CAN DNA SEQUENCES HELP WITH SORTING BIODIVERSITY SAMPLES?

LIM SHIMIN GWYNNE

NATIONAL UNIVERSITY OF SINGAPORE

2009

CAN DNA SEQUENCES HELP WITH SORTING BIODIVERSITY SAMPLES?

LIM SHIMIN GWYNNE
(*B.Sc.(Hons.), NUS*)

A THESIS SUBMITTED

FOR THE DEGREE OF MASTER OF SCIENCE

DEPARTMENT OF BIOLOGICAL SCIENCES

NATIONAL UNIVERSITY OF SINGAPORE

2009

ACKNOWLEDGEMENTS

People I would like to thank from NUS include those from the Department of Biological Science, the Biodiversity Group, and the Evolutionary Biology Laboratory most of all. I would also like to extend my gratitude towards my collaborators at the Universiti Brunei Darussalam (UBD) who were the very soul of graciousness and generosity while hosting me.

Specific thanks to:

Prof. Meier: For discussing, countless editing sessions, brainstorming, nagging, pushing, funding, free Spinelli coffee and food, *etc.* and so on. I can't see myself writing this thesis at all without your guidance.

Dr. Ulmar Grafe: For opening up your house to Yuchen and I, collecting, explaining, brainstorming, guiding, and warning me about the poisonous vipers that lurk in the undergrowth.

Katrin Grafe: For feeding and opening her home to us. I am terribly sorry for tracking swamp through the nice clean floors of your house.

Hanyran: For sorting out the bulk of the specimens to morphotypes, and showing me where the experimental sites and all the good pitchers of *Nepenthes bicalcarata* are.

Yuchen: For chauffering, assisting in fieldwork (such as doing all the heavy lifting), taking pictures of the field site, imaging the specimens, consenting to being a guinea pig, and generally being a good sport whenever bothered for his help and expertise.

Sujatha: For coaching me through the entire process, and lending me your thesis, for encouraging me during the last few hours, and the innumerable Spinelli lunches! =)

Michael: For being scarily amazing at PCR and sequencing, and providing us with vast amounts of incredible sepsid material that will keep us busy for a good long time, and for hosting me in Munich. ESPECIALLY for introducing me to resealable sequencing plate lids

Denise: For working on the *Allosepsis indica*, pictures, being a nurturing

goddess who keeps the flies alive and breeding. It's okay, your time will come too!

Huifang: For dragging me off the Bangkok after this, and letting me vent about my thesis like 1000x

Kathy and Wei Song: I am *still* messing with the sepsid *COI* dataset. Can you imagine?

Yujie, Laura, Andrea and Amrita: For letting me interrupt their work with those long and meaningful chats I employ as a means of procrastination and feeding me when I demanded that it be so.

Patrick: For showing me the Dolichopodidae (still my favourite dipteran family!), encouraging me all this while, all the way from Belgium, and for showing me an unforgettable time while I was there. Mussels and beer, yum!

Parents: For making sure I don't have to deal with public transport, packing me off to school with nice things to eat, listening to me complain about why everything and everybody else is wrong.

Sibling: For being super nice about having her vacation interrupted by my minging.

The people I forgot: I swear I'd have thanked you if I weren't writing this at 1 a.m. in the morning. I'll treat you all to coffee sometime.

Green tea: A haiku
The lingering taste
Of green tea gone cold again
Can't end soon enough

TABLE OF CONTENTS

Acknowledgements i

Table of Contents iii

Summary vii

List of Tables x

List of Figures xi

List of Publications xii

General Introduction 1

Chapter 1: Use of the *COI* barcode for species richness estimation

1.1 Introduction 7

1.2 Materials and Methods 12

1.2.1 Taxon and character sampling 12

1.2.2 Alignment and analysis 15

1.3 Results 16

1.3.1 Congruence between DNA and taxonomic species estimates 16

1.3.2 Congruence between taxonomic species and *COI* clusters 18

1.4 Discussion 19

1.4.1 The relative performance of DNA and parataxonomy 19

1.4.1 Congruence between DNA cluster content and species 22

1.5 Conclusion 24

Chapter 2: The Corethrellidae of Borneo: Species richness and acoustic specificity

2.1 Introduction *27*

2.1.1 Biogeography and life history 27

2.1.2 Acoustic specificity and Southeast Asian species diversity 29

2.2. Materials and Methods *32*

2.2.1 Sampling habitat and localities 32

2.2.2 Acoustic lures 33

2.2.3 Collecting off frogs 34

2.2.4 DNA amplification and sequencing 35

2.2.5 Sequence alignment and analysis 38

2.3 Results *40*

2.3.1 α- and β- diversity of *COI* and morphotypes 41

2.3.2 Estimates of species richness and species turnover 43

2.4 Discussion *46*

2.4.1 Corethrella species diversity 46

2.4.1 COI and morphotype conflict 47

1.5.1 Hearing capacity and specificity in *Corethrella* 48

1.5.1 Ecological interactions and the extinction crisis 50

2.5 Conclusion *51*

Chapter 3: Do sepsid species with wide distributions in Southeast Asia contain cryptic species?

3.1 Introduction *53*

3.2. Materials and Methods *56*

3.2.1 Collection and identification 56

3.2.2 DNA extraction, amplification, sequencing and alignment 57

3.2.2 Pairwise and phylogenetic analysis 59

3.3 Results *61*

3.3.1 Dataset 61

3.3.2 Sepsid population tree 63

3.4 Discussion *73*

3.4.1 Cryptic species and reporting bias 73

3.4.2 Widespread species and population structure 76

3.4.3 Synanthropic introduction alongside domesticated ruminants 77

3.4.4 Recolonisation and genetic drift 78

3.5 Conclusion *80*

Chapter 4: From 'cryptic species' to integrative taxonomy: sequences, morphology and behaviour support the resurrection of *Sepsis pyrrhosoma* (Diptera: Sepsidae)

4.1 Introduction *82*

4.2. Materials and Methods *85*

4.2.1 Collection, rearing and morphology 85

4.2.2 DNA sequences 86

4.2.3 Phylogenetic analyses 87

4.2.4 Observations of mating behaviour 88

4.2.5 Determination of reproductive isolation 88

4.3 Results *89*

4.3.1 Morphology 89

4.3.2 Molecular data 92

4.3.3 Behavioural observations and reproductive isolation trials 94

4.3.4 Taxonomic conclusion 96

4.3.5 Species re-description 97

4.4 Discussion *102*

4.5 Conclusion *106*

Chapter 5: Morphology and DNA sequences confirm the first neotropical record for the holarctic sepsid species *Themira leachi* Meigen, 1826 (Diptera: Sepsidae)

5.1 Introduction *108*

5.2. Materials and Methods *108*

5.3 Results *109*

5.4 Discussion *111*

Overall Conclusions **114**

References **119**

Appendix **138**

SUMMARY

In my thesis, I test and demonstrate the utility and limitations of DNA sequences in species richness estimation, the identification of cryptic species, and the confirmation of widespread species.

In my first chapter, four datasets of differing taxonomic groups and hierarchical rank are used to test the congruence and consistency of *COI* sequence-based species richness estimation. Two datasets came from coleopteran families, 1 from the dipteran Sepsidae, and 1 large dataset for all Metazoa was downloaded from Genbank. Species richness estimation based on DNA sequences and identification by taxonomic experts yielded very similar results while richness estimates usually differ greatly when parataxonomists and taxonomists are asked to evaluate the same samples. The boundaries of DNA distance-based delimitation and traditional species are often in conflict.

In the second chapter, I use the techniques validated in the first chapter to estimate the species diversity of the Corethrellidae in Borneo. I test for species specificity in the phonotacic response of the flies towards synthetic pulsed tones and frog calls, but find no evidence for host specificity. The sampled and estimated α-diversity of corethrellid flies are both very high for the main field site and exceeds the species diversity of all studies of corethrellid diversity in the Neotropics.

In the third chapter, I use *COI* to test for cryptic species in eight sepsid species with wide distributions in Asia. The species were sampled from 37 localities in 14 countries. I determine that all but one species are likely to be genuinely widespread with low intraspecific variation between populations. The exception, *Allosepsis indica* (Wiedemann, 1824) is likely to consist of at least six species, although the morphological differences between the species is continuous. In the other seven species, I determine population structure and rule out the hypothesis that movement of domesticated cattle secondarily introduced sepsids throughout Southeast Asia.

In the fourth and fifth chapter, I use *COI* as supplementary information for taxonomic problems that remained unresolved after morphological study. I contributed to the discovery of a cryptic species by detecting an unexpected pattern of pairwise distance in specimens of *Sepsis flavimana* Meigen, 1826 that was indicative of two species. Further investigation revealed a cryptic species, *Sepsis pyrrhosoma* Melander & Spuler, 1917, which was previously synonymised with *S. flavimana*. The species status was further substantiated with reproductive isolation and behavioural data. In the fifth and final chapter, I use *COI* to confirm a surprising new record for the sepsid species *Themira leachi* (Meigen, 1826). Specimens of what turned out to be *T. leachi* were collected from Sierra Cristal National Park, Cuba, 3,500 kilometres away from their previously known southernmost locality of Newfoundland, Canada. *COI* provided an independent source of data to confirm the species and

identification and to rule out the existence of a cryptic species at the Neotropical locality.

I generated 819 sequences of mt-*COI* in total for all analyses in two families of Diptera, the Sepsidae and Corethrellidae, at an average of 548 bases per sequence.

LIST OF FIGURES

Figure	Description	Page
2.1	♀, morphotype I *COI* Cluster K (Table 2.4), darkfield image taken with the Visionary Digital Imaging System, courtesy Yuchen Ang.	41
2.2	*Corethrella* species accumulation curves for Belait district	44
3.1	Consensus maximum parsimony tree for *A. indica*. Clusters A-F are denoted with corresponding forelegs of male *A. indica*, showing the morphological continuum	63
3.2	Consensus maximum parsimony tree for *A. frontalis*	64
3.3	Consensus maximum parsimony tree for *A. niveipennis*	65
3.4	Consensus maximum parsimony tree for *M. fasciculatus*	66
3.5	Consensus maximum parsimony tree for *P. plebeia*	67
3.6	Consensus maximum parsimony tree for *S. coprophila*	68
3.7	Consensus maximum parsimony tree for *S. dissimilis*	69
3.8	Consensus maximum parsimony tree for *S. nitens*	70
3.9	*Sepsis pyrrhosoma* (♂ unless otherwise noted).	91
4.2	Consensus tree of *Sepsis flavimana* group.	93
5.1	Morphology of *Themira leachi* from Cuba	110

LIST OF TABLES

Table	Description	Page
1.1	Relative performance of *COI* clusters to identified species in *Trigonopterus* weevils	17
1.2	Relative performance of *COI* clusters to identified species in the Sepsidae	17
1.3	Relative performance of *COI* clusters to identified species in the Australian Dysticidae	17
1.4	Relative performance of *COI* clusters to identified species in the Metazoan sequences from Genbank	17
2.1	Sampled localities in Brunei	33
2.2	Frequency of morphotypes sorted	35
2.3	List of primers used for amplifying *COI* in this study	38
2.4	Morphotypes and 3%-delimited *COI* clusters. Species in bold denotes collection off the frog. The symbol 'X' represents a pulsed pure tone.	42
2.5	Threshold distances and the clumped/split clusters.	43
2.6	The number and geographical uniqueness of *COI* 3% distance-delimited clusters, which approximate species.	45
3.1	The three datasets of widespread species with their outgroups, which were selected from sister clades according to the phylogeny by (Su et al. 2008)	60
3.2	List of species, the number of specimens sampled, the maximum pairwise distance and the number of clusters for each species at the defined thresholds.	61
3.3	The number of *A. indica* clusters delimited from 2-7%. The number in brackets denotes the number of clusters. Clades A-F refer to the distinct monophyletic *A. indica* groups in Fig 3.1.	62
4.1	Uncorrected pairwise genetic distances between and within and between *Sepsis flavimana* and *S. pyrrhosoma* morphotypes.	86
4.2	Qualitative comparison of behavioural elements observed in *S. flavimana* and *S. pyrrhosoma* (virgin) mating trials.	95
4.3	Results of the hybridisation experiments	96

LIST OF PUBLICATIONS

1. Ang, Y., **Lim, G.S.**, & Meier, R., 2008. Morphology and DNA sequences confirm the first Neotropical record for the Holarctic sepsid species *Themira leachi* (Meigen) (Diptera: Sepsidae). Zootaxa 1933, 63-65

2. Meier, R. & **Lim, G.S.**, 2009. Conflict, convergent evolution, and the relative importance of immature and adult characters in endopterygote phylogenetics. The Annual Review of Entomolology 54, 85-104.

3. Ang, Y., Tan, D.S.H., **Lim, G.S.**, Meier, R., 2009. From DNA barcoding to integrative taxonomy: an iterative process involving DNA sequences, morphology, and behaviour leads to the resurrection of *Sepsis pyrrhosoma* Melander & Spuler 1917 (Sepsidae: Diptera). Zoologica Scripta 39, 51-61.

4. **Lim, G.S.**, Hwang, W.S., Kutty, S.N., Meier, R. & Grootaert, P., 2010. Mitochondrial and nuclear markers support the monophyly of Dolichopodidae and suggest a rapid origin of the subfamilies (Diptera). Systematic Entomology 35, 59-70.

GENERAL INTRODUCTION

In a reply that was published in *Nature*, William T. Astbury reiterated his vision of a molecular biology as “an approach from the viewpoint of the so-called basic sciences with the leading idea of searching below the large-scale manifestations of classical biology for the corresponding molecular plan.” (Astbury 1961). Although primarily focused on the understanding of biology at the cellular level, the molecular biology has indirectly also brought about a revolution in the field of organismic biology. DNA sequencing is the most prominent among the various molecular techniques co-opted by organismic biologists. DNA sequence information has proved useful for phylogenetic inference and population studies, but is now also increasingly used in taxonomy and biodiversity research.

The taxonomic crisis has contributed to the adoption of molecular information for phylogenetic inference, species identification, and species delimitation. Some authors argue that morphological analysis is unprofitable due to reasons such as the slow pace of taxonomic research (Janzen 2004; Tautz et al. 2003; Waugh 2007), chronic underfunding (Lee 2000; Wheeler 2004), systematic marginalisation of taxonomists and taxonomic practice (Giangrande 2003). Furthermore, the urgency brought about by the extinction crisis has engendered broad acceptance of perfunctory alternatives in ecological and conservation studies, such as parataxonomy and taxonomic sufficiency (Maurer 2000; Terlizzi et al.

2003). To this end, DNA barcoding and DNA taxonomy have been proposed as a panacea to these problems. Proponents claim that a ca. 650-base piece of the mitochondrial cytochrome oxidase *c* subunit 1 (*COI*) can solve many problems with species delimitation and identification. This was initially met with considerable scepticism (DeSalle et al. 2005; Hickerson et al. 2006; Lambert et al. 2005; Will et al. 2005; Will and Rubinoff 2004). However, there is now broad consensus that *COI* has great utility in helping to resolve some of the more pressing issues facing organismic biologists today (Moritz and Cicero 2004; Rubinoff 2006; Rubinoff and Holland 2005).

Mitochondrial DNA has emerged as the workhorse of the molecular laboratory, particularly for studies of Metazoa. There are some prosaic reasons for this: mitochondrial sequences are far easier to obtain than nuclear sequences; mt-DNA exists in multiple copies per cell, there are few problems with heterozygosity, mt-DNA evolves faster, the accumulated mutations are largely neutral and can be used for dating (Rubinoff and Holland 2005). Although Roe and Sperling (2007) recommend that *COI* sequence length should be maximised for the purposes of DNA barcoding, Zhang (2007) shows that beyond 200 base pairs, *COI* delimitation success does not improve significantly, a view echoed by (Hajibabaei et al. 2006), making collection of *COI* data from even museum specimens potentially useful.

Here, I explore the use of COI for estimating the species richness of biodiversity samples and for helping to identify and provide support for the diagnosis of cryptic and widespread species.

The first chapter focuses on the ability of *COI* to estimate the species richness in a sample of specimens. I compare the estimate based on of *COI* with the estimate from taxonomic experts. The datasets that are used in this test included aligned *COI* sequences of dipteran Sepsidae, coleopteran Dytiscidae and Curculionidae, as well as the Metazoa. I collaborated with Dr. Michael Balke to generate the sepsid dataset and was responsible for sequencing two-thirds of the 603 sequences. Information on the number of species in a habitat is important for conservation biology but the slow pace of identifying speciemens based on traditional techniques creates many problems. This has created the need for reasonably quick, accurate and cross-comparable way to estimating species richness (Blaxter 2004; Smith et al. 2005; Sodhi et al. 2004). Should *COI*-based estimates compare well to those based on identification by taxonomists, conservation biologists will no longer have to face the taxonomic impediment (Giangrande 2003), especially when dealing with hyperdiverse, understudied taxa.

The second chapter is on the Corethrellidae of Borneo. I generated 356 *COI* sequences from specimens collected in multiple field sites on Borneo. The first chapter revealed that DNA sequences could be used for species richness estimation. In this chapter I use this technique for estimating the species richness of this particularly hyperdiverse and

understudied family of parasitoid Diptera that specialises on feeding on frog blood (Borkent 2008). In the course of my laboratory work, I also devised two alternative methods for rapidly and efficiently extracting DNA from these very small and fragile insects (<2mm) without causing damage or discoloration to the voucher specimen. This is important because my genetic study will have to be followed up with morphological work and all too often voucher specimens are lost during DNA extraction. This is problematic because subsequent visits to the collecting localities often reveal that the habitat has been lost or modified, and new specimens can no longer be collected at the original locality.

In chapter three, I test the prevalence of cryptic species in the widespread Southeast Asian members of the Sepsidae, and demonstrate the dangers of over-generalisation when discussing the prevalence of cryptic speciation. Mitochondrial DNA sequence information can be used to detect plastic, homoplastic or conserved morphology that may confound the identification of species. This has, in part, led to the rapid explosion of studies into cryptic species and speciation, as pointed out by (Bickford et al. 2007). Widespread species are usually suspected of harbouring multiple cryptic species due to potentially long periods of geographical isolation that increase intraspecific morphological and molecular variability (Wiens 1999), possibly to the point where speciation may have occurred. However, in this chapter I reveal that only one out of eight tested widespread species of Sepsidae contains cryptic species.

In chapters three and four, I discover the existence of cryptic species *Allosepsis indica* (Wiedemann, 1824) and *Sepsis pyrrhosoma* Melander & Spuler 1917 through the use of *COI* sequences, illustrating the benefits of collecting and maintaining a comprehensive molecular library for any taxonomic group. By sequencing new specimens as they arrive in the laboratory, I contributed to the taxonomic refinement of Sepsidae by earmarking those specimens that have unexpected genetic signatures. In chapter four, I collaborated with Y. Ang (who described and illustrated the morphology), T. S. H. Denise and M. R. Bin Ismail (who performed the behavioural and reproductive isolation experiments) to identify and resurrect the cryptic sepsid species *Sepsis pyrrhosoma*. The initial observation that led to the resurrection was observing that the *COI* sequences for specimens identified as *Sepsis flavimana* Meigen, 1826 belonged to two distinctly different lineages. This manuscript is now in press in *Zoologica Scripta* (Tan et al. 2009). In chapter three, which has been published in *Zootaxa*, the *COI* sequences I generated for sepsid specimens collected from Sierra Cristal National Park in Cuba revealed that *Themira leachi* Meigen, 1826 is found in the Neotropical region, nearly 3,500 kilometres south of its previously known southern limit of distribution in the New World. I worked with Y. Ang to publish this surprising finding (Ang et al. 2008). Both chapters demonstrate how DNA sequence and morphological information can complement each other.

For this thesis, I sequenced a total of 819 *COI* sequences from two dipteran families, Corethrellidae and Sepsidae in total, performing all the

alignments, pairwise comparisons and phylogenetic analyses on these sequences for all chapters. The format of this thesis will be as follows: In chapters one, two, four and five, I use multiple first-person pronouns where appropriate, as the research and results were performed collaboratively. All chapters address independent issues in biodiversity studies and are intended to evolve into independent publications. I have therefore not written this thesis as a continuous narrative.

CHAPTER 1

Use of the *COI* barcode for species richness estimation

1.1 INTRODUCTION

Charismatic taxonomic groups such as birds, butterflies, mammals and now amphibians have traditionally dominated the study of conservation biology. Although the aesthetic appeal of charismatic groups works better for conservation aims, conservation biologists recognise the vital ecosystem functions that understudied and hyperdiverse groups play (Smith et al. 2009). One of the more visible examples of such groups is the invertebrates, which contain more than 97% of multicellular animal species diversity (Groombridge 1992), and are increasingly becoming conservation priorities because of the high extinction risks (Gaston and O'Neill 2004; Thomas et al. 2004) they face, not least due to the fact that we have very little idea of the true diversity and magnitude of ecological roles they play. Incorporating invertebrate data in biodiversity research is thus one of the most important challenges of modern conservation biology (Myers and Mittermeier 2003; Myers et al. 2000), particularly in conservation management and resource allocation, where species/habitat priorities have to be set and ranked in varying levels of priority. However, the obstacles faced in obtaining useful data on invertebrates are formidable, given that such taxa are often species-rich (Groombridge 1992; Myers and Mittermeier 2003) and have small ranges (Gaston et al. 1998; Trontelj et al. 2009; Zaksek et al. 2009). Specimens need to be collected and prepared before they can be identified. This can be a particularly labour-intensive process. The number of taxonomic experts for most invertebrate groups is small and getting smaller (Wheeler 2004). Given these

problems, high-quality data that can be used confidently to guide conservation priorities are rare and there is a premium on finding novel ways to sort ecological samples to species.

A novel source of data has become available in the form of DNA sequences, which is getting cheaper and faster to produce due to technological advances in sequencing technology. There is broad consensus that DNA will play a major role in how specimens are sorted and described, but to what extent does it replace the traditional taxonomic process is still a matter of some debate (Meier et al. 2008; Vogler and Monaghan 2007). Some authors promote the use of sequences for identifying described species only, *i.e.* DNA barcoding as proposed by (Hebert et al. 2003), while others envision a more significant role such as species identification as well as the determination of species boundaries (Tautz et al. 2003). Many studies have tested the efficacy of DNA sequences against morphology and usually find conflict between the signal provided by DNA and traditional data (Elias et al. 2007; Hickerson et al. 2006; Meier et al. 2006; Meyer and Paulay 2005; Monaghan et al. 2006; Rubinoff and Holland 2005; Wahlberg et al. 2003).

However, there is a distinction between the problems of using DNA (most commonly the mitochondrial cytochrome oxidase *c* subunit I (*COI*)) to identify species, and using it to estimate species richness in biodiversity samples. Does DNA do equally well (or badly) at both? Here, in order to answer this question, we compare the performance of *COI* in species richness estimates with those based on taxonomic expert identification.

In order to be adopted as a new tool for processing and analysing biodiversity samples, the new technique has to be able to outperform traditional methods in terms of equality, speed and cost, or any combination of the three. Currently, the most commonly used technique for determining species richnesss in biodiversity samples is parataxonomic sorting to 'morphospecies', *i.e.* by workers who are not taxonomic experts for the group in question, and may have varying levels of skill and ability in sorting (Basset et al. 2000). Several studies have compared the species richess estimates by taxonomists and parataxonomists for the same samples so as to quantify the quality of sorting by parataxonomists. The most comprehensive review of this by (Krell 2004) analysed 80 studies across a wide variety of invertebrates and found that the mean deviation between expert and parataxonomic species richness estimates was 32%, with a median of 22%. However, the cause for concern should be the extremely high variance in estimate congruence. Species richness estimates between experts and parataxonomists can range from identical to a difference of up to 117%. The accuracy of the estimates is hence unreliable (Abadie et al. 2008; Krell 2004). For 11 of studies, the morphospecies of parataxonomists were also compared to the species sorted by taxonomists. On average, only 69% (the median was 80%) of all species-level specimens were identical. These are the standards that DNA sequence-based sorting must surpass in order to be competitive.

Here, we empirically test four datasets of different hierarchical levels and structure for the utility of *COI* in rapid assessments of species

diversity. For all datasets used in this study, taxonomic experts have already identified the specimens to species before their DNA was extracted and *COI* sequenced. The first dataset consists of 603 sequences for 76 species of Sepsidae (Diptera), sampled from across the distribution of this cosmopolitan family. The second dataset consists of 226 sequences for 50 species of *Trigonopterus* weevils (Coleoptera) collected from one field site in Papua New Guinea. The third dataset consists of 1 140 sequences of Australian representatives of the Dytiscidae diving beetles, covering their endemic Australian distribution almost completely. Lastly, we use a large Metazoan dataset with 35 371 sequences obtained from GenBank to test the generality of our findings.

Various authors have proposed many analytical techniques for delimiting putative species based on their DNA sequences. However, we limit our methods to the objective-clustering algorithm first described in (Meier et al. 2006). This is because of the relatively large size of two of our datasets and the large proportion of singleton species.

The objective-clustering algorithm (part of a DNA pairwise sequence analysis package SpeciesIdentifier (Meier et al. 2006)) uses pairwise distance thresholds to group sequences into clusters. All sequences in a cluster must have at least one sequence in the same cluster with which it has a pairwise distance below the user-defined threshold. Using this technique, we answer four questions in this study. Firstly, can *COI* estimates outdo parataxonomists in terms of quality, speed and/or cost? Secondly, is the species richness of a sample as

determined by a taxonomist similar to the species richness estimate determined by distance-based delimitation of DNA sequences? Thirdly, how congruent are the DNA sequence clusters with those of traditionally recognized species? Finally, we compare the results of different datasets for their consistency in the first two questions.

1.2 MATERIALS AND METHODS

1.2.1 Taxon and character sampling

We use four aligned *COI* datasets in this study: Sepsidae (Diptera), Curculionidae and Dytiscidae (Coleoptera) and Metazoa.

The first dataset consists of 603 sequences for 76 sepsid species, out of the *ca.* 300 described species. The samples came from multiple localities in four continents, excluding Antarctica. 48 species in the dataset had at least one conspecific sequence. All sepsids were identified using morphology by taxonomists R. Meier and Y. Ang. In order to obtain DNA sequences from specimens preserved in varying conditions and periods of time, we used a variety of molecular techniques. For some specimens, we used a direct PCR approach that eliminates formal DNA extraction and purification procedures. We removed flies from the storage tubes where they were preserved in 90-100% ethanol, and blotted them briefly on paper towels just long enough to drain off excess alcohol. The moist specimens were transferred into individual tubes of 8-well strip PCR tubes containing the master mix. In order to improve PCR success, we added 1μL of dissolved bovine serum albumin (Sinopharm Chemical Reagent Co. Ltd.,

Shanghai, China at a concentration of 70μg/mL. BSA neutralizes PCR inhibitors that may have leached out from the tissues of the flies.

For other specimens, we used a direct DNA extraction method. This method was most suitable for sepsids of moderate size (most *Sepsis* species). The flies were placed in 50 – 80 μL of TE buffer (10mM TRIS-CL, 0.5mM EDTA pH 9.0) that have been eluted into 8-well strip PCR tubes or 96-well plates. The tubes/plates were placed into a thermocycler and heated to 95^{o}C for 15 minutes. During the heating, cells break down and release sufficient genomic DNA that it can be used as template for PCR or genomic amplification. The latter ensures that template DNA remains stable for long periods of time. I used 2 – 3 μL of DNA-enriched buffer to each reaction. The relative content and concentration of the reagents in the master mix are identical to those used in direct PCR. For particularly large species (*Themira*) and orange-coloured specimens (*e.g. Australosepsis* males, *Allosepsis indica*, *Sepsis nitens*), we extracted DNA from the left hind leg instead of the entire specimen.

The PCR reactions were prepared in 25μL reactions containing 0.1μL TaKaRa ExTaq (Kyoto, Japan), 2.5μL 10X buffer and 2μL 2mM dNTP mixture, which were also provided by TaKaRa, 1.22μL 10μM primer for both forward and reverse direction, and 16μL DNase-free sterile RO water (1st Base Pte. Ltd, Singapore, Singapore). The primers used are specified in (Lim 2007), with shorter primers being designed and used when the specimen was old and/or had been stored in suboptimal conditions. Cycling temperatures were: 95°C for three minutes to activate

the hot start polymerase, followed by 34 cycles of 95°C for 30 seconds (denaturation), 50°C for 30 seconds (annealing), and 72°C for one minute (extension). The amplification products were kept on hold at 15°C until they were retrieved for gel electrophoresis to confirm that the COI fragment had been successfully amplified. Five μL of the reaction mix was loaded into 1% agarose gel for this purpose. Amplified products were purified using Bioline SureClean (Randolph, MA) and suspended in DNase-free water (1st BASE Pte. Ltd., Singapore, Singapore). Terminator sequencing reactions were then performed in both forward and reverse directions in 10μL volumes, using BigDye ver. 3.1 (Applied Biosystems, Foster, CA) used according to manufacturer specifications. A final purification was performed with Agencourt® CleanSEQ® kit (Agencourt Bioscience Corporation, Beverly, MA) before carrying out direct sequencing in an ABI PRISM® 3100 Genetic Analyzer (Perkin Elmer Applied Biosystems, Norwalk, CT). Sequences were edited and concatenated in Sequencher, before being aligned in ClustalX 2.01 (Thompson et al. 1997).

The second dataset of *Trigonopterus* weevils was published by (Riedel et al. 2009) and comprises 226 sequences from specimens collected off foliage and leaf litter along a transect (300-1520m) of the Cyclops Mountains in Papua New Guinea. These have been identified using morphological techniques by A. Riedel.

The third dataset comprises of 1 140 sequences for 195 species of Australian Dytiscidae, representing the epigean species diversity that have

been sampled as part of a continent-wide, comprehensive study of several endemic radiations The specimens were identified using traditional techniques by taxonomists L. Hendrich and M Balke. Two fragments of *COI* were amplified, the front (5') half with forward primer LCO-1460 5' GGT CAA CAA ATC ATA AAG ATA TTG G 3' and reverse primer HCO-2198 5' TAA ACT TCA GGG TGA CCA AAA AAT CA 3' from (Folmer et al. 1994), using a PCR annealing temperature of 50 - 55ºC. The back (3') half was sequenced using primers C1-J-2183 5' CAA CAT TTA TTT TGA TTT TTT GG 3' (forward) and L2-N-3014 5' TCC AAT GCA CTA ATC TGC CAT ATT A3' (reverse) (Simon et al. 1994).

The last and biggest dataset originally comprised 49 000 metazoan *COI* sequences downloaded from GenBank and aligned (details in (Meier et al. 2008)). Selecting for all conspecific sequences with < 300 bp overlap yielded a final dataset of 35 371 sequences representing 10 772 metazoan species, with 4 599 species having at least one conspecific sequence.

1.2.2 Alignment and analysis

Different techniques were used to align the sequences in the different datasets, but all alignments were protein-translatable and gap-free. The Metazoan dataset was the sole exception. In this dataset, sequences were aligned based on their amino-acid translations (Meier et al. 2008).

All datasets were analysed in SpeciesIdentifier, part of the TaxonDNA ver. 1.5 alpha12 package

(http://code.google.com/p/taxondna/downloads/list) (Meier et al. 2006). All datasets were analysed at four different user-defined distance thresholds, from 1 – 4%. After each clustering analysis, SpeciesIdentifier provides a summary containing the number of clusters, specifications of the sequences within each cluster and their pairwise distances relative to all other sequences in the same cluster, as well as three output files that contain 1) The clusters that contain all the sequences of one species, *i.e.* congruent clusters in agreement with traditional taxonomy. 2) Multiple clusters where sequences for the same species has been split, *i.e.* split clusters. 3) Clusters that contain sequences of more than one species, *i.e.* lumped clusters. Some clusters were both split and lumped, with some of the sequences from a species A clustering together with sequences of another species B. In this scenario, species A has been split into multiple clusters, while species B has been lumped together with species A.

1.3 RESULTS

1.3.1 Congruence between DNA and taxonomic species richness estimates

There was a very high level of congruence in species numbers determined by taxonomic experts and the number of *COI* clusters, especially at the commonly utilised thresholds of 2% and 3%. These two thresholds resulted in species estimates deviating less than 10% from the number of species quantified with taxonomic methods (Tables 1.1-1.4).

Curculionidae	**1%**	**2%**	**3%**	**4%**
Cluster / Species	129.6%	114.8%	109.3%	109.3%
Perfect / Species	60.0%	77.4%	84.7%	84.7%
Lumped / Species	0.0%	0.0%	0.0%	0.0%
Split / Species	40.0%	22.6%	15.3%	15.3%

Table 1.1: Relative performance of *COI* clusters to identified species in *Trigonopterus* weevils

Sepsidae	**1%**	**2%**	**3%**	**4%**
Cluster / Species	132.9%	101.3%	97.4%	93.4%
Perfect / Species	46.5%	60.5%	60.5%	60.5%
Lumped / Species	13.2%	15.8%	14.5%	13.2%
Split / Species	57.9%	25.0%	22.4%	19.7%

Table 1.2: Relative performance of *COI* clusters to identified species in the Sepsidae

Dytiscidae	**1%**	**2%**	**3%**	**4%**
Cluster / Species	132.0%	96.4%	92.8%	87.2%
Perfect / Species	60.3%	72.3%	71.3%	68.7%
Lumped / Species	7.9%	8.2%	9.2%	10.2%
Split / Species	64.5%	15.9%	12.3%	8.2%

Table 1.3: Relative performance of *COI* clusters to identified species in the Australian Dysticidae

Metazoa	**1%**	**2%**	**3%**	**4%**
Cluster / Species	118.7%	106.6%	99.9%	94.5%
Perfect / Species	77.9%	80.9%	79.2%	75.9%
Lumped / Species	4.9%	4.9%	5.8%	6.7%
Split / Species	36.0%	20.8%	14.9%	11.8%

Table 1.4: Relative performance of *COI* clusters to identified species in the Metazoan sequences from Genbank

Different datasets attained estimation optima at different thresholds, with Sepsidae (Table 1.2) and Dytiscidae (Table 1.3) having greatest congruence at 2% threshold, while the Curculionidae (Table 1.1) were still oversplit by *COI* at 4%. The Metazoan dataset (Table 1.4) showed very close matching between cluster and taxonomic species number at 3%, at 99.9% the estimates were close to identical. In general, with a 2%

threshold, species richness estimates based on cluster differ by 1.3-14.8%, with a mean deviation of 6.7% from those based on taxonomic identifications. At 3%, the congruence is 0.1-3%, with a mean deviation of 4.79% from taxonomic identification.

The deviation grew much more severe when either higher (4%) or lower (1%) thresholds were set (Tables 1.1-1.4) Predictably, setting a high threshold depressed cluster numbers by causing more sequences to lump together into single clusters, while setting a very low threshold inflated them by splitting up sequences within a cluster into multiple clusters.

1.3.2 Congruence between taxonomic species and COI clusters

There was much higher conflict between *COI* and taxonomic experts when it came to agreement of identity. At 2% and 3% thresholds, only 60.5-80.9% of the sequence clusters were congruent with species as circumscribed by taxonomists. Problematic clusters either did not contain all the sequences for a taxonomic species thus generating split clusters, and/or contained sequences from multiple species, otherwise known as lumped clusters. A higher threshold caused many closely related species to be lumped together, and a lower threshold caused splitting of clusters that belonged to species with high intraspecific variability. There were more split clusters than lumped clusters in all datasets at all thresholds.

1.4 DISCUSSION

1.4.1 The relative performance of DNA and parataxonomy

Our analyses of four *COI* datasets show that in every case, species richness estimates at a 3% pairwise distance threshold are within 10% of the estimate made by a taxonomist sorting the same samples. This answers the first question we asked about whether *COI* can outdo parataxonomy in terms of quality, speed and/or cost. The difference is obvious, since parataxonomic sorting was found to have a mean deviation of 32% (and a median of 22%) out of 80 studies. Contrast this to the results from our four datasets, with a mean deviation of 5.9% at the 3% threshold, and 4.2% at the 2% threshold. *COI* outperforms parataxonomy by a factor of at least 5.

Krell considers the staggering inconsistency in the quality of sorting to be one of the most serious problems in parataxonomy (2004), as some samples were identical in species richness estimates when compared to taxonomists, while the results differed by up to 117% for other samples. The large range is likely to have led to the large standard deviations summarised in Krell's review (2004). This inconsistency makes generalising or comparing results across parataxonomy-based studies unreliable. Our *COI* datasets however, have a standard deviation of 0.03% to 3.3%, suggesting that DNA sequence-based estimates are much more predictable. However, due to the limited number of datasets in our study, we cannot be sure that our findings are definitive and general across

differing groups. Two studies that address related questions are those conducted by (Smith et al. 2005) and (Borisenko et al. 2008). Smith *et al.* tested the performance of *COI* clusters or molecular operational taxonomic units (MOTUs) in a biodiversity survey of Malagasy ants (2005). Initial sorting to genus level was conducted by parataxonomists, recognising 90 morphospecies from 280 specimens in total. 268 individuals from these morphospecies were successfully sequenced for *COI*, to find 126 MOTUs at 2% clustering, and 117 MOTUs at 3% clustering. Hence, there is 71% and 76% congruence between morphospecies sorting of specimens by parataxonomists and MOTUs at 2% and 3% thresholds respectively. While this may not seem too wildly off the mark, comparisons between collection sites suggested that morphospecies sorting tend to lump similar species and consequently underestimated the β-diversity of species. In the other study, Borisenko *et al.* (2008) trapped mammals in Suriname and compared field identifications with those retrieved by DNA barcoding. The mammal species richness estimates between taxonomic experts and DNA sequences were very similar (74 species versus 73 DNA clusters).

Hence, by making the relative comparisons of performance of parataxonomy versus DNA barcodes and taxonomy versus DNA barcodes, it seems clear that the quality of estimates for species richness is better for sequence-clusters, compared to sorting by parataxonomists. However, there are other more prosaic concerns, such as speed and cost. Unfortunately, these factors are much more difficult to predict across different studies. For instance, some biodiversity samples are

predominantly composed by a few very common species. Molecular analyses of such samples would be very expensive and time-consuming. In such cases, parataxonomists can do the job far more cheaply and efficiently. In other samples, sorting by morphology may be much more labour-intensive and time consuming, making molecular sequencing a more efficient alternative. It is likely that most studies in the future will use some combination of both techniques. In taxa that are more easily identified by morphology and/or difficult to obtain due to CITES regulations (generally the larger charismatic animals), morphology will suffice in their identification; groups that are more intractable in terms of identification by parataxonomists will become the domain of sequence-based sorting. For groups or subsets of samples that are generally unambiguous in their morphology, a small subsample per species should be included for molecular assessment. Sequences from the subsampled specimens can be used to confirm the morphospecies sorting. This strategy of subsampling from pre-sorted samples will likely be necessary for most studies due to the expensive and sometimes time-consuming nature of DNA sequencing (Riedel et al. 2009; Smith et al. 2005). With regard to cost, while technological process has lowered reagent and procedural costs considerably, manpower cost required to handle specimens is still very high (Meier et al. 2008). For instance, the process of vouchering and tissue extraction is difficult to automate, and furthermore, raw sequences produced still need to be processed by trained workers. Thus, we believe that the estimate of USD 10 per specimen will not decrease in the near

future. This implies that processing a biodiversity sample of 10 000 specimens will require a molecular sequencing and analysis budget of at least USD 100 000.

1.4.2 Congruence between DNA cluster content and species

Our analyses show that DNA clusters do not perform nearly as well in determining species identity as recognised by taxonomic experts as they are at estimating species richness. The sobering reality is that a very large proportion of DNA clusters conflict with the species boundaries determined by taxonomic experts (Meier et al. 2006). Congruence is only observed for 60.5-80.9% (an average of 72.8% +/- 8.92%) for all clusters at 2% in our datasets, while at 3%, congruence ranges from 60.5-84.7% (an average of 73.9% +/- 10.52%) for all datasets (Tables 1.1-1.4). The remaining 20-40% of clusters are incongruent with traditional taxonomy because of they either lump multiple species together, or split a single species into multiple clusters, each containing only some of the sequences, or both. Overall, lumping was much less common than splitting, increasing as the defined-distance threshold increased. Given the moderate level of congruence between clusters and species, it is perhaps surprising that *COI* manages to produce such close estimates of species richness. However, it is not difficult to imagine that the 'counting error' of lumped species (underestimation) and split species (overestimation) clusters partially cancel each other out.

Does parataxonomy manage to delimit species boundaries more effectively than *COI*? A review of the evidence shows this is not the case. For 11 samples, Krell was able to provide comparative information on the species units on the species units of parataxonomists and taxonomists (2004). He found that on average, only 69% (a median of 80%) of the units showed one-to-one correspondence. Hence, it appears that both *COI* clustering and parataxonomic sorting give a roughly similar results. Three-way studies where biological samples are sorted by parataxonomists, taxonomic experts and sequenced would provide particularly useful data for a proper study looking at relative conflict and congruence between all three techniques. Unfortunately, they are currently absent in the scientific literature. We can only speculate on the sources of conflict. Other authors have observed widespread incongruence between traditional species and sequence clusters (Ferguson 2002). It is not surprising that using a distance threshold for delimited sequence clusters cannot usually yield taxonomic species units, given that the variability for *COI* is mostly found in neutral positions of the gene (Roe and Sperling 2007). There are two reasons for this: mitochondrial cytochrome oxidases are usually not a direct target of speciation mechanisms, and secondly, they are under strong stabilising selection. *COI* genetic distances will increase between species that have been separated by longer periods, reflecting the time of divergence. However, given that the most important test of any delimitation technique must be able to accurately distinguish between closely related sister species (Meier et al. 2008), *COI* will very likely fail to resolve

relationships in a significant proportion of cases. The only way threshold-based clusters can be congruent with species is if speciation occurs in a regular, clocklike fashion (Meier et al. 2006). This seems very unlikely to reflect biological reality.

1.5 CONCLUSION

In this chapter, we present evidence that DNA sequences can be used to estimate the species richness in biodiversity samples. To do this, we collected four datasets of aligned *COI* sequences from different taxonomic groups and hierarchical levels: one family from Diptera (Sepsidae), two families from Coleoptera (Curculionidae and Dytiscidae), as well as the Metazoa. The estimate is generally within 10% of the estimate that would be provided by a taxonomic expert. DNA sequence-based species richness estimates also outperform parataxonomy by a wide margin on both accuracy and consistency, making DNA estimates of species richness very attractive. However, other factors such as cost and speed must be taken into account, as must the tractability and feasibility of using morphology to reliably identify specimens in the biodiversity sample in question. Furthermore, there was reasonably strong conflict between *COI* distance-based delimitation in identifying specimens and those performed by trained taxonomists (around 20-40%). The conflict is approximately equivalent between parataxonomy and traditional taxonomy. Three-way studies where taxonomy, parataxonomy and DNA information are available for a single biodiversity sample would be

extremely useful in establishing the relative conflict between all three methods of biodiversity sample processing. It appears likely that competent species-level sorting of specimens will remain the sole purview of trained taxonomists. Species richness estimation however, may be a matter of routine DNA sequencing in the future.

CHAPTER 2

The Corethrellidae of Borneo: Species richness and acoustic specificity

2.1 INTRODUCTION

2.1.1 Biogeography and life history

Of the 14 extant haematophagous families within order Diptera (Lukashevich and Mostovski 2003), those within the suborder Culicomorpha (Amorim and Yeates 2006) were the first major group of arthropods to feed on the blood of vertebrates, *i.e* dinosaurs (Grimaldi and Engels 2005). Culicomorph families and members that have retained vertebrate haematophagy remain one of the greatest challenges to human health because they are vectors for many diseases. Most families have hence been the focus of much research. An exception is the family Corethrellidae Wood and Borkent, 1986, or the frog-biting midges.

The Corethrellidae are distributed pan-tropically with the bulk of species diversity in tropical lowlands, although some species can be found as far north as southeast Canada, and as far south as New Zealand (Borkent 2008). Currently, 97 extant species from *Corethrella* Coquillett, 1902 are known, of which more than half have been recently described in (Borkent 2008). The difficulty in describing these species lies in size and their rarity in malaise traps; corethrellid flies are very small, averaging less than two millimetres in length. The best understood fauna for *Corethrella* are the Nearctic and Neotropical regions (specifically Costa Rica) (Borkent 2008) while there are only two described species from the Oriental region. The first is *Corethrella calathicola* Edwards, 1930 whose type locality is Singapore (Edwards 1930) and the second species is *C. brunnea* Borkent,

2008 that has been described from a single female specimen collected from a Malaise trap in Sarawak (Borkent 2008). It is very likely that most of the species in this group have yet to be discovered, given the small number of specimens that are known and the uneven collecting effort. Specifically, the *Corethrella* diversity in the Oriental and Australasian regions is likely to be extremely high given the large number of frog hosts.

Although it was only recently that the Nematocera was abandoned as paraphyletic with respect to the Brachycera, the monophyly of the Culicomorpha has always been supported (Amorim and Yeates 2006). In 1989, Wood and Borkent removed the monogeneric family Corethrellidae from the Chaoboridae in order to retain the monophyly of the latter. This decision was based on larval morphological characters that suggested a sister group relationship between Corethrellidae and the group consisting of both culicids and chaoborids. Like most of the Culicomorpha, this family is relatively old, and has a Gondwanan distribution (Borkent 2008; Borkent and Szadziewski 1991), with the earliest known fossils dating from the early Cretaceous (Lukashevich and Mostovski 2003). The interfamilial relationships have recently been supported by molecular data from nuclear genes (Bertone et al. 2008).

As indicated by their common name, the Corethrellidae are thought to specialise on frogs, where females are nocturnal telmophagous (feeding from blood pools created by tissue laceration) parasitoids that locate their prey based on the vocalisations of male frogs (Borkent 2008; Camp 2006; McKeever and Hartberg 1980). Although there have been reports of blood

meals of avian or reptilian origin (Williams and Edman 1968), the accuracy of this information has been disputed (Borkent 2008, but see Camp 2006) for comments on barcoding host identity through cytochrome oxidase b (*Cytb*) sequences). There are also species known (*c.f. Corethrella mckeeveri*, McKeever & Colless, 1991), or thought to be (*c.f. Corethrella alba* Borkent, 2008) autogenous, based on the reduced character of their mandibular teeth. Laboratory raised adults of some species also demonstrated that some species are capable of facultative autogeny (*c.f. C. appendiculata* Grabham, 1906 and *C. ranapungens* Borkent, 2008) (Borkent 2008). Corethrellid larvae live in small pools or at the margins of larger water bodies with little to no disturbance, preferably vegetation-rich. They can also be found in phytotelmata, stagnant water collected in tires, or within the pitchers of some species of *Nepenthes* (Mogi and Yong 1992). The larvae are predators and cannibals, feeding on rotifers, nematodes and culicomorph larvae (Borkent ; Cresswell 2000; Mogi and Chan 1996). They also exhibit surplus killing behaviour (Lounibos et al. 2008), although it is not clear why. It has been suggested that the practice of killing is a form of kin selection and reduces competition.

2.1.2 Acoustic specificity and Southeast Asian species diversity

Only after the discovery that *Corethrella* exhibits phonotaxis towards their frog prey did a useful collecting method become available to study the group (Borkent 2008; McKeever 1977). Finding phonotactic behaviour in Corethrellidae was initially surprising because such behaviour is particularly challenging for small insects due to the small difference in

time and amplitude between the first incidence of detection of sound in the left and right hearing receptor. This difficulty is compounded with the limitations in rate of neuron signalling, as well as the challenging acoustic environment in which corethrellids live. Some research is now beginning to uncover the extent to which *Corethrella* is able to localise the host based on acoustic signals (Bernal et al. 2006), as well as the evolutionary implications of the costs of predator-eavesdropping (Bernal et al. 2007; Bretman and Tregenza 2007). Unfortunately, an explanation for the mechanism and physiology of the excellent hearing ability of *Corethrella* has remain elusive, but it is likely that the females in this group are using their antennal Johnston's organ to listen for their prey, similar to how male mosquitoes use theirs to locate conspecific females (Gibson and Russell 2006; Gopfert and Robert 2000).

The study of peat swamps echoes that of the Corethrellidae, with both underappreciated and unevenly studied (Ng et al. 1994; Page 2002). In 1992, an IUCN wetlands programme workshop on integrated planning and management of tropical lowland peatlands concluded that the tropical peat swamps of Southeast Asia urgently needed more research (Rieley and Ahmad-Shah). This situation has been addressed to some extent for some parts of Southeast Asia, not so for others (Page 2002). Studies have flatly contradicted prior assumptions about peat swamp habitats being depauperate in species, mostly in blackwater fish species in the peat swamps of Peninsular Malaysia (Ng et al. 1994; Beamish et al. 2003). However, the richness of species assemblages within peat swamps varies

widely (Beamish et al. 2003; Page 2002). Furthermore, no agreement has been reached with regard to how to define peat swamp forests and it is likely that different forests in Southeast Asia that are currently categorised as peat swamp are ecologically distinct.

The two objectives of my study were to 1) Uncover the species diversity and 2) acoustic specificity behind the Corethrellidae in the peat swamps of Brunei and other localities in Borneo.

Due to the very small size of the *Corethrella*, their morphology was relatively difficult to observe. Proper taxonomic work requires that the specimens be dissected and individually slide mounted, which was not possible given the large number of specimens and my lack of expertise in culicomorph dissection and morphology. In order to circumvent the problems of using only parataxonomy to delimit my samples, I chose the barcoding fragment of mitochondrial cytochrome *c* oxidase I (*COI*) as proposed in the previous chapter. The main advantage of *COI* is that sequences are available for many species and that the gene is easily amplified. Furthermore the gene is variable enough to distinguish relatively recently diverged lineages. The challenge was to extract *COI* information from the flies that preserves the specimen for future slide mounting and species description where necessary. While there is a great range of extraction methods available (Nishiguchi et al. 2002), many of them may destroy or damage the specimen. Although non-destructive DNA extraction protocols have been reported, they too were problematic here either because they require the perforation of the insect's cuticle or leg

removal (Rochlin et al. 2007), which is difficult to perform on the *Corethrella* specimens or involve immersing the specimen for hours in a solution of proteolytic enzymes, which renders the specimen fragile, colourless and useless for future taxonomic work. By modifying a protocol by (Grevelding et al. 1996), I devised two non-destructive methods for extracting genomic DNA from individual specimens while preserving the morphological integrity of specimen vouchers for future taxonomic work such as the description of new species, which will be performed by Dr. Art Borkent, who recently published a definitive monograph on the Corethrellidae (Borkent 2008).

2.2 MATERIALS AND METHODS

2.2.1 Sampling habitat and localities

Initial reconnaissance sampling in Singapore (Upper Pierce, behind Singapore Island Country Club) with a *Rana baramica* Boettger, 1901 acoustic lure led to the capture of a single individual. After that, our sampling efforts of the *Corethrella* were focused in Brunei. There, samples of *Corethrella* were trapped with acoustic lures and processed by U. Grafe and H. Ahmadsah from the Universiti Brunei Darussalam. These samples were shipped over preserved in 95% ethanol, already sorted to morphotype. I went to Brunei on two separate occasions for collecting expeditions and to design experiments testing the acoustic specificity of the *Corethrella*, from 17 May to 23 May 2008 and 19 April to 24 April 2009.

While there, I sorted all the specimens I collected to morphotype and preserved them for molecular analysis upon return to Singapore

The habitats sampled were mixed peat swamp (*kerangas* and *kerapa*) forests. Brunei is estimated to have a ground cover of peat land of 13.5% (Rieley and Ahmad-Shah 1992). Sampling of *Corethrella* was conducted in the Belait district along Labi road (114°30'E, 4°35'N), as well as in the Temburong district (115°09'E, 4°33'N) in Brunei. In Sarawak, collection was done in Lanjak Entimau (112° 4'E, 1° 28'N) and Gunung Mulu (113° 55'E, 4° 23'N) forest reserves. Specific localities in Brunei are given in Table 2.1.

District	**Site**
Belait	Labi
Tutong	Tasek Merimbun
Temburong	Sungei Baki, Sungei Esu, Sungei Mata Ikan, Sungei Apan, Ashton trail

Table 2.1: Sampled localities in Brunei

Most of the collecting occurred at the Labi site over a significant period of time. Other sites in Brunei and Sarawak were sampled as single events. All sampling was conducted between the 1900 – 2200 hours. We used two different types of sampling: acoustic lures, as well as collecting flies off the frog hosts.

2.2.2 Acoustic Lures

Prior experiments by Borkent and Camp used acoustic lures of frog calls, (mostly *Hyla gratiosa* LeConte, 1856) (Borkent 2008; Borkent and Belton 2006; McKeever and French 1991) or synthetic calls with dominant

frequencies of 420 Hz (Camp 2006). The upper limit in frequency that *Corethella* can detect thus remains unknown. This is an important issue as many species of frogs have dominant frequencies much higher than those used by Camp (2006). In our study we replayed recorded calls of *Rana baramica*, and *Rana glandulosa* Boulenger, 1882. Synthetic tones of 2.5 kHz were generated in 150ms pulses to test the attraction of *Corethrella* to a simple tone at a frequency that is thought impossible for culicomorphs to detect via their Johnston's organ.

We used Saul Mineroff Electronics (SME–AFS) portable field speakers (Elmont, NY) for acoustic lure trapping, which was set to broadcast at 100dB in a playback loop for 45 minutes per session. These were placed atop BG-Sentinel™ mosquito traps (Regensburg, Germany), right next to the opening. Set-ups were placed in areas of the peat swamp where we could not hear any competing frog calls. Flies that were attracted to the acoustic stimulus would fly towards the speakers, only to be sucked through the opening of the mosquito trap by a vacuum generated by a battery-operated fan. Black mesh bags collected the trapped *Corethrella* and bycatch. The insects were placed into individual plastic screwtop vials.

2.2.3 Collecting off frogs

We also located calling frogs in the undergrowth of the *kerangas*. After finding a frog, we visually inspected it with a dimmed headlamp for *Corethrella*. Once the presence of these parasitoids was confirmed, we

swiftly covered the *Corethrella*-infested frog with a coffee sock or a plastic ziplock bag and swept rapidly around it in a sideways figure-of-eight fashion. Flies were consequently aspirated into plastic vials. All vials and containers were frozen overnight at –20°C before sorting to morphospecies, based on physical variables such as size, shape, colour and the scale patterning on wings and legs.

Sorting to morphospecies was carried out under a light microscope based on physical variables such as size, shape, colour and the scale patterning on winds and legs (Grafe et al. 2008). I obtained 10 morphotypes (I – X) from such sorting of the samples by my collaborators. The morphotypes and quantity sorted are listed in Table 2.2. We consequently labelled and preserved the morphotypes in 95% ethanol, replacing them in -20°C until they were removed for further analysis. In order to test these morphotypes against *COI* information, I subsampled specimens from each type according to their relative abundance for molecular analysis. In total, 432 specimens were chosen for DNA sequencing.

Type	**I**	**II**	**III**	**IV**	**V**	**VI**	**VII**	**VIII**	**IX**	**X**
Freq.	1831	715	74	28	6	3	2	2	2	2

Table 2.2: Frequency of morphotypes sorted

2.2.4 DNA amplification and sequencing

For the DNA sequencing, I developed two molecular techniques that allowed for fast results at low cost, while leaving the specimen intact.

These methods are the direct PCR method and the direct extraction method.

In the direct polymerase chain reaction (PCR) method, PCR reactions are prepared in 22μL reactions containing 0.1μL TaKaRa ExTaq™ (Kyoto, Japan), 2.5μL 10X buffer and 2μL 2mM dNTP mixture (TaKaRa), 1.22μL 10μM primer for both forward and reverse direction, and 13μL DNase-free sterile RO water (1st Base Pte. Ltd, Singapore, Singapore). Instead of template DNA, I removed the flies from storage tubes where they were preserved in 100% ethanol and blotted them briefly on paper towels just long enough to drain off excess alcohol. As fully dried specimens will float on the master mix due to surface tension, they were immediately transferred into 8-well strip PCR tubes containing the master mix. It was not necessary to pierce the cuticule. However, in order to improve PCR success, I also added 1μL of dissolved bovine serum albumin (Sinopharm Chemical Reagent Co. Ltd., Shanghai, China) at a concentration of 70μg/mL. BSA binds with PCR inhibitors that may have remained in the template DNA due to the modified extraction process.

In the direct extraction method, which can also be modified to further extract template DNA from specimens that have been processed via direct PCR, individual *Corethrella* specimens are placed in 15 – 20 μL of TE buffer (10mM TRIS-CL, 0.5mM EDTA pH 9.0) that have been eluted into 8-well strip PCR tubes or 96-well plates. The tubes/plates are placed into a thermocycler and heated to 95°C for 15 minutes. During the heating,

cells break down and release sufficient genomic DNA into the buffer that it can be used as template for PCR or genomic amplification. The latter ensures that template DNA remains stable for long periods of storage. I used 2 – 3 μL of DNA-enriched buffer to each reaction. The relative content and concentration of the reagents in the master mix are identical to those used for direct PCR.

The temperature protocol for PCR was as follows: Three minutes at 95°C followed by 34 cycles of 94°C for 30 seconds, 52°C or 55°C for 30 seconds, and extension of 72°C for one minute. The first cycle of three minutes at 95°C that denatures the antibody that inhibits ExTaq™ also serves to leach sufficient template mtDNA from the flies to produce successful amplifications. Multiple primers were used, as there was differential success between morphotypes. I used a range of primers from different authors, as well as designing my own. The list of primers is provided in Table 2.3. The most useful primers were C1-J-1718/C1-N-2329 (Simon et al. 1994), which amplifies a region of *COI* frame-shifted about a 100 base pairs downstream of the LCO-1460/HCO-2198 (Folmer et al. 1994). After amplification, *Corethrella* specimens were removed from buffer or PCR mix and vouchered in 80% ethanol.

Source	Primer	Sequence	Orientation (set)
(Simon et al. 1994)	C1-J-1718	5' GGA GGA TTT GGA AAT TGA TTA GTT CC 3'	Forward (1)
	C1-N-2329	5' ACT GTA AAT ATA TGA TGA GCT CA 3'	Reverse (1)
	TY-J-1460	5' TAC AAT TTA TCG CCT AAA CTT CAG CC 3'	Forward (2)
	C1-N-2191	5' CCC GGT AAA ATT AAA ATA TAA ACT TC 3'	Reverse (2)
(Folmer et al. 1994)	LCO-1490	5' GGT CAA CAA TCA TAA AGA TAT TGG 3'	Forward (3)
	HCO-2198	5' TAA ACT TCA GGG TGA CCA AAA AAT CA 3'	Reverse (3)
Designed primers	cLCO	5' TTG GAA CTT CTT TAA GTT TA 3'	Forward (4)
	cHCO	5' TCA AAA TAA ATG TTG GTA TAA AAT AGG 3'	Reverse (4)

Table 2.3: List of primers used for amplifying *COI* in this study

Amplification success was checked using agarose gel electrophoresis. Reactions that supplied clear bands were then purified using Bioline SureClean (Randolph, MA) according to manufacturer specifications. The purified product was cycle-sequenced with the same primers for PCR in terminator reactions using BigDye Terminator ver. 3.1 (Applied Biosystems, Foster, CA). A final purification was performed with Agencourt® CleanSEQ® kit (Agencourt Bioscience Corporation, Beverly, MA) before carrying out direct sequencing in an ABI PRISM® 3100 Genetic Analyzer (Perkin Elmer Applied Biosystems, Norwalk, CT).

2.2.5 Sequence alignment and analysis

Sequences were edited, checked for accuracy (e.g., translatable to amino acid sequence), and concatenated in Sequencher 4.0 (Gene Codes Corporation, Ann Arbor, MI). I used ClustalX 2.0.1 (Thompson et al. 1997) with default opening and extension costs for aligning the sequences, which served as my source of molecular data for the study.

As with Chapter 1 of my thesis, in order to meaningfully examine the congruence between morphotype and molecular clusters, I used a range of threshold distances to delimit the *COI* clusters. As elaborated in the previous chapter, this gives an estimate of cluster number that encapsulates the upper and lower bound of species richness. Here I use pairwise distances to delimit clusters, with an objective clustering threshold distance of 3% (Meier et al. 2006). I also delimited clusters at 2%, 4% and 5% to observe how distinct the clusters were from each other.

The idea of objective clustering differs from the tree-based identifications first introduced by Hebert because of its explicit treatment of triangle inequalities (Hebert et al., 2003). Objective clustering allows for threshold violations within clusters as long as every sequence within that cluster possesses a neighbour with pairwise distance below the threshold (Meier et al. 2006). Clusters were consequently used in place of morphotypes to estimate sampling completeness and to discover the α-diversity of *Corethrella* in Labi peat swamp forest.

Sequences were loaded into SpeciesIdentifier (as part of the TaxonDNA package) (Meier et al. 2006), and the objective cluster algorithm used to determine how many clusters there were within the subsampled morphotypes. I used a range of threshold distances (from 2 – 5%) to delimit clusters due to the inherent arbitrariness of the conventional 3% intraspecific distance threshold, as discussed in the previous chapter.

We also conducted a sample-based assessment of species

richness on samples collected from the Belait district (Labi peat swamp), which was subsampled for molecular analysis 77 times, substituting clusters (3% threshold limit) for species. Other areas did not have enough sampling replicates. This was performed in EstimateS ver. 8 (Colwell 2006). The program parameters were set as follows: 50 randomised runs were performed on the dataset without replacement, and the classic formula for bias correction in Chao1 and Chao2 was used based on programme recommendations. All other parameters were set at programme default.

2.3 RESULTS

In addition to collecting *Corethrella* off peat swamp frogs *R. baramica* and *R. glandulosa*, we collected them from rainforest frog species: *Ansonia longidigita*, Inger, 1920 *Megophrys nasuta* Schlegel, 1837, *Metaphrynella sundana* Peters, 1867, *Limnonectes leporinus* Andersson, 1923 and *Philautus hosii* Boulenger, 1895.

Flies were relatively undamaged after DNA extraction and amplification, as can be seen in Figure 2.1.

Figure 2.1: ♀, morphotype I *COI* Cluster K (Table 2.4), darkfield image taken with the Visionary Digital Imaging System, courtesy Yuchen Ang.

Excluding sequencing or PCR failures, I obtained 356 *COI* sequences after subsampling from the 10 morphotypes, with an average of 546 bases per specimen. Aligning these sequences produced a gap-free 654-base alignment. Common morphotypes have been sub-sampled for molecular sequencing more often, in order to test the molecular diversity hidden by morphologically similar specimens.

2.3.1 α- and β- diversity of COI and morphotypes

Clustering at the conventionally established threshold distance of three percent generated 27 clusters, one of which turned out to be a mis-sorted culicid specimen (*Uranotaenia* sp. Lynch Arribálzaga, 1891), also

known to be phonotactically attracted to frog vocalisations (Borkent and Belton 2006). The 26 *COI* clusters have almost doubled the number of morphospecies obtained during the initial sort. Table 2.4 provides information on the morphotypes and genetic clusters.

Cluster	$N_{Individuals}$	Morphotype	Locality	Stimulus
A	1	IV	Labi	*R. glandulosa*
B	1	(Unsorted)	Singapore	*R. baramica*
C	1	VI	Labi	*R. baramica*
D	1	VI	Labi	*R. baramica*
E	1	VI	Labi	*R. baramica*
F	1	VI	Labi	🔉
G	1	VI	Labi	On ***R. baramica***
H	1	VI	Labi	*R. glandulosa*
I	1	X (Culicid)	Labi	*R. baramica*
J	1	II	Lower Apan	On ***M. nasuta***
K	157	I	Labi, Tasek Merimbun	🔉, *R. baramica*, *R. glandulosa*
L	1	(Unsorted)	Sulawesi	*Bufo* sp.?
M	85	II	Labi	🔉, *R. baramica*, *R. glandulosa*
N	1	II	Labi	*R. baramica*
O	13	IV	Labi	🔉, *R. glandulosa*
P	1	IV	Labi	*R. glandulosa*
Q	45	II, III, V, VII	Apan, Ashton Trail, Mulu, Labi, Lanjak Entimau	🔉, *R. glandulosa* On ***L. leporinus***, ***M. sundana***, and ***P. hosii***
R	1	IV	Lanjak Entimau	🔉
S	2	III	Lanjak Entimau	🔉
T	2	III, IV	Labi	🔉
U	16	I, II	Mata Ikan, Lanjak Entimau	🔉 On ***M. nasuta***, and ***A. longidigita***
V	6	II, IX	Baki, Esu, Ashton Trail	On ***M. nasuta***
W	7	I, III	Lower Apan, Baki, Lanjak Entimau	On ***P. hosii*** and ***L. leporinus***
X	6	I	Labi	🔉, *R. glandulosa*
Y	1	VIII	Labi	🔉
Z	1	VIII	Labi	*R. baramica*
AA	1	X	Labi	🔉

Table 2.4: Morphotypes and 3%-delimited *COI* clusters. Species in bold denotes collection off the frog. The symbol '🔉' represents a pulsed pure tone.

Most putative species of corethrellids are rare. Out of the 26 clusters, 10 are represented by single sequences, with two additional clusters being doubletons. The singletons are also conflicting quite strongly with their assigned morphotypes, either being split into multiple distinct clusters, or being lumped in a single cluster. Multiply sampled morphotypes and clusters were equally attracted to the calls of different frog species (K and M), frog calls and pulsed pure tones (K, M, O, Q and X). Hence I did not find *Corethrella* flies exhibiting species specificity in their phonotactic response to different acoustic stimuli.

In order to establish an upper and lower bound for the approximate species diversity of *Corethrella,* I performed objective clustering over a range of threshold distances from 2% - 5%, generating a range of clusters as shown in Table 2.5 below. The relatively lower rate of splitting and clumping is probably due to undersampling.

Threshold	**$N_{Clusters}$**	**Clusters split/lumped relative to 3% threshold**
2%	28	K_1, K_2 & W_1, W_2
4%	24	(U, V), (G, H)
5%	22	(U, V), (G, H) & (Q, S)

Table 2.5: Threshold distances and the clumped/split clusters.

2.3.2 Estimates of species richness and species turnover

All estimators gave us similar patterns, with the trend lines converging very closely to each other. Here we show non-parametric diversity estimators Chao1, Chao2 and ACE and ICE. As discussed in (Colwell 2006), the diversity indices tend to converge as samples

accumulate. The graph (Fig. 2.2) is almost linear, and far from reaching an asymptote, a common phenomenon observed in tropical arthropods (Gotelli and Colwell 2001). Based on the current information, the corethrellid diversity cannot be estimated but will exceed 43 species in the *kerangas* and *kerapa* forests in Labi peat swamp alone.

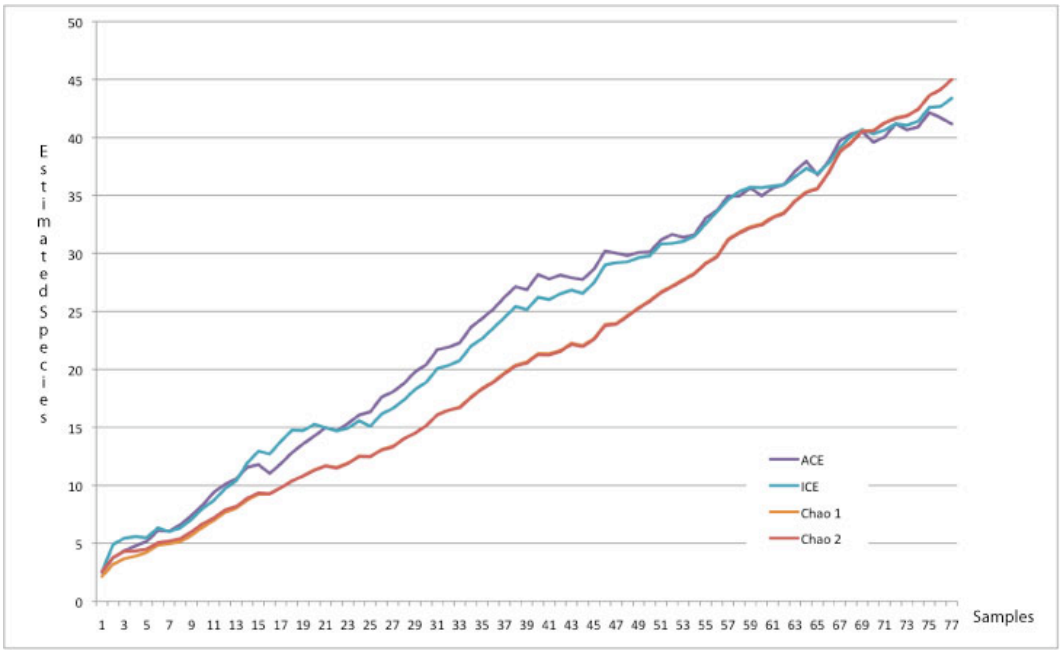

Figure 2.2: *Corethrella* species accumulation curves for Belait district

I could not utilise β- diversity estimators for species turnover due to the very uneven and infrequent sampling events between localities. However, the rate of new clusters obtained per sampling effort was very high, as was the number of unique clusters obtained in each locality. Individuals in cluster Q were the most widespread, having been collected from both Brunei and Sarawak, in two separate localities each (Table 2.6).

Locality	District (No. Sampling Events)	No. of Clusters at 3%	No. of Unique Clusters	Overlaps Cluster?	Shared Locality
Brunei	Belait (77)	18	16	K	Tutong
				Q	Temburong, Lanjak Entimau, Gunung Mulu
	Tutong (3)	1	0	K	Belait
	Temburong (10)	5	3	Q	Belait, Lanjak Entimau, Gunung Mulu
				U	Lanjak Entimau
Sarawak	Lanjak Entimau (10)	5	3	Q	Belait, Temburong, Gunung Mulu
				U	Temburong
	Gunung Mulu (2)	1	0	Q	Belait, Temburong, Lanjak Entimau
Sulawesi	? (1)	1	1	N. A.	
Singapore	Upper Pierce (1)	1	1	N. A.	

Table 2.6: The number and geographical uniqueness of *COI* 3% distance-delimited clusters, which approximate species.

Based on observed species richness, the Belait district of Brunei was the richest. This is most probably due to better sampling, with specimens sequenced from 77 samples. These formed 18 clusters, 16 of them unique to Belait alone (Table 2.6). There is a correlation between sampling effort and diversity, since Temburong and Lanjak Entimau each have specimens in 5 distinct clusters from 10 sampling events, while areas such as Tutong and Gunung Mulu, which were only sampled twice or thrice, obtained specimens that fell into a single cluster. As the Sulawesi and Singaporean specimen were collected opportunistically, they are only significant for being completely distinct from all the Bornean samples.

2.4 DISCUSSION

2.4.1 Corethrella species diversity

Even without thorough taxonomic treatment, the signal from morphology and DNA clearly indicate that the corethrellid species diversity of Borneo, and likely Southeast Asia is orders of magnitude higher than what is currently known. This confirms the preliminary prediction by (Borkent 2008) about Southeast Asia potentially matching the Neotropical region in terms of species diversity. Here I show that the species diversity from acoustic-lure trapping single *kerangas* forest site in Brunei matches that from the most species-rich locality sampled from Costa Rica by (Borkent 2008). He reports trapping 10 species of *Corethrella* from Carara National Park, which is known to have 14 species. We have approximately 18 species from Labi alone (Table 2.4), none of which have been described, although they were recognisably distinct from morphologically similar *Corethrella* species from Japan and the New World, (Borkent 2008; Borkent, pers. comm.). Furthermore, based on α-diversity estimations, there may be twice as many species of *Corethrella* in Labi alone (Fig. 2.2). This works out to at least 43 parasitoid species for a locality that has less than a quarter of this species diversity in anuran host species (Grafe, pers. comm). This is particularly surprising given that estimates of species based on Chao1 and Chao2 are conservative in the face of inadequate sampling (Chao et al. 2009). The degree of species overlap between

habitats is also not very high; the species in the most exhaustively sampled locality of Labi in the Belait district shared only 2 species with other districts in Brunei or Sarawak. This is a further indication of either high species endemism or thoroughly incomplete sampling.

Expanding on anuran diversity outside Brunei, there are currently over 138 anuran species known from Borneo (Inger and Lian 1996) with a flurry of papers describing more species in the past 13 years (Das 2005, 2008; Das and Haas 2003, 2005; Inger et al. 2001; Matsui et al. 2007; Stuebing and Wong 2000). It is thus likely that the anuran diversity of Borneo has not been exhaustively documented. If we extrapolate parasitoid *Corethrella* diversity based on the diversity of their hosts, it is likely that the diversity of *Corethrella* may be even higher than here estimated.

2.4.2 COI and morphotype conflict

Even without sequencing all *Corethrella* specimens, *COI* clusters indicate a much higher level of diversity than suggested by morphospecies sorting. Furthermore, the morphotypes were not reliably sorted because similar morphotypes were split or lumped into different *COI* clusters through the range of selected threshold distances. The unreliability of parataxonomic sorting has been an issue of some concern, especially for biodiversity (Krell 2004) and ecological studies (Bortolus 2008). Inaccuracy may be particularly expected for taxa such as *Corethrella* that are very small with most useful morphological details only being apparent to

taxonomic specialists. This echoes the problems of other groups that are difficult to distinguish because of difficulties in obtaining useful morphological characters (Besansky et al. 2003).

2.4.3 Hearing capacity and specificity in Corethrella

The high rate of unique *COI* sequences led us to hypothesise that this may indicate host-specificity. We therefore modified our sampling methods to include pulsed tones of different frequencies. Although sampling is still incomplete at the time of thesis writing, preliminary experiments on maximally attractive frequencies in Brunei have shown that the *Corethrella* can and do exhibit positive phonotacic behaviour towards pulsed pure tones at 3kHz and higher, although the number of collected specimens is smaller at high frequency. We also cannot detect species specificity in the attraction of *Corethrella* to the range of pulsed tones, partially because most specimens that are collected are singleton. The most abundant corethrellid type that clusters in K and M (Table 2.4) have been collected from Labi in pulsed tone traps spanning the range from 1.5kHz to 5kHz. Furthermore, multiple kinds of acoustic stimuli (calls, pure tones, etc.) can attract individuals belonging to the same cluster (Table 2.4). However, there may be a further host of criteria that a corethrellid female requires before she feeds (Borkent 2008). It is also possible that the *Corethrella* flies may be attracted to frog calls of multiple species, even landing on them, but they refrain from feeding because of the chemosensory cues are wrong (Grafe, pers. comm.). Our observation of non-specificity is in agreement with reports in the literature that indicate

that rock music, zebra finch, and mole cricket songs are effective acoustic lures for neotropical *Corethrella* (Camp 2006).

The closest well-studied example of a dipteran parasitoid that is able to locate her host via acoustic localisation is that of tachinid *Ormia orchracea* Bigot, 1889 (Robert et al. 1996), although it is also known in tsetse ((Tuck et al. 2009) and sarcophagid flies (Robert et al. 1999; Yack 2004). However, these taxa utilise tympanal hearing (similar to the mechanism used by humans), which is currently thought to be the only form of hearing that can reliably detect far-field sounds (Yack 2004). However, *Corethrella*, and all culicomorph midges practice antennal hearing, which is thought to be capable of only detecting low-frequency, near-field sounds (Robert 2005), although it is clear that the call of a *Rana baramica* (dominant frequency ~2.7kHz) is neither low-pitched, nor are the males occurring at a high enough density for their calls to be considered near-field. Furthermore, they are positioned approximately 2 metres above ground (pers. obs.), making the task of localisation more difficult as it becomes a matter of triangulating the location of a small frog approximately 30-70 mm in length in 3-dimensional space. Furthermore, in a 1978 experiment by Mangold, both *Corethrella* and *Ormia* were both attracted to the calls of *Gryllus* crickets (Mangold 1978). A recent paper by Roberts extends the hearing thought possible by mosquitoes over much higher frequencies (Robert 2009), providing some evidence for that the performance of antennal hearing is on par with that of tympanal hearing.

A point of note is the antennal posture of the *Corethrella* as they approach their hosts. Unlike culicids, observations of female *Corethrella* approaching their anuran prey in the field or in field photographs show them to have their antennae extended outwards perpendicular to their body, a phenomenon confirmed by my collaborator (Ulmar Grafe, pers. comm.). This has not been recorded in the literature, but may prove significant with regards to increasing acoustic sensitivity by extending the time differential between the detection of acoustic signal between the left and right antennal receptor.

2.4.4 Ecological interactions and the extinction crisis

Corethrella are also threats to anuran health by being vectors of disease, specifically trypanosomes (Johnson et al. 1993). This is analogous to the human public health crisis of malaria, spread by the disease vector, the culicid *Anopheles gambiae*. This association between *Corethrella*, trypanosomes and anurans is likely to be very old (Borkent 2008; McKeever and French 2000). Sexual signalling costs of anurans include: 1) The exertions of visible display, in this case the energetic expense of sustained vocalisations 2) The threat of a predator 'eavesdropping' on the signal 3) Trypanosome infection by *Corethrella* vectors, which stresses the male immune system, and 4) hypovolemia from a considerable number of *Corethrella* feeding on a small host (see Camp 2006 for a back of envelope estimation). These factors may compound the high level of environmental stress known to accelerate extinction in amphibians (Alford et al. 2007; Pounds et al. 2007; Rohr et al.

2008; Skerratt et al. 2007; Wake and Vredenburg 2008). As specialists on anurans, the Corethrellidae are also in imminent danger of catastrophic extinctions (Dunn et al. 2009; Koh et al. 2004; Sodhi et al. 2004).

2.5 CONCLUSION

The Corethrellidae are a very fascinating taxon with an interesting life history and remarkable phonotactic sensibility. Unfortunately, they remain poorly studied. Here, I report the high diversity of undescribed *Corethrella* species in Borneo, based on sampling in Brunei and to a lesser extent in Sarawak. A large number of undescribed species is confirmed based on morphology and *COI* sequences. The large difference in species numbers based on *COI* and morphotype is an issue of some concern and provides further evidence for the lack of precision in parataxonomic sorting. With preliminary sampling in Sarawak and Brunei generating a surprisingly high species richness estimate (Table 2.6), it can only mean that an exhaustive series of trapping in Borneo will lead to an exponential increase in *Corethrella* species with the species diversity being even higher for Southeast Asia. As a point of natural historic interest, it is a pity that so little is known about a group which seems to be ecologically so varied, as given the current rate of species extinction of anurans, the *Corethrella* are sure to follow suit.

CHAPTER 3

Do sepsid species with wide distributions in Southeast Asia contain cryptic species?

3.1 INTRODUCTION

The phenomenon of cryptic species hiding within species formerly considered to have a widespread distribution has recently attracted much attention (Bickford et al. 2007). Given that most cryptic species are detected based on DNA sequences, technological advances in molecular biology initiated this revolution. Sequences are now routinely used to re-examine species boundaries (see Singh 2003 for review). Sequencing is most frequently used in those cases where morphological information has shown to be inconclusive (Gaubert et al. 2005; Mutanen 2005). It is often the DNA barcode gene region (mitochondrial cytochrome oxidase *c* subunit I (*COI*)) that is used to investigate whether there are deep divergences between groups of sequences formerly thought to belong to one species. Despite all problems with interpreting DNA sequence information for taxonomic purposes, *COI* and other molecular markers are currently the best tool for discovering hidden diversity in taxa that are suspected to include cryptic species (Blaxter et al. 2004; Gomez et al. 2007; Lukhtanov et al. 2009), overwhelmingly diverse taxa (Lin et al. 2009), or reunite life stages and genders of taxa which are extremely dimorphic (Johnson et al. 2009). Chapter 4 will discuss an example of how *COI* helped to discover and resurrect cryptic species, *Sepsis pyrrhosoma* Melander & Spuler 1917 where it was previously synonymised within a morphologically plastic species, *Sepsis flavimana* Meigen 1826. This case is sadly rare, as many published studies that report finding cryptic species do not follow up with formal descriptions.

I will here follow the definition of cryptic species as proposed by Bickford et al. 2007; i.e., that they are multiple species that have been classified as a single nominal species at any point in time because of the difficulty or unreliability in distinguishing them by traditional means. Beyond the inherent value of good taxonomic practice, studies of cryptic species can be used as sources of information for broader scientific purpose. Firstly, cryptic species are useful for phylogeographic and speciation studies (Lohman *et al.*, 2008; Wirta, 2009) because they are closely related enough to trace recent history. Establishing the geographical boundaries of cryptic species originating (Lukhtanov et al. 2009) from within a widespread species with large distributions can also help date the emergence of barriers of reproductively isolate populations (Willett and Ladner, 2009; Zaksek et al., 2009). Secondly, the existence and ecological functions of cryptic species are a significant issue in conservation biology (Garcia-Mudarra et al., 2009; Thum and Harrison, 2009; Vialatte et al., 2008), particularly in the management and preservation of functional ecosystems and genetic diversity (Giangrande, 2003; Milankov *et al.*, 2008). Finally, the detection of cryptic species can be crucial when the species in question are of significance to humans, such as pest species (Collins and Paskewitz, 1996; Vogler et al., 2008), biological control (Beard, 1999; Stouthamer *et al.*, 1999) and medicinal leeches (Siddall *et al.*, 2007). Misidentifying species boundaries can have numerous negative effects.

The frequency of cryptic species in widespread Southeast Asian is

currently poorly undestood. Many studies either exclude Southeast Asia due to lack of data, (such as Moritz et al. 2000, a study on rainforest dynamics), or researchers have resorted to creative methods to obtain specimens from the region; e.g., by sampling ornamental Southeast Asian fish species from the pet trade that were dead-on-arrival (Steinke et al. 2009). Currently, the best-sampled taxa are herptiles, which have been extensively studied (Inger and Voris 2001; Stuart et al. 2006). The lack of study for other groups is unfortunate, given the geological history and events during the Pleistocene and Holocene that have shaped the distribution of Southeast Asia's fauna and flora (Lambeck 2004; Sathiamurthy and Voris 2006; Soares et al. 2008), making it a valuable field site to investigate speciation (Beehler 2001; Holloway 2003). Furthermore, Southeast Asia is a biodiversity hotspot (Brooks et al. 2002), which means that the status and validity of species is an urgent concern for conservation management.

One problem with studying the literature on cryptic species is publication bias. Studies finding new cryptic species are usually published, while research revealing the existence of genuinely widespread tramp species is less likely to be published (see (Balke et al. 2009; Errard et al. 2005; Le Breton et al. 2004) for exceptions). Studies revealing the existence of genuine widespread species lacking cryptic lineages may thus contribute to a 'file drawer' problem (Csada et al. 1996). A publication bias against non-significant results may have implications on our understanding and inference of the mechanisms of speciation and

biogeography. Hence, I propose that regardless of what a study of widespread species reveals, the results should be published.

Here, I study eight species of Sepsidae, a family of synanthropic commensal Diptera. They are small black flies that are commonly found on mammal dung (Pont and Meier 2002). This means that the study of the species boundaries of widespread Southeast Asian members must also compensate for the confounding influence for secondary introduction of sepsid species based on the movement of domesticated cattle, which were domesticated twice independently in Eurasia about 8,000 - 10,000 years ago (Bruford et al. 2003; Loftus et al. 1994). Here, I use DNA barcodes on identified sepsid flies that are widely distributed in this region to test whether they contain cryptic species or are widespread tramp species.

3.2 MATERIALS AND METHODS

3.2.1 Collection and identification

Eight widespread species of sepsids were collected from a total of 37 localities, representing 14 countries in Southeast Asia such as Brunei, Cambodia, Indonesia, Malaysia, Philippines and Thailand as well as Australia, China, India, Pakistan, Papua New Guinea, Taiwan, the Union of the Comoros, and the United Arab Emirates where these widespread species are also found. The species are *Allosepsis indica* (Wiedemann, 1824), *Australosepsis frontalis* Walker, 1860, *Australosepsis niveipennis* Becker, 1903, *Meroplius fasciculatus* Brunetti, 1909, *Parapaleosepsis*

plebeia (Meijere, 1906), *Sepsis coprophila* Meijere, 1906,*Sepsis dissimilis* Brunetti, 1909, and *Sepsis nitens* Wiedemann, 1824 (See Appendix for list of specimens and locality information). Flies were collected off bovine dung by sweep netting before being killed in absolute ethanol. Some females were also retained to set up cultures, in order to perform behavioural observations and mating experiments. The latter was relevant when morphology and *COI* seem to yield conflicting signal. Sepsid taxonomists identified male individuals to species based on morphological diagnostic characters, most commonly found on the sexually dimorphic foreleg.

I also included three unidentified and possibly new species of *Sepsis* Fallen, 1810. In order to test for congruence between morphology and *COI* of the *Sepsis* species, a single-blind experiment was performed on three groups of sepsids with distinct genetic barcodes. A sepsid taxonomist was asked to sort individual specimens preserved in 95% ethanol from five *COI* clusters to species based on morphology. The morphology-based identifications were then compared with the results of clustering.

3.2.2 DNA extraction, amplification, sequencing and alignment

Only individuals of sepsid males or females that could be identified to species (females of genus *Australosepsis* can be identified to species based on the presence of wingspots) were used for molecular analysis. They were extracted using a mix of protocols, including a modified CTAB

protocol (Lim 2007), along with non-destructive methods detailed in chapter 1, based on the protocol I devised in chapter 2. The rest of the specimen was then relabelled and vouchered in 80% ethanol.

The target sequence was the barcoding region of *COI*. Primers were used as in (Lim 2007), with shorter primers being designed and used when the specimen was old and/or had been stored in suboptimal conditions, as they are easier to amplify (Lim 2007). PCRs were performed in 25μL reactions, including 100ng/mL of template DNA and primer, buffer and dNTP concentrations added based on manufacturer recommendations. The only modification was that I halved the amount of TaKaRa ExTaq™ (Kyoto, Japan). Cycling temperatures were: 95°C for three minutes to activate the hotstart polymerase, followed by 34 cycles of 95°C for 30 seconds (denaturation), 50°C for 30 seconds (annealing), and 72°C for one minute (extension). The PCRs were kept at a holding temperature of 15°C until they were retrieved for gel electrophoresis to confirm that the *COI* fragment had been successfully amplified. 5μL of the reaction mix was loaded into 1% agarose gel for this purpose. Amplified products were purified using Bioline SureClean (Randolph, MA) and suspended in DNase-free water (1^{st} BASE Pte. Ltd., Singapore, Singapore). Terminator sequencing reactions were then performed in both forward and reverse directions in 10μL volumes, using BigDye ver. 3.1 (Applied Biosystems, Foster, CA) used according to manufacturer specifications. A final purification was performed with Agencourt® CleanSEQ® kit (Agencourt Bioscience Corporation, Beverly, MA) before

carrying out direct sequencing in an ABI PRISM® 3100 Genetic Analyzer (Perkin Elmer Applied Biosystems, Norwalk, CT). Sequences were edited and concatenated in Sequencher, before they were aligned in ClustalX 2.01 (Thompson et al. 1997). No gaps were observed. The dataset was then used for the following analyses. Balke M contributed half of the sepsid sequences used in this analysis.

3.2.3 Pairwise and phylogenetic analysis

Similar to the method used in the previous chapter on *Corethrella* sequences, I used the objective-clustering algorithm in SpeciesIdentifier to determine how many species were falling into congruent clusters (terminology following (Meier et al. 2006)). The threshold limit set ranged from 2% - 4% for the datasets. I followed up on *A. indica*, which showed abnormally high distances, by defining distances of up to 7%.

For ease of analysis, sequences were split up into three datasets, based on the sepsid phylogeny in (Su et al. 2008) Each dataset also included outgroups that were selected based on Su et al. (2008). The datasets included the species as listed in Table 3.1. The unidentified *Sepsis* species will henceforth be labelled as *Sepsis* sp. A, *Sepsis* sp. B, and *Sepsis* sp. C.

	Dataset #1	Dataset #2	Dataset #3
Outgroup	*Meroplius fukuharai*	*Toxopoda* sp.	*Dicranosepsis emiliae*
Widespread species	*M. fasciculatus*	*A. indica*	*A. frontalis*
		P. plebeia	*A. niveipennis*
			S. nitens
			S. dissimilis
			S. coprophila
Unidentified species			*S.* sp. A
			S. sp. B
			S. sp. C

Table 3.1: The three datasets of widespread species with their outgroups, which were selected from sister clades according to the phylogeny by (Su et al. 2008)

Both Maximum Likelihood and Maximum Parsimony trees (ML and MP trees respectively) can be used to infer whether phylogenetic structure exists between the populations. In conjunction with information on genetic divergence, this would indicate whether the populations of sepsids existed before the introduction of domestic cattle. Parsimony analysis to determine the genetic structure of the populations was performed in TNT ver. 1.1 (Goloboff et al. 2008), under the following settings: New tech search at level 65, best tree to be found 10 times. Support was determined with jackknife resampling under the same settings for 250 replicates. Likelihood analyses were also performed in GARLI (Zwickl 2006), after MrModelTest (Posada and Crandall 1998) considered the GTR + I + Γ to be the best model under AIC and LRT criteria. This model was utilised, along with 10 000 generations. Support was obtained via bootstrap support for 100 replicates under the same settings.

3.3 RESULTS

3.3.1 Dataset

I obtained 270 sequences of all widespread species, with an alignment 733 bases long. The alignment was gap-free and could be translated into amino acid sequence. Most widespread species were split into multiple clusters at 1% (*A. indica*, *A. niveipennis*, *P. plebeia*, *S. coprophila*, and *S. dissimilis*) but fell into single monospecific clusters at thresholds >1% (Table 3.2). *Australosepsis niveipennis* and *S. dissimilis* remained split into 2 clusters at 2%, but both cases were due to very distinct *COI* sequences in specimens from the Union of the Comoros.

Species	**∑ Individuals**	**Max % distance**	**1%**	**2%**	**3%**	**4%**
A. indica	69	16.61	10	6	6	6
A. frontalis	65	2.54	1	1	1	1
A. niveipennis	20	3.33	2	2	1	1
M. fasciculatus	14	1.05	1	1	1	1
P. plebia	15	1.96	2	1	1	1
S. coprophila	13	1.72	2	1	1	1
S. dissimilis	22	3.58	3	2	1	1
S. nitens	29	1.49	1	1	1	1
***S.* sp. A**	6	1.87	2	1	1	1
***S.* sp. B**	6	0.37	1	1	1	1
***S.* sp. C**	11	2.04	1	1	1	1

Table 3.2: List of species, the number of specimens sampled, the maximum pairwise distance and the number of clusters for each species at the defined thresholds.

The groups produced by naïve morphological sorting of the unidentified *Sepsis* specimens were congruent with the results of objective sequence clustering through the range of distances defined, from 1-4%.

The single exception was *S.* sp. A, which was collected in a single locality, but was split into 2 clusters at 1%.

Table 3.3 shows the number of clusters for *A. indica* at a range of thresholds from 2% - 7%. Of the now 11 species, only individuals of *Allosepsis indica* show unexpectedly large pairwise *COI* distances (Table 3.3), at a puzzling magnitude that approaches family-level distances in the Calyptratae (Kutty, pers. comm.). However, there appear to be no distinct morphological differences between the genetically distinct populations, indicating that these individuals may belong to multiple, cryptic species.

Clusters	**2%**	**3%**	**4%**	**5%**	**6%**	**7%**
Split	*A. indica* (6): Clades A, B, C, D, E, F	*A. indica* (6): Clades A, B, C, D, E, F	*A. indica* (6): Clades A, B, C, D, E, F	*A. indica* (6): Clades A, B, C, D, E, F	*A. indica* (5): Clades A+C, B, D, E, F	*A. indica* (3): Clades A+B+C, D+E, F

Table 3.3: The number of *A. indica* clusters delimited from 2-7%. The number in brackets denotes the number of clusters. Clades A-F refer to the distinct monophyletic *A. indica* groups in Fig 3.1.

As both ML and MP trees were congruent in their hypotheses of intraspecific relationships, both likelihood bootstrap and parsimony jackknife support values have been mapped on the strict consensus tree obtained from the parsimony analysis. Support <60 was discarded. These support values, with likelihood bootstrap above the branch, parsimony jackknife below (Fig 3.1 – 3.8) indicate which populations have support. Due to the large size of the datasets #2 and #3, the trees were trimmed to depict each widespread species separately.

3.3.2 Sepsid population trees

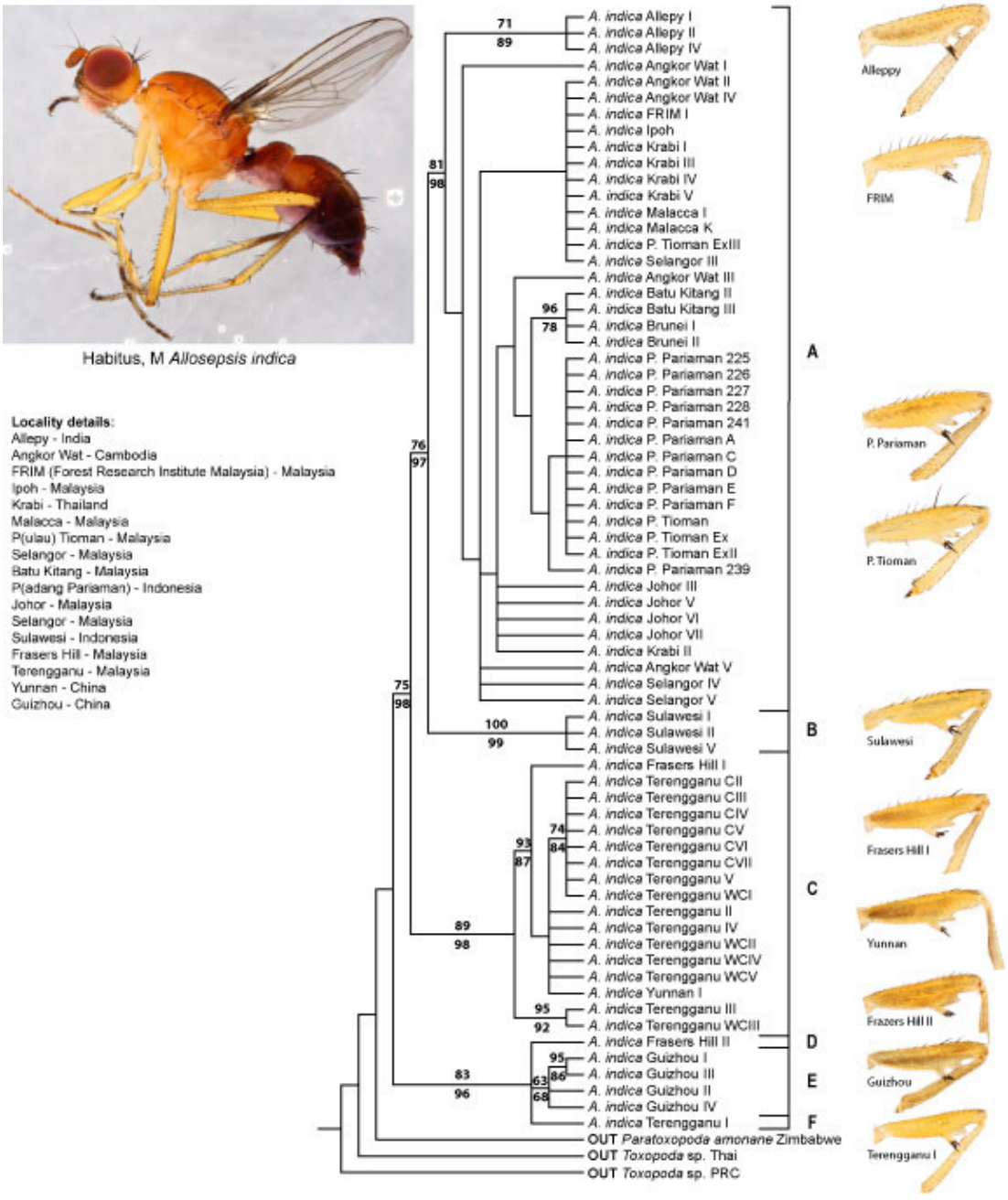

Figure 3.1: Consensus maximum parsimony tree for *A. indica*. Clusters A-F are denoted with corresponding forelegs of male *A. indica*, showing the morphological continuum

Habitus, M *Australosepsis frontalis*

Locality details:
Angkor Wat - Cambodia
Ba tu Kitang - Malaysia
Kalimantan - Indonesia
Bkt.(Bukit) Tingei - Indonesia
GHSNP (Gunung Halimun Salak National Park) - Indonesia
Chiang Mai - Thailand
Ipoh - Malaysia
Malacca - Malaysia
Mt Camagong - Phillipines
Sulawesi - Indonesia
P(ulau) Tioman - Malaysia
Trang - Thailand
Krabi - Thailand
Belumut - Malaysia

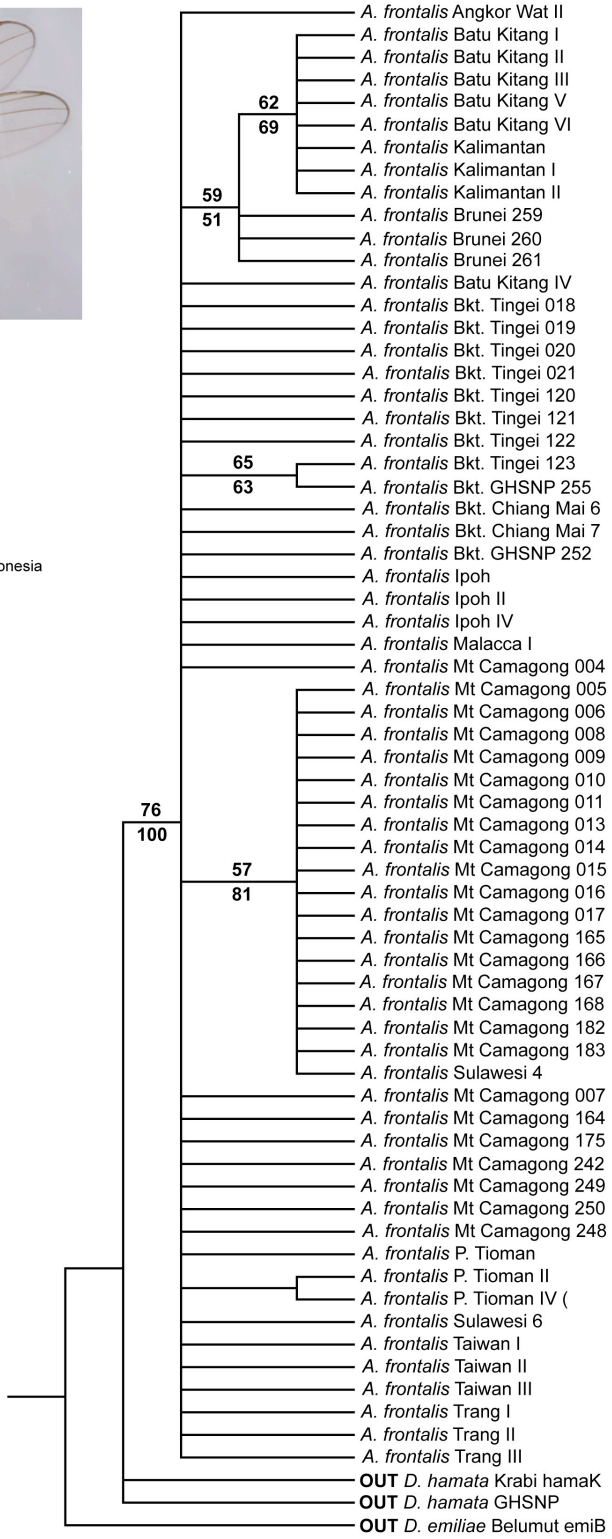

Figure 3.2: Consensus maximum parsimony tree for *A. frontalis*

Habitus, M *Australosepsis niveipennis*

Locality details:
Batu Kitang - Malaysia
NSW (New South Wales) - Australia
Calicut - India
Chiang Mai - Thailand
P(ulau) Tioman - Malaysia
Grand Comoros - Union of the Comoros
GHSNP (Gunung Halimun Salak National Park) - Indonesia
Krabi - Thailand
Belumut - Malaysia

Figure 3.3: Consensus maximum parsimony tree for *A. niveipennis*

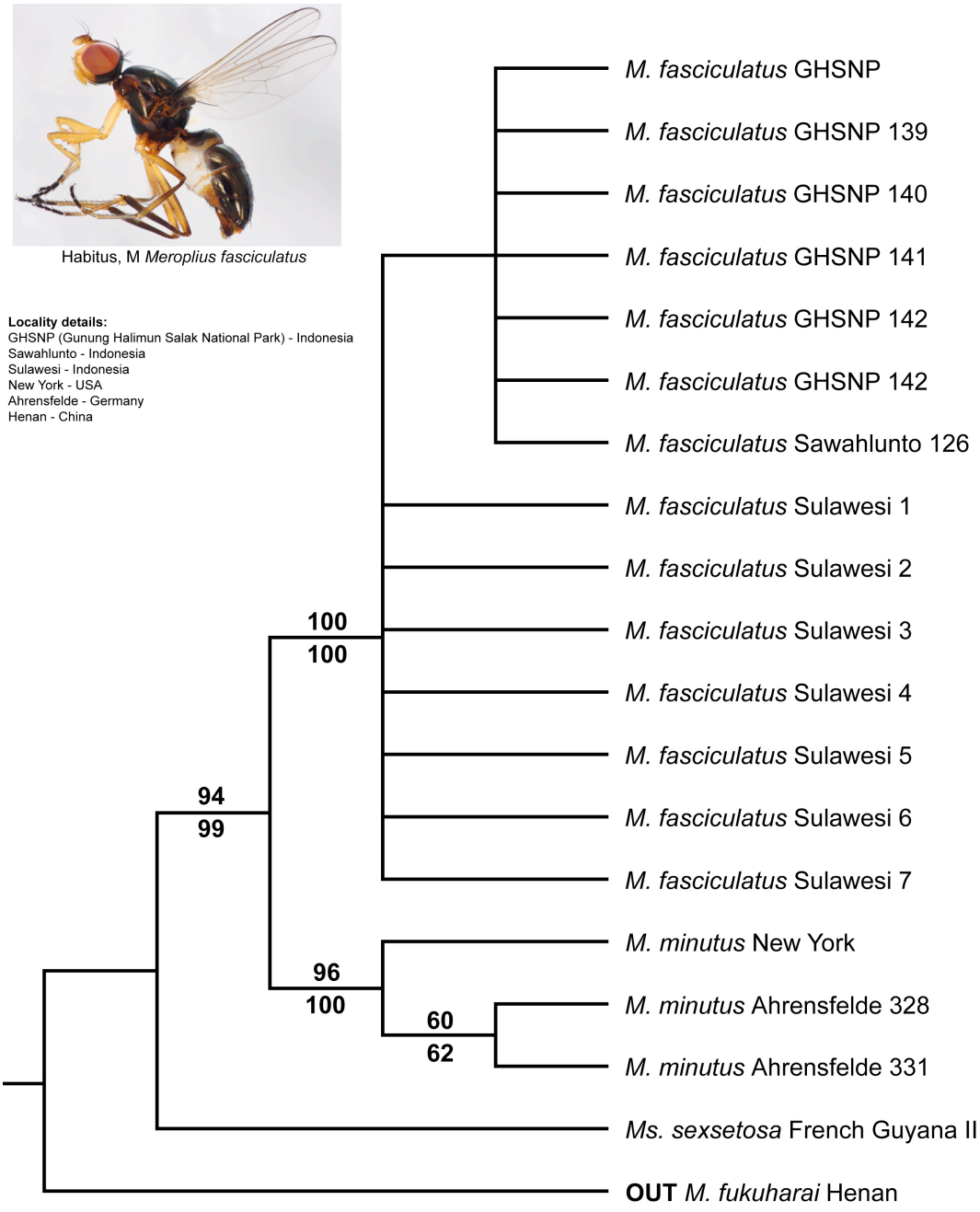

Figure 3.4: Consensus maximum parsimony tree for *M. fasciculatus*

Habitus, M *Parapaleosepsis plebeia*

Locality details:
Goroka - Papua New Guinea
NSW (New South Wales) - Australia
Woollogong - Australia
Perth - Australia

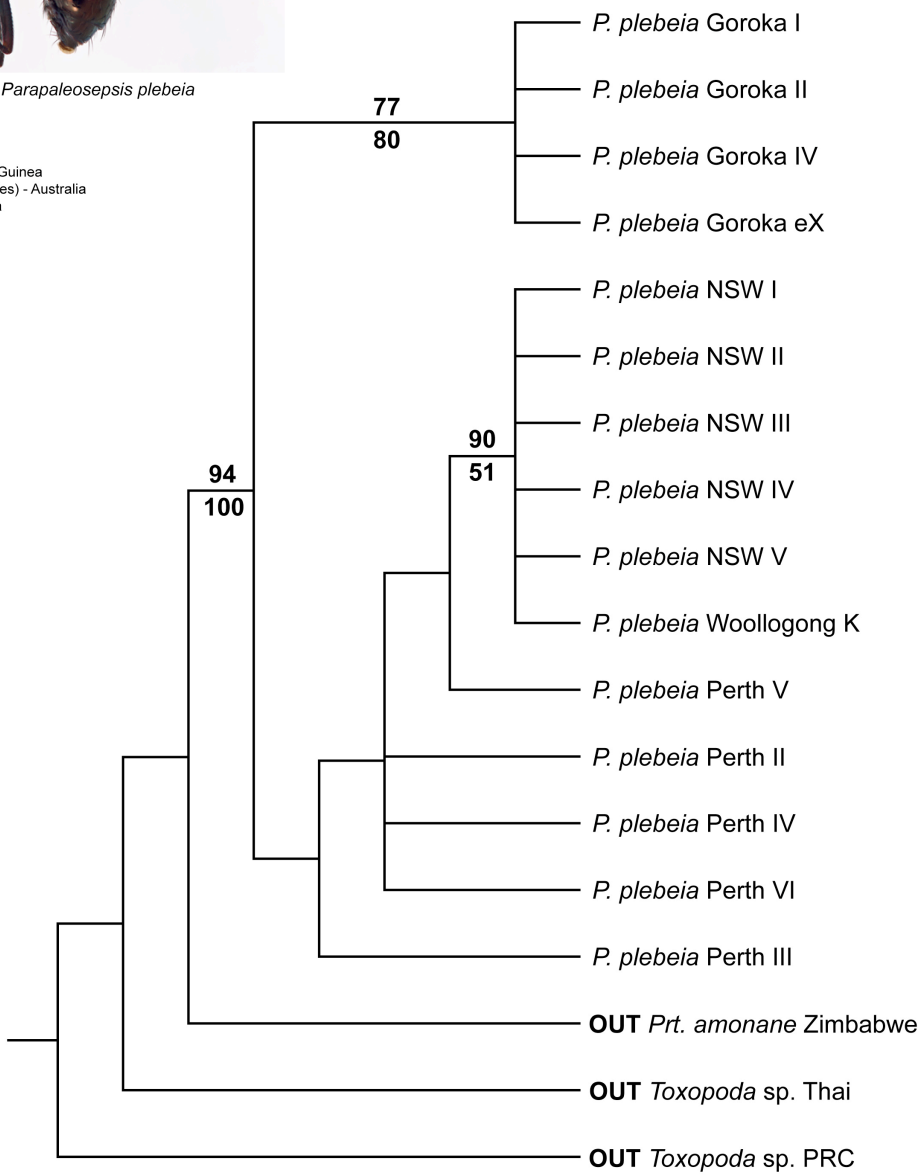

Figure 3.5: Consensus maximum parsimony tree for *P. plebeia*

Habitus, M *Sepsis coprophila*

Locality details:
Bkt.(Bukit) Tingei - Indonesia
Sulawesi - Indonesia
Krabi - Thailand
GHSNP (Gunung Halimun Salak National Park) - Indonesia
Belumut - Malaysia

Figure 3.6: Consensus maximum parsimony tree for *S. coprophila*

Habitus, M *Sepsis dissimilis*

Locality details:
Batu Kitang - Malaysia
Trang - Thailand
Bkt(Bukit) Tingei - Indonesia
P(adang) Pariaman - Indonesia
Chiang Mai - Thailand
Citeureup - Indonesia
Goroka - Papua New Guinea
Ipoh - Malaysia
Malacca - Malaysia
Mt Camagong - Phillipines
Sulawesi - Indonesia
Grand Comoros - Union of the Comoros
Krabi - Thailand
GHSNP (Gunung Halimun Salak National Park) - Indonesia
Belumut - Malaysia

Figure 3.7: Consensus maximum parsimony tree for *S. dissimilis*

Habitus, M *Sepsis nitens*

Locality details:
Bkt.(Bukit) Tingei - Indonesia
Sulawesi - Indonesia
P(adang) Pariaman - Indonesia
Citeureup - Indonesia
Guizhou - China
Tamil Nadu - India
Tanda Saeed - Pakistan
Ipoh - Malaysia
Krabi - Thailand
GHSNP (Gunung Halimun Salak National Park) - Indonesia
Belumut - Malaysia

Figure 3.8: Consensus maximum parsimony tree for *S. nitens*

Refer to Figure 3.1 (page 63): Cluster analysis of *Allosepsis indica* resulted in 6 distinct clades A-F, which were more than 5% apart from each other. This result is recovered in both parsimony and likelihood analyses. There is hence very strong signal from *COI* that multiple cryptic species exist in what is now referred to as *A. indica*. However, as shown by the pictures of male forelegs from representatives of all six groups, there is a morphological continuum. The largest pairwise distance is found between 'Terengganu I' and the Bornean samples (Batu Kitang+Brunei), at 16.61%.

Refer to Figure 3.2 (page 64): *Australosepsis frontalis* has the tree with the poorest resolution. The only geographically isolated group recovered with support is for specimens from Borneo (Sarawak and Brunei). The two most divergent sequences are between 'Sulawesi 6' and Philippine 'Mt Camagong 167', at 2.54%.

Refer to Figure 3.3 (page 65): Unlike its sister species, *A. niveipennis* has much more geographical structure, with support for both Australian (New South Wales) and Taiwanese populations. All *A. niveipennis* from Asia and the Middle East (UAE: United Arab Emirates) were sister to an individual collected from the Comoros islands, off the eastern coast of Africa. This individual was also more than 2% apart from all other specimens causing the split clusters in Table 3.2. The two most divergent sequences are between Australian 'New South Wales A' and the sequence 'Comoros III', at 3.33%

Refer to Figure 3.4 (page 66): Although parsimony analysis retrieved a monophyletic group of *M. fasciculatus*, there was no support for population differentiation between the Sumatra+Java specimens, and those from Sulawesi. The two most divergent sequences are 'Sulawesi 7' and 'GHSNP 141' from West Sumatra, at 1.05%.

Refer to Figure 3.5 (page 67): Populations of *P. plebeia* from two populations from southwest (Perth) and southeast (New South Wales) and Papua New Guinea show very defined structure. Although the separation of haplotypes between the two landmasses was recovered by phylogenetic analyses, there was no support. However, there was support for the isolation of Papuan and New South Wales populations. The sequences with the greatest pairwise distance are New South Wales IV and Goroka I, at 1.96%

Refer to Figure 3.6 (page 68): *Sepsis coprophila* exhibits the strongest population structure between populations, with strong support for unique population haplotypes between Sumatra and Sulawesi, in contrast to *M. fasciculatus*. The two most divergent sequences are 'Sulawesi 1' and all the samples 'Bukit Tingei' from West Sumatra, at 1.72%.

Refer to Figure 3.7 (page 69): Support for *S. dissimilis* can be found for the populations on West Sumatra (Bukit Tingei + Padang Pariaman), as well as the distinctly isolated group from the Comoros islands. This is similar with the results for *A. niveipennis*. The Comoros specimens of *S. dissimilis* were also responsible for the split clusters in Table 3.2 at 2%

threshold limit, as with *A. niveipennis*. There is also support that links haplotypes from India (Tamil Nadu) and southern Thailand (Trang). The two most divergent sequences are 'Grand Comoros I' and West Sumatran 'Bukit Tingei 050', at 3.58%

Refer to Figure 3.8 (page 70): *Sepsis nitens* specimens from China (Guizhou), India (Tamil Nadu) and Pakistan (Tanda Saeed) were recovered together with support. The other supported cluster comes mainly from one locality in West Sumatra (Padang Pariaman), with the inclusion of one specimen from another locality on from the same region. The two most divergent sequences are between 'Sulawesi 1' and 'Padang Pariaman 244'.

In general, while most sepsids exhibit geographical stratification to a greater (*P. plebeia*, *S. coprophila*) and lesser (*A. frontalis*, *M. fasciculatus*) degree, there is very little consistency in haplotype isolation in geographic localities between species.

3.4 DISCUSSION

3.4.1 Cryptic species and reporting bias

In this study, I selected seven species of sepsids that have a species range throughout Southeast Asia, along with *P. plebeia*, which has an Australasian distribution. I find that out of the eight species, only *A. indica* has multiple, cryptic species within what was previously considered to be a single, widespread species. The seven other sepsid species are

morphologically relatively uninform and fall into species-specific clusters when a comprehensive *COI* dataset is analyzed. There is hence no evidence for the presence of cryptic species.

Allosepsis indica is possibly one of the easiest sepsid species from Southeast Asia to identify. Both males and females are usually the largest sepsids found on ruminant dung, with yellowish-orange bodies (although smaller individuals may be brownish-black). The males also possess very distinct foreleg morphology, specifically a large protrusion with 4 spines arranged radially around the tip (see Fig. 3.1). Due to their very distinct appearance, I did not initially suspect that there would be multiple cryptic species within *A. indica*. However, molecular data for populations from throughout Southeast Asia indicate that there may be at least six species, represented by extremely distinct molecular clusters (See Fig. 3.1). Each cluster is at least 5% away from all others (Table 3.3). Could *COI* information be misleading about species boundaries in this species? There is evidence from reproductive isolation experiments using live cultures from representatives of 4 out of the 6 clusters to support the signal from DNA (Denise S. H. T., pers. comm.). This suggests that for *A. indica,* molecular data are more indicative of species status (under the reproductive species concept) than morphology.

DNA provides an independent and discrete source of information for comparative analysis with morphology. While other sources of information, such as reproductive isolation, acoustic information and behaviour may also be useful in distinguishing species, they are generally much more

difficult to obtain. This has been the main factor leading to the dramatic upswing in the number of published studies on multiple cryptic species from single, supposedly widespread species for taxa where non-morphological signals are used for intersexual communication. While valuable, these studies are likely leading to a publication bias that gives the impression that widespread species routinely contain cryptic species.. This may lead to an overrepresentation of the strength of the scientific phenomenon in question. The pervasiveness and potential ramifications of this problem has been discussed in the fields of medical and ecological research, especially with regard to its impact on downstream meta-analyses and systematic reviews that attempt to uncover greater patterns and trends by evaluating published studies (Jennions and Mÿller 2002; Silvertown and McConway 1997).

Two recent meta-analyses that have been published clearly illustrate the pitfalls of attempting to discover trends without careful consideration. In 2007, (Pfenninger and Schwenk 2007) conducted a meta-analysis based on 2207 studies of cryptic metazoan species sourced from the Zoological Record™ database from 1978 – 2006. Based on the patterns revealed by their data, they concluded that cryptic species were likely to be distributed equally in all biogeographical regions and taxonomic groups, directly contradicting the conjectures made by (Bickford et al. 2007). This startling result was refuted recently, by (Trontelj and Fišer 2009), who reveal serious methodological errors in the analyses by Pfenniger and Schwenk (2007). Upon reanalysis after correcting for study

intensity and taxon size, they found that the proportion of cryptic species varied up to two orders of magnitude. They further elaborated that cryptic diversity can arise out of a multitude of factors specific to the genus and environment. Even with reanalysis, the authors cast doubt on the accuracy of the conclusions of meta-analysis, given the very recent increase in the number of studies and the bias toward studies conducted in temperate regions. I would argue that publication bias is another factor that needs to be taken into consideration.

3.4.2 Widespread species and population structure

The widespread sepsids did not show strong or consistent intraspecific population structure. This is likely due to small intraspecific pairwise *COI* distances and the lack of haplotype sorting between geographically distinct populations. For the most part, there was only weak population-level signal within the species, usually with one or two populations being distinct from the rest; *e.g.*, *A. frontalis*: the Bornean population. Two examples of species with very distinct populations are *A. niveipennis* and *S. dissimilis*. The distinct populations are both from the Union of the Comoros, which is off the coast of East Africa. This immediately suggests that the genetic distances are due to the extremely long period of geographic separation between the African and Southeast Asian populations. This contrasts with a study conducted on another widespread synanthropic dipteran species, *Musca domestica*, which exhibits surprising spatial diversity and geographic structure (Cummings and Krafsur 2005).

3.4.3 Synanthropic introduction alongside domesticated ruminants

The widespread sepsid species examined in this study are predominantly found throughout Southeast Asia on pats of dung left by domesticated cattle. The larvae feed on the dung thus providing a potential mechanism of human-assisted dispersal. This is testable by generating a phylogeny based on *COI* sequences to detect whether individuals from geographically distant populations are genetically undifferentiated. This can be observed in the species *A. frontalis* and *S. dissimilis*. Both have been collected from a range of localities and show no recoverable geographical populations with the exception of a Bornean group in *A. frontalis* (Fig. 3.2), and a distinct clade in the Comoros islands for of *S. dissimilis* (Fig. 3.7).

Other species *A. niveipennis*, *M. fasciculatus*, *P. plebeia, S. coprophila* and *S. nitens*, exhibit at least two or more isolated populations with bootstrap/jackknife support (Figs 3.3-3.6, 3.8). This is consistent with long periods of geographic separation and may indicate that there were already pre-existing native populations of these sepsids before the introduction of domesticated ruminants. The earliest record of locally domesticated Southeast Asian cattle is recorded at 3,500 years ago (Mohamad et al. 2009), only later adapting to exploit their dung as a food source. Furthermore, there are native cattle throughout Southeast Asia such as the banteng, gaur, anoa and kouprey (Lenstra and Bradley 1999), which may have been the original source of breeding substrate for these sepsids. Interestingly, there are no native ruminants in *P. plebeia*'s species

range, although it shows strong population segregation between parts of Australia and Papua New Guinea (Fig 3.5). However, this species is known to be a generalist and may have bred in marsupial dung prior to the introduction of ruminants.

3.4.4 Recolonisation and genetic drift

Positing genetic segregation of populations in geographically distinct locations rests on the assumption of independent evolution of *COI* haplotype over long periods of spatial isolation, such as by genetic drift and/or local selection events (Scheiner 1993). Over longer time scales, it is easy to imagine an allopatric speciation event arising from such a scenario, otherwise known as the vicariance model of speciation, as proposed by (Nelson and Platnick 1981).

However, the geological history of the region is complicated and there have been long periods with fluctuating sea levels during the Pleistocene and Holocene (Lambeck 2004; Sathiamurthy and Voris 2006). This implies that land bridges between different islands were formed at every glacial maximum, allowing geographically distinct populations to re-colonise each other. This periodic admixture of populations would eliminate haplotype sorting built up over periods of isolation while sea levels are high. This could also explain the lack of differentiation in most populations, while genetic bottlenecks or strong positive selection events in local habitats would generate a monophyletic *COI* haplotype shared among all or most members of specific populations of sepsid species.

In our study, most widespread sepsid species turn out to be taxonomically sound single species (*A. frontalis*, *A. niveipennis, S. coprophila*, *S. dissimilis*, *S. nitens*, *M. fasciculatus* and *P. plebeia*). Only one species contains multiple cryptic species (*A. indica*). Searching for cryptic species in the Sepsidae is complicated by the fact that many sepsid species also exhibit considerable morphological plasticity based on the environment (Pont and Meier 2002). This indicates that in the case of cryptic species, only integrative and iterative studies incorporating elements of morphology, DNA, reproductive isolation and behaviour are capable of distinguishing between truly cryptic species (Castroviejo-Fisher et al. 2009; Fonseca et al. 2008; Tan et al. 2009) and discovering their biogeographic and evolutionary history (Garcia-Mudarra et al. 2009; Mathews 2006; Padial et al. 2009; Padial and de la Riva 2009; Trontelj et al. 2009; Wirta 2009).

Negative results are often not published. In the case of widespread species, a reverse search of the literature to that done by (Vialatte et al., 2008) in ISI Web of Science (http://scientific.thomson.com/products/wos) and Zoological Record Plus (http://www.csa.com/factsheets/zooclust-set-c.php) using the phrases 'widespread species' or 'tramp species' and excluding 'cryptic species' in title, abstract and keywords found only 1337 references, even when timespan was set to all years. This is a third of the number of cryptic species studies published in the last fifty years.

3.5 CONCLUSION

The rapid rise in publications on cryptic species can be linked to the adoption of DNA sequencing technology by evolutionary biologists (Bickford et al. 2007). Cryptic species are interesting as test cases for biogeographic and speciation hypotheses; however there may be an overemphasis on the existence of cryptic species, possibly due to a publication bias where non-significant findings (i.e., widespread species are truly widespread) are not reported. I test this using *COI* in 7 widespread Southeast Asian and 1 Australasian species in Sepsidae, a synanthropic commensal family of Diptera, and find evidence that only nominal species *A. indica* is actually composed of multiple cryptic species.

CHAPTER 4

From ‘cryptic species’ to integrative taxonomy: sequences, morphology and behaviour support the resurrection of *Sepsis pyrrhosoma* (Diptera: Sepsidae)

4.1 INTRODUCTION

DNA sequence data have recently gained much popularity in taxonomic research and it is generally acknowledged today that they provide important evidence for delimiting species (Meier et al. 2008; Vogler and Monaghan 2007). DNA data can now be generated at a fast rate, with relatively low cost, and by personnel lacking the taxon-specific knowledge required for morphological research (Lee 2000; Scotland et al. 2003; Vogler and Monaghan 2007). However, increasingly the widespread use of DNA sequences has also created problems in the form of so-called 'cryptic species' that are now routinely proposed whenever morphology and DNA sequence evidence – at least initially – yield different inferences about species boundaries (Bickford et al. 2007). The use of the term 'cryptic species' implies that the unit is already properly diagnosed as a species. However, this is rarely so and in most cases a resolution of the conflict between morphology and DNA sequence information is not even attempted. As a consequence, such 'cryptic species' are accumulating in the literature and interfere with a proper classification and the assessment of biodiversity. Here we demonstrate how an iterative process based on multiple sources of data can move a 'cryptic species' from being only a putatively new species-level taxon to being formally recognized as a species based on sufficient evidence (see also: (de Leon et al. 2006; Dorchin et al. 2009; Gomez et al. 2007; Mehdiabadi et al. 2006; Page et al. 2005; Petersen et al. 2007).

It is sometimes assumed that this process only requires enough data, but this is not necessarily the case, as the same data may yield different species inferences under different species concepts (Denise et al. 2008; Laamanen et al. 2003). Many authors avoid this issue – presumably due to the vitriol related to species concept discussions. However, it is precisely when data are in disagreement that it is important to be explicit about species concepts, because in these cases species concepts can matter (Tan *et al.* 2009). Here we suggest that the best solution is applying a two-step process: one can first evaluate the available data based on the species concept that is favoured by the authors. Afterwards, the same data can be discussed under alternative species concepts (Laamanen et al. 2003). This approach will ensure that species are clearly defined given that the authors' opinion based on their species concept will be binding under nomenclatural rules. At the same time the treatment is transparent and allows proponents of alternative species concept to draw their own conclusions.

Most species in entomology are recognized based on morphological characters. Sepsid flies are no exception, but the use of morphology for some species can be problematic because of the bewildering amount of phenotypic variability present in this family (Pont and Meier 2002). In sepsids most of this variability is related to environmental factors, such as the amount of food available to the larvae (Meier 1995). In these cases DNA sequences are particularly useful for clarifying species boundaries, because the sequences are not affected by the environmental variables. In

other cases the observed intraspecific variability is at least partially genetic (Reusch and Blanckenhorn 1998). Here, DNA sequences can still be used as additional evidence, but any observed sequence variability across allopatric populations can be difficult to interpret (Ang et al. 2008; Memon et al. 2006; Petersen et al. 2007) because recently diverged species can share barcodes and may thus be incorrectly lumped into one species (Meier 2008; Meier et al. 2006). Similarly, allopatric populations within old species may have distinctly different sequences and DNA evidence may erroneously suggest that they should be split into multiple species (Meier 2008; Meier et al. 2008).

Here we demonstrate the value of an iterative approach using multiple sources of data by clarifying the species boundaries of *Sepsis flavimana* Meigen, 1826. Such use of multiple data sources for confirming cryptic species remains rare in the literature (but see (de Leon et al. 2006; Dorchin et al. 2009; Gomez et al. 2007; Mehdiabadi et al. 2006; Page et al. 2005; Petersen et al. 2007). As with many similar cases in recent literature (Bickford et al. 2007), our taxonomic problem started with finding unexpectedly high levels of *COI* divergence between what appeared to be allopatric populations that were collected from various locations in North America. Based on recently published identification keys (Ozerov 2000; Pont and Meier 2002) these specimens all keyed out to one species, *S. flavimana*. This particular species is one of the most morphologically variable sepsids, with much of its variability related to size (Mayden 1999; Munari 1983; Pont and Meier 2002). Not surprisingly, this species has

spawned a large number of synonyms (Ozerov 2005). Among others (Ozerov 2000)) synonymised four Nearctic species with *S. flavimana* when revising the North American fauna (*S. vicaria* Walker, 1849, *S. pyrrhosoma* Melander and Spuler, 1917, *S. melanopoda* Duda, 1926 and *S. kertezsi* Duda, 1926).

However, the unexpectedly high level of genetic variability that we found within the North American populations of what appeared to be *S. flavimana* motivated us to re-investigate the morphology in order to test whether these genetically distinct populations may also be morphologically distinct. As additional sources of data, we were also able to study the mating behaviour and test for reproductive isolation based on cultures that we had established for two genetically distinct populations from North America and Europe. We then apply all these data to four species concepts that represent the main categories of concepts in the literature (Wheeler and Meier 2000); i.e., those based on reproductive isolation, monophyly, diagnosability, and on a mixture of criteria. We argue for the use of reproductive isolation as a criterion for determining the number of species. However, we also discuss the number of species that would be obtained under competing concepts.

4.2 MATERIALS AND METHODS

4.2.1 Collection, rearing and morphology

Sepsis '*flavimana*' specimens were collected from six American populations (from Raleigh, NC, New Orleans, LA, New York, Palmyra, VA,

and Athens and Dyar Pasture, GA) and stored in 100% ethanol for subsequent morphological and genetic study (Table 1).

	S. flavimana **(North America)**	***S. flavimana*** **(Europe)**	***S. pyrrhosoma*** **(North America)**
S. flavimana **(North America)**	0.00-0.52%		
S. flavimana **(Europe)**	1.69-2.87%	0.00-1.70%	
S. pyrrhosoma **(North America)**	6.14-7.04%	6.17-7.65%	0.00-1.62%

Table 4.1: Uncorrected pairwise genetic distances between and within and between *Sepsis flavimana* and *S. pyrrhosoma* morphotypes.

In addition, live specimens from New Orleans (LA, USA), Kevelaer (NRW, Germany), and Ahrensfelde (Schleswig-Holstein, Germany) were reared in laboratory cultures using sucrose syrup as a carbohydrate source and cow dung as a breeding substrate. Cow dung was initially frozen at -80˚C for several days to kill any insects infesting the dung prior to collection. Fly cultures were maintained at 25-28˚ C in 2*l* plastic containers. Compound microscopy and high-fidelity microscopic photography (Visionary Digital ™ BK+ system using a Canon EOS D1 Mark III fitted with Infinity Optics K2 Long Distance Microscope on CF4P3 objective settings) were used to study the morphology of specimens from all eight localities in detail.

4.2.2 DNA sequences

We amplified and sequenced a 778bp fragment of cytochrome oxidase c subunit I (*COI*) including the DNA barcoding region from 50

individuals representing multiple populations of five nominal *Sepsis* species (*S. biflexuosa*, *S. duplicata*, *S. flavimana*, *S. fissa*, and *S.* '*pyrrhosoma*') with *S. fissa* designated as outgroup based on (Su et al. 2008). Genomic DNA was extracted from tissues using a modified CTAB extraction protocol (Shajahan 1995). The cycling conditions for the PCR started with an initial denaturation at 95^{o}C (3 mins), followed by 5 cycles of 1 min at 95^{o}C, 1 min annealing at 44^{o}C, 1.5 min extension at 72^{o}C and 30 cycles utilizing an annealing temperature of 48^{o}C. As primers we used mtd8: 5' CCA CAT TTA TTT TGA TTT TTT GG 3' and mtd12: 5' TCC AAT GCA CTA ATC TGC CAT ATT A 3'. All sequences were aligned with CLUSTALX 2.01 (Thompson et al. 1997) and the alignment was free of indels.

4.2.3 Phylogenetic analyses

Maximum likelihood and maximum parsimony were used to infer the gene-tree for the *COI* of *S.* '*flavimana*' populations and related species. A new technology parsimony search was implemented in TNT 1.1 (Goloboff et al. 2008)with search level 55; the minimum tree length was found 10 times. Node support was assessed through jackknife resampling, with absolute frequency differences and 36% character deletion for 250 replicates. A maximum likelihood bootstrap tree was obtained with GARLI 0.951 (Zwickl 2006). Using the Akaike Information Criterion (AIC), MrModeltest (Nylander 2004) selected the GTR + Γ + I model for *COI*. The analysis was automatically terminated if the log likelihood did not improve by 0.01 or more after 50,000 generations. Support was obtained as

maximum likelihood bootstrap with 250 replicates under the same settings.

4.2.4 Observations of mating behaviour

Virgin flies were obtained from each culture by isolating a petri dish of larvae-infested dung from the laboratory colony in an empty container and segregating males and females within six hours of ecclosion. Sepsid flies, at least in the *flavimana* group, acquire sexual maturity after two to five days (pers. obs.). Flies were thus assumed to be sexually mature after five days as adults. To examine and compare behavioural elements between populations, one virgin male was introduced to a 3.5cm plastic petri dish containing a single virgin female, and the behaviour of both flies was recorded at 7X – 15X magnification with an analogue video recorder attached to a trinocular Leica MZ16A microscope. Recordings began upon introduction of both flies and ended either after successful copulation or after 60 minutes if copulation did not occur. The analogue recordings were then digitised and analysed frame-by-frame (25 frames per second) using the video editing software Final Cut Pro (Apple Computer, Inc. 2005). Behavioural elements were then recorded to facilitate comparisons among populations. Ten and 12 mating trials were replicated and recorded for the two populations from North America and Europe respectively.

4.2.5 Determination of reproductive isolation

To examine the reproductive compatibility between populations of *S.* '*flavimana*', we attempted to mate males and females from different continents. Five sexually mature virgin flies of each sex were placed in

rearing containers under conditions identical to those in which cultures from individual collection localities live and breed successfully. Male and female flies originated from different continents, and the following reciprocal pairings were attempted: Ahrensfelde ♂ × New Orleans ♀; Ahrensfelde ♀ × New Orleans ♂. We also examined reproductive compatibility between the two European populations: Ahrensfelde ♂ × Kevelaer ♀ and Ahrensfelde ♀ × Kevelaer ♂. No flies died during the course of these trials. Each of these five male × female pairings was thrice replicated. The breeding substrate in each container (a 7cm petri-dish containing cow dung) was examined every other day for the presence of fertilized eggs or larvae. Substrate with fertilized eggs or larvae was removed and placed in separate containers for pupation of larvae. Where hybrid flies were obtained, they were again segregated by sex within one day of ecclosion to maintain their virginity. We then attempted to back-cross these hybrids with virgin flies from their parental cultures. To ascertain whether flies from failed back-crossing trials were fertile, we attempted to mate them with other flies from their own respective populations.

4.3 RESULTS

4.3.1 Morphology

We found two discrete morphotypes among the North American specimens that could be distinguished by a suite of morphological characters. One morphotype was indistinguishable from all European

specimens of *Sepsis flavimana*, while the other morphotype has the following distinguishing features (Fig. 4.1): (1) the flies are consistently lighter in colour (especially on the thoracic pleura, face, gena and legs cf. Fig 2. A, G vs. H); (2) the male fore-tibia lacks a distinct ventro-basal bump on the tibia (C) as compared to the European morphotype (J); (3) the epandrium and base of the surstylus of the male is light in colour and only the tip is dark (lateral view A, dorsal view F); (4) the surstylus has a sub-medial tooth (D). Features 1 and 2 are consistent with the description of *S. pyrrhosoma* by Melander and Spuler (1917) which mentions that the species is "largely reddish along the sides" with "face and cheek yellowish," and with a male fore tibia "slightly decreasing in size towards the tip and bearing a very weak and setulose tubercle on the underside near the base." Features 1–3 are also visible on the holotype of *S. pyrrhosoma* (W. Mathis, pers. comm.). For convenience, we refer to this as the '*pyrrhosoma*' morphotype. Some North American specimens from Palmyra, VA, and Dyar Pasture, GA, are morphologically indistinguishable from European *S. flavimana*, and are henceforth referred to as the '*flavimana*' morphotype.

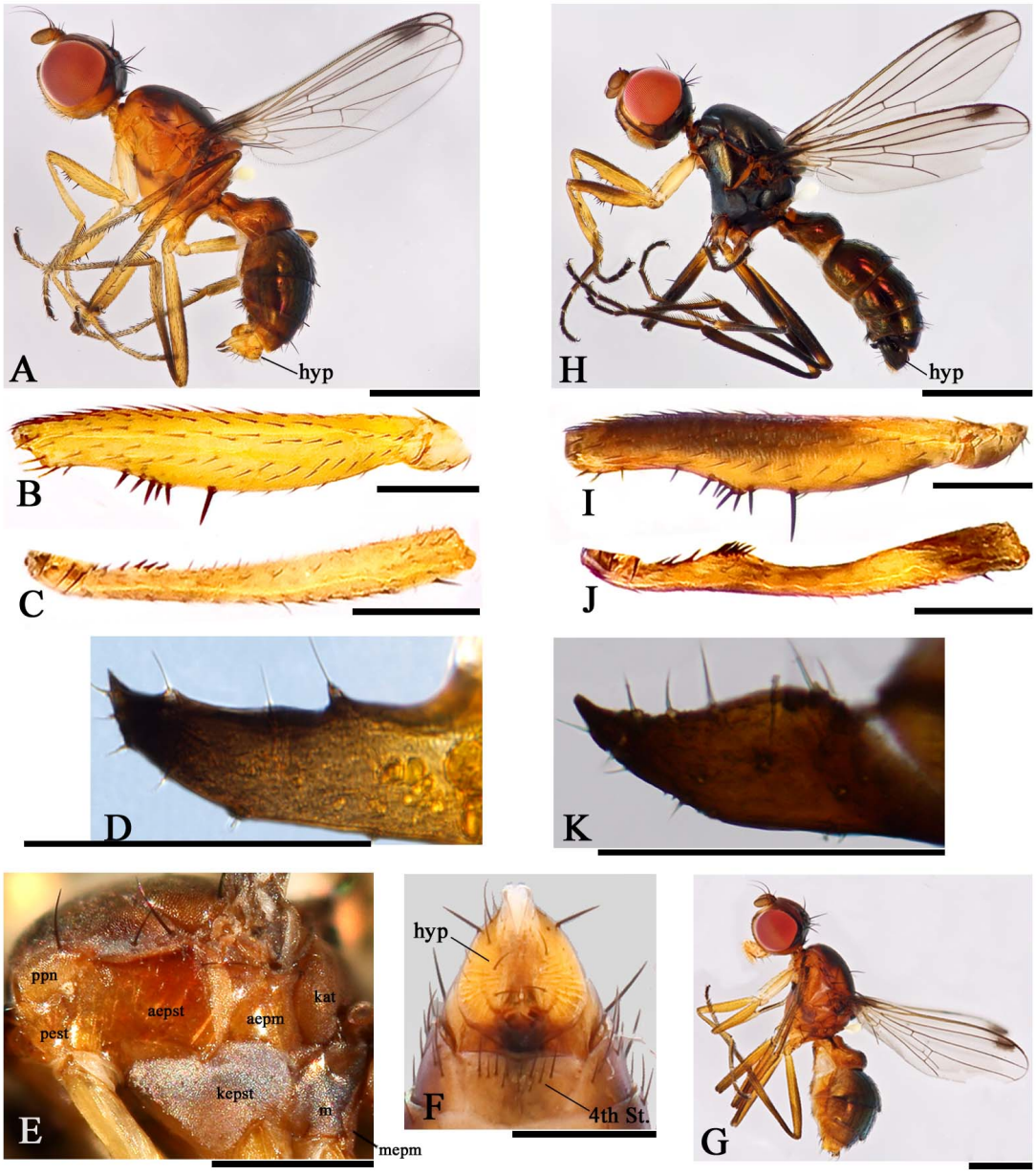

Figure 4.1: A–G. *Sepsis pyrrhosoma* (♂ unless otherwise noted). A. Habitus, lateral view, showing hypopygium (hyp). B. Fore-femur, posterior view. C. Fore-tibia, posterior view. D. Surstylus, dorsal view. E. Thorax, lateral view, showing pruinosity pattern on postprotonotum (ppn), preepisternum (pest), anepisternum (aepst), ketepisternum (kepst), anepimeron (aepm), katatergite (kat), meron (m) and metepimeron (mepm). F. Postabdomen, ventral view, showing 4th sternite (4th st.) and hypopygium. G. ♀ habitus, lateral view. H–K. *Sepsis flavimana* (♂). H. Habitus, lateral view, showing hypopygium. I. Fore-femur, posterior view. J. Fore-tibia, posterior view. K. Surstylus, dorsal view. Scale bars (A, E, G, H): 1mm; (B, C, F, I, J): 0.5mm; (D, K): 0.1mm

4.3.2 Molecular data

We obtained *COI* barcode sequences of ca. 778 bp from 50 specimens (GenBank accession numbers EU435804, EU435807, EU435808, EU435818, GQ354410, GQ388730-GQ388774). The parsimony analysis of these data found 125 trees with a length of 211 steps and the sequences for all species formed strongly supported monophyletic clusters. The two morphotypes were monophyletic sistergroups with strong support; this result was mirrored in the maximum likelihood analysis (see Fig. 4.2).

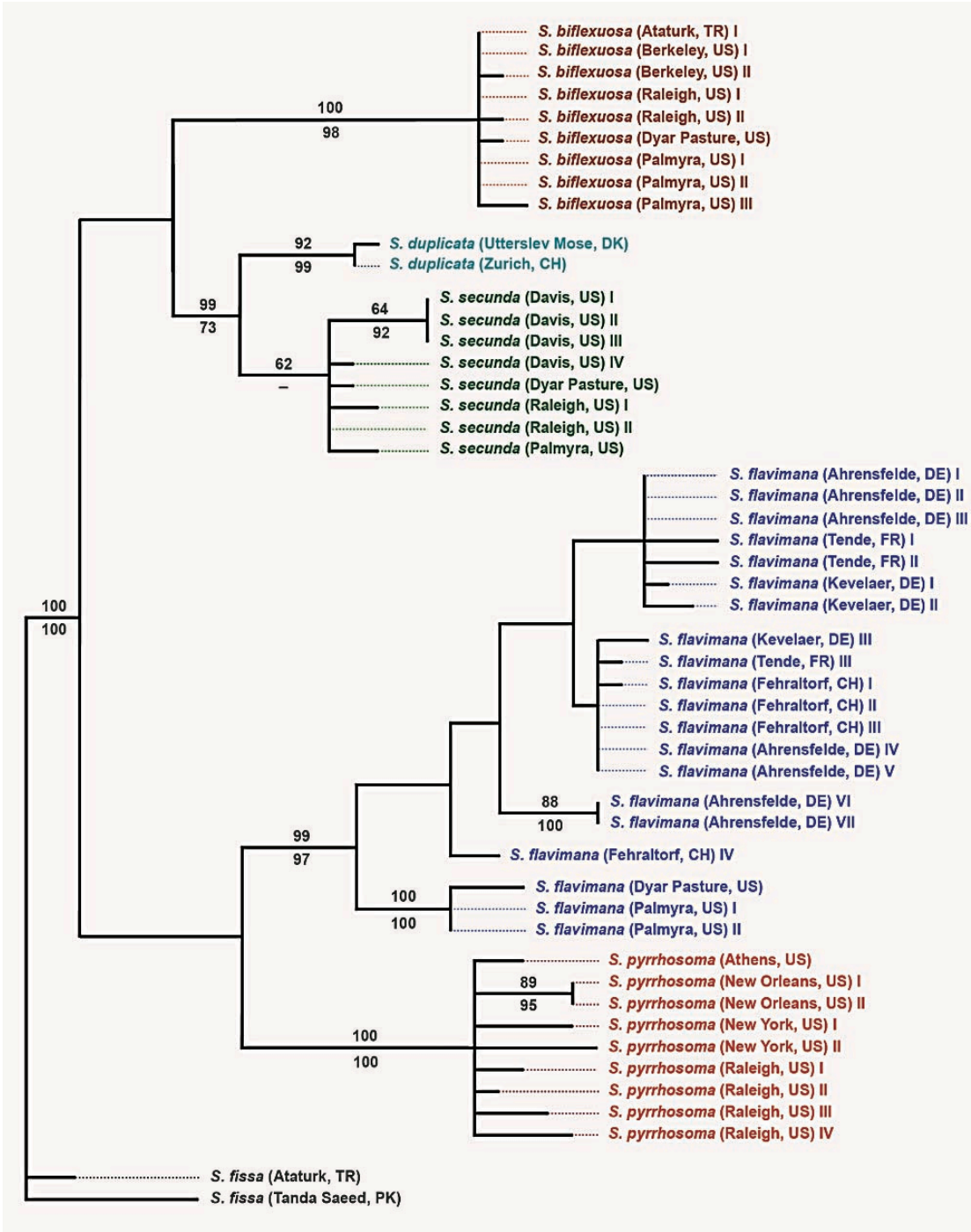

Figure 4.2: Consensus tree of *Sepsis flavimana* group. Parsimony jackknife percentiles are given above branches and maximum likelihood bootstrap percentiles below.

Uncorrected pairwise distances between the two morphotypes from North America ranged from 6.1–7.0%. However, distances within morphotypes were considerably smaller: 0.0–0.5% in the '*flavimana*' morphotype and

0.0–1.6% in the '*pyrrhosoma*' morphotype. Distances between the European *S. flavimana* and American '*flavimana*' morphotypes were 1.7–2.9%, while distances between '*pyrrhosoma*' morphotypes from North America and all other '*flavimana*' morphotypes (from both Europe and North America) were 6.2–7.7%. The distances between European *S. flavimana* populations were 0.0–1.7% (Table 1).

4.3.3 Behavioural observations and reproductive isolation trials

The mating behaviour of the *pyrrhosoma* morphotype differs from that of European *S. flavimana* in several respects. In *S. flavimana*, the male proboscis only touches the female on the dorsal region of the thorax, while in the *pyrrhosoma* morphotype the male proboscis is used to stimulate the female ocelli instead (Table 2). The video evidence for these differences can be viewed from http://evolution.science.nus.edu.sg/pyrrhosoma_clips.html or http://www.youtube.com/watch?v=0ypmqN8t_Xw http://www.youtube.com/watch?v=eO4ZZAuacRI).

Behavioural Elements	***S. flavimana***	***S. pyrrhosoma***
male mid-leg tarsi curl	observed in both species	
male hind-leg tap of female abdomen		
substance transfer (from male hind-leg to female thorax)		
male mid-leg rub of female thorax		
degree of female resistance (shaking)	violent and persistent	mild and sporadic
separation after copulation	prolonged struggle to break genital contact	rapid
precopulatory surstylus stimulation	absent	present and prolonged
location of male proboscis contact with female	dorsal part of female thorax	female ocelli
Mating success (virgin trials)	33.3 %[a]	100%[b]

Table 4.2: Qualitative comparison of behavioural elements observed in *S. flavimana* and *S. pyrrhosoma* (virgin) mating trials. $N^a = 12$; $N^b = 10$

Lower resolution video-clips have also been uploaded as supplementary information. In addition, males of the *pyrrhosoma* morphotype were observed to stimulate the postabdomen of females with their surstylus prior to copulation, but this behaviour was absent in *S. flavimana*. During copulation, *S. flavimana* females constantly shook their bodies, but this apparent resistance to mating was not as violent or obvious in the *pyrrhosoma* morphotype, where female body shakes were sporadic and less energetic. Prolonged struggles lasting five or more seconds were

consistently observed during genital decoupling between *S. flavimana* individuals. In contrast, separation was prompt in the *pyrrhosoma* morphotype, with males typically requiring only two quick about-turns to disengage from the female. Finally, the mating success rates of virgin the *pyrrhosoma* morphotype were much higher than in *S. flavimana*.

Mating trials reveal no potential for gene flow between the two species, since hybrid offspring were produced in only one direction (Ahrensfelde *S. flavimana* ♀ × *pyrrhosoma* morphotype from New Orleans ♂) and these hybrids failed to produce viable offspring regardless of whether they were mated with other hybrids or with either parent species (backcrosses; Table 3).

	Mass Crossings	**Back Crossing**
Ahrensfelde ♂ × New Orleans ♀	×	(N/A)
Ahrensfelde ♀ × New Orleans ♂	✔	×
Ahrensfelde ♂ × Kevelaer ♀	✔	✔
Ahrensfelde ♀ × Kevelaer ♂	✔	✔

Table 4.3: Results of the hybridisation experiments

To demonstrate that our laboratory conditions were sufficient to foster mating between reproductively compatible flies, we successfully crossed *S. flavimana* from the two European populations.

4.3.4 Taxonomic conclusion

Morphology, genetic data, behavioural differences, and reproductive isolation support the presence of two morphologically and biologically

distinct North American taxa. There are four available North American names that are currently under synonymy with *S. flavimana*. The oldest name is *Sepsis vicaria* Walker, 1849. However, the identity of this species remains uncertain because the type material consists of females only (Ozerov 2000) and only the weakest morphological character that distinguishes the *pyrrhosoma* morphotype from *S. flavimana* could potentially be studied (lighter colour than in *S. flavimana*). The second oldest name is *Sepsis pyrrhosoma* Melander and Spuler 1917. A male lectotype was designated by Zuska in 1967 (Ozerov 2000) and the curator of the Smithsonian, Dr. W. Mathis, confirmed that the type is of the *pyrrhosoma* morphotype with regard to color, the male fore-tibia, and surstylus morphology. We therefore here resurrect *S. pyrrhosoma* from synonymy and re-describe the species.

4.3.5 Species re-description

Sepsis pyrrhosoma Melander and Spuler, 1917

Family SEPSIDAE

Sepsis pyrrhosoma Melander & Spuler 1917: Fig. 14.

Holotype in National Museum of Natural History (NMNH), Washington, DC, USA. Ozerov (2000: 116) provides the following informations on the type: ♂ of *Sepsis pyrrhosoma* (designated by Ozerov 1998: 87), labelled “Lafayette Ind[iana].[,] Jul[y]”, “A[.]L[.] Melander Collection 1961”, “Type Sepsis pyrr[h]osoma Mel[ander]. & Sp[uler]. J.

Zuska 1967", "Lectotypus ♂ Sepsis pyrrhosoma Melander, Spuler, 1917: 25. design. A. Ozerov, 1994", "Sepsis flavimana Mg. ♂ A.L. Ozerov det., January, 1994.

Other examined material. ♂♂♀♀ ex culture established from ♀♀ from grassland along Leake Avenue near Mississippi River, New Orleans, LA), ca. 5m ASL, 29˚ 55' 48.34" N 90˚ 8' 4.17" W 2008 (Coll. R. Meier); in Raffles Museum of Biodiversity Research, Singapore (RMBR). Additional specimens were obtained from Raleigh, NC, New York and Athens, GA.

Etymology. The specific name first given by Melander and Spuler in their original description of the species (Melander & Spuler 1917), is derived from the combination of the Greek πυρο (pyro; fire) and σώμα (soma; body), an indication towards the reddish hue of the fly's body. The gender is neutral.

Distribution. Apparently limited to the South-eastern regions of North America, Indiana, Louisiana, Georgia, North Carolina, Virginia, and Pennsylvania.

Diagnosis. Adult *Sepsis pyrrhosoma* resemble lightly coloured specimens of *S. flavimana*. However, *S. pyrrhosoma* can be consistently distinguished from the latter by the following characters. While *S. flavimana* (H) is always black to dark brown in thorax and head colour, *S. pyrrhosoma* (A, G) is mostly reddish to yellow on the pleura and abdominal sections as well as on the face and gena. Fore-femora of *S. pyrrhosoma* are also consistently light yellow (B), while *S. flavimana* invariably retains a

dark brown region dorsally (I). Colour is the only way to potentially distinguish ♀ *S. pyrrhosoma* (G) from ♀ *S. flavimana* morphologically, but this character is unreliable for old and/or alcohol-preserved specimens. Additional characters in the male are: (1) the fore-tibial ventro-basal bump is always slight or non-existent in *S. pyrrhosoma* (C), bearing small, weak bristles, while *S. flavimana* (J) has a distinct bump with longer and thicker bristles; (2) The hypopygium of *S. flavimana* (H) is entirely black with a smooth, beak-like surstylus (K) while *S. pyrrhosoma* possesses a yellow hypopygium with only the surstylus darkened apically (A, F). (3) The *S. pyrrhosoma* surstylus bears sub-medial inward-facing protrusions not present in *S. flavimana* (cf. D & K).

The original description of *Sepsis pyrrhosoma* by (Melander and Spuler 1917) and Spuler (1917) was brief. The following is a more detailed description of the adult based on the specimens recently collected from North America.

Colour (A–D, G). Similar in both sexes. Vertex and occipital region black, frons and facial ridge dark brown. Parafacial, facial carina and gena light brown to yellow. Pedicel and 1st flagellomere brown, arista black. Clypeal margins black. Scutum and subscutellum black. Postprotonotum and pleural areas mostly yellowish red, except for the dorsal margin of the anepisternum, pleural wing process, meron, metepisternum and dorsal half of katepisternum, which are dark brown. All coxae and trochanters light yellow, as are fore-femora and tibiae. All tarsi are light yellow except for the 4th and 5th tarsomeres, which are black. Mid and rear femora

infuscate on the dorsal and ventral side medio-distally, while tibiae are brown to dark brown basal-medially. Abdomen with a cupreous tinge, yellowish red except for dorsal regions of tergites and all sternites, which are dark brown. Epandria and cerci yellow, surstylus yellow but black apically.

Pruinosity (E). Similar in both sexes. Head glossy except for occipital region, gena and face, which is moderately pruinose with macrotrichia. Scutum, pronotum and scutellum also moderately pruinose. Subscutellum and anatergite glossy except for sparse microtrichia near margins. Proepisternum similarly glossy with microtrichia limited to dorsal and ventral margins. Microtrichia also present on posterior margin of anepisternum, anterior, dorsal and posterior areas of anepimeron. Katepisternum and meron heavily pruinose, while katatergite medium pruinose. Metepimeron with a shiny patch ventrally.

Head (A, G). Similar in both sexes. Roundish, facial carina short and shallow. Parafacial and gena narrow. With two subvibrissal bristles and numerous short setae along lower genal margin. Numerous supracervical setae present. Eyes maroon, roundish but posteriorly compressed on dorsal and ventral sides. Postcellar setae ¾ of ocellar setae, both divergent. Outer vertical setae ½ the size of the inner vertical setae. Pedicel bearing setae along apical margin with 1 dorsal bristle. Flagellomere in profile long-oval, rounded apically, almost twice as long as wide. Aristae dorsal and bare. Larger specimens tend to have a disproportionately larger head compared to other specimens.

Thorax (A, E, G). With the following paired setae: 1 postprotonotal setae, 2 notopleural setae, 1 supraalar setae, 2 postsutural dorsocentral setae, 1 anepisternal setae and 1 apical scutellar setae. Anepisternal setulae absent. Scutellum compressed, more than twice as wide as is long.

Abdomen (A, D, F, G). Tergites (t) similar in both sexes; all with relatively long setulae at discal and marginal regions. Syntergite 1+2 with 1-2 pairs lateral marginal bristles, proceeding t3 – 5 with 1 pair, t6 with none. Spiracles 1 & 2 in intersegmental membrane, spiracles 3 – 5 on margin of tergite plate, spiracles 6 and 7 within t7. Abdomen is slightly constricted after syntergite (synt) 1+2. Sternites (st) well defined, with s4 bearing 2–3 rows of strong setae posteriorly (F). Bristles are more prominent in males than females. ♂ terminalia – Symmetrical surstyli short and angulate, decussating and fused to epandrium (D, F); with inward protrusion medially (D). Cercal lobes fused, each with 1 translucent apical seta.

Legs (A–C, G). ♂ fore-legs: slightly enlarged femur bearing one large ventral (v) bristle at the middle and a slight tubercle bearing four to six shorter bristles on posterio-ventral side (B). Fore-tibia slim with a very slight basal bump bearing a row of weak bristles (C). Mid-femur with one large and long anterior-ventral (av) bristle in center. Mid-tibia with two smaller bristles av, centrally and preapically, one small dorsal (d) bristle preapically; apex with bristles except on d region. Hind-femur without distinct bristles; hind-tibia with one small d bristle preapically, one to two

small av bristles apically. Hind-tibia bearing very faint region of osmoterium on anterior-dorsal region medially. ♀ fore-legs simple and unmodified. Mid- and rear-femur without bristles. Mid-tibia with one small v bristle on median, one small av bristle preapically, with apice similar to ♂. Rear-tibia similar to ♂ but without osmoterial region.

Wing (A, G). Elongate, longer than abdomen. Veins bare except for a few minute setulae on ventro-basal side of stem vein. Wing entirely covered with microtrichia, with oblongish pterostigma at tip of R2+3. Anal lobe well developed, A2 not reaching wing margin. Upper calypter brown with long thin setae on margin. Lower calypter absent. Halter creamy to yellow.

4.4 DISCUSSION

Our study of the species boundaries of *Sepsis pyrrhosoma* demonstrates how multiple sources of data can be used to resolve the status of so-called cryptic species that have been suggested by unexpectedly large genetic distances within a single nominal species. Our approach is iterative in that unexpected genetic variability prompted renewed morphological evaluation. This re-evaluation uncovered consistent morphological characters that distinguish *S. flavimana* and *S. pyrrhosoma*. However, this morphological evidence initially appeared weak, because many *Sepsis* species exhibit considerable size variability that is known to be correlated with differences in body colour and other important diagnostic features such as fore-legs, and claspers (Pont &

Meier 2002). *Sepsis pyrrhosoma* could therefore be easily mistaken as a *S. flavimana* which probably explains why the former had been synonymised. In order to further strengthen the morphological and genetic evidence for our hypothesis that *S. pyrrhosoma* is a valid species, we studied the mating behaviour and reproductive isolation and the morphological evidence corroborates the presence of two distinct taxa. Overall, a case of initial conflict between morphology and DNA sequences turned into a case of concordance that was further strengthened with additional data. Note that we are not proposing that such an extensive repertoire of data needs to be collected for all cases. We believe that such detailed study will only be needed for the relatively small number of taxa where different data sources initially appear to be in conflict (e.g., Laamanen *et al.* 2003; Petersen *et al.* 2007).

As pointed out earlier, ultimately decisions on species boundaries depend on which species concept is used, and as pointed out by many authors, there are a large number of species concepts. For example, (Mayden 1999; Paterson 1985) lists 22 different concepts, but fortunately this bewildering diversity can be pared down by either grouping similar concepts into categories and/or only considering concepts that are used regularly. We would argue that the four main categories of species concepts are covered in Wheeler and Meier (2000): (1) concepts based on reproductive isolation or cohesion typified by the Biological Species Concept (Mayr 2000) and the Hennigian Species Concept (Meier and Willmann 2000); (2) concepts based on monophyly, as represented by the

Phylogenetic Species Concept *sensu* (Mishler and Theriot 2000); (3) concepts based on the diagnosability of populations, such as the Phylogenetic Species Concept *sensu* Wheeler and Platnick (Wheeler and Platnick 2000); and (4) concepts using a mixture of criteria, such as the Evolutionary Species Concept *sensu* Wiley & Mayden (Wiley & Mayden 2000).

Reproductive isolation is the core criterion for both Biological and Hennigian species concepts, and all evidence suggests that *S. flavimana* and *S. pyrrhosoma* are reproductively isolated. Furthermore, these species are likely to be sympatric. *Sepsis flavimana* were collected at Dyar Pasture, GA, which is only 28km south of Athens, GA, where *S. pyrrhosoma* were collected, and it is likely that there is appropriate breeding substrate (dung) between these two localities although the two species appear to prefer different substrates. *Sepsis flavimana* is predominantly found on cow dung (Pont & Meier 2002) while S. *pyrrhosoma* has only been collected on dog dung (Raleigh, NC; Athens, GA) or in localities where dog dung is the most likely breeding substrate (New Orleans, LA).

The phylogenetic species concept *sensu* Wheeler & Platnick (2000) defines species as populations with a unique combination of characters. If *S. pyrrhosoma* and *S. flavimana* are considered separate populations then each has a unique combination of characters as well as distinct *COI* barcodes. We can thus defend *S. pyrrhosoma* and *S. flavimana* as separate phylogenetic species. However, it can be argued that this

conclusion is based on *a priori* decisions on which taxa form populations because population-aggregation analyses require such *a-priori* decisions on population boundaries. Alternatively, one could have treated all North American specimens in the *flavimana/pyrrhosoma* complex as one population; i.e., the characters that are here interpreted as being species species-specific would have been treated as traits with population-level variability and only one species would have been recognized (Laamanen *et al.* 2003; Tan *et al.* 2008). The phylogenetic species concept *sensu* Mishler and Theriot is also likely to recognize *S. pyrrhosoma* as a separate species because it forms a biologically distinct, reproductively isolated monophyletic unit. These features — distinct biology and reproductive isolation — also likely render *S. pyrrhosoma* a distinct species under the evolutionary concept. We believe that the application of various species concepts to a dataset similar to ours will often support the same conclusion. Furthermore, although we believe that all authors should have a preferred species concept, proponents of different species concepts may often come to the same conclusion; *i.e.*, those authors that are afraid of criticism when applying a particular species concept may have less to fear than they may think.

The only species concept that would have to come to a conflicting conclusion is the recognition concept, which defines species as units that share a common fertilization system (Paterson 1985). The decisive step in this species concept is the recognition of the other specimens as being mating partners. As such *S. pyrrhosoma* and *S. flavimana* would belong to

the same species because they can successfully mate and produce (albeit infertile) offspring. Note that this species concept would also lead to the synonymisation of numerous other sepsid species, because males of many species are known to initiate mating with all females of approximately right size.

4.5 CONCLUSION

We here demonstrate how 'cryptic species' proposed based on genetic evidence can be resolved using multiple sources of data. We argue that these units either have to be rejected or formally recognized, or else 'cryptic species' will overwhelm the systematic literature. We also demonstrate that systematists can treat the 'species-concept problem' without having to fear the vitriol that is often related to discussing competing concepts. We believe that for most species many concepts are likely to arrive at the same conclusion. Finally, we have to acknowledge that in collecting the data for resurrecting *S. pyrrhosoma* the North American *S. flavimana* emerged as a potential new 'cryptic species' based on the genetic evidence. We believe that with the widespread use of DNA sequences such cases will become very common. As one taxonomic problem is resolved another appears based on the newly gathered data. In this sense, DNA sequences will not speed-up taxonomic research, but will lead to the estimation of more accurate species boundaries based on a more satisfactory amount of data.

CHAPTER 5

Morphology and DNA sequences confirm the first neotropical record for the holarctic sepsid species *Themira leachi* Meigen, 1826 (Diptera: Sepsidae)

5.1 INTRODUCTION

Even for the most cosmopolitan of species, climate frequently presents effective barriers for dispersal. Many eurytopic and synanthropic species go extinct when introduced into a new climatic zone. For example, translocated ants remain in sheltered environments reminiscent of their home climate (McGlynn 1999). Here we report the occurrence of a primarily Holarctic dipteran species, *Themira leachi* Meigen, 1826, in Neotropical Cuba. This discovery suggests that the species may have a large disjunct distribution, as the next closest record lies almost 3,500 km to the north in Nearctic Newfoundland, Canada (Ozerov 1998).

The genus *Themira* comprises 35 species and belongs to the relatively small clade of the cosmopolitan dung-fly family Sepsidae (Ozerov 2005). The genus is primarily distributed in the Holarctic, with only four species bordering on other biogeographic regions (Meier 2007; Ozerov 1998; Pont and Meier 2002). *Themira leachi* has been recorded throughout Northern Europe, spanning eastwards through Asiatic Russia and Mongolia. (Ozerov 1998) added the species to the Nearctic fauna by reporting specimens from Northern Canada.

5.2 MATERIALS AND METHODS

Recently, five specimens (four males, one female) were collected from dung in Cuba (2002: Pinares de Mayarí pine forest, Sierra Cristal National Park, *ca.* 650m ASL). The morphology of the males suggested that they are *Themira leachi*, but since this record is so far beyond the known range of the

species, we used detailed morphological study and DNA sequencing to confirm this preliminary identification. Line drawings were prepared for the Cuban specimen in order to compare them to drawings for European specimens. In addition, we generated high-resolution color-photographs of the habitus and important diagnostic structures for European and Cuban specimens with a Visionary Digital™ Plus Lab System, using a Canon EOS 40D with a mounted Infinite K2 Long Distance Microscope (CF4 objective at position 1 and 3). For the images at the highest magnification, a 10X Olympus objective was used (position 3).

A *ca.* 660 bp piece of the *COI* gene was sequenced for four Cuban specimens using the DNA extraction, amplification, and sequencing protocols described in (Su et al. 2008). These sequences were submitted to Genbank (EU831274 – EU831277) and compared to a known sequence of *T. leachi* from Europe (Genbank: EU435823) as well as *COI* sequences for ten other *Themira* species (Su et al. 2008).

5.3 RESULTS

The pairwise distances between the European specimens and those we sequenced from Cuba ranged from 0.5% to 0.8%.

Detailed morphological investigations reveal that the Cuban specimens are indeed very similar to specimens from Europe and consistent with Ozerov's (1998) and Meier and Pont's (2002) diagnoses. Forelegs, sternites and hypopygiums were used for comparison; Cuban and European specimens are shown in Figure 5.1.

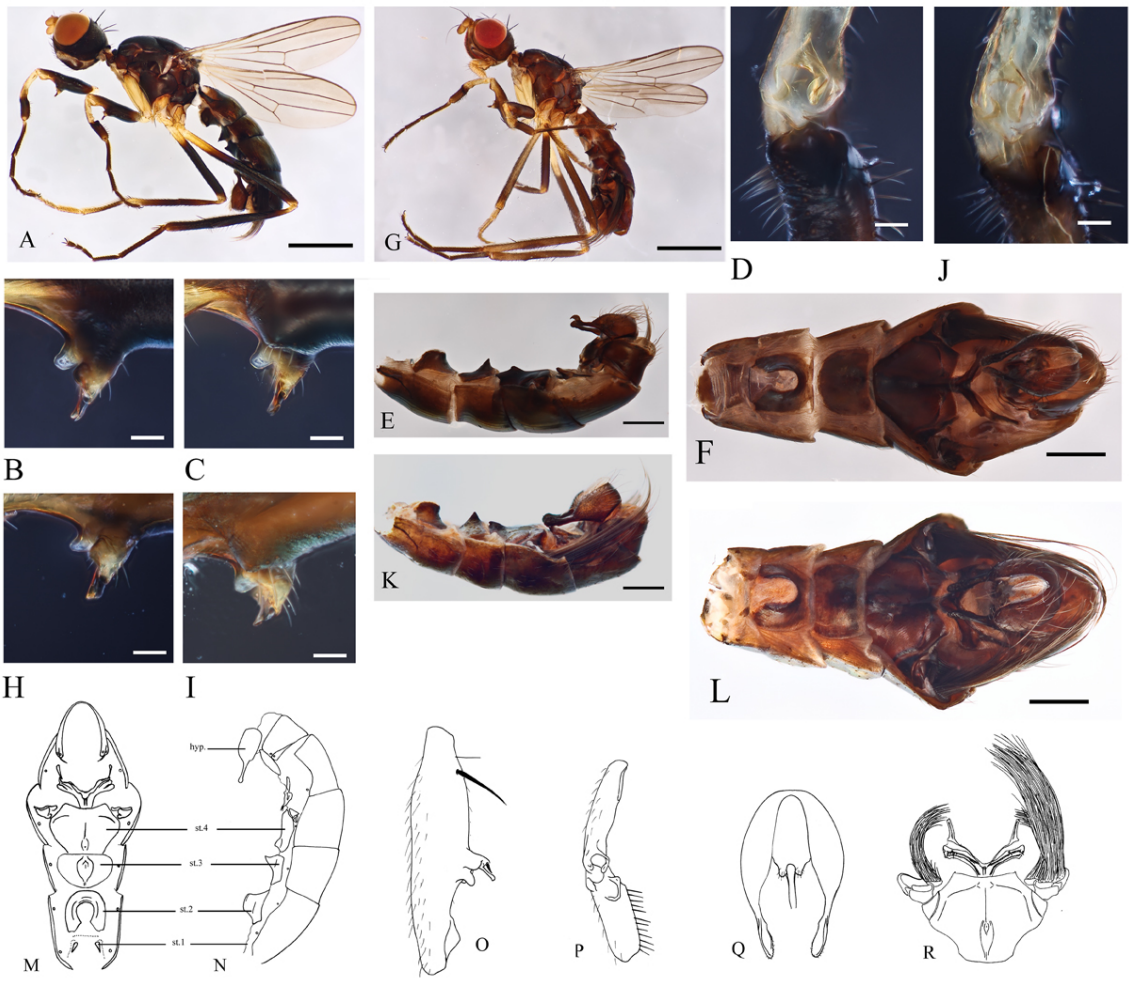

Figure 5.1: Morphology of *Themira leachi* from Cuba (photographed A-F; drawn M-R) and Europe (photographed G-L). Habitus: A, G; fore-femoral modifications (anterior view): B, H; fore-femoral modifications (posterior view): C, I; fore-tibial modifications (anterior view): D, J; abdomen (lateral view, sternite bristles removed): E, K, N; abdomen(ventral view, sternite bristles removed): F, L, M; fore-femur (anterior view): O; fore-tibia (anterior view): P; hypopygium (dorsal view, setulation omitted): Q; 4th sternite (dorsal view): R. Scale bars for A, G: 1mm; B-D and H-J: 0.1mm; E, F, K, L: 0.5mm

The fore femora and tibiae of both specimens possess similar modifications whose function and co-evolution with female wings have been discussed in the recent literature [on femur: *c.f.* B & H (anterior view), C & I (posterior view); on tibia: *c.f.* D & J; (Ang et al. 2008; Ingram et al. 2008;

Puniamoorthy et al. 2008). The 4th sternite and hypopygium can be seen on the abdomen and are also very similar in structure and diagnostic for *T. leachi* (lateral view: *c.f.* E & K; ventral view *c.f.* F & L). Even more striking are the 2nd and 3rd sternites, which are well developed and have been modified into a raised, anteriorly open crater on the 2nd sternite and a pronounced protrusion on the 3rd sternite. These sternite modifications are unique to *T. leachi* and the only difference between the specimens is minor (a more pronounced and hook-like ventral protrusion on the 4th sternite of the European specimen).

5.4 DISCUSSION

Overall, the foreleg, sternite, and hypopygium morphology are very similar between the European and Cuban specimens of *T. leachi* and suggest the presence of only one species.

Recent studies of morphologically uniform species with wide distributions have suggested that such species frequently contain 'cryptic' species that can be discovered once DNA sequence data become available (Bickford et al. 2007). We thus compared the Cuban and European specimens with regard to the mitochondrial gene *COI*. Pairwise distances between the European and Cuban sequences were 0.5% to 0.8%. Whether such distances are typical for inter- or intraspecific variability can be judged when they are compared to a distribution of distances for closely related species in *Themira* ((Meier et al. 2006; Memon et al. 2006; Petersen et al. 2007). Based on the known sequences for ten *Themira* species, the mean interspecific distance for closest relatives is 6.2% and only one species pair [*T. lucida* (Staeger in

Schiødte) vs. *T. flavicoxa* Melander & Spuler] has distances below those observed between the Cuban and European *T. leachi*. However, *T. lucida* and *T. flavicoxa* are morphologically distinct (Ozerov 1998) while we did not find any significant morphological differences between the Cuban and European specimens of *T. leachi*.

Given that the Cuban specimens are almost certainly *T. leachi*, the lack of North American records south of Newfoundland is surprising. The Nearctic Sepsidae have been extensively sampled (Meier, 2007) and Ozerov (1998) identified all material from major museum collections for his revision of *Themira*, which focused on the North American fauna. He did not find any *T. leachi* south of Newfoundland. One may speculate that the species was introduced to Cuba as a synanthrophic commensal given that humans commonly transport arthropods to new areas (Jenkins 1996; Kobelt and Nentwig 2008; Smith et al. 2007) and Pinares de Mayarí is only 10km south of Nipe bay. *Themira leachi*, however, is not particularly common (Meier, pers. comm.) and trade volumes would favor an introduction to the USA instead of Cuba.

The alternative to introduction is that *Themira leachi* genuinely occurs in Cuba, which may surprise given the relatively few species that are shared between the Holarctic Region and Neotropical Cuba. It may also surprise because the majority of *Themira* species require relatively cool temperatures (Meier and Pont 2000). However, *T. leachi* is also known from some subtropical/Mediterranean localities (e.g., Hungary, Italy, Russian South Primore; Ozerov, 2005) and the climate of Pinares de Mayarí which consists

of a large ridge differs significantly from that of the surrounding lowlands. Observations by (Carabia 1945) suggest that it is similar to a cloud forest with relatively low temperatures and continuous moisture even during the traditionally dry Cuban winter months. These cooler temperatures could explain why *T. leachi* may be able to survive in the tropics. Also, while a number of *Themira* species rely on specific media such as waterfowl or cow dung for breeding, *T. leachi* has less specific substrate requirements and can survive on a variety of decaying material, from decomposing vegetation to various types of excrement (Pont and Meier 2002). The occurrence of *T. leachi* in Cuba nevertheless remains puzzling and more information should be collected.

OVERALL CONCLUSIONS

This thesis explores the ways *COI* can be used to estimate species richness or complement to complement traditional taxonomic research by providing an independent source of information. The latter is particularly important when traditional methods fail to provide conclusive results.

In the first and second chapter, *COI* was used to estimate species richness. My first chapter used four datasets from 3 arthropod families, namely the Sepsidae (Diptera), Curculionidae and Dytiscidae (Coleoptera) to test the congruence between taxonomic identification and sequence-based species estimation. Species richness estimates for *COI* were very similar to those from taxonomic experts, with a deviation of less than 10% in all cases at both 2% and 3% thresholds. Contrast this with the performance of parataxonomy, with a mean deviation of 32% from the estimates made by taxonomic experts. $Krell. However, the use of *COI* for biodiversity studies comes with a caveat; it delimits specimens in a manner incongruent with the species as determined by taxonomic expertise. Only 60-80% of all DNA-delimited clusters are also species based on traditional techniques (Meier et al. 2006). However, the reliable species richness estimation of *COI* means that groups that are hyperdiverse and receive very little taxonomic attention can still be studied, with useful results.

I utilised this in the second chapter in order to investigate the species diversity of Bornean Corethrellidae. The parasitoid dipteransare a

very poorly understood monogeneric tropical clade. While the Corethrellidae have very interesting life history and adaptations to help them find their anuran hosts, they are very poorly studied outside of the Nearctic and Neotropical regions. Here, I used *COI* to test the validity of morphospecies sorting, finding strong conflict between DNA and parataxonomy. I estimated the number of *Corethrella* species sampled in my main study site, Labi peat swamp in Brunei Darussalam, and found that it was richer in *Corethrella* species than the current most *Corethrella*-diverse locality in the Neotropical Carara National Park, Costa Rica. Specimens sampled from the Bornean locality form 18 clusters at 3%, while Carara National Park is known to have 14 species. I further used the information from *COI* clusters to estimate the α-diversity based on non-parametric species richness estimators and found that there leas are more than 43 species in Labi peat swamp alone.

My third and fourth chapters involve *COI* as an exploratory tool and an additional source of data to analyse sepsid species with known widespread distributions. The goal was to test whether these widespread species contain cryptic species. I use eight species of widespread sepsids in Southeast Asia and Australasia in the third chapter, and determine that most widespread species in the Sepsidae are truly widespread, with low intraspecific distances between even very geographically distant populations. I find one species, *Allosepsis indica*, to have large disjunct distances, although morphological characters are continuous between populations which have very high pairwise distances for *COI*.. Data from

reproductive isolation seems to agree with the results of *COI.* I also discuss and discard the hypothesis that the observed distribution of these species is due to secondary introduction by the movement of domesticated cattle, based on the population structure observed in my analyses of *COI*. Instead it is more likely that the observed results are due to natural vicariance and genetic drift.

In the fourth chapter, I discover an unexpected pattern of *COI* pairwise distances within a sample of *Sepsis flavimana*. Upon further investigation and using information from morphology, behaviour, and reproductive isolation, a cryptic species, *Sepsis pyrrhosoma,* is resurrected from synonymy.

In the fifth chapter, I use *COI* as a tool to confirm a species hypothesis made based on morphology. The specimens of the species in question, *Themira leachi,* had been collected in Sierra Cristal National Park, Cube, approximately 3500km south of its known distribution in the New World (Newfoundland, Canada). Here, *COI* was used to confirm that the Cuban specimens is not a cryptic species with similar morphology.

The third, fourth and fifth chapters illustrate how DNA can be of great value to even morphologically-trained taxonomic experts. Through comparing and analysing sequences of *COI*, I have helped to filter hundreds of specimens and to earmark individuals whose morphology and molecular signal are in conflictand needed additional study. This approach is exemplified in *Allosepsis indica*, *Sepsis pyrrhosoma* and *Themira leachi*.

Going forward, there remains much work to be done with regard to species richness estimation via the use of *COI*. Our analysis of the performance of *COI* clusters needs confirmation from additional *COI* datasets. The datasets can be used to further test the consistency of *COI*-based estimates of species richness to those of taxonomic experts for different taxonomic groups. Three-way studies where parataxonomists and taxonomists study specimens before the samples are sequenced for *COI* would be particularly useful, as they would allow for a comparison of the congruence and conflict between different techniques for estimating species richness. A good candidate for this would be the samples of *Corethrella* that I described in the second chapter. These specimens have already been sorted to morphotypes by parataxonomists and sequenced, and are awaiting formal description by the taxonomic expert, A. Borkent. After this is accomplished, the Bornean *Corethrella* dataset could be used to test whether the species richness estimates and species boundaries delimited by parataxonomy or *COI* are more or less congruent with the conclusions made by taxonomic experts.

With regards to the Corethrellidae, a systematic, concerted effort to inventory their diversity in Southeast Asia is urgently needed. The danger is that the rate of species description and collection may not catch up to the rate of extinction, especially since there is only one active taxonomist for this enigmatic family and many frog species are in the process of going extinct. The two cosmopolitan groups of Diptera examined in this paper are likely to face very different fates. While the synanthropic saprophagous

Sepsidae and the anuran parasitoid Corethrellidae are strongly associated to vertebrate groups Bovinae and Anura respectively, it is not likely that sepsid species will find themselves with a lack of substrate in the future. The same cannot be said of the *Corethrella* species found in the peat swamps of Borneo, as both their hosts (Wake and Vredenburg 2008) and their habitats (Rieley and Ahmad-Shah 1992) are severely threatened. As parasitoids that feed on endangered hosts in threatened habitats, the Corethrellidae exemplify a group of invertebrates that is particularly vulnerable to extinction (Dunn et al. 2009; Koh et al. 2004). This is the reason why every effort has to be made to set rigorous standards and field collection expeditions must attempt to collect as much information about species and localities as possible (Dayrat 2005; Valdecasas et al. 2008).. This is now happening with bio-imaging and non-invasive DNA extraction and sequencing.

REFERENCES

Abadie, J.C., Andrade, C., Machon, N., Porcher, E., 2008. On the use of parataxonomy in biodiversity monitoring: a case study on wild flora. Biodiversity and Conservation 17, 3485-3500.

Alford, R.A., Bradfield, K.S., Richards, S.J., 2007. Ecology: Global warming and amphibian losses. Nature 447, E3-E4.

Amorim, D.d.S., Yeates, D.K., 2006. Pesky gnats: Ridding dipteran classification of the Nematocera. Studia Dipterologica 13, 555-666.

Ang, Y., Lim, G.S., Meier, R., 2008. Morphology and DNA sequences confirm the first Neotropical record for the Holarctic sepsid species Themira leachi (Meigen) (Diptera: Sepsidae). Zootaxa 1933, 63-65.

Ang, Y., Puniamoorthy, N., Meier, R., 2008. Secondarily reduced foreleg armature in *Perochaeta dikowi* sp.n. (Diptera : Cyclorrhapha : Sepsidae) due to a novel mounting technique. Systematic Entomology 33, 552-559.

Astbury, W.T., 1961. Molecular Biology or Ultrastructural Biology ? Nature 190, 1124-1124.

Balke, M., Ribera, I., Hendrich, L., Miller, M.A., Sagata, K., Posman, A., Vogler, A.P., Meier, R., 2009. New Guinea highland origin of a widespread arthropod supertramp. Proceedings of the Royal Society B-Biological Sciences 276, 2359-2367.

Beamish, F.W.H., Beamish, R.B., Lim, S.L.H., 2003. Fish assemblages and habitat in a Malaysian blackwater peat swamp. Environmental Biology of Fishes 68, 1-13.

Basset, Y., Novotny, V., Miller, S.E., Pyle, R., 2000. Quantifying biodiversity: Experience with parataxonomists and digital photography in Papua New Guinea and Guyana. Bioscience 50, 899-908.

Beehler, B.M., 2001. ECOLOGY: Island Biogeography Redux, with a Speciation Twist. Science 294, 1007-1008.

Bernal, X.E., Page, R.A., Rand, A.S., Ryan, M.J., 2007. Cues for Eavesdroppers: Do Frog Calls Indicate Prey Density and Quality? The American Naturalist 169, 409-415.

Bernal, X.E., Rand, A.S., Ryan, M.J., 2006. Acoustic preferences and localization performance of blood-sucking flies (Corethrella Coquillett) to tungara frog calls. Behavioral Ecology 17, 709-715.

Bertone, M.A., Courtney, G.W., Wiegmann, B.M., 2008. Phylogenetics and temporal diversification of the earliest true flies (Insecta: Diptera) based on multiple nuclear genes. Systematic Entomology 33, 668-687.

Besansky, N.J., Severson, D.W., Ferdig, M.T., 2003. DNA barcoding of parasites and invertebrate disease vectors: What you don't know can hurt you. Trends in Parasitology 19, 545-546.

Bickford, D., Lohman, D.J., Sodhi, N.S., Ng, P.K.L., Meier, R., Winker, K., Ingram, K.K., Das, I., 2007. Cryptic species as a window on diversity and conservation. Trends in Ecology & Evolution 22, 148-155.

Blaxter, M., 2004. The promise of a DNA taxonomy. Philosophical Transactions of the Royal Society B: Biological Sciences 359, 669-679.

Blaxter, M., Elsworth, B., Daub, J., 2004. DNA taxonomy of a neglected animal phylum: an unexpected diversity of tardigrades. Proceedings of the Royal Society B-Biological Sciences 271, S189-SS192.

Borisenko, A.V., Lim, K.B., Ivanova, N.V., Hanner, R., Herbert, P.D.N., 2008. DNA barcoding in surveys of small mammal communities: a field study in Suriname. Molecular Ecology Ressources 8, 471–479.

Borkent, A., 2008. The frog-biting midges of the world (Corethrellidae : Diptera). Zootaxa 1804, 3-456.

Borkent, A., Belton, P., 2006. Attraction of female *Uranotaenia lowii* (Diptera : Culicidae) to frog calls in Costa Rica. Canadian Entomologist 138, 91-94.

Borkent, A., Szadziewski, R., 1991. The first records of fossil Corethrellidae (Diptera). Insect Systematics & Evolution 22, 457-463.

Borkent, A., The Corethrellidae of Costa Rica. http://www.inbio.ac.cr/papers/Ceratopogonidae/index1.htm

Bortolus, A., 2008. Error cascades in the biological sciences: The unwanted consequences of using bad taxonomy in ecology. Ambio 37, 114-118.

Bretman, A., Tregenza, T., 2007. Strong, silent types: the rapid, adaptive disappearance of a sexual signal. Trends in Ecology & Evolution 22, 226-228.

Brooks, T.M., Mittermeier, R.A., Mittermeier, C.G., da Fonseca, G.A.B., Rylands, A.B., Konstant, W.R., Flick, P., Pilgrim, J.,

Oldfield, S., Magin, G., Hilton-Taylor, C., 2002. Habitat loss and extinction in the hotspots of biodiversity. Conservation Biology 16, 909-923.

Bruford, M.W., Bradley, D.G., Luikart, G., 2003. DNA markers reveal the complexity of livestock domestication. Nature Reviews Genetics 4, 900-910.

Camp, J.V., 2006. Host attraction and host selection in the family Corethrellidae (Wood and Borkent) (Diptera). In Graduate Faculty. p. 69. Georgia Southern University, Statesboro.

Carabia, J.P., 1945. The Vegetation of Sierra de Nipe, Cuba. Ecological Monographs 15, 321-341.

Castroviejo-Fisher, S., Guayasamin, J.M., Kok, P.J.R., 2009. Species status of Centrolene lema Duellman and Senaris, 2003 (Amphibia: Centrolenidae) revealed by Integrative Taxonomy. Zootaxa, 16-28.

Chao, A., Colwell, R.K., Lin, C.W., Gotelli, N.J., 2009. Sufficient sampling for asymptotic minimum species richness estimators. Ecology 90, 1125-1133.

Colwell, R.K., 2006. *EstimateS*: Statistical estimation of species richness and shared species from samples. Version 8.

Cresswell, J.E., 2000. Resource input and the community structure of larval infaunas of an eastern tropical pitcher plant *Nepenthes bicalcarata*. Ecological Entomology 25, 362-366.

Csada, R.D., James, P.C., Espie, R.H.M., 1996. The "File Drawer Problem" of Non-Significant Results: Does It Apply to Biological Research? Oikos 76, 591-593.

Cummings, M.A., Krafsur, E.S., 2005. Spatial diversity in mitochondrial cytochrome c oxidase in house flies. Medical and Veterinary Entomology 19, 53-59.

Das, I., 2005. A new species of *Polypedates* (Anura : Rhacophoridae) from Gunung Murud, Sarawak (Northwestern Borneo). Raffles Bulletin of Zoology 53, 265-270.

Das, I., 2008. Two new species of *Pelophryne* (Anura : Bufoidae) from Gunung Murud, Sarawak (Northwestern Borneo). Raffles Bulletin of Zoology 56, 435-443.

Das, I., Haas, A., 2003. A new species of *Kalophrynus* (Anura : Microhylidae) from the Highlands of north-central Borneo. Raffles Bulletin of Zoology 51, 109-113.

Das, I., Haas, A., 2005. A new species of *Rhacophorus* (Anura : Rhacophoridae) from Gunung Gading, Sarawak. Raffles Bulletin of Zoology 53, 257-263.

Dayrat, B., 2005. Towards integrative taxonomy. Biological Journal of the Linnean Society 85, 407-415.

de Leon, J.H., Jones, W.A., Samou, M., Morgan, D.J.W., 2006. Genetic and hybridisation evidence confirms that a geographic population of *Gonatocerus morrilli* (Hymenoptera : Mymaridae) from California is a new species: Egg parasitoids of the glassy-winged sharpshooter *Homalodisca coagulata* (Homoptera : Cicadellidae). Biological Control 38, 282-293.

Denise, T.S.H., Ali, F., Kutty, S.N., Meier, R., 2008. The need for specifying species concepts: How many species of silvered langurs (*Trachypithecus cristatus* group) should be recognized? Molecular Phylogenetics and Evolution 49, 688-689.

DeSalle, R., Egan, M., Siddall, M., 2005. The unholy trinity: taxonomy, species delimitation and DNA barcoding. Philosophical Transactions of the Royal Society B: Biological Sciences 360, 1905-1916.

Dorchin, N., Scott, E.R., Clarkin, C.E., Luongo, M.P., Jordan, S., Abrahamson, W.G., 2009. Behavioural, ecological and genetic evidence confirm the occurrence of host-associated differentiation in goldenrod gall-midges. Journal of Evolutionary Biology 22, 729-739.

Dunn, R.R., Harris, N.C., Colwell, R.K., Koh, L.P., Sodhi, N.S., 2009. The sixth mass coextinction: are most endangered species parasites and mutualists? Proceedings of the Royal Society B-Biological Sciences 276, 3037-3045.

Edwards, F.W., 1930. Notes on exotic Chaoborinae, with descriptions of new species (Diptera, Culicidae). Annals and Magazine of Natural History 10, 528–540.

Elias, M., Hill, R.I., Willmott, K.R., Dasmahapatra, K.K., Brower, A.V.Z., Malllet, J., Jiggins, C.D., 2007. Limited performance of DNA barcoding in a diverse community of tropical butterflies. Proceedings of the Royal Society B-Biological Sciences 274, 2881-2889.

Errard, C., Delabie, J., Jourdan, H., Hefetz, A., 2005. Intercontinental chemical variation in the invasive ant *Wasmannia auropunctata* (Roger) (Hymenoptera Formicidae): a key to the invasive success of a tramp species. Naturwissenschaften 92, 319-323.

Ferguson, J.W.H., 2002. On the use of genetic divergence for identifying species. Biological Journal of the Linnean Society 75, 509-516.

Folmer, O., Black, M., Hoeh, W., Lutz, R., Vrijenhoek, R., 1994. DNA primers for amplification of mitochondrial cytochrome *c* oxidase subunit I from diverse metazoan invertebrates. Molecular Marine Biology and Biotechnology 3, 294-299.

Fonseca, G., Derycke, S., Moens, T., 2008. Integrative taxonomy in two free-living nematode species complexes. Biological Journal of the Linnean Society 94, 737-753.

Garcia-Mudarra, J.L., Ibanez, C., Juste, J., 2009. The Straits of Gibraltar: barrier or bridge to Ibero-Moroccan bat diversity? Biological Journal of the Linnean Society 96, 434-450.

Gaston, K.J., O'Neill, M.A., 2004. Automated species identification: why not? Philosophical Transactions of the Royal Society of London Series B-Biological Sciences 359, 655-667.

Gaston, K.J., Quinn, R.M., Blackburn, T.M., Eversham, B.C., 1998. Species-range size distributions in Britain. Ecography 21, 361-370.

Gaubert, P., Wozencraft, W., Cordeiro-Estrela, P., Veron, G., raldine, 2005. Mosaics of Convergences and Noise in Morphological Phylogenies: What's in a Viverrid-Like Carnivoran? Systematic Biology 54, 865-894.

Giangrande, A., 2003. Biodiversity, Conservation, and the "Taxonomic Impediment". Aquatic Conservation: Marine and Freshwater Ecosystems 13, 451-459.

Gibson, G., Russell, I., 2006. Flying in tune: Sexual recognition in mosquitoes. Current Biology 16, 1311-1316.

Goloboff, P.A., Farris, J.S., Nixon, K.C., 2008. TNT, a free program for phylogenetic analysis. Cladistics 24, 774-786.

Gomez, A., Wright, P.J., Lunt, D.H., Cancino, J.M., Carvalho, G.R., Hughes, R.N., 2007. Mating trials validate the use of DNA barcoding to reveal cryptic speciation of a marine bryozoan taxon. Proceedings of the Royal Society B-Biological Sciences 274, 199-207.

Gopfert, M.C., Robert, D., 2000. Nanometre-range acoustic sensitivity in male and female mosquitoes. Proceedings of the Royal Society of London Series B-Biological Sciences 267, 453-457.

Gotelli, N.J., Colwell, R.K., 2001. Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. Ecology Letters 4, 379-391.

Grafe, T.U., Saat, H.B.M., Hagen, N., Kaluza, B., Berudin, Z.B.H., Wahab, M.A.B.A., 2008. Acoustic localisation of frog hosts by blood-sucking flies *Corethrella* Coquillet (Diptera: Corethrellidae) in Borneo. Australian Journal of Entomology 47, 350-354.

Grevelding, C.G., Kampkotter, A., Hollmann, M., Schafer, U., Kunz, W., 1996. Direct PCR on fruitflies and blood flukes without prior DNA isolation. Nucleic Acids Research 24, 4100-4101.

Grimaldi, D., Engels, M.S., 2005. Evolution of the Insects. Cambridge University Press., New York.

Groombridge, B. ed., 1992. Global Biodiversity: Status of the Earth's Living Resources. Chapman & Hall, London.

Hajibabaei, M., Smit, M.A., Janzen, D.H., Rodriguez, J.J., Whitfield, J.B., Hebert, P.D.N., 2006. A minimalist barcode can identify a specimen whose DNA is degraded. Molecular Ecology Notes 6, 959-964.

Hebert, P.D.N., Cywinska, A., Ball, S.L., deWaard, J.R., 2003. Biological identifications through DNA barcodes. Proceedings of the Royal Society B-Biological Sciences 270, 313-321.

Hickerson, M.J., Meyer, C.P., Moritz, C., 2006. DNA barcoding will often fail to discover new animal species over broad parameter space. Systematic Biology 55, 729-739.

Holloway, J.D., 2003. An addiction to Southeast Asian biogeography. Journal of Biogeography 30, 161-163.

Inger, R.F., Lian, T.F., 1996. Checklist of the frogs of Borneo. Raffles Bulletin of Zoology 44, 551-574.

Inger, R.F., Lian, T.F., Yambun, P., 2001. New species of toad of the genus *Ansonia* (Anura : Bufonidae) from Borneo. Raffles Bulletin of Zoology 49, 35-37.

Inger, R.F., Voris, H.K., 2001. The biogeographical relations of the frogs and snakes of Sundaland. Journal of Biogeography 28, 863-891.

Ingram, K.K., Laamanen, T.R., Puniamoorthy, N., Meier, R., 2008. Lack of morphological coevolution between male forelegs and female wings in *Themira* (Sepsidae: Diptera: Insecta). Biological Journal of the Linnean Society 93, 227-238.

Janzen, D., 2004. Now is the time. Philosophical Transactions of the Royal Society B: Biological Sciences 359, 731-732.

Jenkins, P.T., 1996. Free trade and exotic species introductions. Conservation Biology 10, 300-302.

Jennions, M.D., Mÿller, A.P., 2002. Publication bias in ecology and evolution: an empirical assessment using the 'trim and fill' method. Biological Reviews 77, 211-222.

Johnson, G.D., Paxton, J.R., Sutton, T.T., Satoh, T.P., Sado, T., Nishida, M., Miya, M., 2009. Deep-sea mystery solved: astonishing larval transformations and extreme sexual dimorphism unite three fish families. Biology Letters 5, 235-239.

Johnson, R.N., Young, D.G., Butler, J.F., 1993. Trypanosome transmission by *Corethrella wirthi* (Diptera, Chaoboridae) to the green treefrog, *Hyla cinerea* (Anura, Hylidae). Journal of Medical Entomology 30, 918-921.

Kobelt, M., Nentwig, W., 2008. Alien spider introductions to Europe supported by global trade. Diversity and Distributions 14, 273-280.

Koh, L.P., Dunn, R.R., Sodhi, N.S., Colwell, R.K., Proctor, H.C., Smith, V.S., 2004. Species Coextinctions and the Biodiversity Crisis. Science (Washington D C) 305, 1632-1634.

Koh, L.P., Dunn, R.R., Sodhi, N.S., Colwell, R.K., Proctor, H.C., Smith, V.S., 2004. Species Coextinctions and the Biodiversity Crisis. Science (Washington D C) 305, 1632-1634.

Krell, F.T., 2004. Parataxonomy vs. taxonomy in biodiversity studies-pitfalls and applicability of 'morphospecies' sorting. Biodiversity and Conservation 13, 795-812.

Laamanen, T.R., Petersen, F.T., Meier, R., 2003. Kelp flies and species concepts: The case of *Coelopa frigida* (Fabricius, 1805) and *C. nebularum* Aldrich, 1929 (Diptera: Coelopidae). Journal of Zoological Systematics and Evolutionary Research 41, 127-136.

Lambeck, K., 2004. Sea-level change through the last glacial cycle: geophysical, glaciological and palaeogeographic consequences. Comptes Rendus Geosciences 336, 677-689.

Lambert, D.M., Baker, A., Huynen, L., Haddrath, O., Hebert, P.D.N., Millar, C.D., 2005. Is a Large-Scale DNA-Based Inventory of Ancient Life Possible?, In Journal of Heredity 96, 279-284.

Le Breton, J., Delabie, J.H.C., Chazeau, J., Dejean, A., Jourdan, H., 2004. Experimental evidence of large-scale unicoloniality in the tramp ant *Wasmannia auropunctata* (Roger). Journal of Insect Behavior 17, 263-271.

Lee, C.E., 2000. Global phylogeography of a cryptic copepod species complex and reproductive isolation between genetically proximate "populations". Evolution 54, 2014-2027.

Lee, M.S.Y., 2000. A worrying systematic decline. Trends in Ecology & Evolution 15, 346-444.

Lenstra, J.A., Bradley, D.G., 1999. Systematics and Phylogeny of Cattle, In The Genetics of Cattle. eds R. Fries, A. Ruvinsky, pp. 1-11. CAB International, New York.

Lim, G.S., 2007. Conflict and Congruence between DNA and morphology in insect systematics., In Department of Biological Sciences. p. 67. National University of Singapore, Singapore.

Lin, S., Zhang, H., Hou, Y., Zhuang, Y., Miranda, L., 2009. High-Level Diversity of Dinoflagellates in the Natural Environment, Revealed by Assessment of Mitochondrial *cox1* and *cob* Genes for Dinoflagellate DNA Barcoding. Applied and Environmental Microbiology 75, 1279-1290.

Loftus, R.T., MacHugh, D.E., Bradley, D.G., Sharp, P.M., Cunningham, P., 1994. Evidence for two independent domestications of cattle. Proceedings of the National Academy of Sciences of the United States of America 91, 2757-2761.

Lohman, D.J., Peggie, D., Pierce, N.E., Meier, R., 2008. Phylogeography and genetic diversity of a widespread Old World butterfly, *Lampides boeticus* (Lepidoptera: Lycaenidae). BMC Evolutionary Biology 8, 301.

Lounibos, L., Makhni, S., Alto, B., Kesavaraju, B., 2008. Surplus Killing by Predatory Larvae of Corethrella appendiculata : Prepupal Timing and Site-Specific Attack on Mosquito Prey. Journal of Insect Behavior 21, 47-54.

Lukashevich, E.D., Mostovski, M.B., 2003. Haematophagous insects in the fossil record. Paleontological Journal 37, 153-161.

Lukhtanov, V.A., Sourakov, A., Zakharov, E.V., Hebert, P.D.N., 2009. DNA barcoding Central Asian butterflies: increasing geographical dimension does not significantly reduce the success of species identification. Molecular Ecology Resources 9, 1302-1310.

Mangold, J.R., 1978. Attraction of *Euphasiopteryx ochracea*, Corethrella sp. and Gryllids to Broadcast Songs of the Southern Mole Cricket. The Florida Entomologist 61, 57-61.

Mathews, L.M., 2006. Cryptic biodiversity and phylogeographical patterns in a snapping shrimp species complex. Molecular Ecology 15, 4049-4063.

Matsui, M., Mohamed, M., Shimada, T., Sudin, A., 2007. Resurrection of *Staurois parvus* from *S. tuberilinguis* from Borneo (Amphibia, Ranidae). Zoological Science 24, 101-106.

Maurer, D., 2000. The Dark Side of Taxonomic Sufficiency (TS). Marine Pollution Bulletin 40, 98-101.

Mayden, R.L., 1999. Consilience and a hierarchy of species concepts: Advances toward closure on the species puzzle. Journal of Nematology. June, 1999; 31, 95-116.

Mayr, E., 2000. The Biological Species Concept, In Species Concepts and Phylogenetic Theory: a Debate. eds Q.D. Wheeler, R. Meier, pp. 17-29. Columbia University Press, New York.

McGlynn, T.P., 1999. The worldwide transfer of ants: geographical distribution and ecological invasions. Journal of Biogeography 26, 535-548.

McKeever, S., 1977. Observations of Corethrella feeding on tree frogs (*Hyla*). Mosquito News 37, 522-523.

McKeever, S., French, F.E., 1991. *Corethrella* (Diptera: Corethrellidae) of eastern North America: laboratory life history and field responses to anuran calls. Annals of the Entomological Society of America 84, 493-497.

McKeever, S., French, F.E., 2000. Corethrellidae (Diptera), Vectors of Present and Perhaps Some of the Earliest Anuran Trypanosomes, In International Virtual Conferences in Veterinary Medicine. http://www.vet.uga.edu/vpp/ivcvm/index.php

McKeever, S., Hartberg, W.K., 1980. An effective method for trapping adult female *Corethrella* (Diptera, Chaoboridae). Mosquito News 40, 111-112.

Mehdiabadi, N.J., Jack, C.N., Farnham, T.T., Platt, T.G., Kalla, S.E., Shaulsky, G., Queller, D.C., Strassmann, J.E., 2006. Social evolution: Kin preference in a social microbe. Nature 442, 881-882.

Meier, R., 1995. Cladistic analysis of the Sepsidae (Cyclorrhapha: Diptera) based on a comparative scanning electron microscopic study of larvae. Systematic Entomology 20, 99-128.

Meier, R., 2007. *Themira biloba* Andersson 1975 (Diptera: Sepsidae), a species from Manhattan's Central Park that is new to the Nearctic region. Journal of the New York Entomological Society 114, 176-177.

Meier, R., 2008. DNA sequences in taxonomy - Opportunities and challenges. New Taxonomy 76, 95-127.

Meier, R., Kwong, S., Vaidya, G., Ng, P.K.L., 2006. DNA barcoding and taxonomy in Diptera: a tale of high intraspecific variability and low identification success. Systematic Biology 55, 715-728.

Meier, R., Kwong, S., Vaidya, G., Ng, P.K.L., 2006. DNA barcoding and taxonomy in Diptera: a tale of high intraspecific variability and low identification success. Systematic Biology 55, 715-728.

Meier, R., Pont, A.C., 2000. Family Sepsidae, In Manual of Palaearctic Diptera, Appendix,. eds L. Papp, B. Darvas, pp. 367-386, Science Herald, Budapest.

Meier, R., Willmann, R., 2000. A defense of the Hennigian species concept, In Species Concepts and Phylogenetic Theory: a Debate. eds Q.D. Wheeler, R. Meier, pp. 167-178. Columbia University Press, New York.

Meier, R., Zhang, G.Y., Ali, F., 2008. The Use of Mean Instead of Smallest Interspecific Distances Exaggerates the Size of the "Barcoding Gap" and Leads to Misidentification. Systematic Biology 57, 809-813.

Melander, A.L., Spuler, A., 1917. The dipterous families Sepsidae and Piophilidae. Washington Agricultural Experiment Station Technical Bulletin 143, 1-103.

Memon, N., Meier, R., Mannan, A., Su Feng-Yi, K., 2006. On the use of DNA sequences for determining the species limits of a polymorphic new species in the stink bug genus *Halys* (Heteroptera: Pentatomidae) from Pakistan. Systematic Entomology 31, 703-710.

Memon, N., Meier, R., Mannan, A., Su Feng-Yi, K., 2006. On the use of DNA sequences for determining the species limits of a polymorphic new species in the stink bug genus *Halys* (Heteroptera: Pentatomidae) from Pakistan. Systematic Entomology 31, 703-710.

Meyer, C.P., Paulay, G., 2005. DNA barcoding: Error rates based on comprehensive sampling. PloS Biology 3, 2229-2238.

Mishler, B., Theriot, E., 2000. The Phylogenetic Species Concept sensu Mishler and Theriot: Monophyly, Apomorphy, and Phylogenetic Species Concepts, In Species Concepts and Phylogenetic Theory: a Debate. eds Q.D. Wheeler, R. Meier, pp. 44-54. Columbia University Press, New York.

Mogi, M., Chan, K.L., 1996. Predatory habits of dipteran larvae inhabiting *Nepenthes* pitchers. Raffles Bulletin of Zoology 44, 233-245.

Mogi, M., Yong, H.S., 1992. Aquatic arthropod communities in Nepenthes pitchers: the role of niche differentiation, aggregation, predation and competition in community organization. Oecologia 90, 172-184.

Mohamad, K., Olsson, M., van Tol, H.T.A., Mikko, S., Vlamings, B.H., Andersson, G.r., Rodríguez-Martínez, H., Purwantara, B., Paling, R.W., Colenbrander, B., Lenstra, J.A., 2009. On the Origin of Indonesian Cattle. PLoS ONE 4, e5490.

Monaghan, M.T., Balke, M., Pons, J., Vogler, A.P., 2006. Beyond barcodes: complex DNA taxonomy of a South Pacific Island radiation. Proceedings of the Royal Society B: Biological Sciences 273, 887-893.

Moritz, C., Cicero, C., 2004. DNA Barcoding: Promise and Pitfalls, In PLoS Biology 2. e354.

Moritz, C., Patton, J.L., Schneider, C.J., Smith, T.B., 2000. Diversification of rainforest faunas: An integrated molecular approach. Annual Review of Ecology and Systematics 31, 533-563.

Munari, L., 1983. Study of the Sciomyzidae (Diptera, Cyclorrhapha) for the biological control of mollusks, hosts of the parasitic helminths, etiological agents of bilharziasis and distomiasis of man and cattle. Societa Veneziana di Scienze Naturali Lavori 8, 9-30.

Mutanen, M., 2005. Delimitation difficulties in species splits: a morphometric case study on the *Euxoa tritici* complex (Lepidoptera, Noctuidae). Systematic Entomology 30, 632-643.

Myers, N., Mittermeier, R.A., 2003. Impact and acceptance of the hotspots strategy: Response to Ovadia and to Brummitt and Lughadha. Conservation Biology 17, 1449-1450.

Myers, N., Mittermeier, R.A., Mittermeier, C.G., da Fonseca, G.A.B., Kent, J., 2000. Biodiversity hotspots for conservation priorities. Nature (London) 403, 853-858.

Ng, P.K.L., Tay, J.B., Lim, K.K.P., 1994. Diversity and conservation of blackwater fishes in Peninsular Malaysia, particularly in the North Selangor peat swamp forest. Hydrobiologia 285, 203-218.

Nelson, G.J., Platnick, N., 1981. Systematics and biogeography: cladistics and vicariance. Columbia University Press, New York.

Nishiguchi, M.K., Doukakis, P., Egan, M., Kizirian, D., Phillips, A., Prendini, L., Rosenbaum, H.C., Torres, E., Wyner, Y., DeSalle, R., Giribet, G., 2002. DNA Isolation Procedures, In Techniques in Molecular Systematics and Evolution. eds R. DeSalle, G. Giribet, W. Wheeler, p. 450. Birkhäuser Verlag, Basel.

Ozerov, A.L., 1998. A review of the genus *Themira* Robineau-Desvoidy, 1830 (Diptera: Sepsidae) of the world, with a revision of the North American species. Russian Entomology Journal 7, 169-208.

Ozerov, A.L., 2000. Revision of the Nearctic species of the genus *Sepsis* Fallen, 1810 (Diptera: Sepsidae). Russian Entomological Journal 9, 161-176.

Ozerov, A.L., 2005. World catalogue of the family Sepsidae (Insecta: Diptera). Zoologicheskie issledovania (Zoological Studies) 8, 1-74.

Padial, J.M., Castroviejo-Fisher, S., Kohler, J., Vila, C., Chaparro, J.C., De la Riva, I., 2009. Deciphering the products of evolution at the species level: the need for an integrative taxonomy. Zoologica Scripta 38, 431-447.

Padial, J.M., de la Riva, I., 2009. Integrative taxonomy reveals cryptic Amazonian species of *Pristimantis* (Anura: Strabomantidae). Zoological Journal of the Linnean Society 155, 97-122.

Page, S. 2002. Biodiversity information on peat swamp forest in Southeast Asia. In STRAPEAT status report: Strategies for implementing sustainable management of peatlands in Borneo. http://www.strapeat.alterra.nl/

Page, T.J., Choy, S.C., Hughes, J.M., 2005. The taxonomic feedback loop: symbiosis of morphology and molecules. Biology Letters 1, 139-142.

Paterson, H.E.H., 1985. The recognition species concept of species, In Transvaal Museum Monograph. ed. E. Vrba, pp. 21-29. Transvaal Museum, Pretoria.

Petersen, F.T., Damgaard, J., Meier, R., 2007. How many DNA sequences are needed for solving a taxonomic problem? The case of two parapatric species of louse flies (Diptera: Hippoboscidae: *Ornithomya* (Latreille, 1802). Arthropod Systematics & Phylogeny 65, 111– 117.

Pfenninger, M., Schwenk, K., 2007. Cryptic animal species are homogeneously distributed among taxa and biogeographical regions. BMC Evolutionary Biology 7, 121.

Pont, A.C., Meier, R., 2002. The Sepsidae (Diptera) of Europe. Brill Academic Publishers, Leiden, the Netherlands.

Posada, D., Crandall, K.A., 1998. MODELTEST: testing the model of DNA substitution. Bioinformatics 14, 817-818.

Pounds, J.A., Bustamante, M.R., Coloma, L.A., Consuegra, J.A., Fogden, M.P.L., Foster, P.N., La Marca, E., Masters, K.L., Merino-Viteri, A., Puschendorf, R., Ron, S.R., Sanchez-Azofeifa, G.A., Still, C.J., Young, B.E., 2007. Global warming and amphibian losses; The proximate cause of frog declines? (Reply). Nature 447, E5-E6.

Puniamoorthy, N., Su, K., Meier, R., 2008. Bending for love: losses and gains of sexual dimorphisms are strictly correlated with changes in the mounting position of sepsid flies (Sepsidae: Diptera). BMC Evolutionary Biology 8, 155.

Reusch, T., Blanckenhorn, W.U., 1998. Quantitative genetics of the dung fly *Sepsis cynipsea*: Cheverud's conjecture revisited. Heredity 81, 111-119.

Riedel, A., Daawia, D., Balke, M., 2009. Deep cox1 divergence and hyperdiversity of *Trigonopterus*) weevils in a New Guinea mountain range (Coleoptera, Curculionidae). Zoologica Scripta 39, 63-74.

Rieley, J.O., Ahmad-Shah, A.A., 1992. Tropical Lowland Peatlands of Southeast Asia, In Intergrating Planning and Management of Tropical Lowland Peatlands. eds E. Maltby, C.P. Immirzi, R.J. Safford. Samara Publishing Limited, Cardigan, Dyfed, UK, Cisarua, Indonesia.

Robert, D., 2005. Directional Hearing in Insects, In Sound Source Localization. pp. 6-35.

Robert, D., 2009. Insect Bioacoustics: Mosquitoes Make an Effort to Listen to Each Other. Current Biology 19, R446-R449.

Robert, D., Miles, R.N., Hoy, R.R., 1996. Directional hearing by mechanical coupling in the parasitoid fly *Ormia ochracea*. Journal of Comparative Physiology a-Sensory Neural and Behavioral Physiology 179, 29-44.

Robert, D., Miles, R.N., Hoy, R.R., 1999. Tympanal hearing in the sarcophagid parasitoid fly *Emblemasoma* sp.: The biomechanics of directional hearing. Journal of Experimental Biology 202, 1865-1876.

Rochlin, I., Santoriello, M.P., Mayer, R.T., Campbell, S.R., 2007. Improved high-throughput method for molecular identification of *Culex* mosquitoes. Journal of the American Mosquito Control Association 23, 488-491.

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. Molecular Phylogenetics and Evolution 44, 325-345.

Roe, A.D., Sperling, F.A.H., 2007. Population structure and species boundary delimitation of cryptic *Dioryctria* moths: an integrative approach. Molecular Ecology 16, 3617–3633.

Rohr, J.R., Raffel, T.R., Romansic, J.M., McCallum, H., Hudson, P.J., 2008. Evaluating the links between climate, disease spread, and amphibian declines. Proceedings of the National Academy of Sciences of the United States of America 105, 17436-17441.

Rubinoff, D., 2006. Utility of Mitochondrial DNA Barcodes in Species Conservation. Conservation Biology 20, 1026-1033.

Rubinoff, D., Holland, B., 2005. Between Two Extremes: Mitochondrial DNA is neither the Panacea nor the Nemesis of Phylogenetic and Taxonomic Inference. Systematic Biology 54, 952-961.

Sathiamurthy, E., Voris, H.K., 2006. Maps of Holocene Sea Level Transgression and Submerged Lakes on the Sunda Shelf. The Natural History Journal of Chulalongkorn University Supplement 2, 1-44.

Scheiner, S.M., 1993. Genetics and evolution of phenotypic plasticity. Annual Review of Ecology and Systematics 24, 35-68.

Scotland, R.W., Olmstead, R.G., Bennett, J.R., 2003. Phylogeny reconstruction: The role of morphology. Systematic Biology 52, 539-548.

Silvertown, J., McConway, K.J., 1997. Does "Publication Bias" Lead to Biased Science? Oikos 79, 167-168.

Simon, C., Frati, F., Beckenbach, A., Crespi, A., Liu, H., Flook, P., 1994. Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. Annals of the Entomological Society of America 87, 651-696.

Simon, C., Frati, F., Beckenbach, A., Crespi, A., Liu, H., Flook, P., 1994. Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. Annals of the Entomological Society of America 87, 651-696.

Singh, R.S., 2003. Darwin to DNA, molecules to morphology: the end of classical population genetics and the road ahead. Genome 46, 938-942.

Skerratt, L., Berger, L., Speare, R., Cashins, S., McDonald, K., Phillott, A., Hines, H., Kenyon, N., 2007. Spread of

Chytridiomycosis Has Caused the Rapid Global Decline and Extinction of Frogs. EcoHealth 4, 125-134.

Smith, M., Fisher, B., Hebert, P., 2005. DNA barcoding for effective biodiversity assessment of a hyperdiverse arthropod group: the ants of Madagascar. Philosophical Transactions of the Royal Society B: Biological Sciences 360, 1825-1834.

Smith, M.A., Fernandez-Triana, J., Roughley, R., Hebert, P.D.N., 2009. DNA barcode accumulation curves for understudied taxa and areas. Molecular Ecology Resources 9, 208-216.

Smith, M.A., Fisher, B.L., Hebert, P.D.N., 2005. DNA barcoding for effective biodiversity assessment of a hyperdiverse arthropod group: the ants of Madagascar. Philosophical Transactions of the Royal Society B-Biological Sciences 360, 1825-1834.

Smith, R.M., Baker, R.H.A., Malumphy, C.P., Hockland, S., Hammon, R.P., Ostoj·-Starzewski, J.C., Collins, D.W., 2007. Recent non-native invertebrate plant pest establishments in Great Britain: origins, pathways, and trends. Agricultural and Forest Entomology 9, 307-326.

Soares, P., Trejaut, J.A., Loo, J.-H., Hill, C., Mormina, M., Lee, C.-L., Chen, Y.-M., Hudjashov, G., Forster, P., Macaulay, V., Bulbeck, D., Oppenheimer, S., Lin, M., Richards, M.B., 2008. Climate Change and Postglacial Human Dispersals in Southeast Asia. Molecular Biology and Evolution 25, 1209-1218.

Sodhi, N.S., Koh, L.P., Brook, B.W., Ng, P.K.L., 2004. Southeast Asian biodiversity: an impending disaster. Trends in Ecology & Evolution 19, 654-660.

Sodhi, N.S., Koh, L.P., Brook, B.W., Ng, P.K.L., 2004. Southeast Asian biodiversity: an impending disaster. Trends in Ecology & Evolution 19, 654-660.

Steinke, D., Zemlak, T.S., Hebert, P.D.N., 2009. Barcoding Nemo: DNA-Based Identifications for the Ornamental Fish Trade. PLoS ONE 4, e6300.

Stuart, B.L., Inger, R.F., Voris, H.K., 2006. High level of cryptic species diversity revealed by sympatric lineages of Southeast Asian forest frogs. Biology Letters 2, 470-474.

Stuebing, R.B., Wong, A., 2000. A new species of frog, *Philautus erythrophthalmus* (Rhacophoridae) from Southwestern Sabah, Malaysia. Raffles Bulletin of Zoology 48, 293-296.

Su, K.F.Y., Kutty, S.N., Meier, R., 2008. Morphology versus molecules: the phylogenetic relationships of Sepsidae (Diptera:

Cyclorrhapha) based on morphology and DNA sequence data from ten genes. Cladistics 24, 902-916.

Tan, D.S.H., Ang, Y., Lim, G.S., Meier, R., 2009. From DNA barcoding to integrative taxonomy: an iterative process involving DNA sequences, morphology, and behaviour leads to the resurrection of *Sepsis pyrrhosoma* (Sepsidae: Diptera). Zoologica Scripta 39, 51-61.

Tautz, D., Arctander, P., Minelli, A., Thomas, R.H., Vogler, A.P., 2003. A plea for DNA taxonomy. Trends in Ecology and Evolution 18, 70-74.

Terlizzi, A., Bevilacqua, S., Fraschetti, S., Boero, F., 2003. Taxonomic sufficiency and the increasing insufficiency of taxonomic expertise. Marine Pollution Bulletin 46, 556-561.

Thomas, J.A., Telfer, M.G., Roy, D.B., Preston, C.D., Greenwood, J.J.D., Asher, J., Fox, R., Clarke, R.T., Lawton, J.H., 2004. Comparative losses of British butterflies, birds, and plants and the global extinction crisis. Science (Washington D C) 303, 1879-1881.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G., 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucl. Acids Res. 25, 4876-4882.

Trontelj, P., Douady, C.J., Fiser, C., Gibert, J., Goricki, S., Lefebure, T., Sket, B., Zaksek, V., 2009. A molecular test for cryptic diversity in ground water: how large are the ranges of macro-stygobionts? Freshwater Biology 54, 727-744.

Trontelj, P., Fišer, C., 2009. Cryptic species diversity should not be trivialised. Systematics and Biodiversity 7, 1-3.

Tuck, E.J., Windmill, J.F.C., Robert, D., 2009. Hearing in tsetse flies? Morphology and mechanics of a putative auditory organ. Bulletin of Entomological Research 99, 107-119.

Valdecasas, A.G., Williams, D., Wheeler, Q.D., 2008. 'Integrative taxonomy' then and now: a response to Dayrat (2005). Biological Journal of the Linnean Society 93, 211-216.

Vogler, A.P., Monaghan, M.T., 2007. Recent advances in DNA taxonomy. Journal of Zoological Systematics and Evolutionary Research 45, 1-10.

Wahlberg, N., Oliveira, R., Scott, J.A., 2003. Phylogenetic relationships of *Phyciodes* butterfly species (Lepidoptera : Nymphalidae): complex mtDNA variation and species delimitations. Systematic Entomology 28, 257-273.

Wake, D.B., Vredenburg, V.T., 2008. Are we in the midst of the sixth mass extinction? A view from the world of amphibians. Proceedings of the National Academy of Sciences 105, 11466-11473.

Waugh, J., 2007. DNA barcoding in animal species: progress, potential and pitfalls. BioEssays 29, 188-197.

Wheeler, Q., 2004. Taxonomic triage and the poverty of phylogeny. Philosophical Transactions of the Royal Society B: Biological Sciences 359, 571-583.

Wheeler, Q.D., Meier, R., 2000. Species Concepts and Phylogenetic Theory. A Debate. Columbia University Press, New York.

Wheeler, Q.D., Platnick , N.I., 2000. The Phylogenetic Species Concept sensu Wheeler and Platnick, In Species Concepts and Phylogenetic Theory: a Debate. eds Q.D. Wheeler, R. Meier, pp. 55-69. Columbia University Press, New York.

Wiens, J.J., 1999. Polymorphism in systematics and comparative biology. Annual Reviews of Ecology and Systematics 30, 327-362.

Will, K., Mishler, B., Wheeler, Q., 2005. The Perils of DNA Barcoding and the Need for Integrative Taxonomy. Systematic Biology 54, 844-851.

Will, K.W., Rubinoff, D., 2004. Myth of the molecule: DNA barcodes for species cannot replace morphology for identification and classification. Cladistics 20, 47-55.

Williams, J.A., Edman, J.D., 1968. Occurrence of blood meals in two species of *Corethrella* in Florida. Annals of the Entomological Society of America 61, 1336-1338

Wirta, H., 2009. Complex phylogeographical patterns, introgression and cryptic species in a lineage of Malagasy dung beetles (Coleoptera: Scarabaeidae). Biological Journal of the Linnean Society 96, 942-955.

Yack, J.E., 2004. The structure and function of auditory chordotonal organs in insects. Microscopy Research and Technique 63, 315-337.

Zaksek, V., Sket, B., Gottstein, S., Franjevic, D., Trontelj, P., 2009. The limits of cryptic diversity in groundwater: phylogeography of the cave shrimp *Troglocaris anophthalmus* (Crustacea: Decapoda: Atyidae). Molecular Ecology 18, 931-946.

Zhang, G.Y., 2007. DNA Barcoding and DNA Taxonomy of Animal Life-A Quantitative Assessment, In Department of Biological Sciences. p. 44. National University of Singapore, Singapore.

Zwickl, D.J., 2006. Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. Ph.D. dissertation, The University of Texas at Austin.

APPENDIX

Taxa	ID	Locality	Country	Bases
Allosepsis indica	I	Brunei	Brunei	713
Allosepsis indica	II	Brunei	Brunei	712
Allosepsis indica	I	Angkor Wat	Cambodia	531
Allosepsis indica	II	Angkor Wat	Cambodia	662
Allosepsis indica	III	Angkor Wat	Cambodia	662
Allosepsis indica	IV	Angkor Wat	Cambodia	681
Allosepsis indica	V	Angkor Wat	Cambodia	681
Allosepsis indica	I	Guizhou	China	530
Allosepsis indica	II	Guizhou	China	620
Allosepsis indica	III	Guizhou	China	594
Allosepsis indica	IV	Guizhou	China	488
Allosepsis indica	I	Yunnan	China	646
Allosepsis indica	I	Alleppy	India	624
Allosepsis indica	II	Alleppy	India	595
Allosepsis indica	IV	Alleppy	India	584
Allosepsis indica	225	Padang Pariaman	Indonesia	554
Allosepsis indica	226	Padang Pariaman	Indonesia	557
Allosepsis indica	227	Padang Pariaman	Indonesia	452
Allosepsis indica	228	Padang Pariaman	Indonesia	568
Allosepsis indica	239	Padang Pariaman	Indonesia	503
Allosepsis indica	241	Padang Pariaman	Indonesia	509
Allosepsis indica	A	PadangPariaman	Indonesia	631
Allosepsis indica	C	Padang Pariaman	Indonesia	657
Allosepsis indica	D	Padang Pariaman	Indonesia	684

Allosepsis indica	E	Padang Pariaman	Indonesia	504
Allosepsis indica	F	Padang Pariaman	Indonesia	653
Allosepsis indica	I	Sulawesi	Indonesia	534
Allosepsis indica	II	Sulawesi	Indonesia	354
Allosepsis indica	V	Sulawesi	Indonesia	587
Allosepsis indica	II	Batu Kitang	Malaysia	679
Allosepsis indica	III	Batu Kitang	Malaysia	679
Allosepsis indica	I	Forest Research Institute Malaysia	Malaysia	561
Allosepsis indica	I	Fraser Hill	Malaysia	463
Allosepsis indica	II	Fraser Hill	Malaysia	447
Allosepsis indica	I	Ipoh	Malaysia	417
Allosepsis indica	III	Johor	Malaysia	516
Allosepsis indica	V	Johor	Malaysia	559
Allosepsis indica	VI	Johor	Malaysia	582
Allosepsis indica	VII	Johor	Malaysia	497
Allosepsis indica	I	Malacca	Malaysia	537
Allosepsis indica	K	Malacca	Malaysia	664
Allosepsis indica		Pulau Tioman	Malaysia	639
Allosepsis indica	Ex	Pulau Tioman	Malaysia	647
Allosepsis indica	ExII	Pulau Tioman	Malaysia	662
Allosepsis indica	ExIII	Pulau Tioman	Malaysia	646
Allosepsis indica	III	Selangor	Malaysia	669
Allosepsis indica	IV	Selangor	Malaysia	682
Allosepsis indica	V	Selangor	Malaysia	619
Allosepsis indica	CII	Terengganu	Malaysia	628
Allosepsis indica	CIII	Terengganu	Malaysia	631
Allosepsis indica	CIV	Terengganu	Malaysia	674
Allosepsis indica	CV	Terengganu	Malaysia	659
Allosepsis indica	CVI	Terengganu	Malaysia	670

Allosepsis indica	CVII	Terengganu	Malaysia	644
Allosepsis indica	I	Terengganu	Malaysia	566
Allosepsis indica	II	Terengganu	Malaysia	596
Allosepsis indica	III	Terengganu	Malaysia	606
Allosepsis indica	IV	Terengganu	Malaysia	672
Allosepsis indica	V	Terengganu	Malaysia	683
Allosepsis indica	WCI	Terengganu	Malaysia	664
Allosepsis indica	WCII	Terengganu	Malaysia	686
Allosepsis indica	WCIII	Terengganu	Malaysia	641
Allosepsis indica	WCIV	Terengganu	Malaysia	633
Allosepsis indica	WCV	Terengganu	Malaysia	608
Allosepsis indica	I	Krabi	Thailand	599
Allosepsis indica	II	Krabi	Thailand	617
Allosepsis indica	III	Krabi	Thailand	646
Allosepsis indica	IV	Krabi	Thailand	663
Allosepsis indica	V	Krabi	Thailand	631
Australosepsis frontalis	259	Brunei	Brunei	508
Australosepsis frontalis	260	Brunei	Brunei	385
Australosepsis frontalis	261	Brunei	Brunei	508
Australosepsis frontalis	II	Angkor Wat	Cambodia	532
Australosepsis frontalis	18	Bukit Tingei	Indonesia	575
Australosepsis frontalis	19	Bukit Tingei	Indonesia	575
Australosepsis frontalis	20	Bukit Tingei	Indonesia	575
Australosepsis frontalis	21	Bukit Tingei	Indonesia	575
Australosepsis frontalis	120	Bukit Tingei	Indonesia	712

Australosepsis frontalis	121	Bukit Tingei	Indonesia	712
Australosepsis frontalis	122	Bukit Tingei	Indonesia	712
Australosepsis frontalis	123	Bukit Tingei	Indonesia	712
Australosepsis frontalis	252	Gunung Halimun Salak National Park	Indonesia	508
Australosepsis frontalis	255	Gunung Halimun Salak National Park	Indonesia	575
Australosepsis frontalis		Kalimantan	Indonesia	628
Australosepsis frontalis	I	Kalimantan	Indonesia	619
Australosepsis frontalis	II	Kalimantan	Indonesia	592
Australosepsis frontalis	175	Padang Pariaman	Indonesia	641
Australosepsis frontalis	242	Padang Pariaman	Indonesia	506
Australosepsis frontalis	249	Padang Pariaman	Indonesia	475
Australosepsis frontalis	250	Padang Pariaman	Indonesia	536
Australosepsis frontalis	248	Padang Pariaman	Indonesia	533
Australosepsis frontalis	4	Sulawesi	Indonesia	530
Australosepsis frontalis	6	Sulawesi	Indonesia	658
Australosepsis frontalis	I	Batu Kitang	Malaysia	649
Australosepsis frontalis	II	Batu Kitang	Malaysia	678
Australosepsis frontalis	III	Batu Kitang	Malaysia	679
Australosepsis frontalis	IV	Batu Kitang	Malaysia	682

Australosepsis frontalis	V	Batu Kitang	Malaysia	678
Australosepsis frontalis	VI	Batu Kitang	Malaysia	632
Australosepsis frontalis		Ipoh	Malaysia	568
Australosepsis frontalis	II	Ipoh	Malaysia	604
Australosepsis frontalis	IV	Ipoh	Malaysia	659
Australosepsis frontalis	I	Malacca	Malaysia	594
Australosepsis frontalis		Pulau Tioman	Malaysia	561
Australosepsis frontalis	II	Pulau Tioman	Malaysia	519
Australosepsis frontalis	IV	Pulau Tioman	Malaysia	659
Australosepsis frontalis	4	Mt. Camagong	Philippines	620
Australosepsis frontalis	5	Mt. Camagong	Philippines	620
Australosepsis frontalis	6	Mt. Camagong	Philippines	620
Australosepsis frontalis	7	Mt. Camagong	Philippines	620
Australosepsis frontalis	8	Mt. Camagong	Philippines	638
Australosepsis frontalis	9	Mt. Camagong	Philippines	638
Australosepsis frontalis	10	Mt. Camagong	Philippines	634
Australosepsis frontalis	11	Mt. Camagong	Philippines	634
Australosepsis frontalis	13	Mt. Camagong	Philippines	634
Australosepsis frontalis	14	Mt. Camagong	Philippines	634

Australosepsis frontalis	15	Mt. Camagong	Philippines	543
Australosepsis frontalis	16	Mt. Camagong	Philippines	627
Australosepsis frontalis	17	Mt. Camagong	Philippines	636
Australosepsis frontalis	164	Mt. Camagong	Philippines	541
Australosepsis frontalis	165	Mt. Camagong	Philippines	622
Australosepsis frontalis	166	Mt. Camagong	Philippines	622
Australosepsis frontalis	167	Mt. Camagong	Philippines	552
Australosepsis frontalis	168	Mt. Camagong	Philippines	622
Australosepsis frontalis	182	Mt. Camagong	Philippines	641
Australosepsis frontalis	183	Mt. Camagong	Philippines	565
Australosepsis frontalis	I	Taiwan	China	631
Australosepsis frontalis	II	Taiwan	China	546
Australosepsis frontalis	III	Taiwan	China	529
Australosepsis frontalis	6	Chiang Mai	Thailand	607
Australosepsis frontalis	7	Chiang Mai	Thailand	548
Australosepsis frontalis	I	Trang	Thailand	546
Australosepsis frontalis	II	Trang	Thailand	566
Australosepsis frontalis	III	Trang	Thailand	626
Australosepsis niveipennis	A	New South Wales	Australia	609

Australosepsis niveipennis	B	New South Wales	Australia	614
Australosepsis niveipennis	C	New South Wales	Australia	684
Australosepsis niveipennis	D	New South Wales	Australia	691
Australosepsis niveipennis	290	Brunei	Brunei	567
Australosepsis niveipennis	I	Calicut	India	604
Australosepsis niveipennis	II	Calicut	India	535
Australosepsis niveipennis	III	Calicut	India	421
Australosepsis niveipennis	I	Batu Kitang	Malaysia	677
Australosepsis niveipennis		Pulau Tioman	Malaysia	655
Australosepsis niveipennis	II	Pulau Tioman	Malaysia	660
Australosepsis niveipennis	III	Pulau Tioman	Malaysia	624
Australosepsis niveipennis	III	Taiwan	China	658
Australosepsis niveipennis	V	Taiwan	China	642
Australosepsis niveipennis	VI	Taiwan	China	547
Australosepsis niveipennis	VII	Taiwan	China	509
Australosepsis niveipennis	4	Chiang Mai	Thailand	627
Australosepsis niveipennis	5	Chiang Mai	Thailand	678
Australosepsis niveipennis	I	Grand Comoros	Union of the Comoros	512
Australosepsis niveipennis		United Arab Emirates	United Arab Emirates	682
Meroplius		Gunung Halimun-Salak	Indonesia	666

fasciculatus		National Park		
Meroplius fasciculatus	139	Gunung Halimun-Salak National Park	Indonesia	640
Meroplius fasciculatus	140	Gunung Halimun-Salak National Park	Indonesia	707
Meroplius fasciculatus	141	Gunung Halimun-Salak National Park	Indonesia	707
Meroplius fasciculatus	142	Gunung Halimun-Salak National Park	Indonesia	707
Meroplius fasciculatus	143	Gunung Halimun-Salak National Park	Indonesia	707
Meroplius fasciculatus	126	Sawahlunto	Indonesia	627
Meroplius fasciculatus	1	Sulawesi	Indonesia	597
Meroplius fasciculatus	2	Sulawesi	Indonesia	589
Meroplius fasciculatus	3	Sulawesi	Indonesia	574
Meroplius fasciculatus	4	Sulawesi	Indonesia	586
Meroplius fasciculatus	5	Sulawesi	Indonesia	455
Meroplius fasciculatus	6	Sulawesi	Indonesia	568
Meroplius fasciculatus	7	Sulawesi	Indonesia	573
Parapaleosepsis plebeia	A	New South Wales	Australia	696
Parapaleosepsis plebeia	B	New South Wales	Australia	703
Parapaleosepsis plebeia	C	New South Wales	Australia	707
Parapaleosepsis plebeia	D	New South Wales	Australia	693
Parapaleosepsis plebeia	III	New South Wales	Australia	706

Parapaleosepsis plebeia	II	Perth	Australia	560
Parapaleosepsis plebeia	III	Perth	Australia	607
Parapaleosepsis plebeia	IV	Perth	Australia	638
Parapaleosepsis plebeia	V	Perth	Australia	639
Parapaleosepsis plebeia	VI	Perth	Australia	626
Parapaleosepsis plebeia	K	Woollogong	Australia	665
Parapaleosepsis plebeia	I	Goroka	Papua New Guinea	572
Parapaleosepsis plebeia	II	Goroka	Papua New Guinea	589
Parapaleosepsis plebeia	IV	Goroka	Papua New Guinea	554
Parapaleosepsis plebeia	eX	Goroka	Papua New Guinea	512
Sepsis coprophila	55	Bukit Tingei	Indonesia	597
Sepsis coprophila	56	Bukit Tingei	Indonesia	627
Sepsis coprophila	57	Bukit Tingei	Indonesia	627
Sepsis coprophila	58	Bukit Tingei	Indonesia	627
Sepsis coprophila	59	Bukit Tingei	Indonesia	721
Sepsis coprophila	60	Bukit Tingei	Indonesia	721
Sepsis coprophila	61	Bukit Tingei	Indonesia	720
Sepsis coprophila	62	Bukit Tingei	Indonesia	720
Sepsis coprophila	63	Bukit Tingei	Indonesia	727
Sepsis coprophila	65	Bukit Tingei	Indonesia	727
Sepsis coprophila	1	Sulawesi	Indonesia	718
Sepsis coprophila	2	Sulawesi	Indonesia	579
Sepsis coprophila	3	Sulawesi	Indonesia	563
Sepsis dissimilis	I	Tamil Nadu	India	544

Sepsis dissimilis	50	Bukit Tingei	Indonesia	441
Sepsis dissimilis	178	Citeureup	Indonesia	622
Sepsis dissimilis	179	Citeureup	Indonesia	625
Sepsis dissimilis	176	Pariaman	Indonesia	624
Sepsis dissimilis	177	Pariaman	Indonesia	624
Sepsis dissimilis	3	Sulawesi	Indonesia	578
Sepsis dissimilis	B07	Batu Kitang	Malaysia	552
Sepsis dissimilis		Ipoh	Malaysia	532
Sepsis dissimilis	I	Malacca	Malaysia	412
Sepsis dissimilis	I	Goroka	Papua New Guinea	494
Sepsis dissimilis	172	Mt. Camagong	Philippines	585
Sepsis dissimilis	173	Mt. Camagong	Philippines	585
Sepsis dissimilis	174	Mt. Camagong	Philippines	585
Sepsis dissimilis	180	Mt. Camagong	Philippines	622
Sepsis dissimilis	181	Mt. Camagong	Philippines	623
Sepsis dissimilis	184	Mt. Camagong	Philippines	623
Sepsis dissimilis	2	Chiang Mai	Thailand	543
Sepsis dissimilis	I	Trang	Thailand	547
Sepsis dissimilis	III	Trang	Thailand	618
Sepsis dissimilis	I	Grand Comoros	Union of the Comoros	533
Sepsis dissimilis	IV	Grand Comoros	Union of the Comoros	560
Sepsis nitens	I	Guizhou	China	552
Sepsis nitens	II	Tamil Nadu	India	571
Sepsis nitens	III	Tamil Nadu	India	627
Sepsis nitens	144	Bukit Tingei	Indonesia	707
Sepsis nitens	145	Bukit Tingei	Indonesia	707
Sepsis nitens	146	Bukit Tingei	Indonesia	677
Sepsis nitens	147	Bukit Tingei	Indonesia	707

Sepsis nitens	148	Bukit Tingei	Indonesia	707
Sepsis nitens	205	Bukit Tingei	Indonesia	581
Sepsis nitens	206	Bukit Tingei	Indonesia	530
Sepsis nitens	207	Bukit Tingei	Indonesia	581
Sepsis nitens	211	Bukit Tingei	Indonesia	580
Sepsis nitens	69	Citeureup	Indonesia	546
Sepsis nitens	70	Citeureup	Indonesia	547
Sepsis nitens	282	Citeureup	Indonesia	464
Sepsis nitens	283	Citeureup	Indonesia	464
Sepsis nitens	284	Citeureup	Indonesia	464
Sepsis nitens	169	Pariaman	Indonesia	641
Sepsis nitens	170	Pariaman	Indonesia	641
Sepsis nitens	171	Pariaman	Indonesia	641
Sepsis nitens	202	Pariaman	Indonesia	580
Sepsis nitens	244	Pariaman	Indonesia	495
Sepsis nitens	1	Sulawesi	Indonesia	566
Sepsis nitens	2	Sulawesi	Indonesia	566
Sepsis nitens	3	Sulawesi	Indonesia	567
Sepsis nitens	4	Sulawesi	Indonesia	567
Sepsis nitens	5	Sulawesi	Indonesia	572
Sepsis nitens	I	Ipoh	Malaysia	601
Sepsis nitens		Tanda Saeed	Pakistan	664
Sepsis sp. A	157	Bukit Tingei	Indonesia	539
Sepsis sp. A	158	Bukit Tingei	Indonesia	529
Sepsis sp. A	207	Bukit Tingei	Indonesia	560
Sepsis sp. A	213	Bukit Tingei	Indonesia	594
Sepsis sp. A	214	Bukit Tingei	Indonesia	553
Sepsis sp. A	I	Sukabumi	Indonesia	594
Sepsis sp. B	125	Sawahlunto	Indonesia	594

Sepsis sp. B	127	Sawahlunto	Indonesia	594
Sepsis sp. B	278	Sawahlunto	Indonesia	616
Sepsis sp. B	279	Sawahlunto	Indonesia	616
Sepsis sp. B	280	Sawahlunto	Indonesia	618
Sepsis sp. B	281	Sawahlunto	Indonesia	541
Sepsis sp. C	81	Citeureup	Indonesia	643
Sepsis sp. C	82	Citeureup	Indonesia	644
Sepsis sp. C	83	Citeureup	Indonesia	581
Sepsis sp. C	84	Citeureup	Indonesia	577
Sepsis sp. C	87	Citeureup	Indonesia	579
Sepsis sp. C	159	Citeureup	Indonesia	675
Sepsis sp. C	160	Citeureup	Indonesia	627
Sepsis sp. C	161	Citeureup	Indonesia	627
Sepsis sp. C	163	Citeureup	Indonesia	539
Sepsis sp. C	201	Citeureup	Indonesia	529
Sepsis sp. C	1	Chiang Mai	Thailand	627

“Well Johnnie why don't you go poke your barcoder into it and find out.”[1]

[1] Stephen Cameron, Daniel Rubinoff and Kipling Will. 2006. “Who will actually use DNA barcoding and what will it cost?” Systematic Biology 55(4): 855 – 847.