which included third codon positions. Previous phylogenetic analyses of insect mt genomes have found that the inclusion of third codon positions lead to reduced nodal support in some groups e.g. polyneopterans (Cameron et al., 2006a), or resulted in artefactual relationships in others e.g. within beetles (Cameron et al., 2009) or termites (Cameron et al., 2012).

Family-level relationships did not support previously proposed classification schemes within the Oestroidea. In all analyses, the Calliphoridae are not monophyletic as *Pollenia* was consistently found to be the sister group of the tachinid *Rutilia*, although nodal support for this relationship was not significant for any but the BA-ALL analysis. The Tachinidae were also rendered paraphyletic by the inclusion of members of the Oestridae, although this is now a well accepted relationship (McAlpine, 1989). The sister group of Calliphoridae (excluding *Pollenia*) was consistently found to be the Sarcophagidae with nodal supports significant for three of the four analyses (0.878 posterior probability in the BA-NO3RD analysis). Similar relationships were found by Kutty et al. (2010), where pollenine and helicoboscine calliphorids and the Rhinophoridae formed a paraphyletic grade at the base of Tachinidae, while mesembrinelline calliphorids and the Oestridae were nested within Tachinidae. Our finding of Calliphoridae + Sarcophagidae to the exclusion of Tachinidae (including *Pollenia* and Oestridae) was not supported by Kutty et al. (2010) or Wiegmann et al. (2011), however nodal support was not significant for any interfamily relationships within the Oestroidea in either study. A complication in comparing between these studies is the number of additional oestroid families for which whole mt genome data have yet to be collected: the Rhinophoridae (included in both previous studies), Rhiniidae and Mystacinobiidae (included in Kutty et al., 2010). While the inclusion of members of these families could potentially break up long branches in the mt genome tree, it should be noted that the branch connecting Sarcophagidae to Calliphoridae (excluding *Pollenia*) is not the longest found in this analysis, and is less than half the length

of the branches found within the Oestridae (between *Dermatobia* and *Hypoderma*) or the Tachinidae *sensu stricto* (between *Exorista* and *Rutilia*).

Within the Calliphoridae (excluding *Pollenia*) there was strong support for the monophyly of the Luciliinae and Chrysomyinae and for the sister-group relationship between Calliphorinae and Luciliinae (both datasets and inference methods). This sister-grouping has been consistently found in molecular systematic studies of this group (Wells et al., 2002; Stevens, 2003; Wallman et al., 2005; Harvey et al., 2008; McDonagh & Stevens, 2011). Within the Chrysomyinae, the tribe Chrysomyini was polyphyletic in all analyses as *Cochliomyia* never grouped with *Chrysomya* to the exclusion of *Protophormia* (tribe Phormini). The paraphyly of Chrysomyini has been found repeatedly (Wells & Sperling, 2001; Stevens, 2003; Singh et al., 2011; McDonagh & Stevens, 2011). *Protophormia* was either the sister-group of a monophyletic *Chrysomya* (RAxML-NO3RD and BA-ALL analyses) or grouped within *Chrysomya* (RAxML-ALL and BA-NO3RD analyses). Most previous analyses have supported the monophyly of *Chrysomya* (Wells et al., 2001; McDonagh & Stevens, 2011) although Stevens (2003) is an exception. However, the presence of the duplicated gene block *trnI* – partial *trnQ* in all *Chrysomya* species examined to date (Lessinger et al., 2004; present study) and its absence from *Protophormia*, is strong evidence for the monophyly of *Chrysomya*. The failure of two of our analyses to recover *Chrysomya* monophyly suggests that the support from mt genome sequence data is, however, weak for this portion of the tree.

As in several previous analyses, the deepest split within *Chrysomya* is between a clade containing *ruffifacies* and one containing *megacephala* (Wells & Sperling, 2001; Harvey et al., 2003b; Stevens, 2003; Chen et al., 2004; Wallman et al., 2005; Nelson et al., 2008). Nelson et al. (2007) and Singh et al. (2011) identified a third major clade, consisting of *latifrons* + *semimetallica* (the subgenus *Eucompsomyia*); however, given the lack of complete mt genome data for either of these species, this phylogenetic hypothesis could not be tested here. In the remaining studies, relationships within *Chrysomya* are not robustly resolved, with the genus consisting of three or more clades without significant nodal support for relationships between these clades (Harvey et al. 2008; McDonagh &

Stevens, 2011). The remaining relationships between *Chrysomya* species are highly variable across the published studies. Some of the relationships which we find have been widely supported, e.g. *ruffifacies* + *albiceps* or *megacephala/saffranae* + *bezziana* were supported in all and all but one, respectively, of the studies which included these species (Wells & Sperling, 2001; Singh et al., 2011; McDonagh & Stevens, 2011; Harvey et al., 2008 does not support *meg/saf* + *bezz*). Conversely, the position of *putoria* is highly variable across the published studies and only Wells & Sperling (2001) have previously found, as we do here, that it is the sister-group of the clade *megacephala/saffranae* + *bezziana*. Finally, both Bayesian and likelihood analyses of the ALL dataset found that *megacephala* was paraphyletic with respect to *saffranae*, whereas both NO3RD analyses found a monophyletic *megacephala* with *saffranae* as its sister. Similar results have been reported before by Harvey et al. (2008), and while other studies have supported reciprocal monophyly, the molecular divergences between these two species have been vanishingly small, e.g. 0.403% (Wallman et al. 2005) and 0.48% (Nelson et al., 2007), the lowest interspecific differences recorded in this genus. Morphologically, the two species are very similar but diagnosable by specialists (Wallman, 2001). Further investigation is needed to determine if these low molecular divergences and possible paraphyly are the result of recent speciation, hybrid introgression or reflect a single species.

The Luciliinae were monophyletic, with high support for a monophyletic *Lucilia*. Some previous studies have failed to resolve relationships between *Hemipyrellia* and *Lucilia*, due to non-monophyly of *Lucilia* (Stevens, 2003; McDonagh & Stevens, 2011). In the latter study, relationships between these two genera varied depending between gene partitions, EF-1 α supported reciprocal monophyly, whereas in both 28S and *cox1*, *Hemipyrellia* rendered *Lucilia* paraphyletic. Furthermore, the species of *Lucilia* which were more closely related to *Hemipyrellia* than congeners differed between each partition and a combined analysis of all three was unresolved; Lucilinae consisted of a polytomy of 6 subclades. The present study is consistent with the McDonagh & Stevens (2011) *cox1* results, as is to be expected given our restriction to mt genes and that neither of the *Lucilia* species, *mexicana* and *cluvia*, which fall outside their main *Lucilia* clade is included here.

Lucilia cuprina formed two distinct clades with very short branch lengths that were collectively paraphyletic with respect to *L. sericata*. The culture of *cuprina* flies from the University of Melbourne (DI190.1-5) plus one of the wild *cuprina* (DI213.1) formed a clade sister to the remaining *cuprina* specimens (DI213.2-5) plus all *sericata* specimens. The non-monophyly of *cuprina* with respect to *sericata* is consistent with previous studies (Stevens & Wall, 1996; 1997; Stevens et al., 2002; Wells et al., 2002; Wallman et al., 2005; Wells et al., 2007; Harvey et al., 2008; Toule et al., 2009; McDonagh & Stevens, 2011). The present study strongly suggests that this paraphyly is reflective of the whole mt genome, and is not an artefact restricted to a single mt gene. Originally these paraphyletic *cuprina* flies were thought to be confined to Hawaii (Stevens et al., 1996; Stevens & Wall, 1997; Stevens et al., 2002; Wells et al., 2002) however they have subsequently been found in Australia (Wallman et al., 2005; present study), southeast Asia (Wells et al., 2007; Harvey et al. 2008), southern Africa (Toule et al., 2009) and North America (DeBry et al., 2010). Additionally, there has been parallel work which suggests that the paraphyly of *cuprina* with respect to *sericata* is limited to mt genes as both RAPD (Stevens & Wall, 1997) and nuclear sequence data (Stevens et al., 2002; Toule et al., 2009; McDonagh & Stevens, 2011) find reciprocal monophyly for each species.

These combined results have been interpreted in several different ways. Stevens & Wall (1996; 1997) and Wallman et al. (2005) hypothesised that paraphyly reflected subspecific differences within *cuprina* with specimens more closely related to *sericata* (hereafter ‘the *cuprina*-haplotypes’) representing the subspecies *cuprina cuprina*, whereas the remaining specimens represented *cuprina dorsalis* (Norris, 1990). Stevens & Wall (1996; 1997) however did not support separate subspecies as specimens morphologically consistent with *cuprina cuprina* were not monophyletic in either study. In the present study the two *cuprina* clades do not correspond to the two subspecies as they are all morphologically *cuprina dorsalis* and members of both clades were caught at the same location at the same time (DI213.1 vs. DI213.2-5).

A second possible explanation for the *cuprina*-haplotypes raised by Toule et al. (2009) is that they represented nuclear pseudogenes (numts). Numts can be co-amplified by the universal primers typically used for genes such as *cox1* (Song et al., 2008). Differences between the numt and real mt gene sequence can result in erroneous phylogenetic reconstructions (Rubinoff et al. 2006). In the present example a numt which incorporated into the nuclear genome in the common ancestor of *cuprina* and *sericata* could produce the pattern observed here. The absence of in frame stop codons in *cox1* was taken by Toule et al. (2009) as evidence that the *cuprina*-haplotype sequences were not numts, however this is a comparatively weak test for numts (Moulton et al., 2010). The whole mt genomes generated in this study are definitive proof that the *cuprina*-haplotypes are not numts. While there are examples of major portions of the mt genome being incorporated into the nuclear genome [e.g. the largest human numt covers 90% of the mt genome (Mourier et al., 2001; Hazkani-Covo et al., 2010)], they cannot incorporate as circular molecules and the method of overlapping long PCRs used in the present study would only work with circular targets.

The final possible explanations for the paraphyly of *cuprina* with respect to *sericata* are that there has been hybridisation between the two species, or that the paraphyletic *cuprina* haplotypes are the result of incomplete lineage sorting. Hybridisation between the two species could result in the introgression of *sericata* mt genome haplotypes into *cuprina*. This was initially considered by Stevens & Wall (1996) to be less plausible than *cuprina* subspecies, however, it has recently gained favour (McDonagh & Stevens, 2011). Laboratory evidence indicates that *sericata* and *cuprina* hybridise readily and that hybrids consistently display *cuprina*-like morphologies (Ullyett, 1945, Waterhouse & Paramonov, 1950). Ongoing field hybridisation would be expected to yield a pattern of multiple unrelated haplotypes and, assuming it was not directional (e.g. only *cuprina* ♂, *sericata* ♀ crosses are viable), there would also be *sericata* morphology flies with *cuprina* mt haplotypes. Despite a global sampling of *sericata* (three continents are represented in Toule et al., 2009, McDonagh & Stevens, 2011 and the present analysis; four continents are represented in Harvey et al., 2008) no shared haplotypes have yet been identified. Rather, in all studies, including the present one, the *cuprina*-haplotypes formed a single, highly distinct

clade. Stevens et al. (2002) and Harvey et al. (2008) respectively found that the *cuprina*-haplotypes were 0.9 and 0.93% diverged from the *sericata* haplotypes in the *cox1* barcoding region, significantly more than the variation within *sericata* (0.26%). Across the whole mt genome, the *cuprina*-haplotypes were between 0.128 and 1.227% (*rrnS* and *atp8*, respectively) diverged from *sericata*, whereas divergences within *sericata* ranged from 0 to 0.610% (*nad4l* and *atp8*, respectively) (Table 5). While not a large “barcoding gap” (sensu Meyer & Paulay, 2005), the distinctiveness of the *cuprina*-haplotypes across the entire genome is a strong argument against recent hybridisation.

Incomplete lineage sorting (Degnan & Rosenberg, 2009), i.e. that mitochondrial haplotypes within *cuprina* have yet to coalesce into a monophyletic lineage, is difficult analytically to distinguish from ancient hybridisation events as both can result in the pattern observed here (Funk & Omland, 2003; Joly et al., 2009). Incomplete lineage sorting was assessed by Toule et al. (2009) using simulation and was rejected for all but unrealistically small population sizes i.e. $N_E = 10^6$ or less. Further, proposing incomplete lineage sorting for this system would require nuclear genes to have coalesced quicker than the mt ones, which is unexpected given the general lower rate of substitution in nuclear genes (Ballard & Whitlock, 2004; Lin & Danforth, 2004; Haag-Liautard et al., 2008). Positive selection could have led to rapid fixation of particular alleles of individual nuclear genes, and thus apparent rapid coalescence, however this is unlikely to have broadly affected the nuclear genome to give the RAPD results from Stevens & Wall (1997). A hybridisation event prior to the last common ancestor of contemporary *sericata* populations is thus the most likely explanation for the observed mitochondrial phylogenetic pattern. For this reason, concerns about the introgression of insecticide resistance and the viability of a female killing system developed for *cuprina* (Scott et al., 2004) expressed by McDonagh & Stevens (2011) may be misplaced as the genetic evidence does not support the ongoing or recent field hybridisation necessary for either scenarios to occur. It is possible that, despite global sampling of *sericata* in this and previous studies, we have yet to identify the most divergent lineages within *sericata* and thus the source of the *cuprina* introgression may yet be extant. Invasive species frequently show low levels of genetic variability outside of their native range (e.g. Tsutsui et al., 2001; Shufan & Payton,

2009), and it is possible that, despite global sampling, we have still only sequenced *sericata* from a limited, “invasive” haplotype pool. Additional sampling of *sericata* within its home range is necessary to completely resolve the history of genetic flow between it and *cuprina*.

Conclusions

The present study demonstrates the comparative ease and economy with which a large number of mitochondrial genomes can be sequenced for a closely related group of species. For mt genomes sequenced via traditional Sanger methods, as here, the largest cost is designing novel primers for each new genome via primer walking. For example, at between US\$6 and US\$8 per primer, the 44 new primers designed for the first *Lucilia cuprina* mt genome sequenced cost approx. US\$300 against approximately US\$250 in sequencing charges (US\$4.50 per lane); with the inclusion of extraction kits, polymerases, general laboratory consumables the cost per mt genome is between US\$600 and US\$700, depending on the number of failed reactions etc. The high rates of primer reuse reported here equate to the ability to resequence mt genomes, or sequence closely related species, for approximately half the cost of sequencing the first mt genome. As this cost per sample greatly exceeds that of single-gene barcode-style approaches (US\$300 vs US\$5 – 10; Cameron et al., 2006b), we do not propose the routine sequencing of whole mt genomes for diagnostics but rather as a valuable trial-phase in developing diagnostic markers. For example, following the sequencing of mt genomes for nine *Chrysomya* specimens (cost approx. \$3000) we can identify 11 genes with a higher proportion of variable sites than the *coxI*-barcode; *nad6* has almost twice the proportion of variable sites, 12.31% vs. 6.99%, and is a similar size to the barcode region, 528 bp vs. 658 bp, compatible with single pass Sanger sequencing. Choosing diagnostics markers by an experimental approach, i.e. “which of several genes is most informative?”, as opposed to the verification approach applied by most DNA barcoding studies, i.e. “is this set of species diagnosable by *coxI*?”, has great potential to increase the efficacy and precision of molecular diagnostic protocols at modest cost.

Next-generation sequencing (NGS) methods have recently been applied to mt genomics, offering an alternative approach to eliminating the time and expense of primer walking: long PCR products are randomly shredded to create a library of short fragments which are directly sequenced by one of the various NGS technologies, principally pyrosequencing (Jex et al., 2010; Timmermans et al., 2010). However, the consumables cost of even the cheapest NGS platforms are much higher per unit run than the primer recycling approach outlined here: approx. US\$1000/run for Ion Torrent or Illumina MiSeq, US\$2-3000/lane for 454 FLX or Illumina Hi-Seq (Glenn, 2011). The use of coded DNA reference tags (often termed barcodes, but having no relation to the *coxI*-barcode discussed elsewhere in this paper) enables the mixing of multiple samples within a single NGS run (Parameswaran et al., 2007). While these tags can be cheaply applied to PCR amplicons by targeting the primers (Binladen et al., 2007; Bybee et al., 2011), creating individual DNA libraries for each sample is necessary to label shredded long PCR products and greatly increases costs. Timmermans et al. (2010) were able to successfully reconstruct whole mt genomes via NGS without labelling individual specimens for a set of 28 beetle species (21 nearly complete and seven partial mt genomes) at a cost of approx. US\$100/genome by pooling PCR products prior to library construction. This method, unfortunately, cannot be applied to such a closely related group as the calliphorids examined in the present study due to the low rates of sequence divergence between conspecific and congeneric specimens. In contrast, the 28 species sequenced by Timmermans et al. (2010) are representatives of 28 different beetle families. The application of their method to deep-level phylogenetic analyses is clear; however it is likely that if applied to a dataset such as the present one, multiple specimens, if not multiple species, would be joined as a single contig. Such a contig could potentially be used to identify SNPs or levels of gene by gene variability for the group of species, as discussed above, but would not yield single mt genomes which could be directly analysed for phylogenetic or biogeographic questions. The rapid improvements to NGS technology over the last decade suggest, however, that the question of the most economical method of collecting whole mt genome data from a selection of closely related taxa will likely change substantively in the near future.

Irrespective of the sequencing approach used to generate them, the comparison of whole mt genomes across a large number of species allows the analysis of genetic variability on a gene by gene basis, at multiple taxonomic levels, and thus allows the identification of potentially more informative markers for diagnostics. Within the Calliphoridae, the targets identified in this study need to be verified for very large samples of flies from each forensically significant genus if they are to match the data which are available for the *coxI* barcode (c.f. Harvey et al., 2008, 119 barcodes vs the 29 mt genomes here). Molecular diagnostic projects could conceivably use the entire mt genome as the marker, one approximately 25 times the size of current *coxI* barcodes, however existing sequencing technology, both Sanger and pyrosequencing methods, makes this uneconomic at the present time. Most recent developments in sequencing technology have emphasised depth of coverage over the simultaneous handling of large numbers of samples necessary for efficient molecular diagnostics but advances in physical separation of samples through gaskets could conceivably change this.

Whole mt genomes produce significantly better nodal support for phylogenetic relationships across the Oestroidea, than most of the previous analyses, especially at the family and subfamily levels (Kutty et al., 2010). Due to low variability and documented introgression events in this family, caution should be taken in inferring species-level phylogenetic patterns from mt genes alone. The inclusion of whole mt genome data does not alter the need for concern as low variability occurs across the entire genome and may be indicative of either recent speciation events or a low background rate of nucleotide substitution. Introgression following hybridisation also affects the whole mt genome, as demonstrated for the *sericata-cuprina* group but cannot be detected without nuclear data to compare against. A combination of whole mt genomes, as used in the present study, with additional nuclear markers (as in Kutty et al., 2010 & McDonagh & Stevens, 2011) is likely to be the best way to accurately resolve phylogenetic relationships within the Calliphoridae.

Acknowledgements

The authors thank Jessica Morgan & Peter James (Queensland Dept of Primary Industries & Fisheries, Australia) and Olivier Roux (Université Paul Sabatier, France) for donating specimens. This study was supported by the US National Science Foundation (DEB0444972: MFW, SLC), the Australian Research Council, Linkage (LP0883711: JFW, MD, SLC) and Future Fellowships programs (FT120100746), the Schlinger Foundation (LAN, DYK) and the Queensland University of Technology, Vice Chancellor's Research Fellowship (SLC).

References

- Altschul, S.F., Madden, T.L. Schäffer, A.A. Zhang, J. Zhang, Z. Miller, W., Lipman, D.J. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389-3402.
- Amendt, J., Krettek, R., Zehner, R. 2004. Forensic entomology. *Naturwissenschaften* 91, 51-65.
- Ballard, J.W.O., Whitlock, M.C. 2004. The incomplete natural history of mitochondria. *Mol. Ecol.* 13, 729-744.
- Beckenbach, A.T. 2011. Mitochondrial genome of Nematocera (Lower Diptera): Evidence of rearrangement following a complete genome duplication in a winter crane fly. *Genome Biol. Evol.* 4, 89-101.
- Binladen, J., Gilbert, M.T.P., Bolback, J.P., Panitz, F., Bendixen, C., Nielsen, R., Willerslev, E. 2007. The use of coded PCR primers enables high-throughput sequencing of multiple homolog amplification products by 454 parallel sequencing. *PLoS ONE* 2, e197.
- Bybee, S.M., Taylor, S.D., Nelson, C.R., Whiting, M.F. 2004. A phylogeny of robber flies (Diptera: Asilidae) at the subfamilial level: molecular evidence. *Mol. Phylogen. Evol.* 30, 789-797.

Bybee S.M., Bracken-Grissom, H., Haynes, B.D., Hermansen, R.A., Byers, R.L., Clement, M.J., Udall, J.A., Wilcox, E.R., Crandall, K.A. 2011. Targeted amplicon sequencing (TAS): A scalable next-gen approach to multilocus multitaxa phylogenetics. *Genome Biol. Evol.* 3, 1312-1323.

Cameron, S.L., Whiting, M.F. 2007. Mitochondrial genomic comparisons of the subterranean termites from the Genus *Reticulitermes* (Insecta: Isoptera: Rhinotermitidae). *Genome* 50, 188-202.

Cameron, S.L., Whiting, M.F. 2008. The complete mitochondrial genome of the tobacco hornworm, *Manduca sexta*, (Insecta: Lepidoptera: Sphingidae), and an examination of mitochondrial gene variability within butterflies and moths. *Gene* 408, 112-123.

Cameron, S.L., Barker, S.C., Whiting, M.F. 2006a. Mitochondrial genomics and the relationships and validity of the new insect order Mantophasmatodea. *Mol. Phylogen. Evol.* 38, 274-279.

Cameron, S.L., Rubinoff, D., Will, K. 2006b. Who will actually use DNA barcoding and what will it cost? *Syst. Biol.* 55, 844-847.

Cameron, S.L., Lambkin, C.L., Barker, S.C., Whiting, M.F. 2007. Utility of mitochondrial genomes as phylogenetic markers for insect intraordinal relationships – A case study from flies (Diptera). *Syst. Entomol.* 32, 40-59.

Cameron, S.L., Sullivan, J., Song, H., Miller, K.B., Whiting, M.F. 2009. A mitochondrial genome phylogeny of the Neuropterida (lace-wings, alderflies and snakeflies) and their relationship to the other holometabolous insect orders. *Zool. Scr.* 38, 575-590.

- Cameron, S.L., Yoshizawa, K., Mizukoshi, A., Whiting, M.F., Johnson, K.P. 2011. Mitochondrial genome deletions and mini-circles are common in lice (Insecta: Phthiraptera). *BMC Genomics* 12, 394.
- Cameron, S.L., Lo, N., Bourguignon, T., Svenson, G.J., Evans, T.A. 2012. A mitochondrial genome phylogeny of termites (Insecta: Termitoidae): Robust support for interfamilial relationships and molecular synapomorphies define major clades. *Mol. Phylogen. Evol.* 65, 163-173.
- Catts, E.P., Goff, M.L. 1992. Forensic entomology in criminal investigations. *Annu. Rev. Entomol.* 37: 253-272.
- Chen, W.-Y., Hung, T.-H., Shiao, S.-F. 2004. Molecular identification of forensically important blow fly species (Diptera: Calliphoridae) in Taiwan. *J. Med. Entomol.* 41, 47-57.
- Collins, R.A., Boykin, L.M., Cruickshank, R.H. & Armstrong, K.F. 2012. Barcoding's next top model: an evaluation of nucleotide substitution models for specimen identification. *Methods Ecol. Evol.* 3, 457-465.
- Cummings, M.P., Handley, S.A., Myers, D.S., Reed, D.L., Rokas, A., Winka, K. 2003. Comparing bootstrap and posterior probability values in the four-taxon case. *Syst. Biol.* 52, 477-487.
- DeBry, R.W., Timm, A.E., Dahlem, G.A., Stamper, T. 2010. mtDNA-based identification of *Lucilia cuprina* (Wiedemann) and *Lucilia sericata* (Meigen) (Diptera: Calliphoridae) in the continental United States. *Forensic Sci. Int.* 202, 102-109.
- Degnan, J.H., Rosenberg, N.A. 2009. Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends Ecol. Evol.* 24, 332-340.

- Dowton, M., Cameron, S.L., Dowavic, J.I., Austin, A.D., Whiting, M.F. 2009a. Characterisation of 67 mitochondrial gene rearrangements in the Hymenoptera reveal underlying trends in mitochondrial genome evolution. *Mol. Biol. Evol.* 26, 1607-1617.
- Dowton, M., Cameron, S.L., Austin, A.D., Whiting, M.F. 2009b. Phylogenetic approaches for the analysis of mitochondrial genome sequence data in the Hymenoptera – a lineage with both rapidly and slowly evolving mitochondrial genomes. *Mol. Phylogen. Evol.* 52, 512-519.
- Edgar, R. 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5: 113.
- Funk, D.J., Omland, K.E. 2003. Species-level paraphyly and polyphyly: frequency, causes and consequences with insights from animal mitochondrial DNA. *Annu. Rev. Ecol. Evol. S.* 34, 397-423.
- Glenn, T.C. 2011. Field guide to next-generation DNA sequencers. *Mol. Ecol. Resour.* 11, 759-769.
- Greenberg, B. 1991. Flies as forensic indicators. *J. Med. Entomol.* 28, 565-577.
- Haag-Liautard, C., Coffey, N., Houle, D., Lynch, M., Charlesworth, B., Keightley, P.D. 2008. Direct estimation of the mitochondrial DNA mutation rate in *Drosophila melanogaster*. *PLoS Biol.* 8, 2204.
- Harvey, M.L., Dadour, I.R., Gaudieri, S. 2003a. Mitochondrial DNA cytochrome oxidase I gene: potential for distinction between immature stages of some forensically important fly species (Diptera) in western Australia. *Forensic Sci. Int.* 131, 134-139.

- Harvey, M.L., Mansell, M.W., Villet, M.H., Dadour, I.R. 2003b. Molecular identification of some forensically important blowflies of southern Africa and Australia. *Med. Vet. Entomol.* 17, 363-369.
- Harvey, M.L., Gaudieri, S., Villet, M.H., Dadour, I.R. 2008. A global study of forensically significant calliphorids: Implications for identification. *Forensic Sci. Int.* 177, 66-76.
- Hazkani-Covo, E., Zeller, R.M., Martin, W. 2010. Molecular poltergeists: mitochondrial DNA copies (*numts*) in sequenced nuclear genomes. *PLoS Genet.* 6, e1000834.
- Hebert, P.D.N., Cywinska, A., Ball, S.L., deWaard, J. R. 2003. Biological identifications through DNA barcodes. *P. Roy. Soc. Lond. B. Bio.* 270, 313–321.
- Jex, A.R., Hall, S.R., Littlewood, D.T.J., Gasser, R.B. 2010. An integrated pipeline for next-generation sequencing and annotation of mitochondrial genomes. *Nucleic Acids Res.* 38, 522-533.
- Joly, S., McLenachan, P.A., Lockhart, P.J. 2009. A statistical approach for distinguishing hybridization and incomplete lineage sorting. *Am Nat.* 174, E54-E70.
- Junqueira, A.C.M., Lessinger, A.C., Teixeira Torres, T., Rodrigues da Silva, F., Vettore, A.L., Arruda, P., Azeredo Espin, A.M.L. 2004. The mitochondrial genome of the blowfly *Chrysomya chloropyga* (Diptera: Calliphoridae). *Gene* 339, 7-15.
- Kilpert, F., Held, C., Podsiadlowski, L. 2012. Multiple rearrangements in mitochondrial genomes of Isopoda and phylogenetic implications. *Mol. Phylogen. Evol.* 64, 106-117.
- Knights, G., Urech, R., Green, P. 2008. Sheep parasites: The LuciTrap sheep blowfly trapping system. DPI&F Note. <http://www2.dpi.qld.gov.au/sheep/8507.html>.

- Kutty, S.N., Pape, T., Wiegmann, B.M., Meier, R. 2010. Molecular phylogeny of the Calyptratae (Diptera: Cyclorrhapha) with an emphasis on the superfamily Oestroidea and the position of Mystacinobiidae and McAlpine's fly. *Syst. Entomol.* 35, 614-635.
- Lessard, B.D., Wallman, J.F., Dowton, M. 2009. Incorrect report of cryptic species with *Chrysomya rufifacies* (Diptera: Calliphoridae). *Invert. Syst.* 23, 507-514.
- Lessinger, A.C., Junqueira, A.C.M., Lemos, T.A., Kemper, E.L., Rodrigues da Silva, F., Vettore A.L., Arruda, P., Azeredo Espin, A.M.L. 2000. The mitochondrial genome of the primary screwworm fly *Cochliomyia hominivorax* (Diptera: Calliphoridae). *Insect Mol. Biol.* 9, 521-529.
- Lessinger, A.C., Junqueira, A.C.M., Conte, F.F., Azeredo-Espin, A.M.L. 2004. Analysis of a conserved duplicated tRNA gene in the mitochondrial genome of blowflies. *Gene* 339, 1-6.
- Levot, G. 2003. Insect fauna used to estimate the post-mortem interval of deceased persons. *Gen. App. Entomol.* 32, 31-39.
- Lin, C.P., Danforth, B.N. 2004. How do insect nuclear and mitochondrial gene substitution patterns differ? Insights from Bayesian analysis of combined datasets. *Mol. Phylogen. Evol.* 30, 686-702.
- Lowe, T.M., Eddy, S.R. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* 25, 955-964.
- Luo, A., Zhang, A., Ho, S.Y.W., Xu, W., Zhang, Y., Shi, W., Cameron, S.L., Zhu, C.D. 2011. Potential efficacy of mitochondrial genes for animal barcoding: a case study using eutherian mammals. *BMC Genomics* 12, 84.

Maddison, W., Maddison, D. 2003. MacClade ver 4.06. Sinauer Associates. Sunderland, Mass.

Malgorn, Y., Coquoz, R. 1999. DNA typing for identification of some species of Calliphoridae: An interest in forensic entomology. *Forensic Science Int.* 102, 111-119.

McAlpine, J.F. 1989. Phylogeny and classification of the Muscomorpha. In: McAlpine, J.F., Wood, D.V. (Eds) *Manual of Nearctic Diptera* vol. 3. Research Branch, Agriculture Canada, Ottawa, pp. 1397-1518.

McDonagh, L.M., Stevens, J.R. 2011. The molecular systematics of blowflies and screwworm flies (Diptera: Calliphoridae) using *28S rRNA*, *COX1* and *EF-1 α* : insights into the evolution of dipteran parasitism. *Parasitology* 138, 1760-1777.

Meyer, C.P., Paulay, G. 2005. DNA barcoding: Error rates based on comprehensive sampling. *PLoS Biol.* 3: e422.

Moulton, M.J., Song, H., Whiting, M.F. 2010. Assessing the effects of primer specificity on eliminating numt coamplification in DNA barcoding: a case study from Orthoptera (Arthropoda: Insecta). *Mol. Ecol. Resour.* 10, 615-627.

Mourier, T., Hansen, A.J., Willerslev, E., Arctander, P. 2001. The human genome project reveals a continuous transfer of large mitochondrial fragments to the nucleus. *Mol. Biol. Evol.* 18: 1833-1837.

Negrisol, E., Babbucci, M., Patarnello, T. 2011. The mitochondrial genome of the ascalaphid owlfly *Libelloides macaronius* and comparative evolutionary mitochondriomics of neuropterid insects. *BMC Genomics* 12, 221.

Nelson, L.A., Wallman, J.F., Dowton, M. 2007. Using CO1 barcodes to identify forensically and medically important blowflies. *Med. Vet. Entomol.* 21, 44-52.

Nelson, L.A., Wallman, J.F., Dowton, M. 2008. Identification of forensically important *Chrysomya* (Diptera: Calliphoridae) species using the second ribosomal internal transcribed spacer (ITS2). *Forensic Sci. Int.* 177, 238-247.

Nelson, L.A., Cameron, S.L., Yeates, D.K. 2012. The complete mitochondrial genome of the flesh fly, *Sarcophaga impatiens* Walker (Diptera: Sarcophagidae). *Mitochondr. DNA* 23, 42-43.

Norris, K.R. 1990. Evidence for the multiple exotic origin of Australian populations of the sheep blowfly, *Lucilia cuprina* (Wiedemann) (Diptera: Calliphoridae). *Aust J Zool.* 38, 635-648.

Oliviera, M.T., Grande Barau, J., Junqueira, A.C.M., Feijao, P.C., da Rosa, A.C., Abreu, C.F., Azeredo Espin, A.M.L., Lessinger, A.C. 2008. Structure and evolution of the mitochondrial genomes of *Haematobia irritans* and *Stomoxys calcitrans*: The Muscidae (Diptera: Calyptratae) perspective. *Mol. Phylogen. Evol.* 48, 850-857.

Parameswaran, P., Kalili, R., Tao, L., Shokrall, S., Gharizadeh, B., Ronaghi, M., Fire, A.Z. 2007. A pyrosequencing-tailored nucleotide barcode design unveils opportunities for large-scale multiplexing. *Nucleic Acids Res.* 35, e130.

Park, S.H., Zhang, Y., Piao, H., Yu, D.H., Jeong, H.J., Yoo, G.Y., Chung, U., Jo, T.-H., Hwang, J.J. 2009. Use of cytochrome c oxidase subunit I (*COI*) nucleotide sequences for identification of the Korean Lucilinae fly species (Diptera: Calliphoridae) in forensic investigations. *J. Korean Med. Sci.* 24, 1058-1063.

Preativatanyou, K., Sirisup, N., Payungporn, S., Poovorawan, Y., Thavara, U., Tawatsin, A., Sungpradit, S., Siriyasatien, P. 2010. Mitochondrial DNA-based identification of some forensically important blowflies in Thailand. *Forensic Sci. Int.* 202, 97-101.

Posada, D., Crandall, K.A. 1998. ModelTest: Testing the best-fit model of nucleotide substitution. *Bioinformatics* 14, 817-818.

Rambaut, A., Drummond, A.J. 2009. Tracer v1.5. <http://beast.bio.ed.ac.uk/Tracer>.

Ratcliffe, S.T., Webb, D.W., Weinzievr, R.A., Robertson, H.M. 2003. PCR-RFLP identification of Diptera (Calliphoridae, Muscidae and Sarcophagidae) – a generally applicable method. *J. Forensic Sci.* 48, 783-785.

Roe, A.D., Sperling, F.A.H. 2007. Patterns of evolution of mitochondrial cytochrome *c* oxidase I and II DNA and implications for DNA barcoding. *Mol. Phylogen. Evol.* 44, 325-345.

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M.A., Huelsenbeck, J.P. 2012. MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* 61, 539-542.

Rubinoff, D., Cameron, S.L., Will, K. 2006. A genomic perspective on the shortcomings of mitochondrial DNA for barcoding and DNA taxonomy. *J Hered* 97, 581-594.

Scott, M.J., Heinrich, J.C., Li, X.L. 2004. Progress towards the development of a transgenic strain of the Australian sheep blowfly (*Lucilia cuprina*) suitable for a male-only sterile release program. *Insect Biochem. Molec.* 34, 185-192.

- Shufran, K.A., Payton, T.L. 2009. Limited genetic variation within and between Russian wheat aphid (Hemiptera: Aphididae) biotypes in the United States. *J. Econ. Entomol.* 102, 440–445.
- Simon, C., Frati, F., Beckenbach, A., Crespi, B., Liu, H., Flook, P.K. 1994. Evolution, weighting and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. *Ann. Entomol. Soc. Am.* 87, 651-701.
- Singh, B., Kurahashi, H., Wells, J.D. 2011. Molecular phylogeny of the blowfly genus *Chrysomya*. *Med. Vet. Entomol.* 25, 126-134.
- Skerratt, L.F., Campbell, N.J.H., Murrell, A., Walton, S., Kemp, D., Barker, S.C. 2002. The mitochondrial 12S gene is a suitable marker of populations of *Sarcoptes scabiei* from wombats, dogs and humans in Australia. *Parasitol. Res.* 88, 376-379.
- Song, H., Buhay, J.E., Whiting, M.F., Crandall, K.A. 2008. Many species in one: DNA barcoding overestimates the number of species when nuclear mitochondrial pseudogenes are coamplified. *P. Natl. Acad. Sci. USA* 105, 13486-13491.
- Song, H., Sheffield, N.C., Cameron, S.L., Miller, K.B., Whiting, M.F. 2010. What happens when the phylogenetic assumptions are violated?: The effect of base compositional heterogeneity and among-site rate heterogeneity in beetle mitochondrial phylogenomics. *Syst. Entomol.* 35, 429-448.
- Sperling, F.A.H., Anderson, G.S., Hickey, D.A. 1994. A DNA-based approach to the identification of insect species used for post-mortem interval estimation. *J. Forensic Sci.* 39, 418-427.
- Stamatakis, A., Hoover, J., Rougemont, J., 2008. A rapid bootstrap algorithm for the RaxML Web-servers. *Syst. Biol.* 75, 758-771.

- Stevens, J.R. 2003. The evolution of myiasis in blowflies (Calliphoridae). *Intl. J. Parasitol.* 33, 1105-1113.
- Stevens, J.R., Wall, R. 1996. Species, sub-species and hybrid populations of the blowflies *Lucilia cuprina* and *Lucilia sericata* (Diptera: Calliphoridae). *P. Roy. Soc. Lond. B. Bio.* 263, 1335-1341.
- Stevens, J.R., Wall, R. 1997. Genetic variation in populations of the blowflies *Lucilia cuprina* and *Lucilia sericata* (Diptera: Calliphoridae). Random amplified polymorphic DNA analysis and mitochondrial DNA sequences. *Biochem. Syst. Ecol.* 25: 81-97.
- Stevens, J.R., Wall, R. 2001. Genetic relationships between blowflies (Calliphoridae) of forensic importance. *Forensic Sci. Int.* 120, 116-123.
- Stevens, J.R., Wallman, J.F. 2006. The evolution of myiasis in humans and other animals in the Old and New Worlds (part 1): phylogenetic analysis. *Trends Ecol. Evol.* 22, 129-136.
- Stevens, J.R., Wall, R., Wells, J.D. 2002. Paraphyly in Hawaiian hybrid blowfly populations and the evolutionary history of anthropophilic species. *Insect Mol. Biol.* 11, 141-148.
- Stevens, J.R., West, H., Wall, R. 2008. Mitochondrial genomes of the sheep blowfly, *Lucilia sericata*, and the secondary blowfly, *Chrysomya megacephala*. *Med. Vet. Entomol.* 22, 89-91.
- Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S. 2011. MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance and maximum parsimony methods. *Mol. Biol. Evol.* 28, 2731-2739.

- Timmermans, M.J.T.N., Dodsworth, S., Culverwell, C.L., Bocak, L., Ahrens, D., Littlewood, D.T.J., Pons, J., Vogler, A.P. 2010. Why barcode? High-throughput multiplex sequencing of mitochondrial genomes for molecular systematics. *Nucleic Acids Res.* 38, e197.
- Toule, R., Downie, D.A., Villet, M.H. 2009. Flies in the ointment: a morphological and molecular comparison of *Lucilia cuprina* and *Lucilia sericata* (Diptera: Calliphoridae) in South Africa. *Med. Vet. Entomol.* 23, 6-14.
- Tsutsui, N.D., Suarez, A.V., Holway, D.A., Case, T.J. 2001. Relationships among native and introduced populations of the Argentine ant (*Linepithema humile*) and the source of introduced populations. *Mol. Ecol.* 10, 2151-5161.
- Ullyett, G.C. 1945. Species of *Lucilia* attacking sheep in South Africa. *Nature* 155, 636-637.
- Vincent, S., Vian J.M., Carlotti, M.P. 2000. Partial sequencing of the cytochrome coxydase (sic) b subunit gene I: A tool for the identification of European species of blwo flies for post-mortem interval estimation. *J. Forensic Sci.* 45, 820-823.
- Wallman, J. F. 2001. A key to the adults of species of blowflies in southern Australian known of suspected to breed in carrion. *Med. Vet. Entomol.* 15, 433-437.
- Wallman, J.F., Donnellan, S.C. 2001. The utility of mitochondrial DNA sequences for the identification of forensically important blowflies (Diptera: Calliphoridae) in southeastern Australia. *Forensic Sci. Int.* 120, 60-67.
- Wallman, J.F., Leys, R., Hogendoorn, K. 2005. Molecular systematics of Australian carrion-breeding blowflies (Diptera: Calliphoridae) based on mitochondrial DNA. *Invert. Syst.* 19, 1-15.

- Waterhouse, D.F., Paramonov, S.J. 1950. The status of the two species of *Lucilia* (Diptera: Calliphoridae) attacking sheep in Australia. *Aust. J. Sci. Res. Ser B.* 3, 310-336.
- Weigle, S., Testini, G., Parisi, A., Dantas-Torres, F., Traversa, D., Colwell, D.D., Otranto, D. 2010. The mitochondrial genome of the common cattle grub *Hypoderma lineatum*. *Med. Vet. Entomol.* 24: 329-335.
- Wells, J.D., Sperling, F.A.H. 2001. DNA-based identification of forensically important Chrysomyinae (Diptera: Calliphoridae). *Forensic Sci. Int.* 120, 110-115.
- Wells, J.D., Stevens, J.R. 2008. Application of DNA-based methods in forensic entomology. *Annu. Rev. Entomol.* 53, 103-120.
- Wells, J.D., Goff, M.L., Tomberlin, J.K., Kurahashi, H. 2002. Molecular systematics of the endemic Hawaiian blowfly genus *Dyscritomyia* Grimshaw (Diptera: Calliphoridae). *Med. Entomol. Zool.* 53, 231-238.
- Wells, J.D., Lunt, N., Villet, M.H. 2004. Recent African derivation of *Chrysomya putoria* from *C. chloropyga* and mitochondrial DNA paraphyly of cytochrome oxidase subunit one in blowflies of forensic importance. *Med. Vet. Entomol.* 18, 445-448.
- Wells, J.D., Wall, R., Stevens, J.R. 2007. Phylogenetic analysis of forensically important *Lucilia* flies based on cytochrome oxidase I sequence: a cautionary tale for forensic species determination. *Int. J. Legal. Med.* 121, 229-233.

Whitworth, T.L., Dawson, R.D. Magalon, H. & Baudry, E. 2007. DNA barcoding cannot reliably identify species of the blowfly genus *Protophthora* (Diptera: Calliphoridae). P. Roy. Soc. Lond. B. Bio. 274, 1731-1739.

Wiegmann, B.M., Trautwein, M.D., Winkler, I.S., Barr, N.B., Kim, J.-W., Lambkin, C., Bertone, M.A., Cassel, B.K., Bayless, K.M., Heimberg, A.M., Wheeler, B.M., Peterson, K.J., Pape, T., Sinclair, B.J., Skevington, J.H., Blagoderov, V., Caravas, J., Kuttu, S.N., Schmidt-Ott, U., Kampmeier, G.E., Thompson, F.C., Grimaldi, D.A., Beckenbach, A.T., Courtney, G.W., Friedrich, M., Meier, R., Yeates, D.K. 2011. Episodic radiations in the fly tree of life. P. Natl. Acad. Sci. USA 108, 5690-5695.

Yamauchi, M.M., Miya, M.U., Nishida, M. 2004. Use of a PCR-based approach for sequencing whole mitochondrial genomes of insects: two examples (cockroach and dragonfly) based on the method developed for decapods crustaceans. Insect Mol. Biol. 13, 435-442.

Yamauchi, M.M., Miya, M.U., Machida, R.J., Nishida, M. 2005. PCR-based approach for sequencing mitochondrial genomes of decapods crustaceans, with a practical example from Kuruma prawn (*Marsupenaeus japonicas*). Mar. Biotech. 6, 419-429.

Zhang, D.X., Hewitt, G.M. 1997. Insect mitochondrial control region: A review of its structure, evolution and usefulness in evolutionary studies. Biochem. Syst. Ecol. 25, 99-120.

Tables

Table 1. DNA diagnostics/systematics studies of the Calliphoridae. Mitochondrial genes use standard abbreviations; *cox1* and *cox2* refers to full length sequences; *cox1-5'* and *cox1-3'* to sequences of just the 5' or 3' halves respectively; *cox1-250bp* is the region from 1000-1250bp of *cox1* (after Vincent et al., 2000); *cox2-3'* is approx. 90% of the gene excluding the 5' end; *nad4l-nad4* includes 210bp from the 3' end of *nad4l*, the entirety of *nad4* and 12bp of *trnH* (after Wallman et al., 2005); nuclear genes are preceded by the prefix 'nuc.'; *lsu-rRNA* = 28S, *ssu-rRNA* = 18S; CAD = carbomoylphosphate synthase or rudimentary; EF-1 α = elongation factor 1 alpha. Subfamilies included in each study: Ch, Chysomyinae; Ca, Calliphorinae; Lu: Luciliinae; Po: Polleniinae; Be: Bengaliinae; He: Helicoboscinae; Au: Auchmeromyiinae; To, Toxotarsinae.

Reference	Gene Regions	Subfamilies Studied
Sperling et al. 1994	<i>cox1</i> ; <i>cox2</i>	Ch, Lu
Vincent et al. 2000	<i>cox1-250bp</i>	Ch, Ca, Lu
Stevens & Wall, 2001	nuc. <i>lsu-rRNA</i>	Ch, Ca, Lu
Wallman & Donnellan, 2001	<i>cox1-3'</i> ; <i>cox2-3'</i>	Ch, Ca, Lu
Wells & Sperling, 2001	<i>cox1</i> ; <i>cox2</i>	Ch
Stevens et al. 2002	nuc. <i>lsu-rRNA</i> ; <i>cox1</i> ; <i>cox2</i>	Lu
Wells et al. 2002	<i>cox1</i> ; <i>cox2</i>	Lu
Harvey et al. 2003a	<i>cox1-3'</i>	Ch, Ca, Lu
Harvey et al. 2003b	<i>cox1</i>	Ch, Lu
Chen et al. 2004	<i>cox1</i>	Ch, Lu
Wells et al. 2004	<i>cox1-3'</i>	Ch
Wallman et al. 2005	<i>cox1-3'</i> ; <i>cox2-3'</i> ; <i>nad4l-nad4</i>	Ca, Lu
Nelson et al. 2007	<i>cox1-5'</i>	Ch
Wells et al. 2007	<i>cox1</i>	Lu
Whitworth et al. 2007	<i>cox1-3'</i> ; <i>cox2</i>	Ch
Nelson et al. 2008	nuc.ITS2	Ch
Harvey et al. 2008	<i>cox1</i>	Ch, Ca, Lu
Lessard et al. 2009	<i>cox2</i>	Ch
Park et al. 2009	<i>cox1</i>	Lu
Toule et al. 2009	<i>cox1</i> ; nuc. <i>lsu-rRNA</i>	Lu
Singh et al. 2010	<i>cox1</i> , nuc. CAD	Ch
Preativatanyou et al. 2010	<i>cox1-3'</i> ; <i>cox2</i>	Ch, Lu
DeBry et al. 2010	<i>cox1-3'</i> ; <i>cox2</i> ; nuc. <i>lsu-rRNA</i>	Lu
Kutty et al. 2010	<i>cox1-5'</i> ; <i>cob</i> ; <i>rrnS</i> ; <i>rrnL</i> ; nuc. <i>lsu-rRNA</i> ; nuc. <i>ssu-rRNA</i> ; nuc. CAD	Ch, Ca, Lu, Po, Be, He, To
McDonagh & Stevens, 2011	<i>cox1</i> ; nuc. EF-1 α ; nuc. <i>lsu-rRNA</i>	Ch, Ca, Lu, Po, Au, Be, He

Table 2. Species/specimens included in this study, their collection details and accession numbers for sequenced mt genomes.

Species	Classification	Accession #	Voucher	Collection details
Newly sequenced				
<i>Chrysomya albiceps</i> (Wiedemann, 1819)	Calliphoridae: Chrysomyinae	XXXXXX	DI254	Zambia, (G. Svenson)
<i>Chrysomya bezziana</i> Villeneuve, 1914	Calliphoridae: Chrysomyinae	XXXXXX	DI217	QDPIF-ARI field strain
<i>Chrysomya megacephala</i> (Fabricius, 1794)	Calliphoridae: Chrysomyinae	XXXXXX	DI212	University of Queensland campus, St Lucia, Brisbane, Qld, Australia (5 January 2006; S.L. Cameron)
<i>Chrysomya megacephala</i>	Calliphoridae: Chrysomyinae	XXXXXX	DI219	QDPIF-ARI lab strain
<i>Chrysomya rufifacies</i> (Macquart, 1843)	Calliphoridae: Chrysomyinae	XXXXXX	DI215	Emma Gorge, El Questro Resort, WA, Australia; (26 September 2002; M.F. Whiting & S.L. Cameron)
<i>Chrysomya rufifacies</i>	Calliphoridae: Chrysomyinae	XXXXXX	DI218	QDPIF-ARI lab strain
<i>Chrysomya saffrana</i> (Bigot, 1877)	Calliphoridae: Chrysomyinae	XXXXXX	DI216	QDPIF-ARI field strain
<i>Protophormia terraenovae</i> Robineau-Desvoidy, 1830	Calliphoridae: Chrysomyinae	XXXXXX	DI243	France, (O. Roux)
<i>Lucilia cuprina</i> (Wiedemann, 1830)	Calliphoridae: Luciliinae	XXXXXX +5	DI190.1 – DI190.5	University of Melbourne colony; (P. Batterham); 5 replicate specimens
<i>Lucilia cuprina</i>	Calliphoridae: Luciliinae	XXXXXX +5	DI213 – DI213.5	Petrie Terrace, Brisbane, Qld, Australia (December 2005; S.L. Cameron); 5 replicate specimens
<i>Lucilia sericata</i> (Meigen, 1826)	Calliphoridae: Luciliinae	XXXXXX	DI246	Canberra, ACT, Australia (C. Lambkin)
<i>Lucilia sericata</i>	Calliphoridae: Luciliinae	XXXXXX	DI220	QDPIF-ARI lab strain
<i>Lucilia sericata</i>	Calliphoridae: Luciliinae	XXXXXX	DI245	Perth, WA, Australia (E. Sinclair)
<i>Lucilia sericata</i>	Calliphoridae: Luciliinae	XXXXXX	DI257	Brigham Young University campus, Provo, UT, USA
<i>Lucila porphyrina</i> Walker, 1856	Calliphoridae: Luciliinae	XXXXXX	DI211	University of Queensland campus, St Lucia, Brisbane, Qld, Australia (21 September 2001; S.L. Cameron)
<i>Hemipyrellia ligurriens</i> (Wiedemann, 1830)	Calliphoridae: Luciliinae	XXXXXX	DI191	University of Queensland campus, St Lucia, Brisbane, Qld, Australia (October 2001; S.L. Cameron)
<i>Calliphora vicina</i> Robineau-Desvoidy, 1830	Calliphoridae: Calliphorinae	XXXXXX	DI242	France, (O. Roux)
<i>Pollenia rudis</i> (Fabricius, 1794)	Calliphoridae: Polleniinae	XXXXXX	DI210	Brigham Young University campus, Provo, UT, USA; (11 November 2005; S.L. Cameron)
<i>Rutilia georlingiana</i> Enderlein, 1936	Tachinidae: Dexiinae	XXXXXX	DI214	Mary River Roadhouse, Burrundie, NT, Australia (27 September 2002; M.F. Whiting & S.L. Cameron)
Previously Reported				
<i>Cochliomyia hominivorax</i> (Coquerel, 1858)	Calliphoridae: Chrysomyinae	AF260826	Lessinger et al., 2000	

<i>Chrysomya putoria</i> (Wiedemann, 1830)	Calliphoridae: Chrysomyinae	AF352790	Junqueira et al., 2004	
<i>Chrysomya megacephala</i>	Calliphoridae: Chrysomyinae	AJ426041	Stevens et al., 2008	
<i>Lucilia sericata</i>	Calliphoridae: Luciliinae	AJ422212	Stevens et al., 2008	
<i>Exorista sorbillans</i> (Wiedemann, 1830)	Tachinidae: Exoristinae	HQ322500	Shao et al. (unpub)	
<i>Sarcophaga impatiens</i> Walker, 1849	Sarcophagidae:Sarcophaginae	JN859549	Nelson et al., 2012	
<i>Dermatobia hominis</i> (Linnaeus Jr., 1781)	Oestridae: Cuterebrinae	AY463155	Azeredo-Espin et al. (unpub)	
<i>Hypoderma lineatum</i> (Viller, 1789)	Oestridae: Hypodermatinae	GU584123	Weigl et al., 2010	
<i>Haematobia irritans</i> (Linnaeus, 1758)	Muscidae: Muscinae	DQ029097	Oliviera et al., 2008	

QDPIF-ARI, Queensland Department of Primary Industries and Fisheries Agricultural Research Institute.

Table 3. Utility of specific primers in mt genome resequencing of calliphorid and tachinid flies.

Classification	Species	Universal Primers	Novel Primers	Recycled Primers	Percentage
Luciliinae	<i>Lucilia cuprina</i>	7	42	-	-
	<i>Lucilia sericata</i>	7	0	42	100
	<i>Lucilia porphyrina</i>	5	12	32	76
	<i>Hemipyrellia ligurriens</i>	7	9	33	82
Calliphorinae	<i>Calliphora vicina</i>	7	1	41	98
Chrysomyinae	<i>Chrysomya megacephala</i>	7	8	34	84
	<i>Chrysomya saffrana</i>	7	1	41*	98
	<i>Chrysomya bezziana</i>	6	4	39	92
	<i>Chrysomya rufifacies</i>	6	13	30	73
	<i>Chrysomya albiceps</i>	7	1	42**	98
	<i>Protophormia terraenovae</i>	7	0	41**	100
Polleniinae	<i>Pollenia rudis</i>	7	15	27	69
Tachinidae	<i>Rutelia georlingiana</i>	7	19	23	61

* Includes recycled primers designed for *Ch. megacephala*

** Includes recycled primers designed for *Ch. rufifacies*

Table 4. Divergence statistics for each gene within the Calliphoridae (C-idae), Chrysomyinae (Ch-inae), *Chrysomya* (Ch) and the *Chrysomya megacephala/saffranaea* clade (Ch. m). For each taxonomic group, the highest variability gene is highlighted in bold.

Gene	Length (bp)	C-idae			Ch-inae			Ch			Ch. m		
		% inf. sites	max. div. (%)	mean div. (%)	% inf. sites	max. div. (%)	mean div. (%)	% inf. sites	max. div. (%)	mean div. (%)	% inf. sites	max. div. (%)	mean div. (%)
<i>rrnS</i>	795	6.164	5.021	2.582	3.270	4.741	2.314	2.767	3.393	1.921	0.377	0.384	0.256
<i>rrnL</i>	1346	6.389	5.311	2.773	3.715	4.525	2.521	2.972	3.229	2.198	0.074	0.302	0.201
<i>atp6</i>	678	18.879	12.361	6.873	12.242	12.203	7.674	8.555	10.999	6.192	0.000	0.593	0.296
<i>atp8</i>	165	12.727	14.808	4.357	9.091	10.432	5.408	7.273	9.102	4.686	0.000	1.227	0.614
<i>cox1</i>	1539	17.024	11.281	6.977	10.266	11.281	6.406	8.122	8.481	5.087	0.130	0.458	0.316
<i>cox2</i>	690	14.493	12.117	6.070	7.681	8.866	5.174	6.232	6.566	4.275	0.000	0.292	0.146
<i>cox3</i>	789	16.857	12.439	6.450	9.379	12.439	6.369	6.844	7.234	4.591	0.127	0.510	0.339
<i>cob</i>	1137	20.053	13.784	8.055	13.984	13.784	8.845	11.697	11.486	7.582	0.264	1.158	0.755
<i>nad1</i>	942	16.773	10.910	6.541	10.828	10.910	6.787	8.174	8.846	5.593	0.212	0.750	0.464
<i>nad2</i>	1017	18.289	13.640	7.950	12.389	12.160	8.049	10.226	10.077	6.615	0.000	0.495	0.297
<i>nad3</i>	357	17.087	13.385	6.809	12.325	12.998	8.332	9.804	10.702	6.578	0.000	0.283	0.236
<i>nad4</i>	1341	14.616	12.911	5.719	10.291	12.911	6.691	7.532	8.147	5.254	0.298	1.515	0.806
<i>nad4l</i>	297	10.774	8.963	4.322	7.744	8.963	4.923	7.407	8.963	4.848	0.000	0.338	0.169
<i>nad5</i>	1720	18.140	11.590	7.450	12.384	10.748	7.664	10.000	9.614	6.517	0.233	0.761	0.536
<i>nad6</i>	528	24.621	17.755	10.028	16.098	16.343	10.430	12.311	13.633	8.418	0.000	0.382	0.191
<i>cox1-5'</i>	658	15.957	11.887	6.779	8.663	10.639	5.715	6.991	8.433	4.458	0.304	0.612	0.510
<i>cox1-3'</i>	825	17.818	12.062	7.080	11.879	12.062	7.043	9.212	8.857	5.798	0.121	0.490	0.265
<i>cox2-3'</i>	636	15.252	12.307	6.362	8.019	8.950	5.356	6.447	6.962	4.447	0.000	0.316	0.158
<i>nad4-nad4l</i>	1558	14.249	12.099	5.616	8.087	12.099	6.526	7.574	8.237	5.238	0.257	1.301	0.725

Table 5. Divergence statistics for each gene within the Luciliinae (L-inae), *Lucilia* (Lu), *Lucilia cuprina/sericata* clade (Lu c/s) and *Lucilia sericata* (Lu s). For each taxonomic group, the highest variability gene is highlighted in bold.

Gene	Length (bp)	L-inae			Lu			Lu.c/s			Lu. s		
		% inf. sites	max. div. (%)	mean div. (%)	% inf. sites	max. div. (%)	mean div. (%)	% inf. sites	max. div. (%)	mean div. (%)	% inf. sites	max. div. (%)	mean div. (%)
<i>rrnS</i>	795	0.881	2.464	0.613	0.503	1.419	0.407	0.503	0.512	0.268	0.000	0.128	0.051
<i>rrnL</i>	1346	1.486	3.707	0.882	0.520	2.843	0.539	0.520	0.454	0.242	0.149	0.226	0.121
<i>atp6</i>	678	3.687	8.281	2.402	1.917	6.822	1.709	1.917	1.950	1.008	0.000	0.296	0.118
<i>atp8</i>	165	2.424	5.692	1.368	1.818	5.047	1.147	1.212	1.852	0.596	0.000	0.610	0.244
<i>cox1</i>	1539	4.808	8.974	2.947	3.314	7.882	2.377	3.184	2.874	1.633	0.065	0.065	0.039
<i>cox2</i>	690	3.768	7.207	2.231	2.464	6.434	1.781	2.464	1.926	1.165	0.435	0.585	0.380
<i>cox3</i>	789	4.056	9.662	2.548	2.028	8.096	1.865	2.028	1.802	1.052	0.253	0.382	0.203
<i>cob</i>	1137	5.893	10.948	3.612	4.222	9.058	2.948	4.134	3.920	2.103	0.352	0.354	0.195
<i>nad1</i>	942	5.202	8.329	2.883	3.185	7.293	2.217	3.185	2.945	1.551	0.212	0.213	0.128
<i>nad2</i>	1017	4.130	8.919	2.450	1.770	7.403	1.639	1.770	1.700	0.844	0.393	0.395	0.237
<i>nad3</i>	357	2.241	12.652	2.410	1.681	5.622	1.511	1.681	1.720	0.894	0.000	0.283	0.113
<i>nad4</i>	1341	3.878	9.311	2.617	2.685	6.897	1.974	2.685	2.444	1.299	0.298	0.600	0.300
<i>nad4l</i>	297	1.347	3.101	0.777	0.673	2.748	0.570	0.673	0.678	0.316	0.000	0.000	0.000
<i>nad5</i>	1720	5.000	9.901	3.083	3.081	8.183	2.318	3.081	2.978	1.534	0.465	0.467	0.280
<i>nad6</i>	528	6.250	11.092	3.419	3.030	9.586	2.489	3.030	2.932	1.562	0.189	0.191	0.115
<i>cox1-5'</i>	658	4.559	7.377	2.604	3.040	7.377	2.215	3.040	2.796	1.530	0.151	0.152	0.091
<i>cox1-3'</i>	825	5.333	10.104	3.270	3.636	7.764	2.567	3.515	3.268	1.826	0.000	0.000	0.000
<i>cox2-3'</i>	636	4.088	7.679	2.388	2.673	7.015	1.917	2.673	2.094	1.245	0.472	0.475	0.349
<i>nad4-nad4l</i>	1558	3.594	8.516	2.388	2.439	6.322	1.805	2.439	2.163	1.177	0.257	0.516	0.258

Figures

Fig. 1

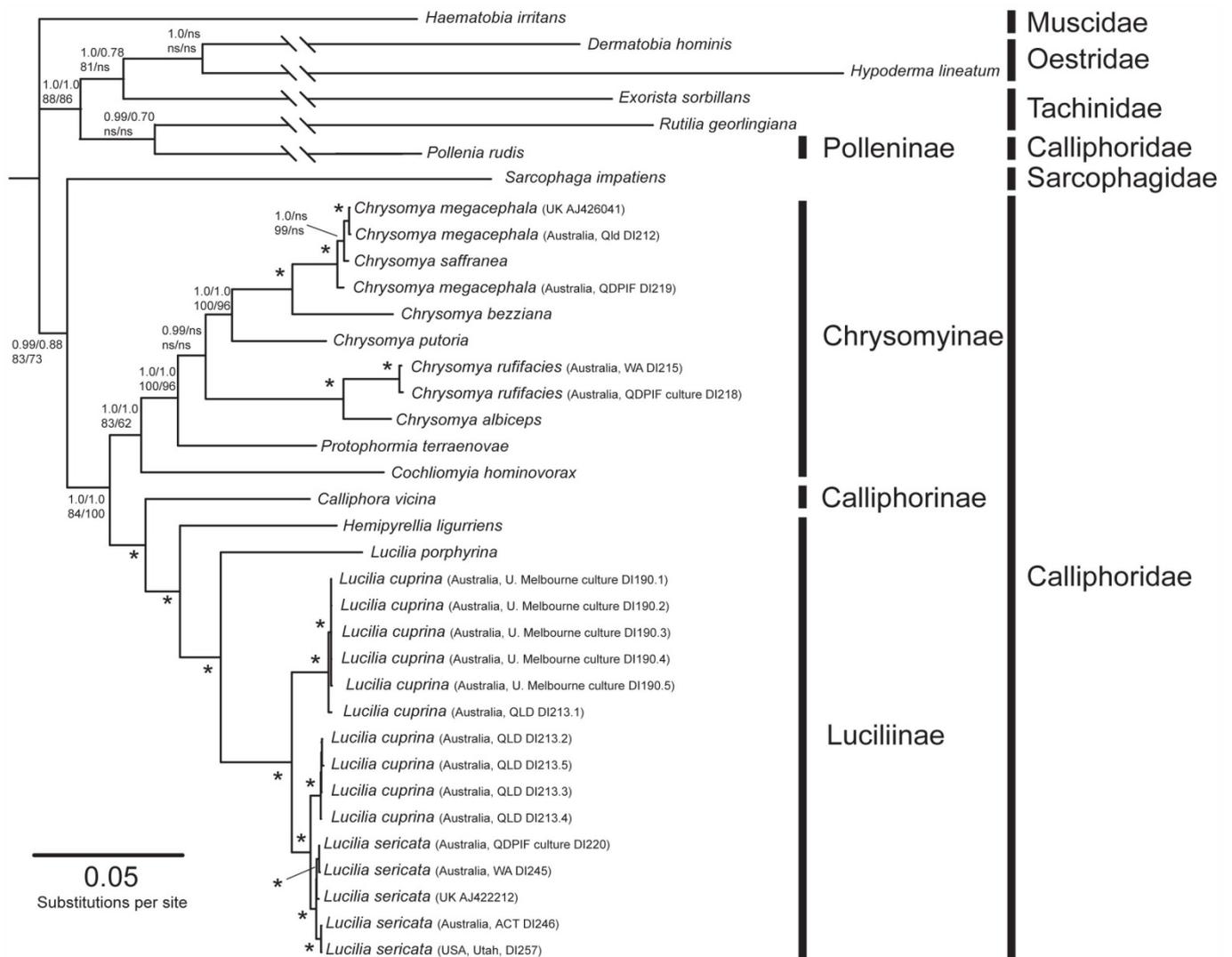


Figure 1. Mitochondrial genome phylogeny of oestroid flies (Diptera: Calyptrata), using the outgroup *Haematobia irritans*. The phylogram shown is that inferred for the Bayes-ALL dataset. Nodal support is given: Bayes-ALL/Bayes-NO3RD posterior probabilities (upper), RAxML-ALL/RAxML-NO3RD bootstrap percentage (lower); nodes marked with a "*" had 1.0 posterior probability and 100% bootstrap support in all four analyses; ns: not significant. Where more than one specimen was included for a given species collection locality and voucher codes or Genbank accession numbers are included for clarity (see Table 2 for full details). Branches marked \\ are drawn 50% of actual inferred length so as to fit the whole tree.

Fig. 2

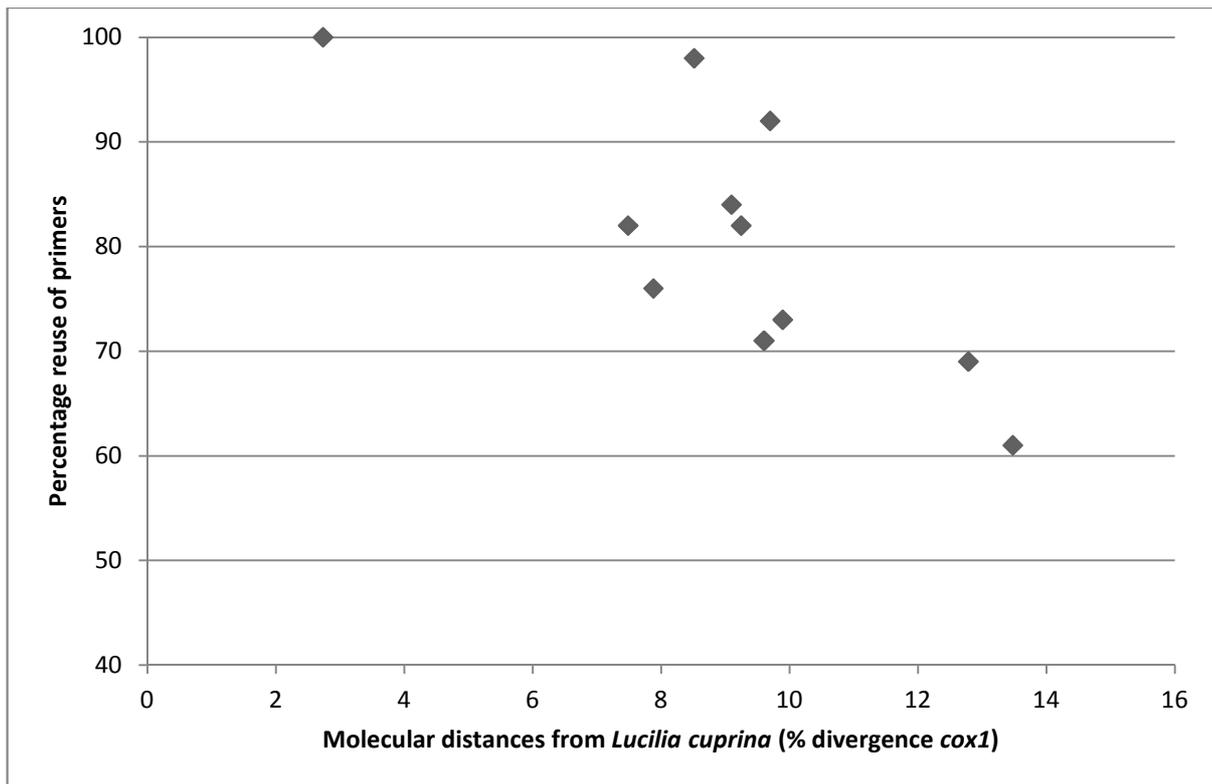


Figure 2. Primer reuse plotted against molecular distance of target species from *L. cuprina*, the species used for original primer design.