The Utility of Standardized DNA Markers in Species Delineation and Inference of the Evolutionary History of Symbiotic Relationships in the Malagasy Ant *Melissotarsus insularis* Santschi, 1911 and its Scale Associate (Diaspididae)

by

Ariel Levitsky

A Thesis
presented to
The University of Guelph

In partial fulfilment of requirements
for the degree of
Master of Science
in
Integrative Biology

Guelph, Ontario, Canada

ABSTRACT

THE UTILITY OF STANDARDIZED DNA MARKERS IN SPECIES DELINEATION AND INFERENCE OF THE EVOLUTIONARY HISTORY OF SYMBIOTIC RELATIONSHIPS IN THE MALAGASY ANT *MELISSOTARSUS INSULARIS* SANTSCHI, 1911 AND ITS SCALE ASSOCIATE (DIASPIDIDAE)

Ariel Levitsky
University of Guelph, 2013

Advisor:
Professor M.A. Smith

A subset of 199 *Melissotarsus insularis* and 130 Diaspididae specimens were analyzed to 1) determine the species status of *M. insularis* and 2) to explore the relative intimacy of the relationship between *M. insularis* and Diaspididae. An analysis of molecular variance and the observed lack of association between clades and distinct habitats on the *M. insularis* phylogeny suggested that while *M. insularis* exhibits isolation by distance, it does not apparently diversify by habitat. When cryptic COI pseudogenes were accounted for, the majority of the genetic diversity exhibited by *M. insularis* was limited to a divergence of 3% or less suggesting that *M. insularis* represents a single, albeit broadly distributed, species. A cophylogenetic reconstruction of the relationship between *M. insularis* and Diaspididae yielded 14 “cospeciation” events but was not significant unlike reconstructions of host-parasite relationships. Analyses of reduced datasets suggested that incomplete taxon sampling may significantly affect cophylogenetic reconstruction results.

Table of Contents

List of Tables

List of Figures

Acknowledgements

First and foremost I would like to thank my advisor Dr. M. Alexander Smith for his undying optimism and encouragement over the course of my degree. My thesis would not have been possible without his insight and expertise. Additionally I would like to thank my committee members Dr. Jinzhong Fu and Dr. Kevin McCann. Dr. Jinzhong Fu played a primary role in the successful completion of my analyses both by providing invaluable guidance and computing resources while Dr. Kevin McCann's background in ecology was instrumental in framing my thesis in a broader context.

I would like to thank Dr. Brian Fisher for providing the *Melissotarsus insularis* and Diaspididae specimens and Dr. Cortland Griswold for his generous donation of computing resources without which the completion of my analyses would not have been possible. I owe my success in the lab to the support of my colleagues, the former and current members of The Smith Lab, Claudia Bertrand and Shannon Eagleton.

I would like to thank my friends for their continued emotional support particularly Rasha Taha, Jonathan Hodgson and Richard Schwarting whose programming expertise was invaluable in mastering challenging software such as the GMYC model.

Lastly, I would like to thank my father who has supported me tenaciously throughout my academic career.

My research was made possible by the National Science Foundation (grants No. DEB-0072713, DEB-0344731 to BLF and DEB-0842395 to BLF and MAS) and a Natural Sciences and Engineering Research Council of Canada Discovery Grant to MAS.

Chapter 1: Introduction

Conservation biology is a crisis discipline DeSalleand Amato (2004) dominated by time sensitive phenomena such as habitat loss and climate change (Heller & Zavaleta 2009; Wilcove *et al.* 1998). Conservation literature consistently refers to the mass loss of biodiversity as a primary agent of ecosystem deterioration (Rogers *et al.* 2010). Although the cumulative effects of biodiversity loss are yet to be documented, the short-term effects are covered extensively in the literature (Diaz *et al.* 2006; Worm *et al.* 2006). The global concern with preserving biodiversity is reflected in a number of international treaties such as the National Biodiversity Strategies and Action Plans established by the Convention on Biological Diversity (CBD) in 2002 (Jóhannsdóttir *et al.* 2010). At the local level, biodiversity conservation programs have expanded and include the designation of areas for conservation, compensation of impoverished communities for the opportunity costs of conservation and manual modification of endangered or invasive populations (Jóhannsdóttir *et al.* 2010; Rands *et al.* 2010). Despite such initiatives, rates of biodiversity loss remain well above background levels and anthropogenic pressure on biodiversity continues to increase (Barnosky *et al.* 2011).

Inadequate funding is consistently emphasized as it is the limiting factor for the accumulation of biodiversity data necessary for informing conservation policy and monitoring the CBD's progress (Balmford *et al.* 2005; Balmford & Whitten 2003; Gardner *et al.* 2008). Biodiversity data is often acquired through surveys that estimate species richness either by surveying biodiversity indicators or a subset of the overall species composition determined by funding and the methodology employed (Kerr *et al.* 2000; Lawton *et al.* 1998). The cost of assessing biodiversity is dependent upon a number of factors including the scale and location of the area of interest, taxa surveyed and sampling effort (Bisevac & Majer 2002; Gardner *et al.* 2008; Mandelik *et al.* 2010; Qi *et al.* 2008; Targetti *et al.* 2011). In most cases, the primary costs are labour and transportation, however, their roles in the survey budget vary depending on whether the

survey is local or remote in relation to the home institution. Equipment and miscellaneous costs, when included, are relatively minor comprising from 4.64% to 20% of the budget (Gardner *et al.* 2008; Mandelik *et al.* 2010; Targetti *et al.* 2011).

Labour expenses fall into two general categories: field work and laboratory work. The ratio of scientist hours devoted to each stage varies between studies depending on the number of specimens sampled, sampling methodology and species diversity (Bisevac & Majer 2002; Gardner *et al.* 2008; Mandelik *et al.* 2010; Targetti *et al.* 2011). The successful identification of the species surveyed is limited by the availability of taxonomic expertise (Bisevac & Majer 2002; Gardner *et al.* 2008; Mandelik *et al.* 2010). For example, a survey of the Brazilian Amazon conducted by Gardner et al. (2008) included a number of groups that were identified to genera only, due to the lack of species level keys and local experts. The specimens that were identified were, in most cases, assigned to a numbered morphospecies (Gardner *et al.* 2008).

Novel surveys are necessary in some cases, particularly if data is required for a poorly known species or area. However, survey costs may be significantly reduced by supplementing the specimens surveyed with natural history collections (NHCs) (Ponder *et al.* 2001). The use of NHCs builds on previous investments reducing and, in some cases, eliminating the costs of field work, transportation and accommodation. The utility of NHCs is often overlooked due to concerns with incomplete sampling, poor preservation of specimens and the risk of damaging type specimens (Gilbert *et al.* 2007; Ponder *et al.* 2001; Wandeler *et al.* 2007). However, techniques aimed at managing the challenges associated with using NHCs are evolving and NHCs continue to grow (Gilbert *et al.* 2007; Graham *et al.* 2004; Horvath *et al.* 2004; Lovejoy *et al.* 2010; Ponder *et al.* 2001; Tingley & Beissinger 2009).

The incorporation of NHCs in biodiversity surveys in combination with DNA analyses addresses the primary deterrents to biodiversity assessment: the lack of funding, time and taxonomic expertise. The DNA sequencing of NHCs can be time and cost effective relative to conventional methods of species identification and may be used to

investigate taxonomic designations (Asher & Hofreiter 2006). The past three decades of advances in DNA sequencing technology (Caterino *et al.* 2000; Hudson 2008; Zhang & Hewitt 2003) and powerful analytical software (Kuhner 2006; Librado & Rozas 2009; Tamura *et al.* 2011) have made it possible to identify polymorphisms, population bottlenecks, migration and phylogenies using genetic data at an unprecedented rate.

My thesis aims to explore the diversity of data that can be derived from an existing museum collection using a combination of genetic data and novel analytical software and the utility of such data in biodiversity conservation. I focus on the ant species *Melissotarsus insularis* and its relationship with its scale symbiont in the family Diaspididae. *M. insularis* is endemic to Madagascar and inhabits a variety of dicotyledonous tree species (Ben-Dov & Fisher 2010). Although the genus, *Melissotarsus*, has been described by several peer-reviewed articles (Ben-Dov 2010; Ben-Dov & Fisher 2010; Delage-Darchen 1972; Fisher & Robertson 1999; Mony *et al.* 2002; Santschi 1911), none focus primarily on *M. insularis*. *M. insularis* appears to tend Diaspididae that occupy galleries excavated by the ant in living trees (Ben-Dov & Fisher 2010), however the nature of the relationship between *M. insularis* and Diaspididae has yet to be investigated (Ben-Dov & Fisher 2010).

The first chapter of my thesis focuses on the challenge of species delineation in the conservation of arthropod biodiversity. I utilize multi-locus genetic data and geographic distribution to explore the species status and phylogeography of *M. insularis*. Preliminary analyses of existing *M. insularis* collections suggested a wide geographic range and divergences in the 5' region of the mitochondrial COI gene or the "DNA barcode" (Hebert *et al.* 2003) of 4% or more. Relative to interspecific species divergences of 2-3% tested against morphological identifications by Smith *et al.* (2005), a divergence of 4% suggests either high intraspecific diversity or hidden diversity. In combination with its wide geographic range and the heterogeneity of habitats it occupies (Ben-Dov & Fisher 2010), the high genetic divergences exhibited by *M. insularis* suggests that it may represent more than one species. If the specimens currently identified as *M. insularis* represent a species complex it is expected that: 1) specimens

will comprise multiple molecular taxonomic units (MOTUs) at a range of divergences and 2) the branching pattern of the *M. insularis* phylogenetic tree would be accounted for with greater accuracy with the inclusion of a species boundary (Monaghan *et al.* 2009). If contemporary habitat distribution drives diversification then *M. insularis* genetic divergence should correlate with distinct habitats. It is expected that an AMOVA test of sequence divergence in response to habitat type will be statistically significant.

The second chapter of my thesis explores the ecosystem processes responsible for maintaining observed biodiversity patterns, namely symbiotic relationships. The intimacy and evolutionary history of the relationship between *M. insularis* and associated Diaspididae will be explored using cophylogenetic reconstruction. The reconstruction will be generated using CoRe-PA 0.5.1 as it is the first cophylogenetic reconstruction program that is capable of processing polytomies (Merkle *et al.* 2010). Although the theory of cophylogeny dates back to 1913 (Fahrenholz 1913), recent advancements in sequencing technology and computing (Caterino *et al.* 2000) have allowed for quantifiable and systematic cophylogenetic reconstruction (Desdevises 2007) and consequently the testing of novel hypotheses. However, reference ranges for phylogenetic congruence have not yet been established making CoRe-PA 0.5.1 results for any given relationship difficult to predict.

All recently published studies using CoRe-PA focused on the relationship between a virus and its host and two of the three concluded that there was significant phylogenetic congruence between the species tested (Dilcher *et al.* 2012; Nemirov *et al.* 2010). However, a mutualistic relationship similar to that shared by *M.insularis* and Diaspididae involving the ant *Crematogaster* and its *Coccus* scale associate yielded a p-value of 0.77 when the number of cospeciations determined by CoRe-PA was compared to a normal distribution. Given that the relationship between *M. insularis* and Diaspididae is presumed to be mutualistic (Ben-Dov & Fisher 2010) it is unlikely to yield results comparable to those achieved with host and virus relationships as species in mutualistic relationships tend to be relatively less constrained by their associates (Bruyndonckx *et al.* 2009; Hoeksema & Bruna 2000). Consequently it is expected that the number of

cospeciations determined by CoRe-PA will fall in the upper tail of the normal distribution, but will not achieve significance.

To provide context for the *M.insularis* and Diaspididae cophylogenetic reconstruction the effect of sample size was tested on the likelihood of achieving a statistically significant number of cospeciations by running CoRe-PA on the original dataset reduced to 75%, 50% and 25% of its initial size. To test the effect of randomization strategy, the original dataset was analyzed using all four randomization strategies available in CoRe-PA: randomizing the host tree, the parasite tree, both the host and parasite tree or the associations between the tips of the two trees (Merkle *et al.* 2010). I expected the significance of the observed cospeciations to vary both positively and negatively with randomization strategy and sample size.

Chapter 2: The Utility of Standardized DNA Markers in the Species Delineation of the Malagasy Ant *Melissotarsus insularis*

Abstract

DNA analysis in combination with natural history collections can be used to supplement and, in some instances replace novel biodiversity surveys reducing the prohibitive costs of biodiversity quantification. The species status of the Malagasy ant *Melissotarsus insularis* was assessed using 127 previously available sequences and three DNA markers, COI, 12S and histone H3, sequenced from an additional 199 museum specimens. Preliminary genetic analyses and geographic distribution suggested that the name *M. insularis* may represent multiple species. However, the majority of the perceived variation resulted from poor sequence quality or the amplification of cryptic pseudogenes. Divergences in the number of synonymous and nonsynonymous changes, nucleotide composition and clustering on a phylogenetic tree identified two to three potential pseudogenes. Most COI sequences formed a single MOTU (molecular operational taxonomic unit) at 2%, 3% and 4% divergence and most 12S sequences formed a single MOTU at 1.6%. All additional MOTUs were comprised of a single sequence each. Histone H3 sequences varied by one character state that did not appear to be monophyletic when plotted on a COI Bayes tree. MOTU results were corroborated by GMYC analyses which concluded that the *M. insularis* phylogenetic tree is unlikely to contain a species boundary. Mantel analyses confirmed isolation by distance with an r of 0.19497 for COI and 0.12665 for 12S but AMOVA analyses did not support diversification by habitat. The species status of *M. insularis* was successfully addressed using DNA sequencing and natural history collections only, precluding the monetary and time investments inherent in novel specimen collection.

Introduction

Quantifying biodiversity at the species level faces several barriers including ambiguous species boundaries, cryptic species and limited resources (funding, time and taxonomic expertise) (DeSalle & Amato 2004; Gardner *et al.* 2008; Kim & Byrne 2006). Prevailing methods of biodiversity quantification such as the surveillance of species composition are ambitious as they require relatively large budgets, may span several years and rely on the availability of trained taxonomists (Balmford & Whitten 2003; Gardner *et al.* 2008; Qi *et al.* 2008). Focussing on species whose biodiversity patterns correlate with those of their habitat (biodiversity indicators) is a common approach to mitigating the costs of biodiversity surveys (Kerr *et al.* 2000; Mandelik *et al.* 2010). However, limiting the number of species surveyed does not guarantee sustainable survey budgets. Surveying as few as four biodiversity indicator species can cost over $150 000 (Targetti *et al.* 2011).

In studies evaluating the costs of biodiversity surveys, labour was consistently identified as the primary expenditure (Bisevac & Majer 2002; Gardner *et al.* 2008; Qi *et al.* 2008; Targetti *et al.* 2011). Labour costs, including species identification, comprised 39% to 88.44% of the survey budget (Gardner *et al.* 2008; Qi *et al.* 2008; Targetti *et al.* 2011). Focus was placed on documenting existing species composition using biodiversity indicators or identifying potential biodiversity indicators for use in future surveys. Known species were identified using morphological taxonomy and attempts at species discovery were not made (Bisevac & Majer 2002; Gardner *et al.* 2008). The cost of transportation and accommodation varied greatly between studies ranging from 11.74% to 41% of the budget depending on the size of the study sites and their proximity to the home institution (Gardner *et al.* 2008; Qi *et al.* 2008; Targetti *et al.* 2011). The costs of transportation and accommodation were particularly high for the biodiversity surveys conducted in the tropics; often considered to be a priority for biodiversity conservation (Gardner *et al.* 2008).

Using natural history collections (NHCs) to supplement and, in some instances (Harper *et al.* 2006; Smith & Fisher 2009; Söller *et al.* 2000; Thomas *et al.* 1990), replace novel biodiversity surveys can reduce the costs of biodiversity quantification both by reducing or eliminating transportation and accommodation costs and by narrowing the survey area or the number of individuals sampled. In cases where the purpose of the survey is to document known species as opposed to species discovery (i.e. the monitoring of biodiversity indicators) NHCs may contribute by serving as a historical reference for the study of species composition over time, inferring species distribution and identifying poorly sampled populations thereby narrowing the study area (Graham *et al.* 2004). Opportunities to narrow the study area using existent species data are critical as reducing the number of transects sampled or the intensity of sampling has been shown to significantly reduce survey costs (Qi *et al.* 2008). If the purpose of the survey is species discovery then NHCs may be used to identify morphological characters unique to the species of interest, conduct a population genetics analysis, sample endangered or elusive individuals or generate a preliminary phylogeny (Asher & Hofreiter 2006; Crandall *et al.* 2009; Suarez & Tsutsui 2004). A preliminary phylogeny may reduce the number of individuals sampled by concentrating future sampling effort on underrepresented clades. Consequently, if enough specimens have been adequately preserved a novel survey may not be necessary (Harper *et al.* 2006; Smith & Fisher 2009; Söller *et al.* 2000; Thomas *et al.* 1990). Furthermore, a preliminary phylogeny can direct research questions and serve as the context for a novel phylogeny.

One of the primary limitations of using NHCs in biodiversity quantification is their taxonomic bias. Groups that are charismatic and easily identified such as vertebrates, birds and plants are often overrepresented in NHCs while arthropods, nematodes and fungi remain underrepresented (Newbold 2010; Spector 2006). The incredible diversity of arthropods complicates specimen collection and identification and is, in part, responsible for their historic underrepresentation in taxonomic studies (Spector 2006). However, arthropods are integral to ecosystem functioning and play a primary role in biodiversity quantification (Spector 2006; Underwood & Fisher 2006). It is, therefore, critical that they are adequately represented in comprehensive estimates of biodiversity

(Newbold 2010; Spector 2006). DNA identification techniques such as DNA barcoding have allowed for the rapid curation of thousands of arthropod specimens making them available for inclusion in biodiversity assessments (Smith *et al.* 2005; Vernooy *et al.* 2010; Wilson 2010).

A combination of diverse methods tends to be the best option for species discovery (DeSalle *et al.* 2005; Padial *et al.* 2010) and contemporary methods of species delineation frequently corroborate species designations determined via morphology and ethology with genetic data (Husseneder & Grace 2001; Perring *et al.* 1993; Smith *et al.* 2006; Tan *et al.* 2009). However, within the context of biodiversity quantification, abstaining from attempts to delineate species prior to acquiring morphological, ethological, and genetic data results in a species delineation process that is, albeit robust, time and resource intensive. Within the past three decades, the analysis of mtDNA in combination with complementary gene fragments has been proposed as a cost and time effective option for species delineation (Vogler & Monaghan 2007). The relatively high rates of mutation and low rates of recombination characteristic of mtDNA yield the high resolution necessary for the inference of species trees on a mass scale (Ballard & Whitlock 2003; Rubinoff & Holland 2005). The use of mtDNA is often criticized for its unique drawbacks; nuclear inserts of mitochondrial DNA (numts) or pseudogenes and variable patterns of inheritance (mtDNA is not always maternally inherited; a common assumption in the generation of mtDNA phylogenies) as they may result in phylogenetic trees that do not accurately represent the species phylogeny (Aitken *et al.* 2004; Galtier *et al.* 2009; Rubinoff & Holland 2005). However, mitochondria play a vital physiological role and the evolution of mtDNA provides characters that can inform species delineation, albeit in a manner different from that of nuclear DNA (Rubinoff & Holland 2005). If analyzed in combination with complementary nuclear markers, geographic distribution and available ecological and climatic data, mtDNA may be used to develop and test species hypotheses (Vogler & Monaghan 2007).

In order to explore the utility of standard DNA markers in maximising the value of NHCs I will focus on elucidating the species status and phylogeography of *Melissotarsus insularis*. *M. insularis* (Fig. 1) is one of four species in the genus *Melissotarsus* and is the only species in the genus endemic to Madagascar (Ben-Dov & Fisher 2010). *Melissotarsus* is unique in that it is the only genus in the order Hymenoptera with adults that possess hypostomal silk glands (Fisher & Robertson 1999). Secreted silk is drawn out using forelegs equipped with a silk brush (Fisher & Robertson 1999). *Melissotarsus* establish colonies by tunnelling in the live wood of 23 different dicotyledonous tree species wherein they maintain armored scale insects presumably as a secondary food source (Ben-Dov & Fisher 2010; Fisher & Robertson 1999). The armored scale insects maintained by *Melissotarsus* have been identified as members of two major subfamilies: Aspidiotinae and Diaspidinae in the family Diaspididae (Ben-Dov & Fisher 2010). Although the genus, *Melissotarsus*, has been the focus of several publications, none address the species status of *M. insularis* (Ben-Dov 2010; Ben-Dov & Fisher 2010; Delage-Darchen 1972; Fisher & Robertson 1999; Mony *et al.* 2002; Santschi 1911). Despite its wide distribution, it is difficult to collect as colonies are concealed in galleries excavated in living trees and workers do not forage outside the galleries (Fisher & Robertson 1999). Its remote location and elusive nature (Fisher & Robertson 1999) make *M. insularis* an excellent candidate for exploring the utility of genetic analyses in maximizing the value of NHCs.

In preliminary analyses of existent *M. insularis* sequences available on the DNA barcoding database BOLD (Ratnasingham & Hebert 2007) over 57% of the uploaded sequences exhibited a divergence of 4% or greater with some sequences diverging as much as 12%. Although no set range for intraspecific and interspecific genetic divergence exists (Cognato 2006) Smith *et al.* (2005) found an average interspecific diversity of 16% between morphologically delineated ant species across Madagascar and when morphological species were tested using molecular data an intraspecific diversity of 2-3% was found. Consequently, Smith *et al*. (2005) suggested that a threshold of 2–3% is an appropriate hypothesis for testing ant species identity. Furthermore, *M. insularis* is widely distributed across Madagascar and occupies a

diverse range of habitats. It is distributed throughout all four of Madagascar's simplified bioclimatic regions (Cornet 1974) and subsists within 15 different families of plants including flowering trees and shrubs (Ben-Dov & Fisher 2010). Considering that habitat heterogeneity has been shown to be a driver of diversification in several other ant species (Pacheco & Vasconcelos 2012; Pfeiffer & Mezger 2012; Smith *et al.* 2005), it is possible that *M. insularis* is comprised of more than one species.

The current study will 1) assess the species status of the Malagasy ant *Melissotarsus insularis* and 2) relate the genetic diversity of *M. insularis* to Madagascar's unique habitats. A map identifying unique habitats was developed by Wilmé *et al.* (2006) using quaternary paleoclimatic patterns, databases mapping Madagascar's rivers and watersheds and the distribution of 35,400 terrestrial vertebrates and will be used as a reference. If *M. insularis* represents a species complex it is expected that specimens comprise multiple molecular taxonomic units (MOTUs) at a range of divergences. Furthermore, the branching pattern of the *M. insularis* phylogenetic tree should be accounted for with greater accuracy with the inclusion of a species boundary. If habitat is the driver of diversification then *M. insularis* genetic divergence should correlate with the unique habitats identified by Wilmé *et al.* (2006). It is expected that an AMOVA test of sequence divergence in response to habitat type will be statistically significant.

Methods

Specimens

My dataset was comprised of a combination of novel and previously available *Melissotarsus insularis* sequences. I amplified the barcoding region of the mitochondrial COI gene, a fragment of the mitochondrial ribosomal tRNA gene 12S and a fragment of the nuclear gene histone H3 from 67 collection events of *M. insularis*. When available, three specimens from each collection event were selected based on their physical integrity for subsampling yielding a total of 199 specimens. Specimens were collected

by Brian Fisher and the Fisher-Griswold Arthropod Team between 2001 and 2009. They were manually extracted from various plant species across Madagascar and stored in ethanol at the California Academy of Sciences.

A total of 129 *M. insularis* COI sequences were uploaded to the Barcode of Life Data Systems (BOLD) (Ratnasingham & Hebert 2007) between 2005 and 2009. One sister from each collection event was included in the phylogenetic analysis. Sisters were selected based on sequence length. In cases where a longer sequence was generated from the same collection event the longer sequence was chosen. *Pheidole megacephala* sequences accessed through GenBank were included as the outgroup for the COI and 12S phylogenies (GenBank accession numbers EF610027.1 and EF518664.1 respectively). For a list of specimen process IDs, sample IDs, GenBank accession numbers and collection details see Appendix 1.

DNA Extraction, Amplification and Sequencing

All specimens were point-mounted with the use of a Leica^{TM} (S8AP0) microscope and photographed with a Dino-Lite Pro2 (AD-4132TA) microscope prior to extraction. The middle right leg of each ant was removed using forceps sterilized with ethanol and then stored in 30 µl of ethanol. In addition to the middle right leg, the abdomen was removed from 95 specimens. DNA was extracted from the middle right leg at the Biodiversity Institute of Ontario as described by Ivanova *et al.* (2006). DNA extraction from the abdomen was conducted using the Macherey-Nagel NucleoSpin® 96 Tissue kit with the following modifications to the manufacturer's instructions. Sub-sampling plates were placed into an incubator for 1 hour and 20 minutes prior to extraction in order to evaporate the ethanol used to store the samples. Instead of Buffer BE, 50 µl of preheated ddH_2O per well was used to elute the DNA and the elution step was not repeated as was suggested in the manual. COI was amplified from *M. insularis* using primers LepF1 and LepR1, C_ANTMR1D and RonMWASPdeg_t1, 12S was amplified using primers 12Sai and 12Sbi-f and H3 was amplified using primers H3F and H3R. For a detailed list of primers see Appendix 2.

Amplification and sequencing was conducted according to standard DNA barcoding protocols based on Hajibabaei *et al.* (2005) and Smith *et al.* (2008). To amplify target DNA, a solution comprised of 625 µl of trehalose, 200 µl of ddH_2O, 125 µl of 10x buffer, 62.5 µl of $MgCl_2$, 12.5 µl each of forward and reverse primer, 6.25 µl of 10 mM dNTPs and 6 µl of Taq polymerase, was aliquoted into a 96 well plate and 1.0 µl of DNA template was added to each well. The plate was then run on an Eppendorf® thermal cycler (6325). PCR temperature profiles used for amplification differed with primer. For a detailed description of PCR temperature profiles by primer see Appendix 3 in the supplementary material. Amplicons were visualized using pre-cast E-Gel® agarose gels by Invitrogen™.

PCR products were not purified prior to sequencing. A standard solution comprised of 550.0 µl of trehalose, 210.83 µl of ddH_2O, 100.83 µl of 5x buffer, 110 µl of 10 µM forward or reverse primer and 18.33 µl of BigDye® was aliquoted into a 96 well plate and 2.0 µl of amplicon were added to each well. The plate was then run on an Eppendorf® thermal cycler using a PCR sequencing program with the following temperature profile: 2:00 minutes at 96°C, 30 cycles of 30 seconds at 96°C, 15 seconds at 55°C and 4:00 minutes at 66°C. Target amplicons were sequenced using the ABI BigDye™ v. 3.1. Cycle Sequencing Kit on an ABI 3730XL (Biodiversity Institute of Ontario) or ABI3730 (Genomics Facility, Advanced Analysis Centre) sequencing platform.

Sequence Editing and Phylogenetic Analysis

Sequences were edited and aligned using Sequencher v. 4.5 and v. 5.0 (Gene Codes Corporation, Ann Arbor, Michigan) and BioEdit (Hall 1999) respectively. Sequences 160 bp or less in length or those that were more than 2% heteroplasmic were excluded from further analysis.

Incongruent amplifications originating from the same specimen were analyzed using the number of synonymous (dS) and nonsynonymous (dN) changes, nucleotide

composition and neighbour-joining trees generated in MEGA 5.0 (Tamura *et al.* 2011) to determine whether they resulted from paternal recombination, heteroplasmy or pseudogene amplification. Departure from expected dS and dN values and nucleotide composition and clustering on a phylogenetic tree are all indications of divergence from the functional gene (Martins Jr *et al.* 2007). Amplifications that exhibited all three characteristics were considered to be likely pseudogenes.

COI and 12S phylogenetic trees were generated using one sister as a representative from each collection event. Sisters were chosen based on sequence length and sequence quality expressed in Sequencher v. 5.0 as the proportion of a sequence comprised of base calls with a high confidence value as defined in the Sequencher v. 5.0 manual (GeneCodesCorporation 2011). A subset of specimens yielding both COI and 12S sequences was utilized to conduct a partition analysis. Neighbour joining and maximum parsimony analyses were conducted in MEGA 5.0 (Tamura *et al.* 2011). Maximum likelihood analysis was conducted in PAUP* version 4.0 (Swofford 2003) while Bayesian and partition analyses were conducted in MrBayes version 3.1.2 (Huelsenbeck & Ronquist 2001). All analyses used evolutionary models determined by the hierarchical likelihood ratio test (hLRT) employed in ModelTest version 3.7 (Posada & Crandall 1998). Analyses conducted in MEGA 5.0 used the Tamura-Nei model of evolution; the closest available model to the best-fit model determined by the hLRT with the gamma parameter specified. The Bayesian analyses were run for 60,000,000-88,000,000 generations depending on the size of the dataset. Burnin was set at 25% and confirmed in Tracer v1.5 (Rambaut & Drummond 2007).

The multiple threshold GMYC (general mixed Yule-coalescent) model (Monaghan *et al.* 2009) a revised version of the GMYC model developed by Pons *et al.* (2006) was utilized to determine a potential species boundary. The ultrametric trees used in the GMYC analysis were generated in BEAST version 1.6.0 (Drummond & Rambaut 2007).

jMOTU (Jones *et al.* 2011) was used to delimit MOTUs (molecular operational taxonomic units) for genetic divergences ranging from 0.001-10% at the default

threshold of 97%. DIVA-GIS v. 7.5 (Hijmans *et al.* 2001) was used to associate sequences with unique habitats previously identified by Wilmé *et al.* (2006) based on their distribution. Mantel tests were conducted in PASSaGE (Rosenberg & Anderson 2011) using genetic distances determined under the Kimura 2-Parameter model of evolution (with the gamma parameter specified based on Modeltest v. 3.7 results) and Euclidean distances determined via the spherical functions implemented in the Geographic Distance Matrix Generator version 1.2.3 (Ersts Internet). AMOVA analyses was conducted in Arlequin v. 3.5 (Excoffier & Lischer 2010).

Results

Sequence Statistics

Of the 199 *M. insularis* specimens subsampled, 184 (92%) yielded COI sequences, 142 (71%) yielded 12S sequences and 85 (43%) yielded H3 sequences that complied with the sequence quality guidelines established in the Methods. COI and 12S exhibited a strong AT bias consistent with most insect mitochondrial sequences (Schwarz *et al.* 2004), while histone H3 was comprised of 24% T, 27% C, 24% A and 25% G. COI and 12S contributed 175 (26%) and 51 (14%) parsimony informative sites respectively. H3 was less variable yielding only one parsimony informative site. A comparison of the 12S and COI genetic distances suggested that the relationship was not significant (Fig. 2). A Mantel test comparing the 12S and COI genetic diversities yielded an r of 0.06246 and a p-value of 0.05123.

Intra-individual Incongruent Amplifications

Incongruence between different amplifications derived from the same DNA extract occurred in 45 samples, 44 of which were from the DNA plate containing extracts ASMEL01-11 to ASMEL95-11 (Plate accession: BIOUG01058) and one of which (ASMEL253-11) was amplified from a different plate (Plate accession: BIOUG01060). In

general, the specimens that were sequenced multiple times due to poor sequence quality exhibited the greatest variation between the two LepF1/LepR1 amplifications whereas specimens that were successfully sequenced on the first attempt yielded congruent amplifications.

The number of synonymous substitutions per synonymous site (dS) exhibited by the sequences that did not agree with either sisters or other amplifications of the same specimen was significantly lower ($p= 1.86 \times 10^{-08}$ when the means were compared using a Student's t-test) than the dS of those that agreed. The sequences that both disagreed and exhibited a low dS were all LepF1/LepR1 or LepF1/C_ANTMR1D amplifications of the 44 extracts from BIOUG01058. Consequently, incongruent amplifications that exhibited a dS of 1.3 or lower were excluded from the dataset utilized to generate the *M. insularis* phylogenies.

The difference in the number of non-synonymous substitutions per non-synonymous site (dN) between the incongruent LepF1/LepR1 and LepF1/C_ANTMR1D amplifications and congruent amplifications was not significant ($p= 0.081$ when the means were compared using a Welch's t-test). However, sequences amplified with RonMWASPdeg_t1 had a significantly higher dN ($p= 2.66 \times 10^{-12}$) than other amplifications of the same specimen.

Nucleotide composition was similar across sequences generated by all primer pairs apart from those amplified with RonMWASPdeg_t1 which had a cytosine composition of 21%, approximately 3% higher than the LepF1/LepR1 amplifications and 5% higher than the LepF1/C_ANTMR1D amplifications. Furthermore, in a neighbour-joining tree that included all amplifications the incongruent RonMWASPdeg_t1 amplifications tended to group together as opposed to grouping with LepF1/LepR1 and LepF1/C_ANTMR1D amplifications of the same specimen (Fig. 4 and Fig. 5). As a result of both the significantly higher dN and the grouping of amplifications on the neighbour-joining tree, the incongruent RonMWASPdeg_t1 amplifications were not included in further analysis as they likely represent a distinct group of pseudogenes.

Phylogenetic Analysis

The best fit model determined for COI and 12S by the hierarchical likelihood ratio tests (hLRTs) in ModelTest v. 3.7 (Posada & Crandall 1998) was TVM+I+G and TrN+I+G respectively. The Tamura-Nei model was used for the neighbour-joining and maximum parsimony analyses conducted in MEGA 5.0 (Tamura *et al.* 2011). The best fit model for both the COI and 12S partitions had gamma distributed among-site rate variation with six substitution types. Partitions were therefore defined but run under the same model.

In general, clades with high bootstrap support or posterior probabilities were conserved across the neighbour-joining, maximum parsimony and Bayes COI trees (see Appendix 4 for all phylogenetic trees). Several specimens exhibited significant variability in all three trees: ASMEL054-11, ASMEL198-11 and ASMEL271-11. Two well supported groups on the COI Bayes tree were not represented by the maximum parsimony tree: ASAM152-05 and ASMEL077-11 which formed the sister clade to ASAM204-05, ASMEL089-11 and ASMEL326-11 with a posterior probability of 0.89 and ASMEL205-11 which formed a clade with ASMEL206-11 and ASMEL211-11 on both the neighbour-joining and Bayes tree with a posterior probability of 1.

Only 30 of the 63 taxa, not including the outgroup, were resolved on the 12S Bayes tree. The majority of resolved taxa were conserved across the 12S neighbour-joining, COI Bayes and Bayes partition trees. ASMEL328-11 tended to group with ASMEL001-11 on the 12S trees but not the COI and partition trees. The representation of ASMEL328-11 on the COI and partition trees is most likely the result of its apparent divergence due to the poor sequence quality of the COI amplification. The 12S maximum parsimony tree differed significantly with only three of all resolved sister groups on the Bayes tree appearing on the maximum parsimony tree.

Both the COI and 12S trees exhibited a topology similar to that of the partition tree. Almost all resolved associations on the COI and 12S phylogenies were present on the partition tree. ASMEL328-11 was represented by a particularly long branch most likely

due to the influence of the low quality COI sequence as opposed to the corresponding 12S sequence which associated with ASMEL001-11 on the 12S phylogeny.

The single parsimony informative site in the H3 dataset has been indicated on the *M. insularis* COI Bayes tree by the colour of the taxa names as either a T (red), a C (blue) or a Y (purple) (Fig. 6). None of the three states appeared to be monophyletic. Furthermore, the tree tips have been colour coded according to the habitats defined by Wilmé *et al.* (2006). In general, specimens tended to group by latitude as opposed to habitat.

Identifying MOTUs

In both the COI and 12S analyses the number of MOTUs (Molecular Operational Taxonomic Units) dropped significantly within the first 2% and reached a plateau at 4% for COI and 1.6% for 12S that extended to a divergence of 10%. MOTU composition for the partition dataset was similar to COI, however, two plateaus were reached one at 1.9% and one at 5.0% (Fig. 7). COI MOTU composition will be reviewed at 2%, 3% and 4%, 12S MOTU composition will be reviewed at 1.6% and the partition MOTU composition will be reviewed at 1.9% and 5.0% respectively.

At a divergence of 2% the COI dataset formed four MOTUs. The majority of sequences were grouped into MOTU001 while an additional three sequences each formed a MOTU: ASMEL250-11, ASMEL328-11 and ASIMB940-09. At 3% divergence all but two sequences were grouped into MOTU001. MOTU002 was comprised of sequences ASIMB940-09 and ASMEL328-11. At 4% divergence the number of MOTUs reached a plateau of one wherein all COI sequences formed a single MOTU. It is unlikely that the composition of MOTUs at 2% and 3% represented genetic variability characteristic of distinct taxonomic groups. The majority of sequences were similar enough to form a single MOTU. None of the additional MOTUs were comprised of more than one sequence distinguishing them as divergent from both the majority of the dataset and one another. ASMEL328-11 was one of the lowest quality sequences selected for

analysis. Although it fell within the rule set used to distinguish suitable sequences, it is possible that much of its variability was the result of coamplification. ASMEL250-11 appeared to be genuinely divergent from the majority of the sequences as its chromatograph exhibited clean, strong peaks. ASIMB940-09 varied from the rest of the dataset, but unlike ASMEL250-11, had four ambiguities.

A similar distribution was observed when the 12S sequences were analyzed (Fig. 7). The number of MOTUs reached a plateau of 4 at a divergence of 1.6%. MOTU002, MOTU003 and MOTU004 were comprised of one sequence each while MOTU001 represented the remainder of the dataset. In this case the sequences comprising the single sequence MOTUs were ASMEL061-11, ASMEL064-11 and ASMEL010-11. ASMEL061-11, ASMEL064-11 and ASMEL010-11 are three of the shortest sequences (147 to 176 bp) in the 12S dataset. All three contain ambiguities that distinguish them from sequences of similar size and are likely responsible for their grouping into individual MOTUs. The results of the partition dataset were almost identical to those of the COI dataset as ASMEL328-11 comprised its own MOTU at 1.9% divergence while MOTU001 represented the remainder of the dataset. At a divergence of 5.0% all partition sequences formed a single MOTU. The lack of overlap between the single sequence MOTUs identified by jMOTU in the COI and 12S datasets and the distribution of the sequences among the MOTUs indicates a possibility that the distinction is an artefact of sequence quality or pseudogene amplification.

GMYC Analysis

Both the single threshold as well as the multiple threshold GMYC (general mixed Yule-coalescent) analysis was performed on the COI and 12S datasets. The alternative hypothesis, that the phylogenetic tree contains a species boundary or multiple potential species boundaries, was tested against the null hypothesis, that the phylogenetic tree does not contain a species boundary. The alternative hypothesis did not account for the data significantly better than the null hypothesis in either case (Table 1).

Mantel Analyses and AMOVA

Mantel analyses of the COI and 12S datasets both yielded a statistically significant, positive r value. The r value for the partition dataset was similar to that of the 12S dataset, however the relationship was not significant (Table 2) suggesting a discrepancy between the COI and 12S sequences. The relationship between genetic divergence and geographic distance has been summarized in Figure 8 for the COI, 12S and partition datasets.

The relationship between genetic divergence and the unique habitats identified by Wilmé *et al.* (2006) was tested using an AMOVA approach. Analyses were run on COI, 12S and partition datasets and in all three cases the majority of the variation occurred within populations (75.88% for COI, 100.98% for 12S and 98.32% for the partition dataset) (Table 3).The variation among unique habitats (regions) comprised a minority of the total variation and was not significant (-7.59% for COI, 7.31% for 12S and 7.60% for the partition dataset). Variation among populations within the same region was significant for COI comprising 31.72% of the total variation likely as a result of the higher genetic diversity exhibited by COI in relation to 12S. Neither the 12S nor the partition dataset exhibited significant variation among populations within the same region. In both cases variation was negative and comprised -8.29% for 12S and -5.92% for the partition dataset of total variation.

Discussion

Prior to the current study, *Melissotarsus insularis* appeared to exhibit a level of genetic diversity that, in combination with its wide distribution, suggested the presence of more than one species. Preliminary analyses of previously amplified COI sequences included intraspecific divergences of 4% and higher. However, in the process of sequencing additional specimens several likely pseudogenes were identified. Following their exclusion from further analysis, a MOTU analysis of the remaining sequences

concluded that the majority of COI and 12S sequences were represented by a single MOTU at a divergence of 2% or higher. The additional MOTUs were represented by one or two specimens. At divergences of 3% or higher the number of additional MOTUs decreased to one in the COI dataset. The additional MOTU in the COI analysis was represented by ASMEL328-11 and ASIMB940-09. ASMEL328-11 stood out in both the MOTU and phylogenetic analysis as a result of its poor sequence quality. The single sequence MOTUs in the COI and 12S datasets did not overlap indicating a low likelihood that they represent evolutionarily divergent specimens. The three additional single sequence MOTUs in the 12S dataset are represented by low quality sequences. In both cases the sequences comprising the additional MOTUs may be pseudogenes particularly with 12S as it had been sequenced only once making it difficult to test for pseudogene amplification. In a study on ant diversity across Madagascar Smith *et al.* (2005) found an average interspecific diversity of 16% between morphologically delineated ant species but suggested that a threshold of 2-3% was appropriate for testing ant species identity. A divergence of 2-3% does not necessarily mark a threshold for species delineation, however, it does provide a context for the current dataset particularly when the source of sequence divergence is suspect even at 2%.

The diversity of the histone H3 gene varies with species and can range from diversity indicative of cryptic species (Steenkamp *et al.* 2002) to minimal diversity both within (Sarnat & Moreau 2011) as well as between species (Bennett & O'Grady 2012). For example, the *Fusarium* (fungus) histone H3 sequences analyzed by Steenkamp *et al.* (2002) yielded 6 parsimony informative sites and the four isolates distinguished by the histone H3 genealogy were loosely congruent with those identified by the β-tubulin, HB9, HB14, and HB26 genealogies. Several species of *Formicidae* in the dataset generated by Sarnat and Moreau (2011) did not exhibit intraspecific variation in the histone H3 gene (i.e. *Pheidole simplispinosa*). A lack of interspecific variation in H3 can be seen in the dataset published by Bennet and O'Grady (2012) between species *Nesophrosyne* sp. 318 (Cicadellidae) and *Nesophrosyne* sp. 48 (Cicadellidae) despite variation between the two species in the COI phylogeny. The histone H3 sequences in the current study yielded one parsimony informative site with three possible character

states: T, C or Y. However, the variation in character states did not distinguish between the MOTUs or clades established by the phylogenetic analysis. Only 32 out of the 126 specimens included in the COI phylogenetic analysis had corresponding histone H3 sequences. Even in cases where a clade appeared to express a singular character state monophyly was difficult to ascertain as the histone H3 character states of the sister sequences were unknown.

The designation of *M. insularis* as a single species was corroborated by the GMYC (general mixed Yule-coalescent) analysis independent of MOTU delimitation. The GMYC model suggests a species boundary by identifying the point within a phylogeny wherein a marked increase in branching rate occurs (Pons *et al.* 2006). It is capable of developing species hypotheses independent of previously delimited groups allowing for a standalone analysis of genetic data as suggested by Brower *et al.* (1996).

The wide distribution of *M. insularis* was addressed by Mantel tests, AMOVA analyses and the relationship between the position of a specimen on the *M. insularis* phylogenetic tree and its collection locality. A Mantel test of the COI and 12S sequences concluded that genetic divergence correlated significantly with geographic distance (see Table 2). Furthermore, when specimens on the COI Bayesian tree were colour coded by locality it was apparent that each clade was associated with a range of latitudes as opposed to the unique habitats identified by Wilmé *et al.* (2006). In combination with Mantel tests, this grouping of specimens by latitude supports the isolation by distance hypothesis wherein a species diverges genetically as a result of geographic distance depending on its dispersal abilities (Wright 1943). The AMOVA analyses concluded that specimen divergence was not significantly associated with the habitats delineated by Wilmé *et al.* (2006). The lack of association is expected given the MOTU results and the lack of association between clades on the phylogenetic tree and locality.

The preliminary hypothesis on the species status of *M. insularis* was successfully addressed using previously generated sequences and museum specimens only. A convergence of several independent tests on genetic divergence and phylogeographic

structure concluded that *M. insularis* is unlikely to represent a species complex but rather a single species. The genetic diversity data obtained from the sequences presumed to represent functional genes will work to inform future surveys of *M. insularis* and caution analyses of speciation against apparent genetic diversity resulting from the amplification of cryptic pseudogenes.

Sequencing Museum Specimens

Approximately 92% of the specimens sub-sampled yielded COI sequences that complied with the sequence quality guidelines established in the Methods in comparison to 71% for 12S sequences and 43% for histone H3 sequences. Specimens ranged in age from 3 to 12 years with approximately 47% of specimens aged at 10 years or more. In a preliminary analysis of 300 Tephritid specimens Van Houdt *et al.* (2009) found that an average of 15% of the sampled specimens aged 10 years or more yielded a DNA barcode when standard protocols were used. The comparatively high sequencing success rate achieved in the current study could be attributed to overall younger specimens, a different set of qualifications for sequence quality and multiple sequencing attempts. Failed specimens were amplified an additional one to three times using degenerate primers (failure tracked) designed to amplify shorter DNA fragments. Museum specimens yield DNA in various stages of degradation depending on the method of collection and preservation and the age of the specimen (Lindahl 1993). They are therefore more likely to yield poor quality sequences and often require additional treatment (Wandeler *et al.* 2007). When comparing the costs of studies that incorporate museum specimens and those that use newly collected specimens it is essential to consider the additional costs necessary to successfully amplify potentially degraded DNA such as designing novel primers and re-amplifying failed specimens. Note that a similar result was not achieved for 12S (amplification efficiency of 71%), in part, because amplification using alternate primers was not attempted due to limited time and resources.

The Role of Cryptic Pseudogenes in Phylogenetic Analysis

Typically, incongruence between amplifications is the result of paternal recombination, mtDNA heteroplasmy or the amplification of nuclear mitochondrial DNA (numts) otherwise known as pseudogenes (Bensasson *et al.* 2001; Magnacca & Brown 2010). Although indirect evidence for paternal recombination has been presented for a number of taxa, the majority of the definitive evidence has been isolated to bivalves, hybrids and humans (Ballard & Whitlock 2003; Barr *et al.* 2005). It is possible but unlikely that the incongruent sequences in my dataset were the result of paternal recombination. The three primer combinations yielded from three to four different amplifications as opposed to two as would be expected in the case of paternal recombination. Much like paternal recombination, heteroplasmy has rarely been recorded in the coding regions of the arthropod genome (Magnacca & Brown 2010). Were the incongruent amplifications the result of paternal recombination or heteroplasmy similar dN (nonsynonymous changes) and dS (synonymous changes) values and nucleotide composition would be expected. However, incongruent amplifications 1) exhibited statistically significant differences in dN and dS values, 2) deviated from the expected nucleotide composition and 3) tended to form groups on a neighbour-joining phylogenetic tree suggesting that they are the result of pseudogene amplification.

Pseudogenes are commonly detected by the presence of stop codons, frameshift mutations and high dN/dS ratios (Bensasson *et al.* 2001; Calvignac *et al.* 2011). However, the method of detection is dependent on whether the pseudogene amplified is cryptic or apparent (Bertheau *et al.* 2011; Martins Jr *et al.* 2007). According to Bertheau *et al.* (2011) cryptic pseudogenes vary from functional genes by approximately 1-3 bp but do not contain stop codons or frameshift mutations. As their apparent divergence from functional genes is minimal, their dN/dS ratios, nucleotide composition and phylogenies are difficult to differentiate from their functional counterparts. I suggest expanding the definition of cryptic pseudogenes to include amplifications that differ from the corresponding functional gene by more than 3 bp but, similarily to the definition proposed by Bertheau *et al.* (2011), do not contain stop codons or frameshift mutations.

The goal of differentiating between cryptic and apparent pseudogenes for the purposes of this study is to a) bring attention to the difficulty associated with identifying cryptic pseudogenes and the challenges they pose to species delineation and b) to determine effective methods for identifying cryptic pseudogenes. My dataset contained a combination of cryptic as well as apparent pseudogenes. Suspected pseudogenes varied from functional genes by 2-30 bp but did not contain stop codons or frameshift mutations. Pseudogenes that varied from functional genes by 1-10 bp behaved similarly and were consequently referred to as cryptic pseudogenes.

The classification of sequences as pseudogenes was not exact as the differences in the sequence statistics were slight. For example, the average dS of congruent and incongruent sequences was 1.51 and 1.26 respectively. Suspect amplifications with a dS of 1.3 or lower were excluded from further analysis. However, a minority of the incongruent amplifications excluded from the analysis had a dS that exceeded 1.3. Although a dS threshold of 1.3 served as a guide it did not conclusively delineate pseudogenes from functional genes. Nuclear pseudogenes accumulate mutations at rates that are 5 to 10 times slower than their functional counterparts (Lü *et al.* 2002). However, pseudogenes are not subject to the evolutionary constraints imposed on functional genes and mutations that would otherwise result in selection against a functional gene may persist in pseudogenes (Saccone *et al.* 2000). Consequently, pseudogenes can exhibit either comparatively ‘slower’ or ‘faster’ rates of evolution. It is possible that the amplifications that exhibited lower dS values were recently diverged pseudogenes that have accumulated mutations at a comparatively slower rate than equivalent functional mitochondrial genes (Lü *et al.* 2002).

The susceptibility of the *M. insularis* specimens to pseudogene amplification was most likely the result of specimen age, the length of the target gene region and the degeneracy of the RonMWASPdeg_t1 and C_ANTMR1D primers. Pseudogenes tend to be shorter than their functional ancestors (D'Errico *et al.* 2004). The specimens used in the current study were between 3 and 12 years old and had likely undergone DNA degradation (Van Houdt *et al.* 2009). Furthermore, RonWASPdeg_t1 and C_ANTMR1D

were designed to amplify shorter regions of the CO1 gene than LepF1/LepR1. Both the age of the specimens and the length of the target gene region resulted in shorter amplicons and may, therefore, have increased the probability of pseudogene amplification.

The LepF1/LepR1 primers were originally designed to amplify the barcode region in insects. In the current study, LepF1/LepR1 amplified both the functional gene and a pseudogene presumed to have a slower rate of evolution. This likely occurred because the pseudogene represented an ancestral state and resembled LepF1/LepR1 more so than the functional gene. RonMWASPdeg_t1 and C_ANTMR1D amplified pseudogenes presumably as a result of their degeneracy. Out of the 41 bases comprising RonMWASPdeg_t1 five (12%) were degenerate while eight (36%) out of 22 bases were degenerate in C_ANTMR1D. The ambiguous bases allowed RonMWASPdeg_t1 and C_ANTMR1D to bind to DNA fragments that were divergent from the target region increasing the probability of pseudogene amplification. The degeneracy of C_ANTMR1D may have accounted for why its apparent amplification target alternated between the functional gene and the pseudogene with greater frequency than RonMWASPdeg_t1 which tended to target either the pseudogene or the functional gene depending on the DNA plate amplified.

Standard DNA barcoding protocols involve the bidirectional sequencing of a 658 bp fragment at the 5’ end of the COI gene (Hebert *et al.* 2003). Multiple amplifications using taxa specific primers are common but do not comply with the standard DNA barcoding ethos (Ratnasingham & Hebert 2007). The minimalism of the DNA barcoding approach reflects its goals: to serve as an economic DNA reference database for species identification (Ratnasingham & Hebert 2007). However, it also increases the chances of undetected pseudogene amplification; a vulnerability that has been criticized in the past (Buhay 2009; Song *et al.* 2009). Pseudogene detection in the current study resulted from the incongruence between multiple amplifications of specimens that were unsuccessfully sequenced on the first attempt. Both the degenerate primers (RonMWASPdeg_t1 and C_ANTMR1D) as well as primers initially used to amplify *M.*

insularis (LepF1/LepR1) amplified pseudogenes. The identification of pseudogenes in the context of DNA barcoding is limited by time and funding. Amplifying and sequencing old or poorly preserved specimens multiple times serves as a relatively economic option to test for the presence of pseudogenes.

In the case of phylogeographic and species revision studies a comprehensive approach is advisable. As was the case with the current study, pseudogenes are not always readily identifiable by the detection of stop codons and frameshift mutations. Resources permitting, pseudogenes can be both avoided and detected by mtDNA purification, the amplification of cDNA, long-range mtDNA amplification and the dilution of the DNA extract (Bensasson *et al.* 2001; Calvignac *et al.* 2011). Calvignac *et al.* (2011) compared all four methods and concluded that diluting the DNA extract was the most efficient approach. Furthermore, dilution of the DNA extract is oftentimes the most economic of the four making it a good candidate for initial attempts at pseudogene exclusion.

Phylogeography of Melissotarsus insularis

The diversification of *M. insularis* in response to latitude as opposed to habitat is inconsistent with the findings of previous studies on the relationship between Hymenopteran diversity and habitat type particularly in tropical regions such as Madagascar (Pacheco & Vasconcelos 2012; Pfeiffer & Mezger 2012; Smith & Fisher 2009). Principle component analysis of Formicidae species composition between collection sites in Brazil concluded that composition varied in response to habitat type but not latitude (Pacheco & Vasconcelos 2012). A similar result in an analysis of 206 ant species was found across four different forest types in Borneo (Pfeiffer & Mezger 2012). The significant variation in diversity across habitat types was attributed to the limited dispersal capabilities of ant queens in contrast to plant specific insect herbivores (Pfeiffer & Mezger 2012).

In comparison to the average interspecific diversity of 16% and the suggested species threshold of 2-3% for Malagasy fauna, *M. insularis* exhibits a low genetic diversity given its wide distribution (Smith *et al.* 2005). Previous studies of Malagasy ant fauna have revealed significant species richness (Smith *et al.* 2005) as well as the presence of cryptic species (Fisher & Smith 2008). In general, Madagascar is considered a biodiversity hot spot for Formicidae as its isolation from mainland Africa, the heterogeneity of its habitats and the absence of army ants serve as drivers of diversification (Fisher 1997; Fisher & Robertson 2002; Fisher & Smith 2008; Vences *et al.* 2009).

M.insularis may have maintained uncharacteristically low levels of genetic divergence due to its ability to disperse and thrive in a variety of habitats. Regular gene flow between proximate populations may have been facilitated by both human mediated as well as natural dispersal. Much like the bark beetle, a recognized economic pest, (Skarpaas & Økland 2009) *Melissotarsus* has been known to occupy economically significant tree species (Mony *et al.* 2002). Consequently its dispersal may be mediated by the timber trade prevalent in Madagascar (Kull 2004; Myers *et al.* 2000; Skarpaas & Økland 2009). Although the natural dispersal capabilities of Formicidae are limited in comparison to other plant specific insect herbivores (Pfeiffer & Mezger 2012), it is worth noting that *M. insularis* reproductives are alates and, on average, have the potential to disperse further than ergatoid species (Peeters *et al.* 2012).

The scale insect with which *M. insularis* maintains a mutualistic relationship, Diaspididae, does not appear to impose limits on its dispersal. Diaspididae is one of the most polyphagous insect families in the world (Andersen *et al.* 2010) allowing it to disperse and adapt to a variety of habitats. All four species in the genus *Melissotarsus* tend to scale insects in the family Diaspididae (Ben-Dov & Fisher 2010). Whether the relationship is obligate or facultative on the part of the ant has not been established, however, a colony lacking Diaspididae individuals has never been found (Ben-Dov 2010; Ben-Dov & Fisher 2010; Delage-Darchen 1972; Fisher & Robertson 1999; Mony *et al.* 2002; Santschi 1911).

In addition to its likely dispersal abilities, *Melissotarsus* appears to be successful at adapting and surviving in a variety of habitats. It inhabits 23 different dicotyledonous tree species (Fisher & Robertson 1999) with colonies reaching sizes of over 1.5 million individuals (Mony *et al.* 2002). *M. insularis*, specifically, was found in all four of Madagascar's simplified bioclimatic regions (Cornet 1974) and 14 out of 28 unique habitats identified by Wilmé *et al.* (2006). It is possible that the apparently broad preferences in host-plant and climate exhibited by *M. insularis* indicate that it is a generalist rather than a specialist (Ali & Agrawal 2012). Its ability to thrive on a variety of hosts in combination with its dispersal potential has likely increased its probability of successful propagation in heterogeneous ecosystems.

Conclusion

Madagascar is known for its species diversity and endemism particularly in the case of Formicidae, a group that tends to yield high resolution phylogeographic data (Underwood & Fisher 2006). One of the primary approaches to quantifying the diversity of Madagascar's ant fauna has been the sequencing of genetic markers (Smith *et al.* 2005). However, genetic diversity can oftentimes be overestimated if the amplification of cryptic pseudogenes is not accounted for (Song *et al.* 2008). The current study demonstrated how cryptic pseudogene amplification inflated the genetic diversity of the Malagasy ant *M. insularis* using previously generated sequences and museum specimens only. The commonly used methods for pseudogene identification such as the detection of stop codons and frameshift mutations are not effective in the case of cryptic pseudogenes which tend to differ from their functional counterparts by 1-3 bp (Bertheau *et al.* 2011). The incorporation of, at minimum, a rudimentary screen for pseudogene amplification is critical in future studies of genetic diversity particularly when working with natural history collections (NHCs). NHCs are more likely to experience DNA degradation compounding pseudogene amplification (Wandeler *et al.* 2007). Despite the challenges associated with sequencing degraded DNA, NHCs

continue to demonstrate the potential for reducing the costs associated with biodiversity quantification both by supplementing novel biodiversity surveys and as part of a standalone analysis.

Table 1: A summary of the GMYC analysis p-values.

Dataset	***Null vs. Single Threshold***	***Null vs. Multiple Threshold***	***Single Threshold vs. Multiple Threshold***
COI	p = 0.184	p = 0.280	p = 0.996
12S	p = 0.698	p = 0.778	p = 0.994

Table 2: A summary of the r and p-values for Mantel tests of the COI, 28S and partition *M. insularis* datasets.

Dataset	***r***	***p-value***
COI	0.195	0.001
12S	0.127	0.022
Partition	0.128	0.067

Table 3: A summary of the hierarchal AMOVA results for the a) COI, b) 12S and c) partition datasets with specimen collection localities tested against unique habitats identified by Wilmé *et al.* (2006).

a)

Source of Variation	***d.f.***	***Sum of Squares***	***Variance Components***	***Percentage of Variation***	***p-value***	***Phi***
Among regions	11	602.284	-2.770	-7.590	0.465±0.005	Φ_{CT}: -0.076
Among populations within regions	26	1440.231	11.571	31.720	0.015±0.001	Φ_{SC}: 0.295
Within populations	88	2435.906	27.681	75.880	0.004±0.001	Φ_{ST}: 0.241
Total	125	4478.421	36.481			

b)

Source of Variation	***d.f.***	***Sum of Squares***	***Variance Components***	***Percentage of Variation***	***p-value***	***Phi***
Among regions	4	116.232	2.123	7.310	0.339±0.005	Φ_{CT}: 0.073
Among populations within regions	1	15.974	-2.406	-8.290	0.677±0.004	Φ_{SC}: -0.089
Within populations	57	1670.206	29.302	100.980	0.492±0.005	Φ_{ST}: -0.010
Total	62	1802.413	29.017			

c)

Source of Variation	***d.f.***	***Sum of Squares***	***Variance Components***	***Percentage of Variation***	***p-value***	***Phi***
Among regions	4	275.313	4.331	7.600	0.206±0.004	Φ_{CT}: 0.076
Among populations within regions	1	37.329	-3.373	-5.920	0.790±0.004	Φ_{SC}: -0.064
Within populations	55	3080.423	56.008	98.320	0.271±0.004	Φ_{ST}: 0.017
Total	60	3393.066	56.966			

Figure 1: A *Melissotarsus insularis* specimen point-mounted at the University of Guelph (2011) (sample ID CASENT0426503-D02).

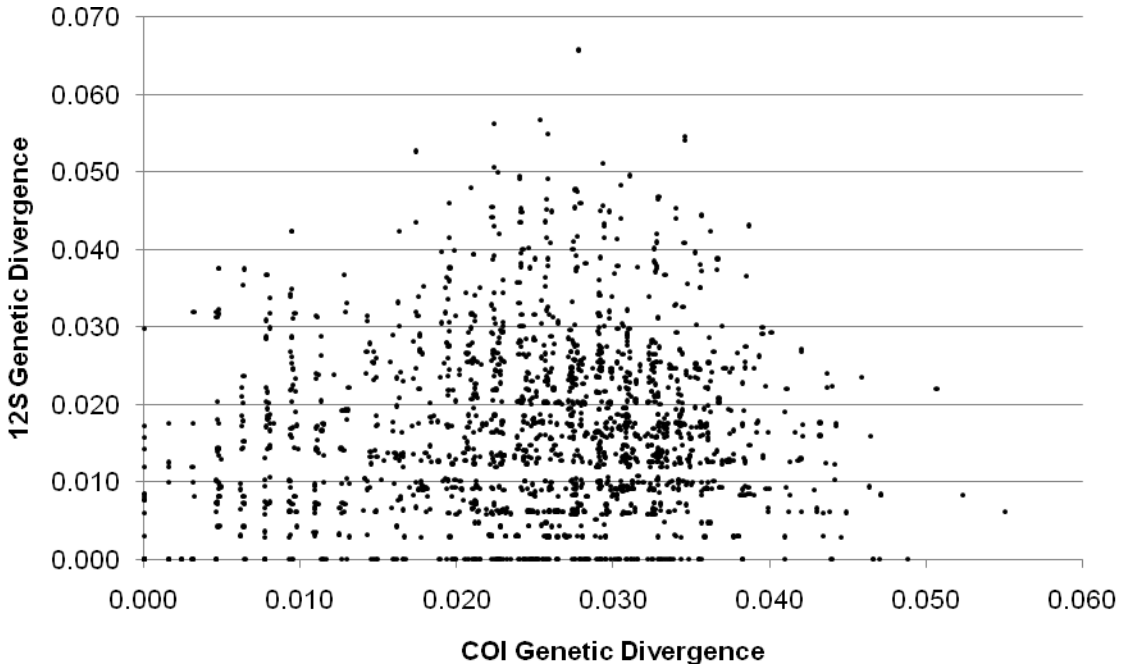


Figure 2: Pairwise sequence divergence for COI and 12S determined in MEGA 5.0 (Tamura *et al.* 2011) using the Kimura 2-Parameter model with the gamma parameter specified based on ModelTest v. 3.7 (Posada & Crandall 1998) results. A Mantel test comparing the 12S and COI matrices yielded an r of 0.06246 and a p-value of 0.05123.

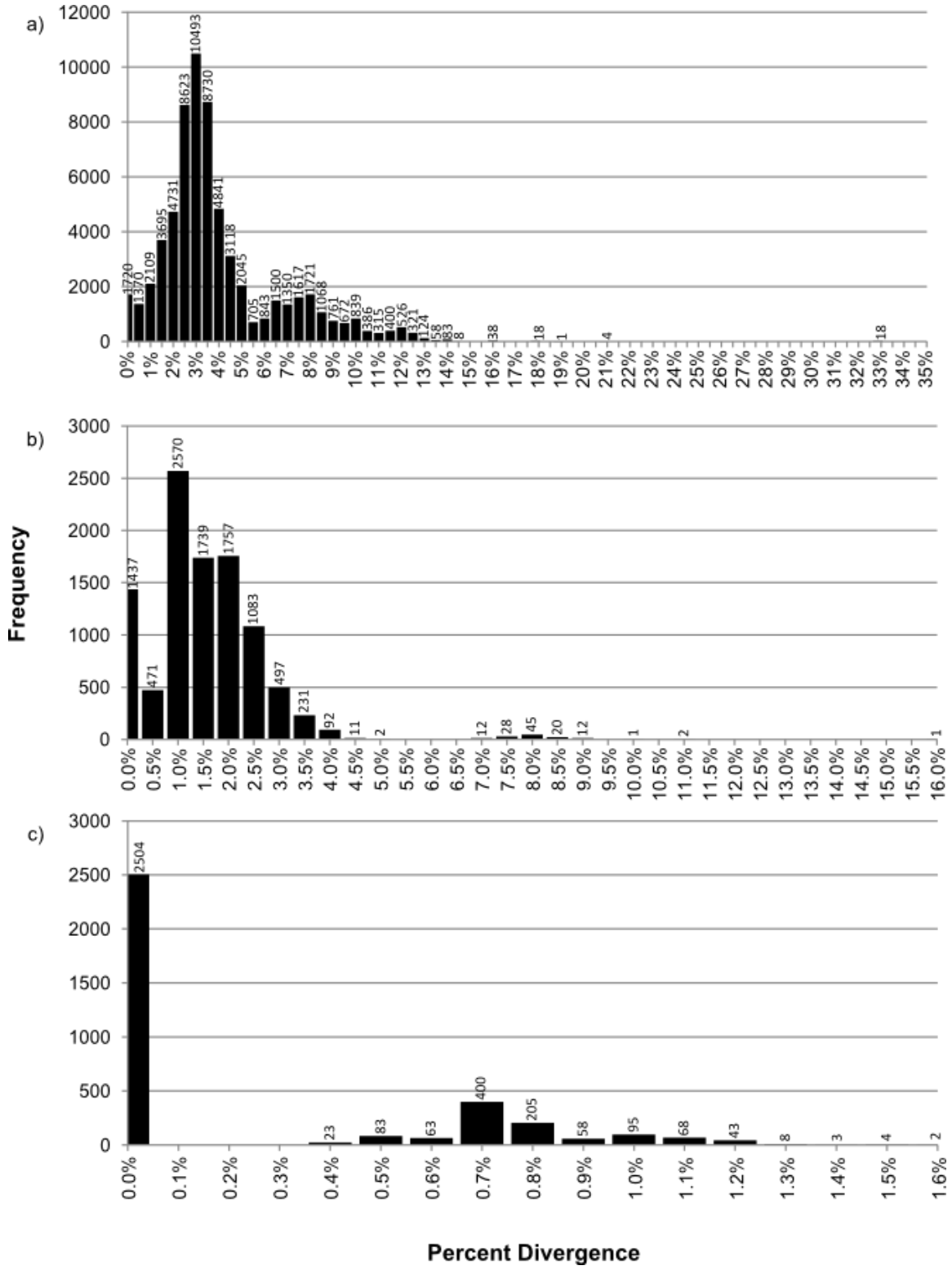


Figure 3: Frequency historgram of genetic percent divergence for the a) COI b) 12S and c) histone H3 datasets generated in MEGA 5.0 (Tamura *et al.* 2011). The frequency of each divergence value is included above the bars. Values of 0 have been omitted.

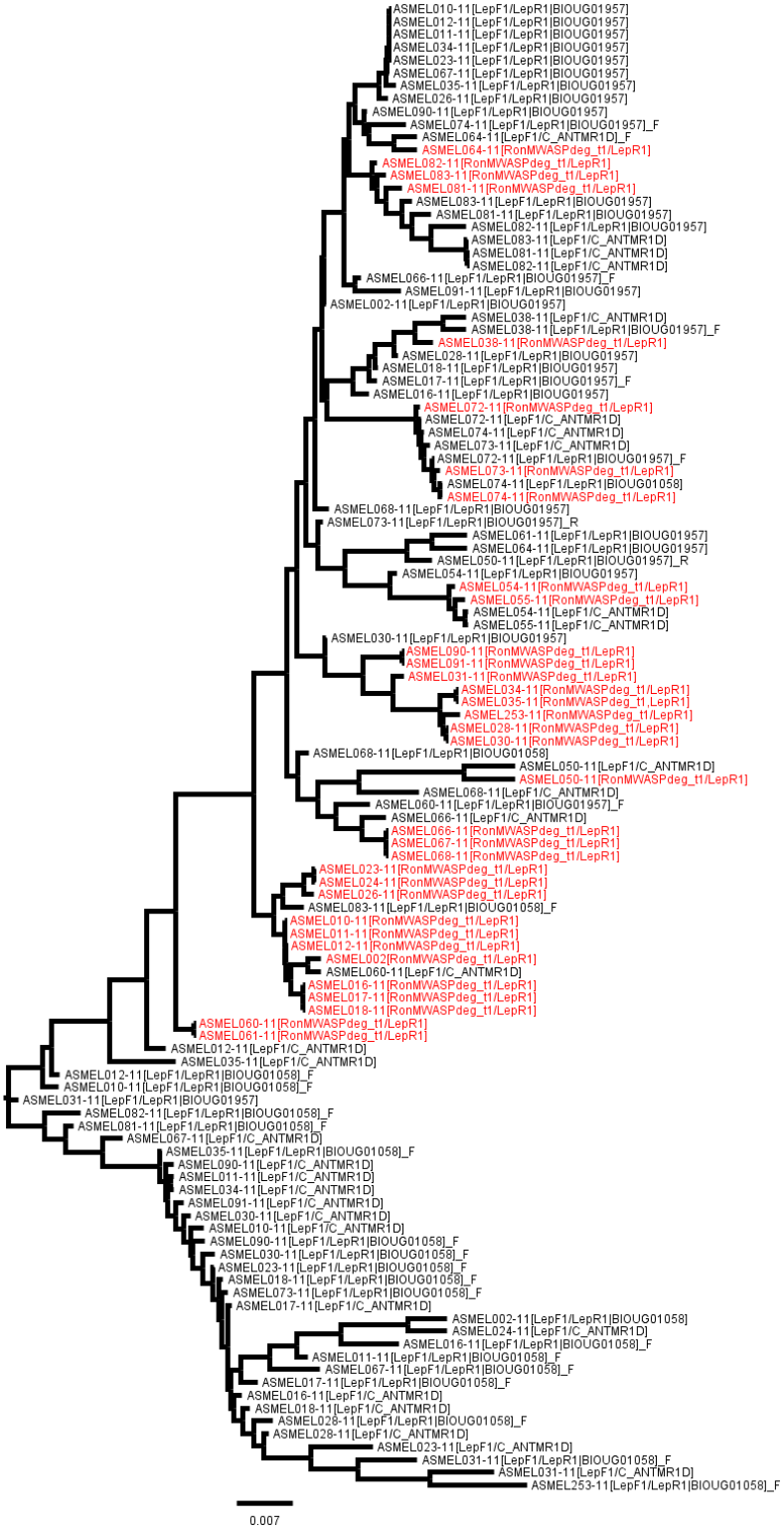


Figure 4: Neighbour-joining phylogenetic tree of all overlapping incongruent amplifications including RonMWASPdeg_t1 generated in MEGA 5.0 (Tamura *et al.* 2011). The RonMWASPdeg_t1 amplifications have been highlighted in red.

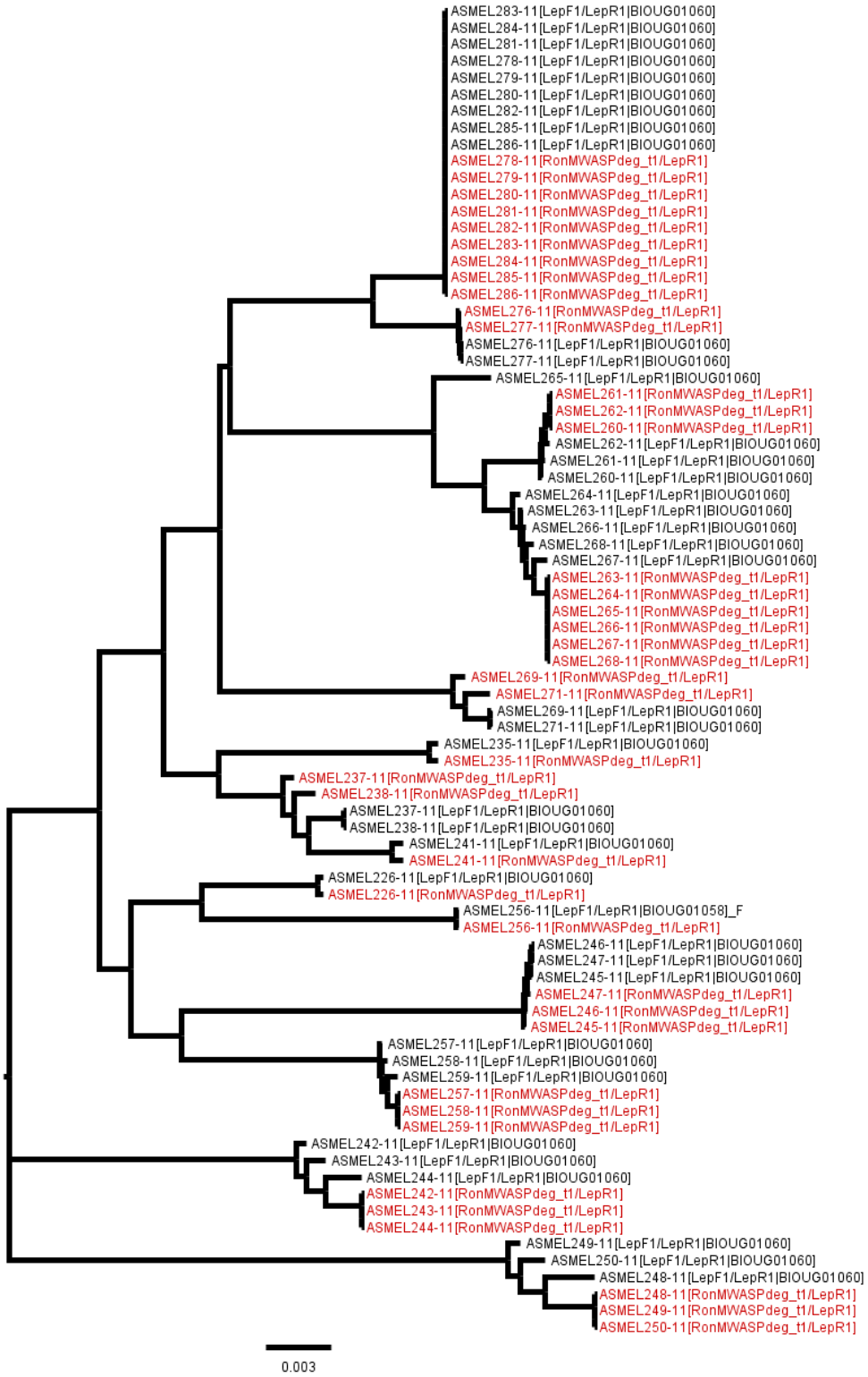


Figure 5: Neighbour-joining phylogenetic trees of all overlapping congruent amplifications including RonMWASPdeg_t1 generated in MEGA 5.0 (Tamura *et al.* 2011). The RonMWASPdeg_t1 amplifications have been highlighted in red.

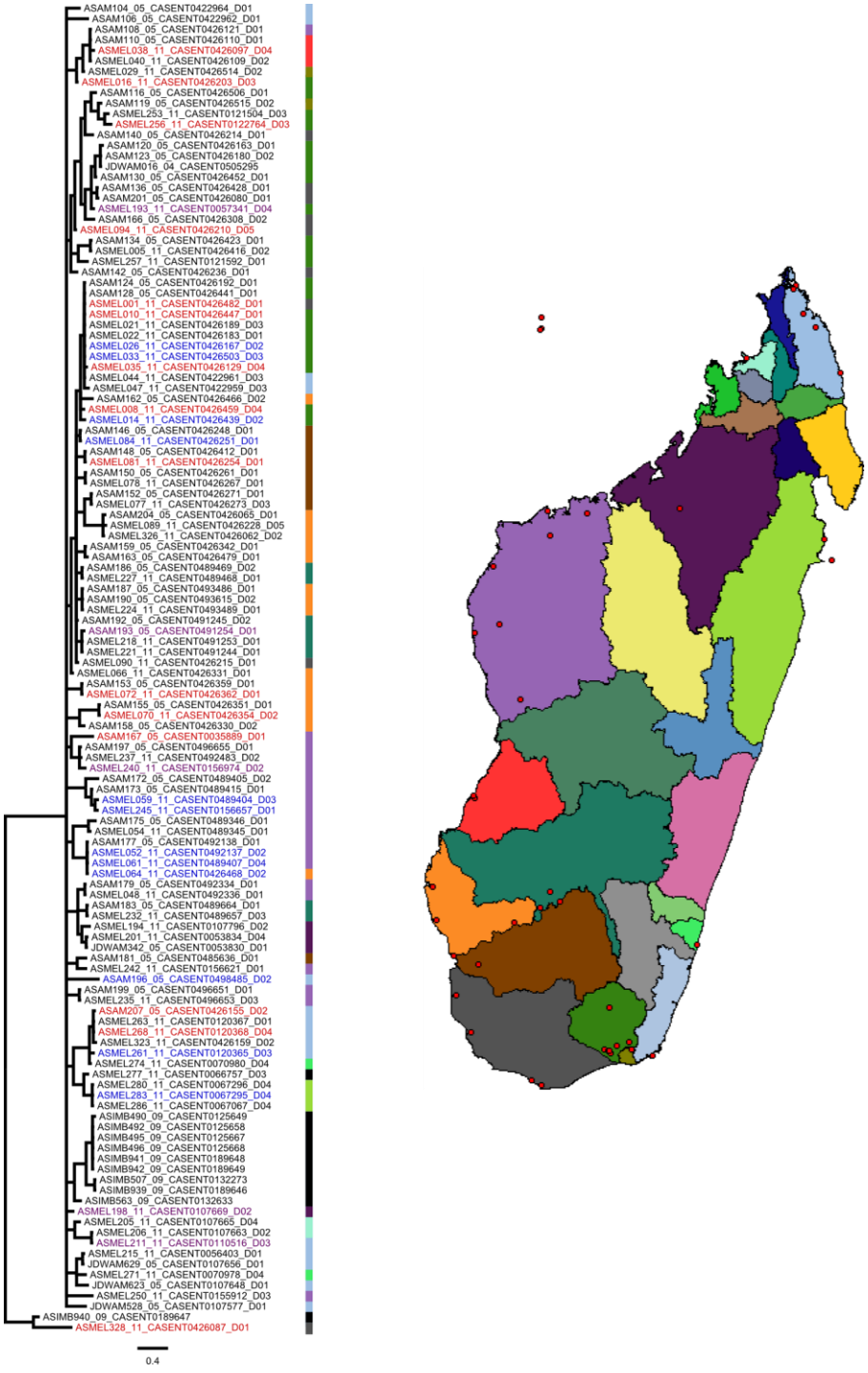


Figure 6: *M. insularis* COI phylogenetic tree generated in MrBayes version 3.1.2 (Huelsenbeck & Ronquist 2001). The parsimony informative site in the H3 dataset is indicated by the colour of the taxa names as a T (red), a C (blue) or a Y (purple). The colour coded bar on the right identifies branches corresponding to unique habitats identified by Wilmé *et al.* (2006). Black bars represent specimens collected on the island of Mayotte off the coast of Madagascar.

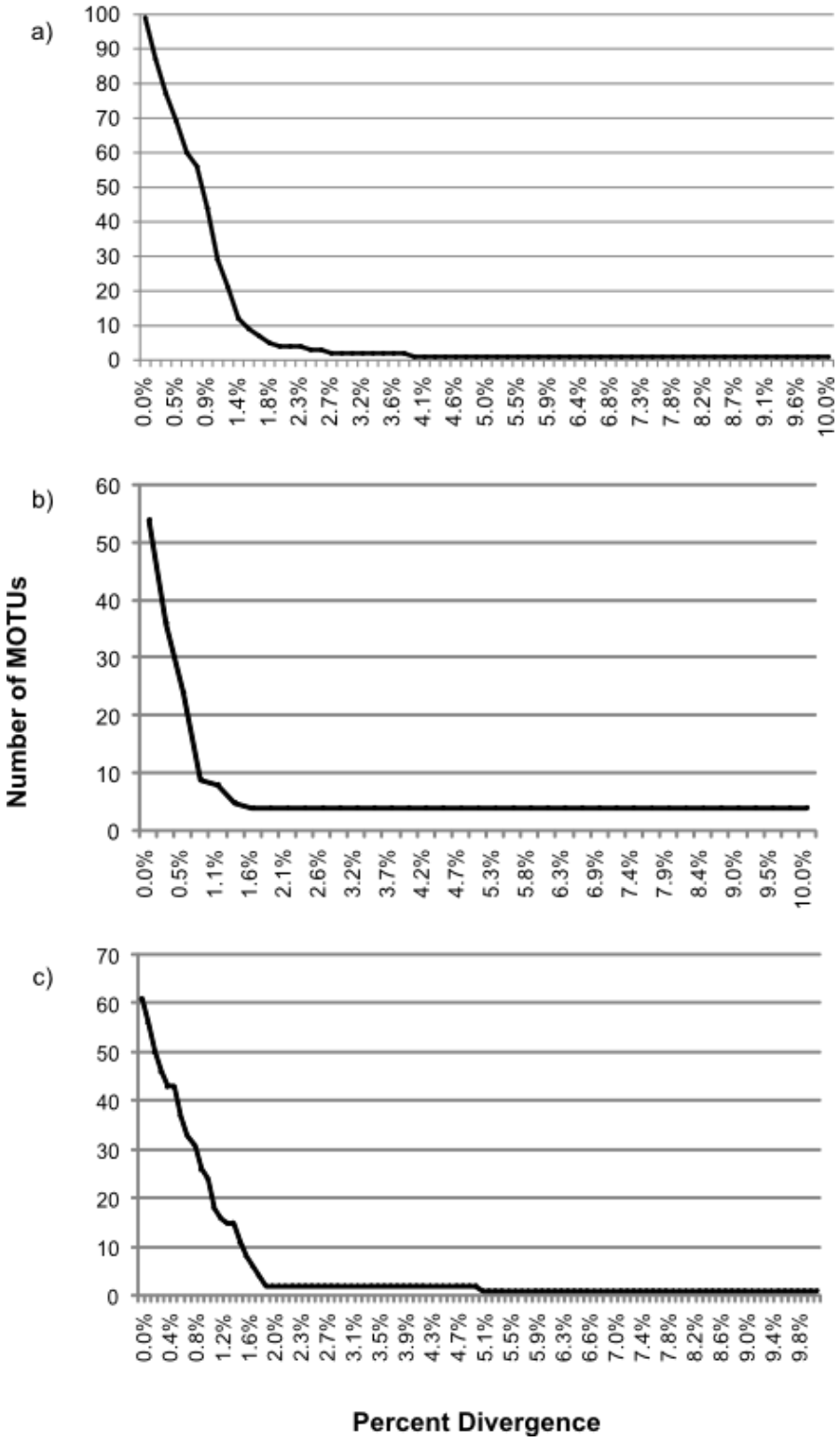


Figure 7: The number of MOTUs identified by jMOTU in relation to percent divergence for the a) COI, b) 12S and c) partition datasets using the default threshold of 97%.

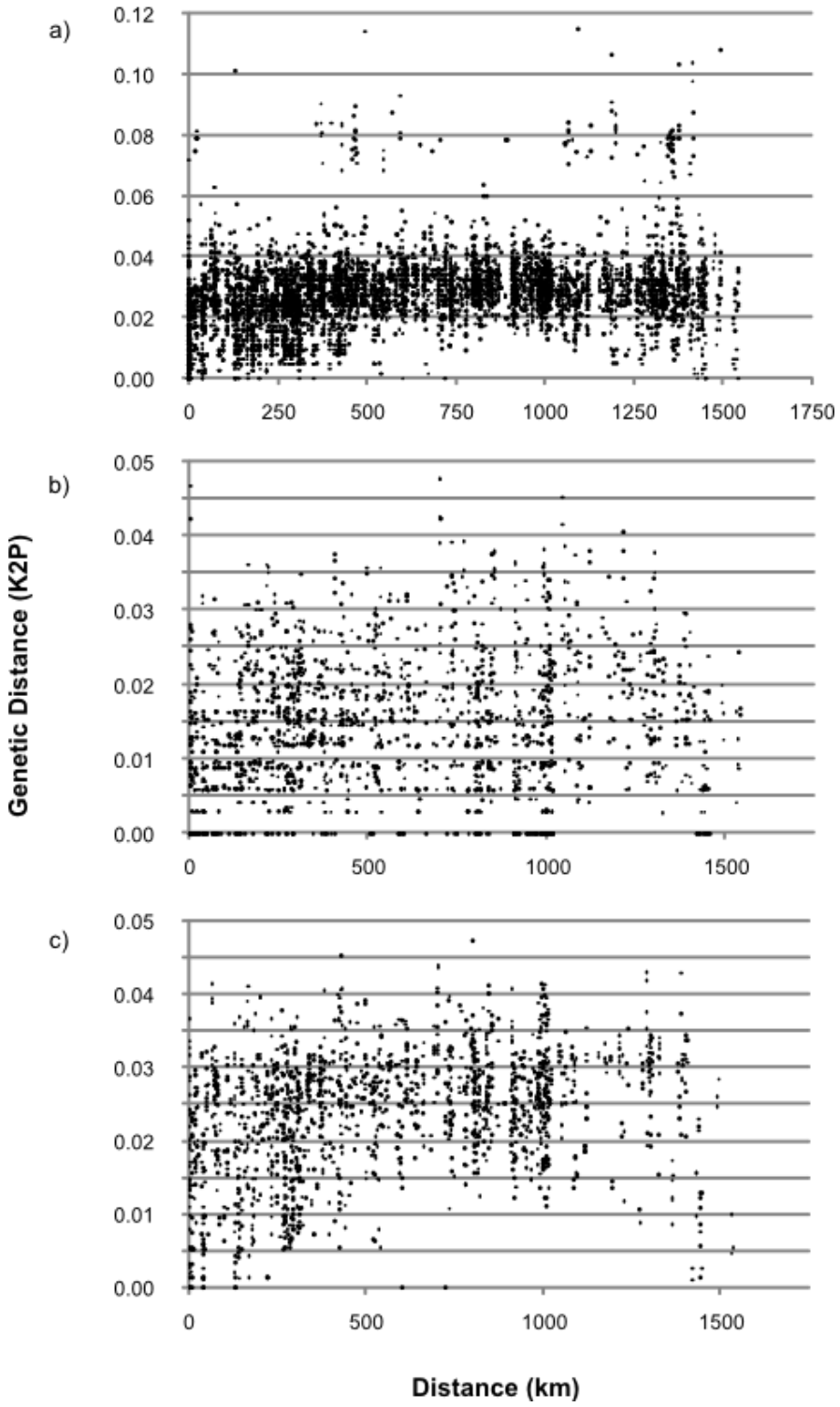


Figure 8: Genetic distance exhibited by *Melissotarsus insularis* in relation to geographic distance by dataset: a) COI, b) 12S and c) partition. Genetic distance was determined in MEGA 5.0 (Tamura *et al.* 2011) using the Kimura 2-Parameter model with the gamma parameter specified according to ModelTest v. 3.7 (Posada & Crandall 1998) results.

Chapter 3: Inferring the evolutionary history of symbiotic relationships via cophylogenetic analysis: *Melissotarsus insularis* and its scale insect (Diaspididae) associate.

Abstract

Symbiotic relationships play a primary role in maintaining sustainable patterns of biodiversity. Advances in cophylogenetic reconstruction software have made it possible to infer the evolutionary history of symbiotic relationships using molecular data. Fragments of the COI and 28S (D2 region) gene were sequenced from 130 specimens of the scale insect Diaspididae. Similarly to its Malagasy ant associate *Melissotarsus insluaris*, Diaspididae did not exhibit diversification by habitat but did show evidence of isolation by distance (Mantel p-values of 0.00100 and 0.27673 for COI and 28S respectively). The relationship between Diaspididae and *M. insularis*, was explored using a previously generated *M. insularis* phylogeny. A novel cophylogenetic reconstruction program, CoRe-PA, was used to infer 14 cospeciations, 11 sortings, 4 duplications and 21 host switches between *M. insularis* and Diaspididae. As expected the number of cospeciations fell in the upper tail of the normal distribution but was not significant unlike with most host and parasite relationships. Analyses of 25%, 50% and 75% reduced datasets yielded p-values ranging from 0.001 to 0.324 suggesting that incomplete taxon sampling may significantly affect cophylogenetic reconstruction results.

Introduction

Biodiversity conservation programs often rely on analyses of species richness derived from the surveillance of predetermined species groups such as biodiversity indicators to monitor the effects of anthropogenic activities on biodiversity loss (Davies & Cadotte 2011). However the predictive value of species richness has been criticized as it represents only one aspect of sustainable biodiversity (McCann 2007; Reyers *et al.* 2002). Biodiversity is maintained by ecosystem processes responsible for sustaining ecosystem structure and function such as migration patterns, seasonal cycles and symbiotic relationships (Bastolla *et al.* 2009; Kiers *et al.* 2010; Maddock & Plessis 1999; Reyers *et al.* 2002). Significant shifts or disturbances within the ecosystem have been shown to affect the expression of these processes (Palmer *et al.* 2008; Parmesan & Yohe 2003; Tylianakis *et al.* 2008). The ability of ecosystem processes to respond to major shifts in ecosystem functioning may implicate them as suitable indicators of sustainable biodiversity particularly when combined with surveys of species richness. Although a potentially predictive ecosystem process, symbiotic relationships are rarely the focus of biodiversity studies potentially due to a lack of a standardized metric much like the one used in documenting species richness (Dyer *et al.* 2010).

Both abiotic and biotic environmental disturbances can alter the structure of symbiotic relationships by affecting the characteristics of the associated species such as behaviour and age of maturation (Palmer *et al.* 2008; Tylianakis *et al.* 2008). Changes in individual species relationships have been shown to compound at the community level (Palmer *et al.* 2008; Tylianakis *et al.* 2008). A review of 688 studies concluded that drivers of global environmental change alter trophic interactions amongst decomposers, weaken mutualisms involving plants and increase herbivory (Tylianakis *et al.* 2008). Although the shift from mutualism to antagonism is seen as a progression common in mutualistic relationships, it has been suggested that the normally gradual process has been accelerated by anthropogenic activities (Kiers *et al.* 2010). Furthermore, symbiotic relationships have been shown to breakdown as the frequency of specialists decreases

and the frequency of generalists or species that succeed in disturbed environments increases in response to climate change (Dukes & Mooney 1999). Some inter-specific relationships have already decoupled due to differing phenological and morphological responses to climate change while many more are threatened as a result of anthropogenic activities (Kiers *et al.* 2010; Moricca & Ragazzi 2008; Visser & Both 2005).

The mechanism responsible for the development of symbiotic relationships, broadly referred to as codiversification, is reflected in the genome allowing us to infer the evolutionary history of symbiotic relationships using phylogenetic analysis (Clark *et al.* 2000; Silvieus *et al.* 2007). The strength of the correlation between the individual genetic diversities of the associated species is dependent upon the nature of their relationship (Clark *et al.* 2000). Species involved in established mutualistic relationships are more likely to correlate in their genetic diversity than newly affiliated species or generalists (Quek *et al.* 2004). However, it is possible for the relationship between two or more species to vary from mutualism to antagonism depending on the community structure and abiotic characteristics of their habitat (Thompson 1999). This variation in the nature of symbiotic relationships is referred to as the geographic mosaic theory of coevolution (Thompson 1994). The geographic mosaic theory of coevolution suggests that symbiotic relationships exist within a selection mosaic that favours different evolutionary trajectories within different populations and that these trajectories experience geographic shifts (Thompson 1999). According to the geographic mosaic theory of coevolution, the lack of an intimate evolutionary history between species involved in a symbiotic relationship is to be expected.

Evidence of codiversification in the phylogenetic trees of two or more associated species is referred to as cophylogeny (Silvieus *et al.* 2007). The study of cophylogeny is relatively new and has experienced the majority of its development within the past decade (Desdevises 2007). The idea that the phylogeny of a parasite should reflect that of its host was pioneered by Farenholz (1913) and thereafter referred to as the Farenholz's Rule (Desdevises 2007). However, it soon became evident that a number

of different evolutionary events could affect the cophylogeny of a parasite and its host resulting in incongruence between phylogenetic trees (Desdevises 2007). This led to the definition of four coevolutionary events: cospeciation, host-switching, duplication and lineage sorting (Desdevises 2007; Shafer *et al.* 2009).

The methods used for cophylogenetic analysis may be broadly grouped into two categories: global-fit methods and event-based methods (Desdevises 2007). Global-fit methods utilize a distance matrix to determine the global congruence between the phylogenetic trees of the host and parasite (Legendre *et al.* 2002). Although the results obtained by global-fit methods are difficult to relate to biological events, they remain in use as they can account for polytomies and inaccuracies in the phylogenetic trees of the host and parasite (Desdevises 2007; Legendre *et al.* 2002). Event-based methods assign costs to each evolutionary event and determine the least costly combination of events that can account for the relationship between the host and parasite phylogenies (Desdevises 2007; Light & Hafner 2007). CoRe-PA 0.5.1, an event-based method developed by Merkle *et al.* (2010) employs the first algorithm capable of processing polytomies. CoRe-PA 0.5.1 estimates the number of cospeciation, sorting, duplication and host switching events given a host and parasite phylogeny and gives the option of running a statistical test that determines the normal distribution for each event type (Merkle *et al.* 2010). The statistical test can be run by either randomizing the host tree, the parasite tree, both the host and parasite tree or the associations between the tips of the two trees (Merkle *et al.* 2010). The results can be analyzed to determine whether the number of observed cospeciation events significantly deviates from the number of expected cospeciation events (Merkle *et al.* 2010). Variation in the normal distribution of the evolutionary events in response to dataset size and the type of randomization selected is yet to be investigated. Analysis of the phylogenetic software, TreeMap v2.0, reveals that sampling strategy plays a role in the number of codivergences deduced using cophylogenetic analysis (Jackson 2004).

The majority of studies that have employed cophylogenetic analysis focused on species involved in well-studied host-parasite relationships: chewing lice and pocket gophers;

Light and Hafner (2007); European bats and ectoparasitic mites; Bruyndonckx *et al.* (2009); the nasal mite *Ptilonyssus sairae* and its passerine bird host; Morelli and Spicer (2007); bees and their microsporidian parasite, *Nosema*; Shafer *et al.* (2009); Puumala virus and bank voles; Nemirov *et al.* (2010); old world arenaviruses and rodents; Coulibaly-N'Golo *et al.* (2011) and Orbiviruses and their vectors; Dilcher *et al.* (2012). Thus far, CoRe-PA 0.5.1 has only been used to reconstruct cophylogenies within host-virus systems (Coulibaly-N'Golo *et al.* 2011; Dilcher *et al.* 2012; Nemirov *et al.* 2010). Few studies have reconstructed the cophylogeny between species involved in mutualistic or poorly understood relationships (Silvieus *et al.* 2007). Highly host-specific mutualisms are rare, most likely due to the inherent fragility of the interaction (Hoeksema & Bruna 2000). Unlike with host-parasite relationships where the host drives speciation by directly influencing the distribution and external environment of the parasite, mutualisms are often facultative (Bruyndonckx *et al.* 2009; Hoeksema & Bruna 2000).

In order to explore the utility of cophylogenetic software in inferring the evolutionary history of mutualistic relationships I will focus on elucidating the symbiotic relationship between the Malagasy ant *Melissotarsus insularis* and its scale associate in the family Diaspididae. The association of Diaspididae with an ant has only been recorded for the genus *Melissotarsus* (Ben-Dov 2010; Ben-Dov & Fisher 2010). The benefit derived by *Melissotarsus* from the maintenance of Diaspididae is unclear as Diaspididae do not produce honeydew unlike *Coccus* scales which commonly maintain mutualisms with ants (Ben-Dov 2010; Mony *et al.* 2002). It has been hypothesized that *Melissotarsus* maintain diaspid populations as a secondary food source (Ben-Dov & Fisher 2010; Mony *et al.* 2002). Both *Melissotarsus* and the scales reside within galleries constructed by *Melissotarsus* in the live bark of 23 dicotyledonous tree species (Fig. 2) (Ben-Dov 2010; Fisher & Robertson 1999; Mony *et al.* 2002). Although the swarming of *Melissotarsus* queens has been observed, no record has been made of *Melissotarsus* queens transporting diaspid symbionts for the purpose of colonizing a novel host plant (Mony *et al.* 2002). Furthermore, preliminary analyses show evidence of deeper

divergences within Diaspididae phylogenies when compared to *M. insularis* suggesting the lack of vertical transmission of Diapididae.

Although little is known about the degree of intimacy inherent in the evolutionary relationship between *Melissotarsus* and Diaspididae (Ben-Dov & Fisher 2010), an analogous relationship between the ant *Crematogaster* and its scale associate *Coccus* has been investigated on several occasions (Ueda *et al.* 2008; Ueda *et al.* 2010). Although Ueda et al. (2008) did not use a cophylogenetic reconstruction program, they did compare the *Crematogaster* and *Coccus* phylogenies manually and consequently hypothesized that they may not share an intimate evolutionary relationship. In order to develop a frame of reference for my hypothesis I generated a cophylogenetic reconstruction of the *Crematogaster-Coccus* symbiosis using CoRe-PA 0.5.1 and sequences made publicly available by Quek *et al.* (2007) and Ueda *et al.* (2010) (Fig. 9). The inferred number of cospeciations given the associations between tree tips was three with a p-value of 0.77 when associations were randomized. The result supports the hypothesis suggested by Ueda *et al.* (2008). The low number of cospeciations is expected in this case as the association between *Crematogaster* and *Coccus* species are not specific. In most cases one *Crematogaster* species tends to up to six *Coccus* species while some *Coccus* species are associated with up to seven ants. It is import to note that, unlike Ueda *et al.* (2008) who focused on species, I will focus on comparing the associations between individual ant colonies.

The current study will 1) analyze the phylogenetic relationship between *Melissotarsus insularis* and its scale symbiont in the family Diaspididae using CoRe-PA 0.5.1 (Merkle *et al.* 2010) and 2) compare the results of the analysis with those derived from datasets 75%, 50% and 25% the size of the initial dataset and those derived using all four options for randomization: randomizing the host tree, the parasite tree, both the host and parasite tree and the associations between the tips of the two trees. Given the predictions established by the geographic mosaic theory of coevolution (Thompson 1999) whether a cophylogenetic reconstruction of individuals residing within the same colony will produce a result indicative of a more or less intimate relationship between *M.*

insularis and Diaspididae is unclear. However, the *Crematogaster* and *Coccus* relationship acts as a model for ant and scale symbioses and thus far has demonstrated a lack of intimacy in comparison to host and parasite cophylogenetic reconstructions (Dilcher *et al.* 2012; Nemirov *et al.* 2010). Consequently I expect the number of observed cospeciations in the *M. insularis* and Diaspididae cophylogenetic reconstruction to fall within the upper tail end of the distribution but not to achieve significance. Furthermore, I expect the significance of the observed cospeciations to vary both positively and negatively with sample size and randomization strategy.

Methods

Specimens

The Diaspididae phylogeny was generated using a dataset comprised of a combination of novel and previously available sequences. I extracted and amplified the barcoding region of the mitochondrial COI gene and the nuclear rRNA gene 28S from 67 collection events of Diaspididae specimens. The specimens were collected by Brian Fisher and the Fisher-Griswold Arthropod Team between 2001 and 2009. They were manually extracted from various plant species across Madagascar and stored in ethanol at the California Academy of Sciences before their shipment to the University of Guelph for analysis in 2010. When available, three specimens from each collection event were selected for subsampling for a total of 130 specimens.

A total of 28 COI and 77 28S (D2 region) Diaspididae sequences were uploaded to the Barcode of Life Data Systems (BOLD) (Ratnasingham & Hebert 2007) between 2005 and 2009. Scale insect specimens were either identified to family only, *Diaspididae*, or as *Melanaspis* mga, *Morganella conspicua* and *Morganella* mga. One specimen from each collection event was included in the phylogenetic analysis. Specimens were selected based on sequence length. In cases where a longer sequence was generated from the same collection event the longer sequence was chosen. *Hemiberlesia lataniae*

sequences accessed through GenBank were included as the outgroup for the COI and 28S phylogenies (GenBank accession numbers HQ179913.1 and FJ040870.1 respectively). For a list of specimen process IDs, sample IDs, GenBank accession numbers and collection details see Appendix 1.

DNA Extraction, Amplification and Sequencing

All specimens were photographed with a Dino-Lite Pro2 (AD-4132TA) microscope prior to extraction. Specimens were transferred to a 96-well plate containing 30 µl of ethanol using forceps sterilized with ethanol. DNA extraction was conducted at the Biodiversity Institute of Ontario as described by Ivanova *et al.* (2006). COI was amplified from Diaspididae using primers PcoF and LepR1 and 28S was amplified using primers D2B and D3Ar. For a detailed list of primers see Appendix 2 in the supplementary material.

Amplification and sequencing was conducted according to standard DNA barcoding protocols based on Hajibabaei *et al.* (2005) and Smith *et al.* (2008). To amplify target DNA, a solution comprised of 625 µl of trehalose, 200 µl of ddH_2O, 125 µl of 10x buffer, 62.5 µl of $MgCl_2$, 12.5 µl each of forward and reverse primer, 6.25 µl of 10 mM dNTPs and 6 µl of Taq polymerase, was aliquoted into a 96 well plate and 1.0 µl of DNA template was added to each well. The plate was then run on an Eppendorf® thermal cycler (6325). PCR temperature profiles used for amplification differed with primer. For a detailed description of PCR temperature profiles by primer see Appendix 3 in the supplementary material. Amplicons were visualized using pre-cast E-Gel® agarose gels by Invitrogen™.

A similar protocol was utilized for sequencing. A solution comprised of 550.0 µl of trehalose, 210.83 µl of ddH_2O, 100.83 µl of 5x buffer, 110 µl of 10 µM forward or reverse primer and 18.33 µl of BigDye® was aliquoted into a 96 well plate and 2.0 µl of amplicon were added to each well. The plate was then run on an Eppendorf® thermal cycler using a PCR sequencing program with the following temperature profile: 2:00 minutes at 96°C, 30 cycles of 30 seconds at 96°C, 15 seconds at 55°C and 4:00

minutes at 66°C. Target amplicons were sequenced using the ABI BigDye™ v. 3.1. Cycle Sequencing Kit on an ABI 3730XL (Biodiversity Institute of Ontario) or ABI3730 (Genomics Facility, Advanced Analysis Centre) sequencing platform.

Sequence Editing and Phylogenetic Analysis

Sequences were edited and aligned using Sequencher v. 4.5 and v. 5.0 (Gene Codes Corporation, Ann Arbor, Michigan) and BioEdit (Hall 1999) respectively. Sequences 160 bp or less in length or those that were more than 2% heteroplasmic were excluded from further analysis.

Incongruence between amplifications originating from the same colony was investigated using the number of synonymous (dS) and nonsynonymous (dN) changes. Sequences that exhibited a marked deviation from the final alignment were excluded from further analysis. COI and 28S phylogenetic trees were generated using one scale as a representative from each collection event. Specimens were chosen based on sequence length and sequence quality expressed in Sequencher version 5.0 as the proportion of a sequence comprised of quality base calls (GeneCodesCorporation 2011). A subset of COI and 28S sequences was utilized to conduct a partition analysis. Neighbour joining and maximum parsimony analyses were conducted in MEGA 5.0 (Tamura *et al.* 2011) while Bayesian and partition analyses were conducted in MrBayes version 3.1.2 (Huelsenbeck & Ronquist 2001). All analyses used evolutionary models determined by the hierarchical likelihood ratio test employed in MrModelTest v. 3.7 (Nylander, 2004). Analyses conducted in MEGA 5.0 used the Tamura-Nei model of evolution; the closest available model to the best-fit model determined by the hLRT with the gamma parameter specified. The Bayesian analyses were run for 60,000,000 million generations each. Burnin was set at 25% and confirmed in Tracer v1.5 (Rambaut & Drummond 2007).

The single and multiple threshold GMYC models (Monaghan *et al.* 2009) were utilized to determine a potential species boundary and jMOTU (Jones *et al.* 2011) was used to

delimit MOTUs across a range of 0.0001-10% divergence at the default threshold of 97%. The ultrametric trees used in the GMYC analysis were generated in BEAST v 1.6.0 (Drummond & Rambaut 2007). DIVA-GIS v. 7.5 (Hijmans *et al.* 2001) was used to associate sequences with unique habitats previously identified by Wilmé *et al.* (2006) based on their distribution. Mantel tests were conducted in PASSaGE (Rosenberg & Anderson 2011) using genetic distances determined under the Kimura 2-Parameter model of evolution (with the gamma parameter specified based on Modeltest v. 3.7 results) and Euclidean distances determined via the spherical functions implemented in the Geographic Distance Matrix Generator v. 1.2.3 (Ersts Internet). AMOVA analyses was conducted in Arlequin v. 3.5 (Excoffier & Lischer 2010).

Cophylogenetic Analysis

The cophylogenetic reconstruction for *M. insularis* and Diaspididae was conducted in CoRe-PA 0.5.1 (Merkle *et al.* 2010). The number of randomizations was set to 1,000 and default costs for evolutionary events were maintained (cospeciation: -2, sorting: 1, duplication: 2, host-switch: 2). Analyses were run on 75%, 50% and 25% of the original dataset. *M. insularis* and Diaspididae pairs were gradually excluded from the host and parasite trees using a random number generator. In order to test whether error in the phylogenetic trees or in the associations between taxa had a significant effect on the number of cospeciations inferred the analysis was run using all four options for randomization: randomizing the associations between taxa, host tree, parasite tree and both trees. Although CoRe-PA 0.5.1 uses the term “cospeciation” to refer to congruence in the branching pattern of the associated species, in the case of this study it is used to refer to codiversification as it was inter-population as opposed to inter-specific interactions that were tested.

Results

Sequence statistics

A total of 130 Diaspididae specimens were amplified yielding 77 (59%) COI sequences and 51 (39%) 28S sequences that complied with the sequence quality guidelines established in the Methods. COI exhibited an AT bias common to insect mitochondrial sequences (Schwarz *et al.* 2004) with a nucleotide composition of 40% T, 13% C, 40% A and 7% G. 28S had a nucleotide composition of 21% T, 27% C, 19% A and 33% G. COI yielded 270 (42%) parsimony informative sites while 28S yielded 64 (15%) parsimony informative sites. All COI sequences had a nine base pair, in frame deletion at position 1864 when aligned with the NCBI reference sequence NC_001709.1 likely occurring in an intermembrane external loop in relation to the *Bos taurus* COI annotation provided by UniProt. It is unlikely that the deletion indicated a pseudogene as it was in frame and occurred in one of the external loops rather than a more functionally conserved trans-membrane helix (Lunt *et al.* 1996)

Incongruent Amplifications

Although all Diaspididae specimens sequenced from the same collection event originated from one *Melissotarsus insularis* colony, there was no *a priori* expectation that they were related or even members of the same species. Therefore, unlike with *M.insularis*, the apparent lack of congruence between sequences from the same collection event did not necessarily indicate the presence of a pseudogene. Six Diaspididae sequences exhibited a marked divergence (exceeding 10%) from the final alignment (i.e. a 28S sequence amplified from ASMEL103-11) resulting in their exclusion from phylogenetic analysis. A greater proportion of the incongruent sequences exhibited a minor divergence (less than 6%). Where possible, a congruent sequence from the same collection event was used.

Phylogenetic Analysis

The best fit model determined for COI and 28S by the hierarchical likelihood ratio tests (hLRTs) in ModelTest v. 3.7 (Posada & Crandall 1998) was GTR+I+G and TrNef+G respectively. The Tamura-Nei model was used for the neighbour-joining and maximum parsimony analyses conducted in MEGA 5.0 (Tamura *et al.* 2011). The best fit model for both the COI and 28S partitions had gamma distributed among-site rate variation with six substitution types. Partitions were therefore defined but run under the same model.

Phylogenetic Trees of Individual Genes

Six major clades were conserved across the Diaspididae COI neighbour-joining, maximum parsimony and Bayes trees. One clade comprised of two sequences, ASAM919-05 and ASMEL184-11, exhibited a deeper divergence on the Bayes and NJ trees than on the MP tree.

The 28S Diaspididae sequences formed five major clades on the neighbour-joining, maximum parsimony and Bayes trees. However two sequences that comprised clades in the maximum parsimony and Bayes trees, ASAMI063-05 and ASMEL151-11, were grouped individually on the neighbour joining tree. Furthermore, ASMEL099-11 paired with ASAM919-05 on the neighbour joining tree likely as a result of long-branch attraction. ASMEL099-11 is included in one of the major clades on both the maximum parsimony and Bayes trees.

The majority of the clades on the 28S phylogenetic tree were unresolved while only five sequences were unresolved on the COI phylogenetic tree. This is likely due to the slower rate of evolution experienced by 28S as a nuclear gene (Saccone *et al.* 2000) resulting in lower, on average, genetic divergences than those exhibited by COI. Because the COI and 28S phylogenetic trees used different sets of sequences only some sequences were included on both trees making them difficult to compare. The major clades comprising the COI phylogeny differ from those comprising the 28S

phylogeny, however some strongly supported sequence pairs, such as ASAMI076-05 and ASAMI080-05, appear on both trees.

The COI partition comprised 60% of the partition dataset biasing the resultant phylogeny. The clades established by the partition phylogeny corresponded to those established by the COI phylogeny. Clades with high bootstrap support or posterior probabilities, usually comprised of two sequences, were present on the partition, COI and 28S phyogenies. See Appendix 5 for all phylogenetic trees.

The clades established on the COI Bayes phylogenetic tree did not appear to group by habitat and only grouped by latitude in some cases. It was not uncommon for specimens from north Madagascar to form clades with specimens from the south (see Fig. 10). The absence of a correlation between specimens originating from the same habitat and position on the phylogenetic tree may be the result of incomplete sampling. The majority of the specimens on the Bayes COI phylogeny exhibit deep divergences apart from the group situated at the bottom of the tree. The bottommost group exhibits relatively shallow genetic divergences and is comprised of specimens from a limited range of latitudes. It is possible that the groups exhibiting deeper divergences represent separate species but lack definition due to underrepresentation in the dataset. Consequently specimens sampled at different latitudes are grouped together on the COI tree despite relatively deep genetic divergences.

Mantel Analyses and AMOVA

Mantel analyses of the COI and 28S datasets both yielded a statistically significant, positive r value (see Table 4). In the case of COI, the relationship between genetic and geographic distance was stronger for Diaspididae than *M. insularis* which had an r of 0.19497. The r value for the partition dataset exceeded that of the COI dataset resulting from stochasticity due to a reduced dataset or correlation between the genetic distances of the COI and 28S sequences. A Mantel test comparing the genetic divergences of COI and 28S yielded a statistically significant r of 0.57476 (see Fig 11). The relationship

between genetic divergence and geographic distance has been summarized in Figure 12 for the COI, 28S and partition datasets.

The relationship between genetic divergence and the unique habitats identified by Wilmé *et al.* (2006) was tested using an AMOVA approach. Analyses were run on COI, 28S and partition datasets. For the COI and 28S datasets the majority of the variation occurred within populations (100.38% for COI and 93.03% for 28S) and variation between unique habitats (regions) (8.43% for COI and 5.63% for 28S) was not significant (see Table 5). Variation among populations (-8.810%) was negative and not significant. The partition dataset exhibited significant variation among regions (p-value=0.011) (23.28%), however, similarly to the COI and 28S datasets, the majority of the total variation occurred within populations (79.41%).

Cophylogenetic reconstruction

The cophylogenetic reconstruction of *M. insularis* and Diaspididae yielded 14 cospeciations, 11 sortings, 4 duplications and 21 host switches (see Fig 13). The number of cospeciations was not significant regardless of the randomization strategy used, however significance varied between methods and is summarized in Table 6.

The effect of sample size was tested by reducing the number of specimens to 75%, 50% and 25% of the original dataset using a random number generator. The significance of the number of observed cospeciations differed with dataset size demonstrating the stochasticity inherent in cophylogenetic reconstruction when species are incompletely sampled. See Table 7 for a summary of coevolutionary events and p-values by dataset size. See Appendix 6 for figures of all cophylogenetic reconstructions.

Discussion

Cophylogenetic reconstruction has traditionally focused on exploring relationships between hosts and their parasites as opposed to species involved in mutualistic relationships with a few exceptions (Silvieus *et al.* 2007). The current study is the first to use the novel cophylogenetic reconstruction software, CoRe-PA 0.5.1 to explore the presumably mutualistic relationship between *Melissotarsus insularis* and Diaspididae.

A total of 130 Diaspididae specimens were sequenced for cophylogenetic reconstruction. As was expected, the Diaspididae sequences exhibited greater genetic diversity than the *M. insularis* sequences likely because several known species of Diaspididae were sampled (Ben-Dov & Fisher 2010). Diaspididae specimens were not identified to species. However, according to Ben-Dov and Fisher (2010) *M. insularis* tends to *Melanaspis madagascariensis* Mamet, *Melissoaspis fisheri* Ben-Dov, *Melissoaspis reticulata* Ben-Dov, *Morganella conspicua* (Brain) and *Morganella formicaria* Ben-Dov.

It is possible that some of the observed variation was the result of pseudogene amplification. Although especially deviant sequences were excluded (i.e. ASMEL103-11) it was impossible to conclusively differentiate between pseudogenes and functional genes. Specimens were only sequenced once using one primer set and, unlike with *M. insularis*, a comparison between sisters was not possible as Diaspididae from the same colony do not necessarily share the same genotype.

ASMEL919-05 consistently diverged from the rest of the sequences forming a long branch on all phylogenetic trees apart from maximum parsimony (see Appendix 5 for all phylogenetic trees). ASMEL919-05 is part of the original specimen collection and was initially included in the dataset as it was sampled from an *M. insularis* colony and conformed to the sequence quality guidelines outlined in the Methods. Following its inclusion in the Diaspididae phylogeny the sequence was run through the Basic Local

Alignment Search Tool (BLAST) and identified as *Pseudaonidia duplex* or the peony scale. Although *Pseudaonidia duplex* belongs to the family Diaspididae it has not, as of yet, been associated with *M. insularis* (Ben-Dov & Fisher 2010). The identification of ASMEL919-05 as *Pseudaonidia duplex* may be the result of a) contamination during either field collection or the amplification process, b) misidentification of the ASMEL919-05 sequence by BLAST or c) the association of a novel Diaspididae species with *M. insularis*.

The Mantel test was employed to test for isolation by distance. All three datasets, COI, 28S and partition, yielded a statistically significant r. As was expected, the relationship between genetic and geographic distance was stronger for Diaspididae than *M. insularis*. The comparatively higher genetic distances exhibited by the Diaspididae sequences were most likely responsible for the strong correlation between genetic and geographic distance as ants and scales were sampled from the same colonies.

AMOVA analyses testing the association between genetic distance and habitat type as defined by Wilmé *et al.* (2006) concluded that the relationship was not significant in the case of COI and 28S and significant in the case of the partition dataset. The association between genetic distance and habitat type exhibited by the partition dataset likely resulted due to either the stochasticity introduced by the reduced dataset or the correlation between the genetic distances of COI and 28S. The partition analysis used specimens for which both the COI and 28S gene fragments were sequenced successfully. The COI and 28S subsets included 41 and 62 specimens respectively whereas the partition subset was comprised of 26 specimens. It is possible that the specimens selected for partition analysis exhibited an overall higher genetic diversity than those included in the COI and 28S subsets. Alternatively, when combined for the partition subset, the COI and 28S sequences increased the genetic diversity and therefore the genetic distance between sequences. Given the lack of a significant relationship between genetic distance and habitat type for the COI and 28S datasets it is unlikely that the significance of the relationship between the genetic distance of the partition dataset and habitat type reflects diversification in response to habitat. Rather

the increase in genetic distance between sequences resulting from the combination of the COI and 28S datasets increased the strength of the relationship between genetic and geographic distance (the partition dataset Mantel test yielded the highest r value) thereby increasing the probability that the sequences would vary by habitat.

The cophylogeny of *M. insularis* and Diaspididae was tested using all four randomization strategies. The stochasticity involved in cophylogenetic reconstruction was explored by reducing the dataset to 75%, 50% and 25% of its original size. The number of observed cospeciations was not significant in all cases apart from the reconstruction of the dataset reduced to 75% of its original size. The predictions derived from cophylogenetic reconstruction rely on the accuracy of the host and parasite phylogenetic trees used. Therefore, cophylogenetic reconstruction is vulnerable to the same challenges associated with the generation of phylogenies such as long-branch attraction, incomplete lineage sorting and incomplete taxon sampling (Heath *et al.* 2008; Maddison & Knowles 2006). In this case, the effect of incomplete taxon sampling was tested by randomly reducing the number of specimens included in the reconstruction. The range of p-values determined for each dataset highlights the stochasticity inherent in reconstructing cophylogenies in cases of incomplete taxon sampling. P-values ranged from 0.001 to 0.324 suggesting that depending on the specimens included the number of cospeciations inferred may or may not have been significant. In comparison, the p-values for the four different randomization strategies ranged from 0.171 to 0.246.

In all cases the number of observed cospeciations given the associations between the host and parasite trees fell in the upper tail end of the normal distribution. Because cophylogenetic reconstruction is a relatively novel discipline (Desdevises 2007) a reference range for the expected number of observed cospeciations in mutualistic relationships has not been estimated. Cophylogenetic reconstruction of host and parasite relationships tend to exhibit signs of a significant congruence between phylogenetic trees (Bruyndonckx *et al.* 2009; Dilcher *et al.* 2012; Light & Hafner 2007; Morelli & Spicer 2007; Nemirov *et al.* 2010; Shafer *et al.* 2009) and in two of the three published studies using CoRe-PA 0.5.1 the relationship between the host and virus was

significant (Dilcher *et al.* 2012; Nemirov *et al.* 2010). The study that did not report a significant number of cospeciations between the host and virus phylogenies generated the virus phylogeny using molecular data while the host phylogeny was compiled from literature data which may have confounded the cophylogenetic reconstruction (Coulibaly-N'Golo *et al.* 2011).

The nature of symbiotic relationships is governed by extrinsic factors that vary geographically such as abiotic environmental characteristics and community composition (Thompson 1999). The geographic theory of coevolution provides a theoretical framework for studying the variation in symbiotic relationships in response to geographic heterogeneity (Thompson 1999). It is a tripartite hypothesis that establishes three major tenets: 1) different evolutionary trajectories are favoured in different populations depending on the environment and species involved, 2) there are populations wherein reciprocal selection is occurring and they are referred to as coevolutionary hotspots and 3) coevolutionary hot spots, gene flow, random genetic drift and extinction of local populations contribute to a continuous shift in the range of coevolving traits (Thompson 1999). According to the geographic theory of coevolution, mutualistic relationships, much like the one between *M. insularis* and Diaspididae may not necessarily exhibit evidence of codiversification in a cophylogenetic reconstruction. Historic variation in the intimacy of the relationship between distinct populations may not have allowed conditions favourable for codiversification to persist sufficiently long for both species to accumulate distinguishing polymorphisms. In host and virus relationships variation in environmental factors and distribution is minimized as the virus resides within the host's body (Bruyndonckx *et al.* 2009). Physical or behavioural changes experienced by the host (i.e. hypothermia, migration) affect the virus increasing the probability of congruence between the host and virus phylogenies. Furthermore, viruses are often transmitted both horizontally as well as vertically (Lipsitch *et al.* 1996). Vertical transmission of a parasite or virus maintains the association with its host increasing the probability of phylogenetic congruence (Bruyndonckx *et al.* 2009; Wilkinson & Sherratt 2001). There has been no evidence of the association between *M. insularis* and Diaspididae persisting by vertical transmission

as *Melissotarsus* queens have never been observed transporting Diaspididae when establishing novel colonies (Mony *et al.* 2002).

I hypothesize that the association between *M. insularis* and Diaspididae arose as a result of continuous contact facilitated by the scale's abundance. According to Stadler and Dixon (2005) the formation of mutualistic relationships between ants and their associates evolved due to their shared cosmopolitan nature. The incidence of mutualistic relationships between ants and aphids in particular, is positively affected by aphid abundance (Stadler & Dixon 2005). Aphids and other ant associates tend to be abundant in fragmented habitats (Stadler & Dixon 2005) much like Madagascar (Kull 2004). It is likely that the progressive fragmentation of Madagascar over the course of the last century (Kull 2004) contributed to sustaining the relationship between *M. insularis* and Diaspididae. Given that the association relies on the abundance of both species, the persistence of the association favours generalists, discouraging intimate codiversification between distinct populations of ant and scale. A cophylogenetic reconstruction with a p-value that falls in the upper tail end of the distribution but not within the range of significance is therefore expected as the association between *M. insularis* and Diaspididae does not exhibit the intimacy characteristic of host and virus relationships. Therefore, a threshold of 0.05 is not an accurate indication of the presence or absence of a symbiotic relationship but only of the intimacy between the evolutionary histories of the species involved.

Conclusion

Estimating the intimacy inherent in symbiotic relationships between ants and their associates is notoriously complex (Heckroth *et al.* 1998). Unlike species richness, symbiotic relationships lack a standardized metric complicating their inclusion in biodiversity quantification initiatives (Dyer *et al.* 2010). Cophylogenetic reconstruction has gained prominence over the course of the last decade as a tool for systematically exploring the evolutionary history of species involved in symbiotic relationships

(Desdevises 2007). However, it relies on the accuracy of its primary input, phylogenetic trees, and is consequently vulnerable to the challenges associated with phylogeny reconstruction.

The current study explored the relative intimacy of the relationship between *M. insularis* and Diaspididae using the novel cophylogenetic reconstruction program, CoRe-PA 0.5.1 (Merkle *et al.* 2010). As expected, the number of cospeciations (codiversifications) fell in the upper tail of the normal distribution but did not achieve significance. Although the phylogenetic congruence between *M. insularis* and Diaspididae is likely influenced by their association, it is possible that similar environmental constraints contributed to the number of observed cospeciations. One possible method for elucidating the influence of environmental constraints would have involved comparing a cophylogenetic reconstruction between Diaspididae and *M. insularis* and Diaspididae and multiple concurrent ant species. Unfortunately insufficient molecular data on concurrent ant species precluded a comparison as using a small dataset (i.e. nine specimens) introduced stochasticity.

Randomization strategy did not appear to have a significant effect on the outcome of the reconstruction, however the size of the dataset and the specimens included in the reconstruction determined whether the result was statistically significant. This suggests that incomplete taxon sampling may play a primary role in confounding cophylogenetic reconstruction results and should be accounted for in future research initiatives. Future development in the field should focus on establishing a metric which can be used to compare cophylogenetic reconstruction results. Comparing results between programs is problematic as unique algorithms are employed and event types are not always standardized.

The challenges inherent in cophylogenetic reconstruction are not unique and are encountered by the majority of algorithms aimed at quantifying species characteristics and interactions. Cophylogenetic reconstruction offers researchers an opportunity to systematically study the evolutionary history of symbiotic relationships and, with further

development, has the potential to significantly contribute to our understanding of species interactions.

Table 4: A summary of the r and p-values for Mantel tests of the COI, 28S and partition Diaspididae datasets.

Dataset	***r***	***p-value***
COI	0.485	0.001
28S	0.277	0.002
Partition	0.566	0.001

Table 5: A summary of the hierarchal AMOVA results for the a) COI, b) 28S and c) partition datasets with specimen collection localities tested against unique habitats identified by Wilmé et al. (2006).

a)

Source of Variation	***d.f.***	***Sum of Squares***	***Variance Components***	***Percentage of Variation***	***p-value***	***Phi***
Among regions	9	546.808	4.892	8.430	0.283+-0.005	Φ_{CT}: 0.084
Among populations within regions	4	199.667	-5.110	-8.810	0.981±0.002	Φ_{SC}: -0.096
Within populations	25	1455.500	58.220	100.380	0.478±0.005	Φ_{ST}: -0.004
Total	38	2201.974	58.002			

b)

Source of Variation	***d.f.***	***Sum of Squares***	***Variance Components***	***Percentage of Variation***	***p-value***	***Phi***
Among regions	11	374.720	1.575	5.63	0.335±0.005	Φ_{CT}: 0.056
Among populations within regions	15	399.639	0.373	1.33	0.333±0.004	Φ_{SC}: 0.014
Within populations	32	832.471	26.015	93.03	0.295±0.005	Φ_{ST}: 0.070
Total	38	2201.974	58.002			

c)

Source of Variation	***d.f.***	***Sum of Squares***	***Variance Components***	***Percentage of Variation***	***p-value***	***Phi***
Among regions	7	848.202	20.692	23.280	0.011±0.001	Φ_{CT}: 0.233
Among populations within regions	4	266.098	-2.393	-2.690	0.712±0.004	Φ_{SC}: -0.035
Within populations	13	917.500	70.577	79.410	0.042±0.002	Φ_{ST}: 0.206
Total	24	2031.800	88.875			

Table 6: A summary of the p-values for all four randomization options available in CoRe-PA 0.5.1.

Randomization Strategy	***p-value***
Randomizing associations	0.172
Randomizing host tree	0.171
Randomizing parasite tree	0.167
Randomizing both trees	0.246

Table 7: A summary of the p-values for datasets 100%, 75%, 50% and 25% the size of the original dataset.

Dataset Size	***p-value***	***Cospeciations***	***Sortings***	***Duplications***	***Host Switches***
100%	0.172	14	11	4	21
75%	0.001	13	17	6	10
50%	0.324	7	7	2	9
25%	0.129	4	3	1	4

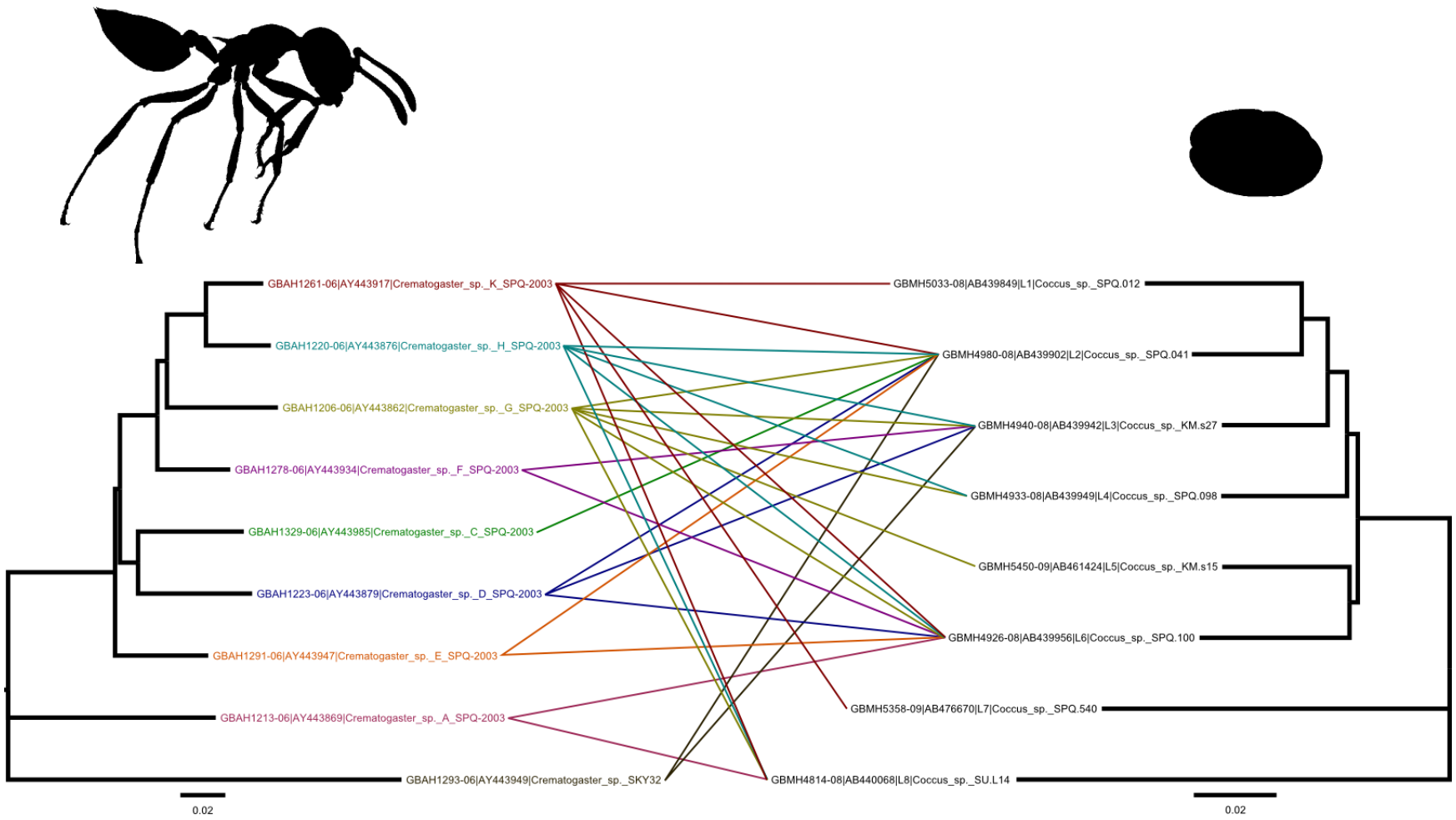


Figure 9: Cophylogenetic reconstruction of Crematogaster and Coccus sequences sourced from Quek (Quek *et al.* 2004) and Ueda (2010). Both phylogenies were generated using the neighbour-joining algorithm in MEGA 5.0 (Tamura *et al.* 2011).

Figure 10: Diaspididae COI Bayes tree generated in MrBayes version 3.1.2 (Huelsenbeck & Ronquist 2001). The parsimony informative site in the H3 dataset is indicated by the colour of the taxa names as a T (red), a C (blue) or a Y (purple). The colour coded bar on the right identifies branches corresponding to unique habitats identified by Wilmé *et al.* (2006). Black bars represent specimens collected on the island of Mayotte off the coast of Madagascar. Phylogenies were generated using the neighbour-joining algorithm in MEGA 5.0 (Tamura *et al.* 2011).

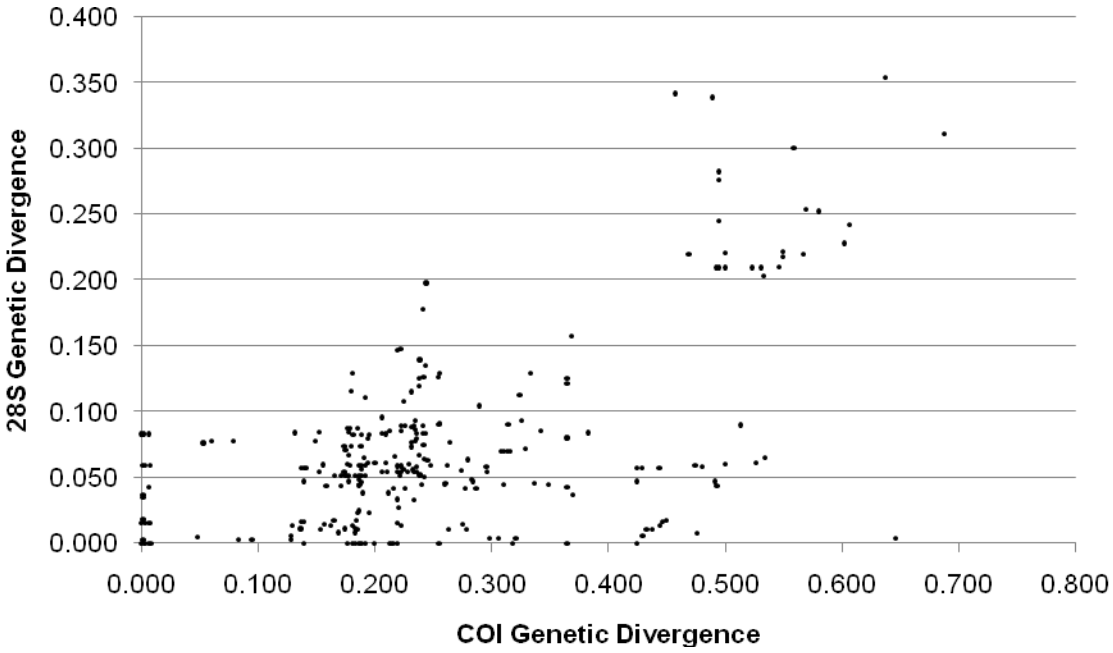


Figure 11: Pairwise sequence divergence for COI and 28S determined in MEGA 5.0 (Tamura *et al.* 2011) using the Kimura 2-Parameter model with the gamma parameter specified based on ModelTest v. 3.7 (Posada & Crandall 1998) results. A Mantel test comparing the 28S and COI matrices yielded an r of 0.57476.

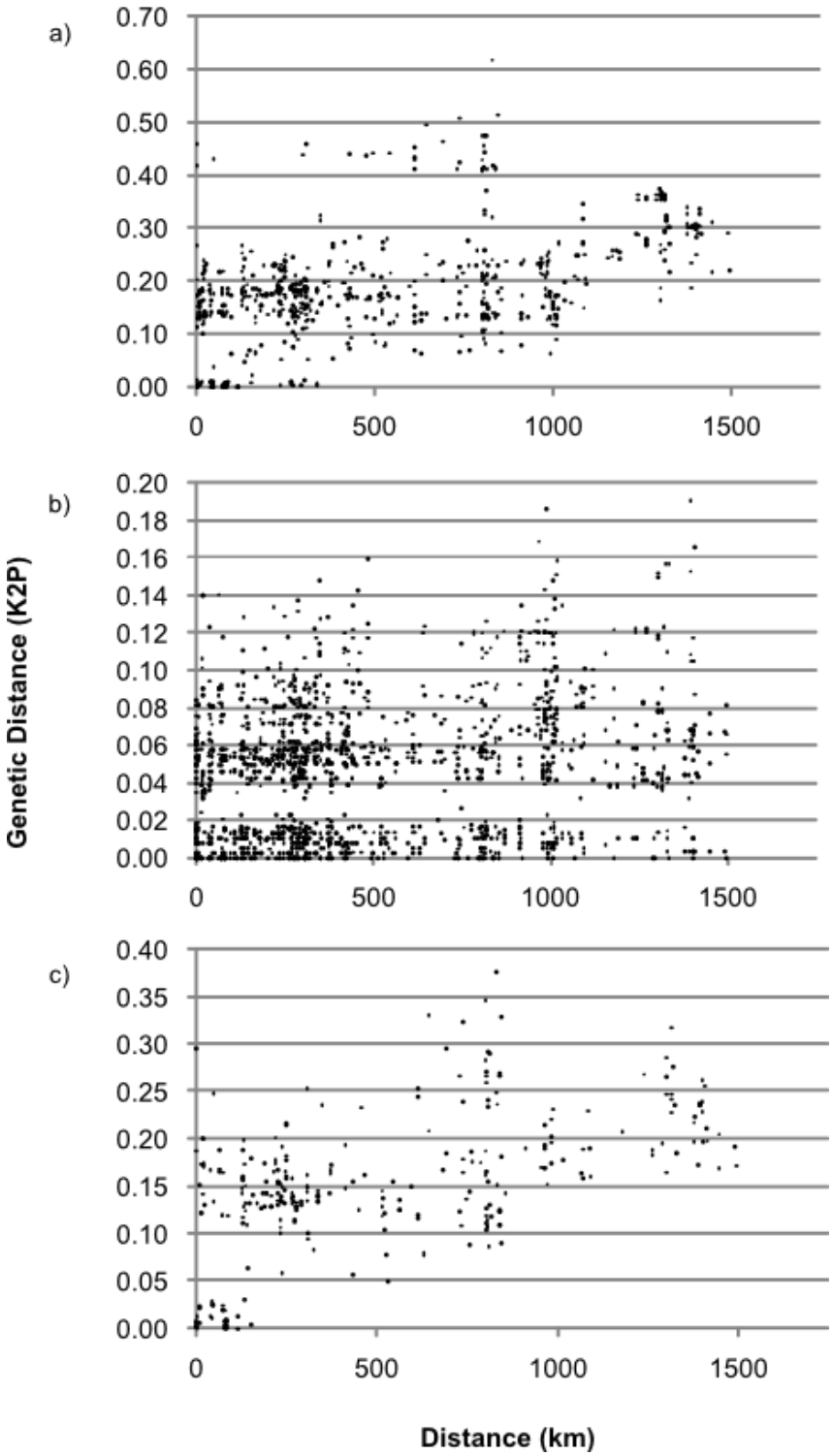


Figure 12: Genetic distance exhibited by Diaspididae in relation to geographic distance by dataset: a) COI, b) 28S and c) partition. Genetic distance was determined in MEGA 5.0 (Tamura *et al.* 2011) using the Kimura 2-Parameter model with the gamma parameter specified according to ModelTest v. 3.7 (Posada & Crandall 1998) results.

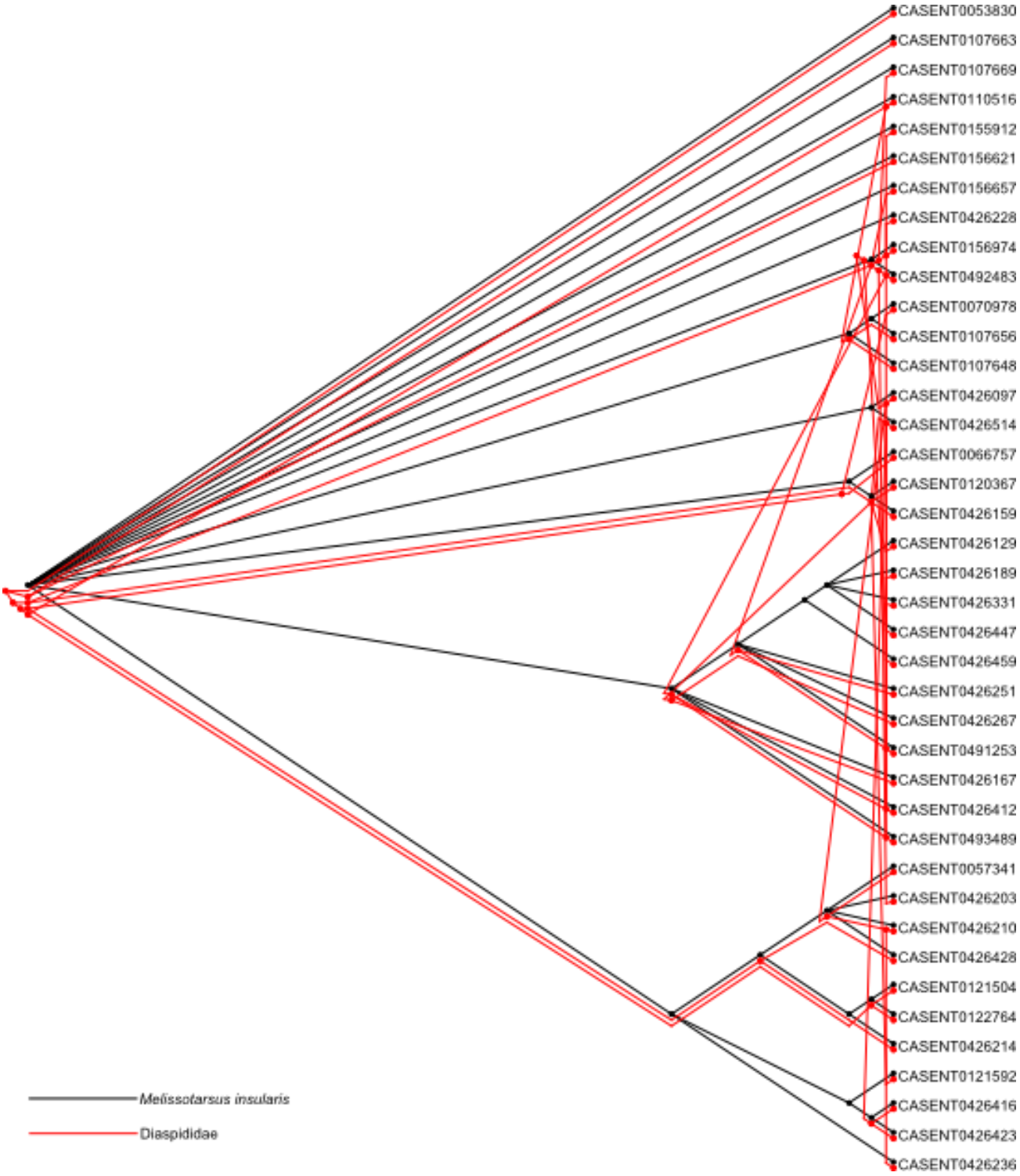


Figure 13: Cophylogenetic reconstruction of Melissotarsus insularis and Diaspididae generated in CoRe-PA 0.5.1 (Merkle *et al.* 2010). Both the ant and scale phylogenies were generated in MrBayes version 3.1.2 (Huelsenbeck & Ronquist 2001) using COI sequences.

Chapter 4: Summary and Future Directions

Biodiversity quantification is inherent in mitigating biodiversity loss and remains a primary challenge for conservation biologists (Balmford *et al.* 2005). In addition to theoretical barriers such as defining species and sustainable biodiversity, biodiversity quantification faces a scarcity of resources as it lacks funding, time and taxonomic expertise (Balmford *et al.* 2005; Balmford & Whitten 2003; Gardner *et al.* 2008). Modern advances in sequencing technology (Caterino *et al.* 2000; Hudson 2008; Zhang & Hewitt 2003) and software (Kuhner 2006; Librado & Rozas 2009; Merkle *et al.* 2010; Monaghan *et al.* 2009; Tamura *et al.* 2011) have actualized economic, fast and accessible options for addressing biodiversity quantification .

My thesis has demonstrated the potential of using museum specimens combined with novel software (two of the programs used were released within the last 4 years) to elucidate the species status of *Melissotarsus insularis* and its relationship with the scale, Diaspididae. The first chapter focussed on the apparent genetic diversity exhibited by *M. insularis* in preliminary analyses suggesting that its current designation may represent more than one species. *M. insularis* specimens were sourced from the California Academy of Sciences and 199 were sequenced at the Biodiversity Institute of Ontario and the Advanced Analysis Center Genomics Facility. Forty-five of the specimens sequenced exhibited incongruence between different amplifications of COI suggesting the presence of pseudogenes. Due to the absence of stop codons and frameshift mutations the suspected pseudogenes were classified as cryptic (Bertheau *et al.* 2011). Once the pseudogenes were accounted for the majority of the genetic diversity was limited to a divergence of 3% or less and, in most cases, could be explained by poor sequence quality. Furthermore, there appeared to be a lack of association between habitat type and population suggested by both the COI phylogenetic tree which appeared to group specimens by latitude as opposed to habitat and the AMOVA analysis. I concluded that the majority of the genetic diversity exhibited by *M. insularis* was an artefact of pseudogene amplification suggesting that *M. insularis*

likely represents a single species. Using museum specimens, molecular data and novel species delineation software, I was able to elucidate the species status of *M.insularis* and explore the role of cryptic pseudogenes in species discovery.

The second chapter explored the intimacy of the relationship between *M. insularis* and Diaspididae using the novel cophylogenetic reconstruction program CoRe-PA 0.5.1 (Merkle *et al.* 2010). CoRe-PA 0.5.1 inferred the number of codiversifications experienced by *M. insularis* and Diaspididae by comparing their phylogenies. Four different randomization strategies and four dataset sizes were tested for their effect on the significance of the number of codiversifications. It was shown that significance varied somewhat with randomization strategy and significantly with the size of the dataset. Depending on the specimens included in the reconstruction the number of codiversifications was either significant ($p=0.001$) or had a p-value of 0.34 implying that incomplete taxon sampling may play a primary role in cophylogenetic reconstruction. The significance of the number of codiversifications fell within the upper range of the normal distribution but not within the range of significance ($p<0.05$). A standard metric for quantifying the degree of intimacy inherent in symbiotic relationships is yet to be developed. However the result adequately reflected the nature of the relationship in relation to that shared by a host and its virus the latter tending to yield a statistically significant number of cospeciations (Dilcher *et al.* 2012; Nemirov *et al.* 2010).

Future Directions

The use of molecular data in biodiversity conservation has experienced considerable growth over the past decade (DeSalle & Amato 2004). In addition to its advantages such as high through-put species identification (Vogler & Monaghan 2007) and cophylogenetic reconstruction (Desdevises 2007), the use of molecular data in biodiversity analysis has several disadvantages including cryptic pseudogene amplification (Bertheau *et al.* 2011) and artefacts resulting from incomplete taxon sampling (Maddison & Knowles 2006).

Unlike apparent pseudogenes which can often be identified by stop codons or frameshift mutations, the difference between cryptic pseudogenes and their functional counterparts is minimally perceptible (Bertheau *et al.* 2011). As of yet, systematic screening of sequences for cryptic pseudogene amplification is not a requirement prior to uploading sequences to GenBank, the European Nucleotide Archive or the DNA Data Bank of Japan. Future research initiatives should focus on developing protocols for identifying cryptic pseudogenes that are targeted towards species delineation and phylogeographic studies as they are particularly vulnerable to the incorporation of unidentified pseudogenes. Economic and time efficient approaches should be prioritized as current research endeavours are trending towards high throughput sequencing (Hudson 2008).

In order to infer the intimacy of a symbiotic relationship using cophylogenetic reconstruction software it is necessary to establish a standard metric. Because cophylogenetic reconstruction programs such as CoRe-PA and TreeMap use unique algorithms and assign arbitrary costs to different evolutionary events, results are difficult to interpret in isolation. Further development in cophylogenetic reconstruction should focus on estimating flexible ranges (referring to either the number of evolutionary events or their significance) that may be used to compare and categorize symbiotic relationships.

Despite its challenges molecular data offers a unique perspective on species richness and inter species dynamics often faster and at a lower cost than alternative methods. Novel developments in bioinformatics have maximized the information derived from molecular data allowing for its use in the study of species delineation, quantification of species richness, symbiotic and trophic interactions (Desdevises 2007; Dyer *et al.* 2010; Vogler & Monaghan 2007). With sequencing technology decreasing in cost and the popularization of mass sequencing endeavours, the role of molecular data in biodiversity analysis will continue to evolve (DeSalle & Amato 2004; Hudson 2008). Although molecular data cannot replace traditional methods of biodiversity analysis the insights it provides are invaluable to mitigating biodiversity loss.

References

Aitken N, Smith S, Schwarz C, Morin PA (2004) Single nucleotide polymorphism (SNP) discovery in mammals: a targeted-gene approach. *Molecular ecology* **13**, 1423-1431.

Ali JG, Agrawal AA (2012) Specialist versus generalist insect herbivores and plant defense. *Trends in plant science* **17**, 293-302.

Andersen JC, Wu J, Gruwell ME*, et al.* (2010) A phylogenetic analysis of armored scale insects (Hemiptera: Diaspididae), based upon nuclear, mitochondrial, and endosymbiont gene sequences. *Molecular phylogenetics and evolution* **57**, 992-1003.

Asher RJ, Hofreiter M (2006) Tenrec phylogeny and the noninvasive extraction of nuclear DNA. *Systematic biology* **55**, 181-194.

Ballard JWO, Whitlock MC (2003) The incomplete natural history of mitochondria. *Molecular ecology* **13**, 729-744.

Balmford A, Crane P, Dobson A, Green RE, Mace GM (2005) The 2010 challenge: data availability, information needs and extraterrestrial insights. *Philosophical Transactions of the Royal Society B: Biological Sciences* **360**, 221-228.

Balmford A, Whitten T (2003) Who should pay for tropical conservation, and how could the costs be met? *Oryx* **37**, 238-250.

Barnosky AD, Matzke N, Tomiya S*, et al.* (2011) Has the Earth's sixth mass extinction already arrived? *Nature* **471**, 51-57.

Barr CM, Neiman M, Taylor DR (2005) Inheritance and recombination of mitochondrial genomes in plants, fungi and animals. *New Phytologist* **168**, 39-50.

Bastolla U, Fortuna MA, Pascual-GarcÃ-a A*, et al.* (2009) The architecture of mutualistic networks minimizes competition and increases biodiversity. *Nature* **458**, 1018-1020.

Ben-Dov Y (2010) On new taxa and some described armoured scale insects (Hemiptera: Diaspididae) living in the galleries of the ant *Melissotarsus insularis* Santschi (Hymenoptera: Formicidae) in Madagascar. *Zootaxa* **2368**, 49-58.

Ben-Dov Y, Fisher BL (2010) The mutualism of *Melissotarsus* ants and armoured scale insects in Africa and Magadascar: distribution, host plants and biology. *Entomologia Hellenica* **19**, 45-53.

Bennett GM, O'Grady PM (2012) Host-plants shape insect diversity: phylogeny, origin, and species diversity of native Hawaiian leafhoppers (Cicadellidae:*Nesophrosyne*). *Molecular phylogenetics and evolution* **65**, 705-717.

Bensasson D, Zhang DX, Hartl DL, Hewitt GM (2001) Mitochondrial pseudogenes: evolution's misplaced witnesses. *Trends in Ecology & Evolution* **16**, 314-321.

Bertheau C, Schuler H, Krumböck S, Arthofer W, Stauffer C (2011) Hit or miss in phylogeographic analyses: the case of the cryptic NUMTs. *Molecular Ecology Resources* **11**, 1056-1059.

Bisevac L, Majer JD (2002) Cost effectiveness and data-yield of biodiversity surveys. *Journal of the Royal Society of Western Australia* **85**, 129-132.

Brower AVZ, DeSalle R, Vogler A (1996) Gene trees, species trees, and systematics: a cladistic perspective. *Annual Review of Ecology and Systematics* **27**, 423-450.
Bruyndonckx N, Dubey S, Ruedi M, Christe P (2009) Molecular cophylogenetic relationships between European bats and their ectoparasitic mites (Acari, Spinturnicidae). *Molecular phylogenetics and evolution* **51**, 227-237.
Buhay JE (2009) "COI-Like" sequences are becoming problematic in molecular systematic and DNA barcoding studies. *Journal of Crustacean Biology* **29**, 96-110.
Calvignac S, Konecny L, Malard F, Douady CJ (2011) Preventing the pollution of mitochondrial datasets with nuclear mitochondrial paralogs (numts). *Mitochondrion* **11**, 246-254.
Caterino MS, Cho S, Sperling FAH (2000) The current state of insect molecular systematics: a thriving Tower of Babel. *Annual Review of Entomology* **45**, 1-54.
Clark MA, Moran NA, Baumann P, Wernegreen JJ (2000) Cospeciation between bacterial endosymbionts (Buchnera) and a recent radiation of aphids (Uroleucon) and pitfalls of testing for phylogenetic congruence. *Evolution* **54**, 517-525.
Cognato AI (2006) Standard percent DNA sequence difference for insects does not predict species boundaries. *Journal of Economic Entomology* **99**, 1037-1045.
Cornet A (1974) Essai de cartographie bioclimatique à Madagascar. ORSTOM, Paris, France.
Coulibaly-N'Golo D, Allali B, Kouassi SK*, et al.* (2011) Novel arenavirus sequences in *Hylomyscus* sp. and *Mus* (*Nannomys*) setulosus from Côte d'Ivoire: implications for evolution of arenaviruses in Africa. *PLoS One* **6**, e20893.
Crandall KA, Robison HW, Buhay JE (2009) Avoidance of extinction through nonexistence: the use of museum specimens and molecular genetics to determine the taxonomic status of an endangered freshwater crayfish. *Conservation genetics* **10**, 177-189.
D'Errico I, Gadaleta G, Saccone C (2004) Pseudogenes in metazoa: origin and features. *Briefings in functional genomics & proteomics* **3**, 157-167.
Davies TJ, Cadotte MW (2011) Quantifying Biodiversity: Does It Matter What We Measure? In: *Biodiversity Hotspots* eds. Zachos FE, Habel JC), pp. 43-60. Springer, Berlin.
Delage-Darchen B (1972) Une fourmi de cote-divoire: *Melissotarsus titubans* Del., n. sp. *Insectes sociaux* **19**, 213-226.
DeSalle R, Amato G (2004) The expansion of conservation genetics. *Nature Reviews Genetics* **5**, 702-712.
DeSalle R, Egan MG, Siddall M (2005) The unholy trinity: taxonomy, species delimitation and DNA barcoding. *Philosophical Transactions of the Royal Society B: Biological Sciences* **360**, 1905-1916.
Desdevises Y (2007) Cophylogeny: insights from fish-parasite systems. *Parassitologia* **49**, 125.
Diaz S, Fargione J, Chapin FS, Tilman D (2006) Biodiversity loss threatens human well-being. *PLoS biology* **4**, e277.
Dilcher M, Hasib L, Lechner M*, et al.* (2012) Genetic characterization of Tribec virus and Kemerovo virus, two tick-transmitted human-pathogenic Orbiviruses. *Virology* **423**, 68-76.

Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC evolutionary biology* **7**, 214.

Dukes JS, Mooney HA (1999) Does global change increase the success of biological invaders? *Trends in Ecology & Evolution* **14**, 135-139.

Dyer LA, Walla TR, Greeney HF, Stireman Iii JO, Hazen RF (2010) Diversity of interactions: A metric for studies of biodiversity. *Biotropica* **42**, 281-289.

Ersts PJ (Internet) Geographic Distance Matrix Generator (version 1.2.3) American Museum of Natural History, Center for Biodiversity and Conservation. Available from http://biodiversityinformatics.amnh.org/open_source/gdmg. Accessed on 2012-11-6.

Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* **10**, 564-567.

Fahrenholz H (1913) Ectoparasitien und abstammungslehre. *Zoologische Anzeiger* **132**, 254–262.

Fisher BL (1997) Biogeography and ecology of the ant fauna of Madagascar (Hymenoptera: Formicidae). *Journal of Natural History* **31**, 269-302.

Fisher BL, Robertson HG (1999) Silk production by adult workers of the ant *Melissotarsus emeryi* (Hymenoptera, Formicidae) in South African fynbos. *Insectes sociaux* **46**, 78-83.

Fisher BL, Robertson HG (2002) Comparison and origin of forest and grassland ant assemblages in the high plateau of Madagascar (Hymenoptera: Formicidae). *Biotropica* **34**, 155-167.

Fisher BL, Smith MA (2008) A revision of Malagasy species of *Anochetus* mayr and *Odontomachus* latreille (Hymenoptera: Formicidae). *PLoS One* **3**, e1787.

Galtier N, Nabholz B, Glémin S, Hurst GDD (2009) Mitochondrial DNA as a marker of molecular diversity: a reappraisal. *Molecular ecology* **18**, 4541-4550.

Gardner TA, Barlow J, Araujo IS*, et al.* (2008) The cost-effectiveness of biodiversity surveys in tropical forests. *Ecology Letters* **11**, 139-150.

GeneCodesCorporation (2011) Quality Scores. In: *Sequencher Tutorial for Windows and Macintosh*, Ann Arbor, MI.

Gilbert MTP, Moore W, Melchior L, Worobey M (2007) DNA extraction from dry museum beetles without conferring external morphological damage. *PLoS One* **2**, e272.

Graham CH, Ferrier S, Huettman F, Moritz C, Peterson AT (2004) New developments in museum-based informatics and applications in biodiversity analysis. *Trends in Ecology & Evolution* **19**, 497-503.

Hajibabaei M, DeWaard JR, Ivanova NV*, et al.* (2005) Critical factors for assembling a high volume of DNA barcodes. *Philosophical Transactions of the Royal Society B: Biological Sciences* **360**, 1959-1967.

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* **41**, 95-98.

Harper GL, Maclean N, Goulson D (2006) Analysis of museum specimens suggests extreme genetic drift in the adonis blue butterfly (*Polyommatus bellargus*). *Biological Journal of the Linnean Society* **88**, 447-452.

Heath TA, Zwickl DJ, Kim J, Hillis DM (2008) Taxon sampling affects inferences of macroevolutionary processes from phylogenetic trees. *Systematic biology* **57**, 160-166.

Hebert P, Ratnasingham S, deWaard J (2003) Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proc. R. Soc. Lond. B (Suppl.)* **270**, S96–S99.

Heckroth HP, Fiala B, Gullan PJ, Idris AHJ, Maschwitz U (1998) The soft scale (Coccidae) associates of Malaysian ant-plants. *Journal of Tropical Ecology* **14**, 427-443.

Heller NE, Zavaleta ES (2009) Biodiversity management in the face of climate change: a review of 22 years of recommendations. *Biological conservation* **142**, 14-32.

Hijmans RJ, Guarino L, Cruz M, Rojas E (2001) Computer tools for spatial analysis of plant genetic resources data: 1. DIVA-GIS. *Plant Genetic Resources Newsletter*, 15-19.

Hoeksema JD, Bruna EM (2000) Pursuing the big questions about interspecific mutualism: a review of theoretical approaches. *Oecologia* **125**, 321-330.

Horvath MB, Martínez-Cruz B, Negro JJ, Kalmar L, Godoy JA (2004) An overlooked DNA source for non-invasive genetic analysis in birds. *Journal of Avian Biology* **36**, 84-88.

Hudson ME (2008) Sequencing breakthroughs for genomic ecology and evolutionary biology. *Molecular Ecology Resources* **8**, 3-17.

Huelsenbeck JP, Ronquist F (2001) MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**, 754-755.

Husseneder C, Grace JK (2001) Evaluation of DNA fingerprinting, aggression tests, and morphometry as tools for colony delineation of the Formosan subterranean termite. *Journal of insect behavior* **14**, 173-186.

Ivanova NV, Dewaard JR, Hebert PDN (2006) An inexpensive, automation-friendly protocol for recovering high-quality DNA. *Molecular ecology notes* **6**, 998-1002.

Jackson AP (2004) Cophylogeny of the Ficus microcosm. *Biological Reviews* **79**, 751-768.

Jóhannsdóttir A, Cresswell I, Bridgewater P (2010) The Current Framework for International Governance of Biodiversity: Is It Doing More Harm Than Good? *Review of European Community & International Environmental Law* **19**, 139-149.

Jones M, Ghoorah A, Blaxter M (2011) jMOTU and taxonerator: turning DNA barcode sequences into annotated operational taxonomic units. *PLoS One* **6**, e19259.

Kerr JT, Sugar A, Packer L (2000) Indicator taxa, rapid biodiversity assessment, and nestedness in an endangered ecosystem. *Conservation Biology* **14**, 1726-1734.

Kiers ET, Palmer TM, Ives AR, Bruno JF, Bronstein JL (2010) Mutualisms in a changing world: an evolutionary perspective. *Ecology Letters* **13**, 1459-1474.

Kim KC, Byrne LB (2006) Biodiversity loss and the taxonomic bottleneck: emerging biodiversity science. *Ecological Research* **21**, 794-810.

Kuhner MK (2006) LAMARC 2.0: maximum likelihood and Bayesian estimation of population parameters. *Bioinformatics* **22**, 768-770.

Kull CA (2004) Isle of Fire: The Political Ecology of Landscape Burning in Madagascar. University of Chicago Press, London.

Lawton JH, Bignell DE, Bolton B*, et al.* (1998) Biodiversity inventories, indicator taxa and effects of habitat modification in tropical forest. *Nature* **391**, 72-75.
Legendre P, Desdevises Y, Bazin E (2002) A statistical test for host-parasite coevolution. *Systematic biology* **51**, 217-234.
Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**, 1451-1452.
Light JE, Hafner MS (2007) Cophylogeny and disparate rates of evolution in sympatric lineages of chewing lice on pocket gophers. *Molecular phylogenetics and evolution* **45**, 997-1013.
Lindahl T (1993) Instability and decay of the primary structure of DNA. *Nature* **362**, 709-715.
Lipsitch M, Siller S, Nowak MA (1996) The evolution of virulence in pathogens with vertical and horizontal transmission. *Evolution*, 1729-1741.
Lovejoy T, Brouillet L, Doolittle WF*, et al.* (2010) Canadian Taxonomy: Exploring Biodiversity, Creating Opportunity. Council of Canadian Academies, Ottawa, Canada.
Lü XM, Fu YX, Zhang YP (2002) Evolution of mitochondrial cytochrome *b* pseudogene in genus *Nycticebus*. *Molecular biology and evolution* **19**, 2337-2341.
Lunt DH, Zhang DX, Szymura JM, Hewltt OM (1996) The insect cytochrome oxidase I gene: evolutionary patterns and conserved primers for phylogenetic studies. *Insect molecular biology* **5**, 153-165.
Maddison WP, Knowles LL (2006) Inferring phylogeny despite incomplete lineage sorting. *Systematic biology* **55**, 21-30.
Maddock A, Plessis MA (1999) Can species data only be appropriately used to conserve biodiversity? *Biodiversity and Conservation* **8**, 603-615.
Magnacca KN, Brown MJF (2010) Mitochondrial heteroplasmy and DNA barcoding in Hawaiian *Hylaeus* (*Nesoprosopis*) bees (Hymenoptera: Colletidae). *BMC evolutionary biology* **10**, 174.
Mandelik Y, Roll U, Fleischer A (2010) Cost-efficiency of biodiversity indicators for Mediterranean ecosystems and the effects of socio-economic factors. *Journal of applied ecology* **47**, 1179-1188.
Martins Jr J, Solomon SE, Mikheyev AS*, et al.* (2007) Nuclear mitochondrial-like sequences in ants: evidence from *Atta cephalotes* (Formicidae: Attini). *Insect molecular biology* **16**, 777-784.
McCann K (2007) Protecting biostructure. *Nature* **446**, 29-29.
Merkle D, Middendorf M, Wieseke N (2010) A parameter-adaptive dynamic programming approach for inferring cophylogenies. *BMC bioinformatics* **11**, S60.
Monaghan MT, Wild R, Elliot M*, et al.* (2009) Accelerated species inventory on Madagascar using coalescent-based models of species delineation. *Systematic biology* **58**, 298-311.
Mony R, Kenne M, Orivel J, Dejean A (2002) Biology and ecology of pest ants of the genus *Melissotarsus* (Formicidae: Myrmicinae), with special reference to tropical fruit tree attacks. *Sociobiology* **40**, 645-654.
Morelli M, Spicer GS (2007) Cospeciation between the nasal mite *Ptilonyssus sairae* (Acari: Rhinonyssidae) and its bird hosts. *Systematic and Applied Acarology* **12**, 179-188.

Moricca S, Ragazzi A (2008) Fungal endophytes in Mediterranean oak forests: a lesson from Discula quercina. *Phytopathology* **98**, 380-386.
Myers N, Mittermeier RA, Mittermeier CG, Da Fonseca GAB, Kent J (2000) Biodiversity hotspots for conservation priorities. *Nature* **403**, 853-858.
Nemirov K, Leirs H, Lundkvist A, Olsson GE (2010) Puumala hantavirus and *Myodes glareolus* in northern Europe: no evidence of co-divergence between genetic lineages of virus and host. *Journal of General Virology* **91**, 1262.
Newbold T (2010) Applications and limitations of museum data for conservation and ecology, with particular attention to species distribution models. *Progress in Physical Geography* **34**, 3-22.
Pacheco R, Vasconcelos HL (2012) Habitat diversity enhances ant diversity in a naturally heterogeneous Brazilian landscape. *Biodiversity and Conservation* **21**, 1-13.
Padial J, Miralles A, De la Riva I, Vences M (2010) The integrative future of taxonomy. *Frontiers in Zoology* **7**.
Palmer TM, Stanton ML, Young TP*, et al.* (2008) Breakdown of an ant-plant mutualism follows the loss of large herbivores from an African savanna. *Science* **319**, 192-195.
Parmesan C, Yohe G (2003) A globally coherent fingerprint of climate change impacts across natural systems. *Nature* **421**, 37-42.
Peeters C, Keller RA, Johnson RA (2012) Selection against aerial dispersal in ants: two non-flying queen phenotypes in *Pogonomyrmex laticeps*. *PLoS One* **7**, e47727.
Perring TM, Cooper AD, Rodriguez RJ, Farrar CA, Bellows Jr TS (1993) Identification of a whitefly species by genomic and behavioral studies. *Science* **259**, 74-77.
Pfeiffer M, Mezger D (2012) Biodiversity assessment in incomplete inventories: leaf litter ant communities in several types of Bornean rain forest. *PLoS One* **7**, e40729.
Ponder WF, Carter GA, Flemons P, Chapman RR (2001) Evaluation of museum collection data for use in biodiversity assessment. *Conservation Biology* **15**, 648-657.
Pons J, Barraclough TG, Gomez-Zurita J*, et al.* (2006) Sequence-based species delimitation for the DNA taxonomy of undescribed insects. *Systematic biology* **55**, 595-609.
Posada D, Crandall KA (1998) Modeltest: testing the model of DNA substitution. *Bioinformatics* **14**, 817-818.
Qi A, Perry JN, Pidgeon JD, Haylock LA, Brooks DR (2008) Cost-efficacy in measuring farmland biodiversity-lessons from the Farm Scale Evaluations of genetically modified herbicide-tolerant crops. *Annals of Applied Biology* **152**, 93-101.
Quek SP, Davies SJ, Ashton PS, Itino T, Pierce NE (2007) The geography of diversification in mutualistic ants: a gene's eye view into the Neogene history of Sundaland rain forests. *Molecular ecology* **16**, 2045-2062.
Quek SP, Davies SJ, Itino T, Pierce NE (2004) Codiversification in an ant-plant mutualism: stem texture and the evolution of host use in *Crematogaster* (Formicidae:Myrmicinae) inhabitants of *Macaranga* (Euphorbiaceae). *Evolution* **58**, 554-570.
Rambaut A, Drummond AJ (2007) Tracer v1.4. Available from http://beast.bio.ed.ac.uk/Tracer

Rands MRW, Adams WM, Bennun L*, et al.* (2010) Biodiversity conservation: challenges beyond 2010. *Science* **329**, 1298-1303.
Ratnasingham S, Hebert PDN (2007) BOLD: The Barcode of Life Data System (http://www. barcodinglife. org). *Molecular ecology notes* **7**, 355-364.
Reyers B, Fairbanks DHK, Wessels KJ, Van Jaarsveld AS (2002) A multicriteria approach to reserve selection: addressing long-term biodiversity maintenance. *Biodiversity and Conservation* **11**, 769-793.
Rogers HM, Glew L, HonzÃ¡k M, Hudson MD (2010) Prioritizing key biodiversity areas in Madagascar by including data on human pressure and ecosystem services. *Landscape and Urban Planning* **96**, 48-56.
Rosenberg MS, Anderson CD (2011) PASSaGE: pattern analysis, spatial statistics and geographic exegesis. Version 2. *Methods in Ecology and Evolution* **2**, 229-232.
Rubinoff D, Holland BS (2005) Between two extremes: mitochondrial DNA is neither the panacea nor the nemesis of phylogenetic and taxonomic inference. *Systematic biology* **54**, 952-961.
Saccone C, Gissi C*, et al.* (2000) Evolution of the mitochondrial genetic system: an overview. *Gene* **261**, 153-159.
Santschi F (1911) Nouvelles fourmis de Madagascar. *Revue Suisse de Zoologie* **19**, 117-134.
Sarnat ELIM, Moreau CS (2011) Biogeography and morphological evolution in a Pacific island ant radiation. *Molecular ecology* **20**, 114-130.
Schwarz MP, Tierney SM, Cooper SJB, Bull NJ (2004) Molecular phylogenetics of the allodapine bee genus *Braunsapis*: A-T bias and heterogeneous substitution parameters. *Molecular phylogenetics and evolution* **32**, 110-122.
Shafer ABA, Williams GR, Shutler D, Rogers REL, Stewart DT (2009) Cophylogeny of *Nosema* (Microsporidia: Nosematidae) and bees (Hymenoptera: Apidae) suggests both cospeciation and a host-switch. *Journal of Parasitology* **95**, 198-203.
Silvieus SI, Clement WL, Weiblen GD (2007) Cophylogeny of figs, pollinators, gallers and parasitoids. In: *Specialization, speciation and radiation: The evolutionary biology of herbivorous insects* (ed. K. T), pp. 225-239. University of California Press, Berkeley, CA.
Skarpaas O, Økland B (2009) Timber import and the risk of forest pest introductions. *Journal of applied ecology* **46**, 55-63.
Smith MA, Fisher BL (2009) Invasions, DNA barcodes, and rapid biodiversity assessment using ants of Mauritius. *Frontiers in Zoology* **6**, 31.
Smith MA, Fisher BL, Hebert PDN (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 MA, Rodriguez JJ, Whitfield JB*, et al.* (2008) Extreme diversity of tropical parasitoid wasps exposed by iterative integration of natural history, DNA barcoding, morphology, and collections. *Proceedings of the National Academy of Sciences* **105**, 12359-12364.
Smith MA, Woodley NE, Janzen DH, Hallwachs W, Hebert PDN (2006) DNA barcodes reveal cryptic host-specificity within the presumed polyphagous members of a

genus of parasitoid flies (Diptera: Tachinidae). *Proceedings of the National Academy of Sciences of the United States of America* **103**, 3657-3662.
Söller R, Warnke K, Saint-Paul U, Blohm D (2000) Sequence divergence of mitochondrial DNA indicates cryptic biodiversity in *Octopus vulgaris* and supports the taxonomic distinctiveness of *Octopus mimus* (Cephalopoda: Octopodidae). *Marine Biology* **136**, 29-35.
Song H, Buhay JE, Whiting MF, Crandall KA (2008) Many species in one: DNA barcoding overestimates the number of species when nuclear mitochondrial pseudogenes are coamplified. *Proceedings of the National Academy of Sciences* **105**, 13486-13491.
Song J, Yao H, Li Y*, et al.* (2009) Authentication of the family Polygonaceae in Chinese pharmacopoeia by DNA barcoding technique. *Journal of Ethnopharmacology* **124**, 434-439.
Spector S (2006) Scarabaeine dung beetles (Coleoptera: Scarabaeidae: Scarabaeinae): An invertebrate focal taxon for biodiversity research and conservation. *The Coleopterists Bulletin* **60**, 71-83.
Stadler B, Dixon AFG (2005) Ecology and evolution of aphid-ant interactions. *Annual review of ecology, evolution, and systematics*, 345-372.
Steenkamp ET, Wingfield BD, Desjardins AE, Marasas WFO, Wingfield MJ (2002) Cryptic speciation in *Fusarium subglutinans*. *Mycologia* **94**, 1032-1043.
Suarez AV, Tsutsui ND (2004) The value of museum collections for research and society. *BioScience* **54**, 66-74.
Swofford DL (2003) PAUP*: Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. .
Tamura K, Peterson D, Peterson N*, et al.* (2011) MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular biology and evolution* **28**, 2731-2739.
Tan DSH, Ang Y, Lim GS, Ismail MRB, Meier R (2009) From 'cryptic species' 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.
Targetti S, Viaggi D, Cuming D, Sarthou JP, Choisis JP (2011) Towards a cost-effectiveness analysis of the measurement of biodiversity indicators.
Thomas WK, Pääbo S, Villablanca FX, Wilson AC (1990) Spatial and temporal continuity of kangaroo rat populations shown by sequencing mitochondrial DNA from museum specimens. *Journal of Molecular Evolution* **31**, 101-112.
Thompson JN (1994) *The coevolutionary process* University of Chicago Press, Chicago, IL.
Thompson JN (1999) Specific hypotheses on the geographic mosaic of coevolution. *the american naturalist* **153**, S1-S14.
Tingley MW, Beissinger SR (2009) Detecting range shifts from historical species occurrences: new perspectives on old data. *Trends in Ecology & Evolution* **24**, 625-633.
Tylianakis JM, Didham RK, Bascompte J, Wardle DA (2008) Global change and species interactions in terrestrial ecosystems. *Ecology Letters* **11**, 1351-1363.

Ueda S, Quek SP, Itioka T*, et al.* (2008) An ancient tripartite symbiosis of plants, ants and scale insects. *Proceedings of the Royal Society B: Biological Sciences* **275**, 2319-2326.

Ueda S, Quek SP, Itioka T, Murase K, Itino T (2010) Phylogeography of the *Coccus* scale insects inhabiting myrmecophytic *Macaranga* plants in Southeast Asia. *Population ecology* **52**, 137-146.

Underwood EC, Fisher BL (2006) The role of ants in conservation monitoring: if, when, and how. *Biological conservation* **132**, 166-182.

Van Houdt JKJ, Breman FC, Virgilio M, De Meyer M (2009) Recovering full DNA barcodes from natural history collections of Tephritid fruitflies (Tephritidae, Diptera) using mini barcodes. *Molecular Ecology Resources* **10**, 459-465.

Vences M, Wollenberg KC, Vieites DR, Lees DC (2009) Madagascar as a model region of species diversification. *Trends in Ecology & Evolution* **24**, 456-465.

Vernooy R, Haribabu E, Muller MR*, et al.* (2010) Barcoding life to conserve biological diversity: beyond the taxonomic imperative. *PLoS biology* **8**, e1000417.

Visser ME, Both C (2005) Shifts in phenology due to global climate change: the need for a yardstick. *Proceedings of the Royal Society B: Biological Sciences* **272**, 2561-2569.

Vogler AP, Monaghan MT (2007) Recent advances in DNA taxonomy. *Journal of Zoological Systematics and Evolutionary Research* **45**, 1-10.

Wandeler P, Hoeck PEA, Keller LF (2007) Back to the future: museum specimens in population genetics. *Trends in Ecology and Evolution* **22**, 634-642.

Wilcove DS, Rothstein D, Dubow J, Phillips A, Losos E (1998) Quantifying threats to imperiled species in the United States. *BioScience* **48**, 607-615.

Wilkinson DM, Sherratt TN (2001) Horizontally acquired mutualisms, an unsolved problem in ecology? *Oikos* **92**, 377-384.

Wilme L, Goodman SM, Ganzhorn JU (2006) Biogeographic evolution of Madagascar's microendemic biota. *Science* **312**, 1063-1065.

Wilson JJ (2010) Assessing the value of DNA barcodes and other priority gene regions for molecular phylogenetics of Lepidoptera. *PLoS One* **5**, e10525.

Worm B, Barbier EB, Beaumont N*, et al.* (2006) Impacts of biodiversity loss on ocean ecosystem services. *Science* **314**, 787-790.

Wright S (1943) Isolation by distance. *Genetics* **28**, 114.

Zhang DX, Hewitt GM (2003) Nuclear DNA analyses in genetic studies of populations: practice, problems and prospects. *Molecular ecology* **12**, 563-584.

Appendix 1: *Melissotarsus insularis* specimen records

Process ID	***Sample ID***	***Collection Date***	***Locality***
ASAM104-05	CASENT0422964-D01	25-Feb-2001	Anosiravo 345
ASAM105-05	CASENT0422964-D02	25-Feb-2001	Anosiravo 345
ASAM106-05	CASENT0422962-D01	22-Feb-2001	Orangea 90
ASAM107-05	CASENT0422962-D02	22-Feb-2001	Orangea 90
ASAM108-05	CASENT0426121-D01	11-Nov-2001	Ankidrodroa 100
ASAM109-05	CASENT0426121-D02	11-Nov-2001	Ankidrodroa 100
ASAM110-05	CASENT0426110-D01	09-Dec-2001	Belo sur Mer BSI21c
ASAM111-05	CASENT0426110-D02	09-Dec-2001	Belo sur Mer BSI21c
ASAM112-05	CASENT0426097-D01	10-Dec-2001	Belo sur Mer BSI21d
ASAM113-05	CASENT0426097-D02	10-Dec-2001	Belo sur Mer BSI21d
ASAM114-05	CASENT0426129-D01	12-Jan-2002	Manantalinjo 150
ASAM115-05	CASENT0426129-D02	12-Jan-2002	Manantalinjo 150
ASAM116-05	CASENT0426506-D01	12-Jan-2002	Manantalinjo 150
ASAM117-05	CASENT0426506-D02	12-Jan-2002	Manantalinjo 150
ASAM118-05	CASENT0426515-D01	16-Jan-2002	Tsimelahy 300
ASAM119-05	CASENT0426515-D02	16-Jan-2002	Tsimelahy 300
ASAM120-05	CASENT0426163-D01	03-Feb-2002	Bealoka 35
ASAM122-05	CASENT0426180-D01	03-Feb-2002	Bealoka 35
ASAM123-05	CASENT0426180-D02	03-Feb-2002	Bealoka 35
ASAM124-05	CASENT0426192-D01	03-Feb-2002	Bealoka 35
ASAM125-05	CASENT0426192-D02	03-Feb-2002	Bealoka 35
ASAM126-05	CASENT0426203-D01	03-Feb-2002	Bealoka 35
ASAM127-05	CASENT0426203-D02	03-Feb-2002	Bealoka 35
ASAM128-05	CASENT0426441-D01	03-Feb-2002	Bealoka 35
ASAM129-05	CASENT0426441-D02	03-Feb-2002	Bealoka 35
ASAM130-05	CASENT0426452-D01	06-Feb-2002	Malaza 40
ASAM131-05	CASENT0426452-D02	06-Feb-2002	Malaza 40
ASAM132-05	CASENT0426459-D01	07-Feb-2002	Anjapolo 65

ASAM133-05	CASENT0426459-D02	07-Feb-2002	Anjapolo 65
ASAM134-05	CASENT0426423-D01	07-Feb-2002	Anjapolo 65
ASAM135-05	CASENT0426423-D02	07-Feb-2002	Anjapolo 65
ASAM136-05	CASENT0426428-D01	17-Feb-2002	Bevazoa 130
ASAM137-05	CASENT0426428-D02	17-Feb-2002	Bevazoa 130
ASAM138-05	CASENT0426210-D01	17-Feb-2002	Soamanitra 150
ASAM139-05	CASENT0426210-D02	17-Feb-2002	Soamanitra 150
ASAM140-05	CASENT0426214-D01	21-Feb-2002	Mahafaly 80
ASAM141-05	CASENT0426214-D02	21-Feb-2002	Mahafaly 80
ASAM142-05	CASENT0426236-D01	21-Feb-2002	Mahafaly 80
ASAM143-05	CASENT0426236-D02	21-Feb-2002	Mahafaly 80
ASAM144-05	CASENT0426228-D01	27-Feb-2002	Andranomite 75
ASAM145-05	CASENT0426228-D02	27-Feb-2002	Andranomite 75
ASAM146-05	CASENT0426248-D01	27-Feb-2002	Andranomite 75
ASAM147-05	CASENT0426248-D02	27-Feb-2002	Andranomite 75
ASAM148-05	CASENT0426412-D01	27-Feb-2002	Andranomite 75
ASAM149-05	CASENT0426412-D02	27-Feb-2002	Andranomite 75
ASAM150-05	CASENT0426261-D01	27-Feb-2002	Andranomite 75
ASAM151-05	CASENT0426261-D02	27-Feb-2002	Andranomite 75
ASAM152-05	CASENT0426271-D01	27-Feb-2002	Andranomite 75
ASAM153-05	CASENT0426359-D01	06-Mar-2002	Tsifota 70
ASAM154-05	CASENT0426359-D02	06-Mar-2002	Tsifota 70
ASAM155-05	CASENT0426351-D01	06-Mar-2002	Tsifota 70
ASAM156-05	CASENT0426351-D02	06-Mar-2002	Tsifota 70
ASAM157-05	CASENT0426330-D01	12-Mar-2002	Beroboka 80
ASAM158-05	CASENT0426330-D02	12-Mar-2002	Beroboka 80
ASAM159-05	CASENT0426342-D01	12-Mar-2002	Beroboka 80
ASAM160-05	CASENT0426342-D02	12-Mar-2002	Beroboka 80
ASAM161-05	CASENT0426466-D01	12-Mar-2002	Beroboka 80
ASAM162-05	CASENT0426466-D02	12-Mar-2002	Beroboka 80
ASAM163-05	CASENT0426479-D01	12-Mar-2002	Beroboka 80
ASAM164-05	CASENT0426479-D02	12-Mar-2002	Beroboka 80
ASAM165-05	CASENT0426308-D01	18-Mar-2002	Mitoho 40
ASAM166-05	CASENT0426308-D02	18-Mar-2002	Mitoho 40

ASAM167-05	CASENT0035889-D01	12-Nov-2002	Andriabe 100
ASAM168-05	CASENT0035889-D02	12-Nov-2002	Andriabe 100
ASAM170-05	CASENT0489407-D02	19-Nov-2002	Androngonibe 30
ASAM171-05	CASENT0489405-D01	19-Nov-2002	Androngonibe 30
ASAM172-05	CASENT0489405-D02	19-Nov-2002	Androngonibe 30
ASAM173-05	CASENT0489415-D01	19-Nov-2002	Androngonibe 30
ASAM174-05	CASENT0489415-D02	19-Nov-2002	Androngonibe 30
ASAM175-05	CASENT0489346-D01	19-Nov-2002	Androngonibe 30
ASAM176-05	CASENT0489346-D02	19-Nov-2002	Androngonibe 30
ASAM177-05	CASENT0492138-D01	26-Nov-2002	Ambinda 10
ASAM178-05	CASENT0492138-D02	26-Nov-2002	Ambinda 10
ASAM179-05	CASENT0492334-D01	26-Nov-2002	Ambinda 10
ASAM180-05	CASENT0492334-D02	26-Nov-2002	Ambinda 10
ASAM181-05	CASENT0485636-D01	27-Jan-2003	Ranohira 725
ASAM182-05	CASENT0485636-D02	27-Jan-2003	Ranohira 725
ASAM183-05	CASENT0489664-D01	01-Feb-2003	Analalava 700
ASAM184-05	CASENT0489664-D02	01-Feb-2003	Analalava 700
ASAM185-05	CASENT0489469-D01	01-Feb-2003	Analalava 700
ASAM186-05	CASENT0489469-D02	01-Feb-2003	Analalava 700
ASAM187-05	CASENT0493486-D01	05-Feb-2003	Zombitse 770
ASAM188-05	CASENT0493486-D02	05-Feb-2003	Zombitse 770
ASAM190-05	CASENT0493615-D02	05-Feb-2003	Zombitse 770
ASAM192-05	CASENT0491245-D02	10-Feb-2003	Sahanafa 500
ASAM193-05	CASENT0491254-D01	10-Feb-2003	Sahanafa 500
ASAM194-05	CASENT0491254-D02	10-Feb-2003	Sahanafa 500
ASAM195-05	CASENT0498485-D01	11-Dec-2003	Ampondrabe 175
ASAM196-05	CASENT0498485-D02	11-Dec-2003	Ampondrabe 175
ASAM197-05	CASENT0496655-D01	01-Dec-2002	Saririaky 20
ASAM199-05	CASENT0496651-D01	01-Dec-2002	Saririaky 20
ASAM200-05	CASENT0496651-D02	01-Dec-2002	Saririaky 20
ASAM201-05	CASENT0426080-D01	12-Dec-2001	Bevazoa 130
ASAM202-05	CASENT0426080-D02	12-Dec-2001	Bevazoa 130
ASAM203-05	CASENT0426080-D03	12-Dec-2001	Bevazoa 130
ASAM204-05	CASENT0426065-D01	07-Jan-2001	Toliara airport

ASAM205-05	CASENT0426065-D02	07-Jan-2001	Toliara airport
ASAM207-05	CASENT0426155-D02	10-Jan-2001	Libanona
ASAM843-05	CASENT0057341-D01	16-Apr-2005	Malaza 40
ASAMX365-06	CASENT0067067-D01	13-Nov-2005	Mandrisy
ASAMX411-06	CASENT0067295-D01	13-Nov-2005	Mandrisy
ASAMX412-06	CASENT0067296-D01	13-Nov-2005	Mandrisy
ASAMX714-06	CASENT0070978-D01	19-Apr-2006	Mahabo 20
ASAMX715-06	CASENT0070980-D01	19-Apr-2006	Mahabo 20
ASAMZ041-07	CASENT0120365-D01	01-Dec-2006	Libanona 2006
ASAMZ043-07	CASENT0120368-D01	01-Dec-2006	Libanona 2006
ASIMB490-09	CASENT0125649	28-Nov-2007	Mont Chongui 550
ASIMB492-09	CASENT0125658	28-Nov-2007	Mont Chongui 550
ASIMB495-09	CASENT0125667	28-Nov-2007	Mont Chongui 550
ASIMB496-09	CASENT0125668	28-Nov-2007	Mont Chongui 550
ASIMB507-09	CASENT0132273	27-Nov-2007	Dapani 1
ASIMB563-09	CASENT0132633	25-Nov-2007	Mont combani 370
ASIMB939-09	CASENT0189646	27-Nov-2007	Dapani 1
ASIMB940-09	CASENT0189647	25-Nov-2007	Mont combani 370
ASIMB941-09	CASENT0189648	28-Nov-2007	Mont Chongui 550
ASIMB942-09	CASENT0189649	28-Nov-2007	Mont Chongui 550
ASMEL001-11	CASENT0426482-D01	17-Feb-2002	Bevazoa 130
ASMEL002-11	CASENT0426482-D02	17-Feb-2002	Bevazoa 130
ASMEL003-11	CASENT0426482-D03	17-Feb-2002	Bevazoa 130
ASMEL004-11	CASENT0426416-D01	07-Feb-2002	Anjapolo 65
ASMEL005-11	CASENT0426416-D02	07-Feb-2002	Anjapolo 65
ASMEL006-11	CASENT0426416-D03	07-Feb-2002	Anjapolo 65
ASMEL007-11	CASENT0426459-D03	07-Feb-2002	Anjapolo 65
ASMEL008-11	CASENT0426459-D04	07-Feb-2002	Anjapolo 65
ASMEL009-11	CASENT0426459-D05	07-Feb-2002	Anjapolo 65
ASMEL010-11	CASENT0426447-D01	06-Feb-2002	Malaza 40
ASMEL011-11	CASENT0426447-D02	06-Feb-2002	Malaza 40
ASMEL012-11	CASENT0426447-D03	06-Feb-2002	Malaza 40
ASMEL013-11	CASENT0426439-D01	03-Feb-2002	Bealoka 35
ASMEL014-11	CASENT0426439-D02	03-Feb-2002	Bealoka 35

ASMEL015-11	CASENT0426439-D03	03-Feb-2002	Bealoka 35
ASMEL016-11	CASENT0426203-D03	03-Feb-2002	Bealoka 35
ASMEL017-11	CASENT0426203-D04	03-Feb-2002	Bealoka 35
ASMEL018-11	CASENT0426203-D05	03-Feb-2002	Bealoka 35
ASMEL019-11	CASENT0426189-D01	03-Feb-2002	Bealoka 35
ASMEL020-11	CASENT0426189-D02	03-Feb-2002	Bealoka 35
ASMEL021-11	CASENT0426189-D03	03-Feb-2002	Bealoka 35
ASMEL022-11	CASENT0426183-D01	03-Feb-2002	Bealoka 35
ASMEL023-11	CASENT0426183-D02	03-Feb-2002	Bealoka 35
ASMEL024-11	CASENT0426183-D03	03-Feb-2002	Bealoka 35
ASMEL025-11	CASENT0426167-D01	03-Feb-2002	Bealoka 35
ASMEL026-11	CASENT0426167-D02	03-Feb-2002	Bealoka 35
ASMEL027-11	CASENT0426167-D03	03-Feb-2002	Bealoka 35
ASMEL028-11	CASENT0426514-D01	16-Jan-2002	Tsimelahy 300
ASMEL029-11	CASENT0426514-D02	16-Jan-2002	Tsimelahy 300
ASMEL030-11	CASENT0426514-D03	16-Jan-2002	Tsimelahy 300
ASMEL031-11	CASENT0426503-D01	12-Jan-2002	Manantalinjo 150
ASMEL032-11	CASENT0426503-D02	12-Jan-2002	Manantalinjo 150
ASMEL033-11	CASENT0426503-D03	12-Jan-2002	Manantalinjo 150
ASMEL034-11	CASENT0426129-D03	12-Jan-2002	Manantalinjo 150
ASMEL035-11	CASENT0426129-D04	12-Jan-2002	Manantalinjo 150
ASMEL036-11	CASENT0426129-D05	12-Jan-2002	Manantalinjo 150
ASMEL037-11	CASENT0426097-D03	10-Dec-2001	Belo sur Mer BSI21d
ASMEL038-11	CASENT0426097-D04	10-Dec-2001	Belo sur Mer BSI21d
ASMEL039-11	CASENT0426109-D01	09-Dec-2001	Belo sur Mer BSI21c
ASMEL040-11	CASENT0426109-D02	09-Dec-2001	Belo sur Mer BSI21c
ASMEL041-11	CASENT0426109-D03	09-Dec-2001	Belo sur Mer BSI21c
ASMEL042-11	CASENT0422961-D01	22-Feb-2001	Orangea 90
ASMEL043-11	CASENT0422961-D02	22-Feb-2001	Orangea 90
ASMEL044-11	CASENT0422961-D03	22-Feb-2001	Orangea 90
ASMEL045-11	CASENT0422959-D01	25-Feb-2001	Anosiravo 345
ASMEL046-11	CASENT0422959-D02	25-Feb-2001	Anosiravo 345
ASMEL047-11	CASENT0422959-D03	25-Feb-2001	Anosiravo 345
ASMEL048-11	CASENT0492336-D01	26-Nov-2002	Ambinda 10

ASMEL049-11	CASENT0492336-D02	26-Nov-2002	Ambinda 10
ASMEL050-11	CASENT0492336-D03	26-Nov-2002	Ambinda 10
ASMEL051-11	CASENT0492137-D01	26-Nov-2002	Ambinda 10
ASMEL052-11	CASENT0492137-D02	26-Nov-2002	Ambinda 10
ASMEL053-11	CASENT0492137-D03	26-Nov-2002	Ambinda 10
ASMEL054-11	CASENT0489345-D01	19-Nov-2002	Androngonibe 30
ASMEL055-11	CASENT0489345-D02	19-Nov-2002	Androngonibe 30
ASMEL056-11	CASENT0489345-D03	19-Nov-2002	Androngonibe 30
ASMEL057-11	CASENT0489404-D01	19-Nov-2002	Androngonibe 30
ASMEL058-11	CASENT0489404-D02	19-Nov-2002	Androngonibe 30
ASMEL059-11	CASENT0489404-D03	19-Nov-2002	Androngonibe 30
ASMEL060-11	CASENT0489407-D03	19-Nov-2002	Androngonibe 30
ASMEL061-11	CASENT0489407-D04	19-Nov-2002	Androngonibe 30
ASMEL062-11	CASENT0489407-D05	19-Nov-2002	Androngonibe 30
ASMEL063-11	CASENT0426468-D01	12-Mar-2002	Beroboka 80
ASMEL064-11	CASENT0426468-D02	12-Mar-2002	Beroboka 80
ASMEL065-11	CASENT0426468-D03	12-Mar-2002	Beroboka 80
ASMEL066-11	CASENT0426331-D01	12-Mar-2002	Beroboka 80
ASMEL067-11	CASENT0426331-D02	12-Mar-2002	Beroboka 80
ASMEL068-11	CASENT0426331-D03	12-Mar-2002	Beroboka 80
ASMEL069-11	CASENT0426354-D01	06-Mar-2002	Tsifota 70
ASMEL070-11	CASENT0426354-D02	06-Mar-2002	Tsifota 70
ASMEL071-11	CASENT0426354-D03	06-Mar-2002	Tsifota 70
ASMEL072-11	CASENT0426362-D01	06-Mar-2002	Tsifota 70
ASMEL073-11	CASENT0426362-D02	06-Mar-2002	Tsifota 70
ASMEL074-11	CASENT0426362-D03	06-Mar-2002	Tsifota 70
ASMEL075-11	CASENT0426273-D01	27-Feb-2002	Andranomite 75
ASMEL076-11	CASENT0426273-D02	27-Feb-2002	Andranomite 75
ASMEL077-11	CASENT0426273-D03	27-Feb-2002	Andranomite 75
ASMEL078-11	CASENT0426267-D01	27-Feb-2002	Andranomite 75
ASMEL079-11	CASENT0426267-D02	27-Feb-2002	Andranomite 75
ASMEL080-11	CASENT0426267-D03	27-Feb-2002	Andranomite 75
ASMEL081-11	CASENT0426254-D01	27-Feb-2002	Andranomite 75
ASMEL082-11	CASENT0426254-D02	27-Feb-2002	Andranomite 75

ASMEL083-11	CASENT0426254-D03	27-Feb-2002	Andranomite 75
ASMEL084-11	CASENT0426251-D01	27-Feb-2002	Andranomite 75
ASMEL085-11	CASENT0426251-D02	27-Feb-2002	Andranomite 75
ASMEL086-11	CASENT0426251-D03	27-Feb-2002	Andranomite 75
ASMEL087-11	CASENT0426228-D03	27-Feb-2002	Andranomite 75
ASMEL088-11	CASENT0426228-D04	27-Feb-2002	Andranomite 75
ASMEL089-11	CASENT0426228-D05	27-Feb-2002	Andranomite 75
ASMEL090-11	CASENT0426215-D01	17-Feb-2002	Soamanitra 150
ASMEL091-11	CASENT0426215-D02	17-Feb-2002	Soamanitra 150
ASMEL092-11	CASENT0426210-D03	17-Feb-2002	Soamanitra 150
ASMEL093-11	CASENT0426210-D04	17-Feb-2002	Soamanitra 150
ASMEL094-11	CASENT0426210-D05	17-Feb-2002	Soamanitra 150
ASMEL095-11	CASENT0057341-D02	16-Apr-2005	Malaza 40
ASMEL192-11	CASENT0057341-D03	16-Apr-2005	Malaza 40
ASMEL193-11	CASENT0057341-D04	16-Apr-2005	Malaza 40
ASMEL194-11	CASENT0107796-D02	15-Dec-2004	Ambohimanga 250
ASMEL195-11	CASENT0107796-D03	15-Dec-2004	Ambohimanga 250
ASMEL196-11	CASENT0107796-D04	15-Dec-2004	Ambohimanga 250
ASMEL197-11	CASENT0107669-D01	14-Dec-2004	Ambohimanga 250
ASMEL198-11	CASENT0107669-D02	14-Dec-2004	Ambohimanga 250
ASMEL199-11	CASENT0107669-D03	14-Dec-2004	Ambohimanga 250
ASMEL200-11	CASENT0053834-D03	14-Dec-2004	Ambohimanga 250
ASMEL201-11	CASENT0053834-D04	14-Dec-2004	Ambohimanga 250
ASMEL202-11	CASENT0053834-D05	14-Dec-2004	Ambohimanga 250
ASMEL203-11	CASENT0107665-D02	09-Dec-2004	Ambato 150
ASMEL204-11	CASENT0107665-D03	09-Dec-2004	Ambato 150
ASMEL205-11	CASENT0107665-D04	09-Dec-2004	Ambato 150
ASMEL206-11	CASENT0107663-D02	09-Dec-2004	Ambato 150
ASMEL207-11	CASENT0107663-D03	09-Dec-2004	Ambato 150
ASMEL208-11	CASENT0107663-D04	09-Dec-2004	Ambato 150
ASMEL209-11	CASENT0110516-D01	06-Dec-2004	Analamerana-Bobakindro
ASMEL210-11	CASENT0110516-D02	06-Dec-2004	Analamerana-Bobakindro
ASMEL211-11	CASENT0110516-D03	06-Dec-2004	Analamerana-Bobakindro
ASMEL212-11	CASENT0107648-D02	01-Dec-2004	Ambondrobe 10

ASMEL213-11	CASENT0107648-D03	01-Dec-2004	Ambondrobe 10
ASMEL214-11	CASENT0107648-D04	01-Dec-2004	Ambondrobe 10
ASMEL215-11	CASENT0056403-D01	01-Dec-2004	Ambondrobe 10
ASMEL216-11	CASENT0056403-D02	01-Dec-2004	Ambondrobe 10
ASMEL217-11	CASENT0056403-D03	01-Dec-2004	Ambondrobe 10
ASMEL218-11	CASENT0491253-D01	10-Feb-2003	Sahanafa 500
ASMEL219-11	CASENT0491253-D02	10-Feb-2003	Sahanafa 500
ASMEL220-11	CASENT0491253-D03	10-Feb-2003	Sahanafa 500
ASMEL221-11	CASENT0491244-D01	10-Feb-2003	Sahanafa 500
ASMEL222-11	CASENT0491244-D02	10-Feb-2003	Sahanafa 500
ASMEL223-11	CASENT0491244-D03	10-Feb-2003	Sahanafa 500
ASMEL224-11	CASENT0493489-D01	05-Feb-2003	Zombitse 770
ASMEL225-11	CASENT0493489-D02	05-Feb-2003	Zombitse 770
ASMEL226-11	CASENT0493489-D03	05-Feb-2003	Zombitse 770
ASMEL227-11	CASENT0489468-D01	01-Dec-2002	Analalava 700
ASMEL228-11	CASENT0489468-D02	01-Dec-2002	Analalava 700
ASMEL229-11	CASENT0489468-D03	01-Dec-2002	Analalava 700
ASMEL230-11	CASENT0489657-D01	01-Dec-2002	Analalava 700
ASMEL231-11	CASENT0489657-D02	01-Dec-2002	Analalava 700
ASMEL232-11	CASENT0489657-D03	01-Dec-2002	Analalava 700
ASMEL233-11	CASENT0496653-D01	01-Dec-2002	Saririaky 20
ASMEL234-11	CASENT0496653-D02	01-Dec-2002	Saririaky 20
ASMEL235-11	CASENT0496653-D03	01-Dec-2002	Saririaky 20
ASMEL236-11	CASENT0492483-D01	01-Dec-2002	Saririaky 20
ASMEL237-11	CASENT0492483-D02	01-Dec-2002	Saririaky 20
ASMEL238-11	CASENT0492483-D03	01-Dec-2002	Saririaky 20
ASMEL239-11	CASENT0156974-D01	31-Oct-2009	Beanka VII 160
ASMEL240-11	CASENT0156974-D02	31-Oct-2009	Beanka VII 160
ASMEL241-11	CASENT0156974-D03	31-Oct-2009	Beanka VII 160
ASMEL242-11	CASENT0156621-D01	21-Oct-2009	Beanka II 250
ASMEL243-11	CASENT0156621-D02	21-Oct-2009	Beanka II 250
ASMEL244-11	CASENT0156621-D03	21-Oct-2009	Beanka II 250
ASMEL245-11	CASENT0156657-D01	21-Oct-2009	Beanka II 250
ASMEL246-11	CASENT0156657-D02	21-Oct-2009	Beanka II 250

ASMEL247-11	CASENT0156657-D03	21-Oct-2009	Beanka II 250
ASMEL248-11	CASENT0155912-D01	19-Oct-2009	Beanka II 250
ASMEL249-11	CASENT0155912-D02	19-Oct-2009	Beanka II 250
ASMEL250-11	CASENT0155912-D03	19-Oct-2009	Beanka II 250
ASMEL251-11	CASENT0121504-D01	09-Dec-2006	Behara Spiny Bush
ASMEL252-11	CASENT0121504-D02	09-Dec-2006	Behara Spiny Bush
ASMEL253-11	CASENT0121504-D03	09-Dec-2006	Behara Spiny Bush
ASMEL254-11	CASENT0122764-D01	09-Dec-2006	Behara Spiny Bush
ASMEL255-11	CASENT0122764-D02	09-Dec-2006	Behara Spiny Bush
ASMEL256-11	CASENT0122764-D03	09-Dec-2006	Behara Spiny Bush
ASMEL257-11	CASENT0121592-D01	08-Dec-2006	Vohidava 850
ASMEL258-11	CASENT0121592-D02	08-Dec-2006	Vohidava 850
ASMEL259-11	CASENT0121592-D03	08-Dec-2006	Vohidava 850
ASMEL260-11	CASENT0120365-D02	01-Dec-2006	Libanona 2006
ASMEL261-11	CASENT0120365-D03	01-Dec-2006	Libanona 2006
ASMEL262-11	CASENT0120365-D04	01-Dec-2006	Libanona 2006
ASMEL263-11	CASENT0120367-D01	01-Dec-2006	Libanona 2006
ASMEL264-11	CASENT0120367-D02	01-Dec-2006	Libanona 2006
ASMEL265-11	CASENT0120367-D03	01-Dec-2006	Libanona 2006
ASMEL266-11	CASENT0120368-D02	01-Dec-2006	Libanona 2006
ASMEL267-11	CASENT0120368-D03	01-Dec-2006	Libanona 2006
ASMEL268-11	CASENT0120368-D04	01-Dec-2006	Libanona 2006
ASMEL269-11	CASENT0070978-D02	19-Apr-2006	Mahabo 20
ASMEL270-11	CASENT0070978-D03	19-Apr-2006	Mahabo 20
ASMEL271-11	CASENT0070978-D04	19-Apr-2006	Mahabo 20
ASMEL272-11	CASENT0070980-D02	19-Apr-2006	Mahabo 20
ASMEL273-11	CASENT0070980-D03	19-Apr-2006	Mahabo 20
ASMEL274-11	CASENT0070980-D04	19-Apr-2006	Mahabo 20
ASMEL275-11	CASENT0066757-D01	23-Nov-2005	Ambohidena 20
ASMEL276-11	CASENT0066757-D02	23-Nov-2005	Ambohidena 20
ASMEL277-11	CASENT0066757-D03	23-Nov-2005	Ambohidena 20
ASMEL278-11	CASENT0067296-D02	13-Nov-2005	Mandrisy
ASMEL279-11	CASENT0067296-D03	13-Nov-2005	Mandrisy
ASMEL280-11	CASENT0067296-D04	13-Nov-2005	Mandrisy

ASMEL281-11	CASENT0067295-D02	13-Nov-2005	Mandrisy
ASMEL282-11	CASENT0067295-D03	13-Nov-2005	Mandrisy
ASMEL283-11	CASENT0067295-D04	13-Nov-2005	Mandrisy
ASMEL284-11	CASENT0067067-D02	13-Nov-2005	Mandrisy
ASMEL285-11	CASENT0067067-D03	13-Nov-2005	Mandrisy
ASMEL286-11	CASENT0067067-D04	13-Nov-2005	Mandrisy
ASMEL322-11	CASENT0426159-D01	10-Jan-2001	Libanona
ASMEL323-11	CASENT0426159-D02	10-Jan-2001	Libanona
ASMEL324-11	CASENT0426159-D03	10-Jan-2001	Libanona
ASMEL325-11	CASENT0426062-D01	07-Jan-2001	Toliara airport
ASMEL326-11	CASENT0426062-D02	07-Jan-2001	Toliara airport
ASMEL327-11	CASENT0426062-D03	07-Jan-2001	Toliara airport
ASMEL328-11	CASENT0426087-D01	12-Dec-2001	Bevazoa 130
ASMEL329-11	CASENT0426087-D02	12-Dec-2001	Bevazoa 130
ASMEL330-11	CASENT0426087-D03	12-Dec-2001	Bevazoa 130
JDWAM016-04	CASENT0505295	03-Feb-2002	Bealoka 35
JDWAM342-05	CASENT0053830-D01	14-Dec-2004	Ambohimanga 250
JDWAM343-05	CASENT0053830-D02	14-Dec-2004	Ambohimanga 250
JDWAM344-05	CASENT0053834-D01	14-Dec-2004	Ambohimanga 250
JDWAM345-05	CASENT0053834-D02	14-Dec-2004	Ambohimanga 250
JDWAM528-05	CASENT0107577-D01	06-Dec-2004	Analamerana-Bobakindro
JDWAM623-05	CASENT0107648-D01	01-Dec-2004	Ambondrobe 10
JDWAM629-05	CASENT0107656-D01	01-Dec-2004	Ambondrobe 10
JDWAM636-05	CASENT0107668-D01	14-Dec-2004	Ambohimanga 250
JDWAM637-05	CASENT0107668-D02	14-Dec-2004	Ambohimanga 250
JDWAM715-05	CASENT0107796-D01	15-Dec-2004	Ambohimanga 250

Appendix 1: Diaspididae specimen records

Process ID	***Sample ID***	***Collection Date***	***Locality***
ASAM898-05	CASENT0426359-D11	06-Mar-2002	Tsifota 70
ASAM899-05	CASENT0426359-D12	06-Mar-2002	Tsifota 70
ASAM900-05	CASENT0426330-D11	12-Mar-2002	Beroboka 80
ASAM901-05	CASENT0426330-D12	12-Mar-2002	Beroboka 80
ASAM902-05	CASENT0426342-D11	12-Mar-2002	Beroboka 80
ASAM903-05	CASENT0426342-D12	12-Mar-2002	Beroboka 80
ASAM904-05	CASENT0426466-D11	12-Mar-2002	Beroboka 80
ASAM905-05	CASENT0426466-D12	12-Mar-2002	Beroboka 80
ASAM906-05	CASENT0426308-D11	18-Mar-2002	Mitoho 40
ASAM907-05	CASENT0426308-D12	18-Mar-2002	Mitoho 40
ASAM908-05	CASENT0489664-D11	01-Feb-2003	Analalava 700
ASAM909-05	CASENT0489664-D12	01-Feb-2003	Analalava 700
ASAM910-05	CASENT0493615-D11	05-Feb-2003	Zombitse 770
ASAM911-05	CASENT0491254-D11	10-Feb-2003	Sahanafa 500
ASAM912-05	CASENT0491254-D12	10-Feb-2003	Sahanafa 500
ASAM913-05	CASENT0107656-D11	01-Dec-2004	Ambondrobe 10
ASAM914-05	CASENT0107656-D12	01-Dec-2004	Ambondrobe 10
ASAM915-05	CASENT0107648-D11	01-Dec-2004	Ambondrobe 10
ASAM916-05	CASENT0107648-D12	01-Dec-2004	Ambondrobe 10
ASAM917-05	CASENT0107577-D11	06-Dec-2004	Analamerana-Bobakindro
ASAM918-05	CASENT0107577-D12	06-Dec-2004	Analamerana-Bobakindro
ASAM919-05	CASENT0107663-D11	09-Dec-2004	Ambato 150
ASAM920-05	CASENT0107663-D12	09-Dec-2004	Ambato 150
ASAM921-05	CASENT0053830-D11	14-Dec-2004	Ambohimanga 250
ASAM922-05	CASENT0053830-D12	14-Dec-2004	Ambohimanga 250
ASAM923-05	CASENT0107796-D11	15-Dec-2004	Ambohimanga 250
ASAM924-05	CASENT0107796-D12	15-Dec-2004	Ambohimanga 250
ASAM925-05	CASENT0057341-D11	16-Apr-2005	Malaza 40
ASAM926-05	CASENT0057341-D12	16-Apr-2005	Malaza 40
ASAM927-05	CASENT0496655-D11	01-Dec-2002	Saririaky 20

ASAM928-05	CASENT0496655-D12	01-Dec-2002	Saririaky 20
ASAM929-05	CASENT0496651-D11	01-Dec-2002	Saririaky 20
ASAM930-05	CASENT0496651-D12	01-Dec-2002	Saririaky 20
ASAM931-05	CASENT0107668-D11	14-Dec-2004	Ambohimanga 250
ASAM932-05	CASENT0107668-D12	14-Dec-2004	Ambohimanga 250
ASAMI049-05	CASENT0426155-D11	10-Jan-2001	Libanona
ASAMI050-05	CASENT0426155-D12	10-Jan-2001	Libanona
ASAMI051-05	CASENT0426065-D11	07-Jan-2001	Toliara airport
ASAMI052-05	CASENT0426065-D12	07-Jan-2001	Toliara airport
ASAMI053-05	CASENT0426618-D11	19-May-2002	Cedarburg 800
ASAMI054-05	CASENT0426618-D12	19-May-2002	Cedarburg 800
ASAMI055-05	CASENT0426097-D11	10-Dec-2001	Belo sur Mer BSI21d
ASAMI056-05	CASENT0426097-D12	10-Dec-2001	Belo sur Mer BSI21d
ASAMI057-05	CASENT0426129-D11	12-Jan-2002	Manantalinjo 150
ASAMI058-05	CASENT0426129-D12	12-Jan-2002	Manantalinjo 150
ASAMI059-05	CASENT0426506-D11	12-Jan-2002	Manantalinjo 150
ASAMI060-05	CASENT0426506-D12	12-Jan-2002	Manantalinjo 150
ASAMI061-05	CASENT0426515-D11	16-Jan-2002	Tsimelahy 300
ASAMI062-05	CASENT0426515-D12	16-Jan-2002	Tsimelahy 300
ASAMI063-05	CASENT0426163-D11	03-Feb-2002	Bealoka 35
ASAMI064-05	CASENT0426163-D12	03-Feb-2002	Bealoka 35
ASAMI065-05	CASENT0426192-D11	03-Feb-2002	Bealoka 35
ASAMI066-05	CASENT0426192-D12	03-Feb-2002	Bealoka 35
ASAMI067-05	CASENT0426203-D11	03-Feb-2002	Bealoka 35
ASAMI068-05	CASENT0426203-D12	03-Feb-2002	Bealoka 35
ASAMI069-05	CASENT0426248-D11	27-Feb-2002	Andranomite 75
ASAMI070-05	CASENT0426248-D12	27-Feb-2002	Andranomite 75
ASAMI071-05	CASENT0426228-D11	27-Feb-2002	Andranomite 75
ASAMI072-05	CASENT0426228-D12	27-Feb-2002	Andranomite 75
ASAMI073-05	CASENT0426236-D11	21-Feb-2002	Mahafaly 80
ASAMI074-05	CASENT0426236-D12	21-Feb-2002	Mahafaly 80
ASAMI075-05	CASENT0426214-D11	21-Feb-2002	Mahafaly 80
ASAMI076-05	CASENT0426214-D12	21-Feb-2002	Mahafaly 80
ASAMI077-05	CASENT0426210-D11	17-Feb-2002	Soamanitra 150

ASAMI078-05	CASENT0426210-D12	17-Feb-2002	Soamanitra 150
ASAMI079-05	CASENT0426428-D11	17-Feb-2002	Bevazoa 130
ASAMI080-05	CASENT0426428-D12	17-Feb-2002	Bevazoa 130
ASAMI081-05	CASENT0426423-D11	07-Feb-2002	Anjapolo 65
ASAMI082-05	CASENT0426423-D12	07-Feb-2002	Anjapolo 65
ASAMI083-05	CASENT0426459-D11	07-Feb-2002	Anjapolo 65
ASAMI084-05	CASENT0426459-D12	07-Feb-2002	Anjapolo 65
ASAMI085-05	CASENT0426452-D11	06-Feb-2002	Malaza 40
ASAMI086-05	CASENT0426452-D12	06-Feb-2002	Malaza 40
ASAMI087-05	CASENT0426441-D11	03-Feb-2002	Bealoka 35
ASAMI088-05	CASENT0426441-D12	03-Feb-2002	Bealoka 35
ASAMI089-05	CASENT0426412-D11	27-Feb-2002	Andranomite 75
ASAMI090-05	CASENT0426412-D12	27-Feb-2002	Andranomite 75
ASAMI091-05	CASENT0426261-D11	27-Feb-2002	Andranomite 75
ASAMI092-05	CASENT0426261-D12	27-Feb-2002	Andranomite 75
ASAMI093-05	CASENT0426271-D11	27-Feb-2002	Andranomite 75
ASAMI094-05	CASENT0426271-D12	27-Feb-2002	Andranomite 75
ASMEL096-11	CASENT0426482-D11	17-Feb-2002	Bevazoa 130
ASMEL097-11	CASENT0426482-D12	17-Feb-2002	Bevazoa 130
ASMEL098-11	CASENT0426482-D13	17-Feb-2002	Bevazoa 130
ASMEL099-11	CASENT0426416-D11	07-Feb-2002	Anjapolo 65
ASMEL100-11	CASENT0426416-D12	07-Feb-2002	Anjapolo 65
ASMEL101-11	CASENT0426459-D13	07-Feb-2002	Anjapolo 65
ASMEL102-11	CASENT0426459-D14	07-Feb-2002	Anjapolo 65
ASMEL103-11	CASENT0426459-D15	07-Feb-2002	Anjapolo 65
ASMEL104-11	CASENT0426447-D11	06-Feb-2002	Malaza 40
ASMEL105-11	CASENT0426447-D12	06-Feb-2002	Malaza 40
ASMEL106-11	CASENT0426447-D13	06-Feb-2002	Malaza 40
ASMEL107-11	CASENT0426203-D13	03-Feb-2002	Bealoka 35
ASMEL108-11	CASENT0426203-D14	03-Feb-2002	Bealoka 35
ASMEL109-11	CASENT0426203-D15	03-Feb-2002	Bealoka 35
ASMEL110-11	CASENT0426189-D11	03-Feb-2002	Bealoka 35
ASMEL111-11	CASENT0426189-D12	03-Feb-2002	Bealoka 35
ASMEL112-11	CASENT0426189-D13	03-Feb-2002	Bealoka 35

ASMEL113-11	CASENT0426167-D11	03-Feb-2002	Bealoka 35
ASMEL114-11	CASENT0426167-D12	03-Feb-2002	Bealoka 35
ASMEL115-11	CASENT0426167-D13	03-Feb-2002	Bealoka 35
ASMEL116-11	CASENT0426514-D11	16-Jan-2002	Tsimelahy 300
ASMEL117-11	CASENT0426514-D12	16-Jan-2002	Tsimelahy 300
ASMEL118-11	CASENT0426514-D13	16-Jan-2002	Tsimelahy 300
ASMEL119-11	CASENT0426503-D11	12-Jan-2002	Manantalinjo 150
ASMEL120-11	CASENT0426503-D12	12-Jan-2002	Manantalinjo 150
ASMEL121-11	CASENT0426503-D13	12-Jan-2002	Manantalinjo 150
ASMEL122-11	CASENT0426129-D13	12-Jan-2002	Manantalinjo 150
ASMEL123-11	CASENT0426129-D14	12-Jan-2002	Manantalinjo 150
ASMEL124-11	CASENT0426129-D15	12-Jan-2002	Manantalinjo 150
ASMEL125-11	CASENT0426097-D13	10-Dec-2001	Belo sur Mer BSI21d
ASMEL126-11	CASENT0426097-D14	10-Dec-2001	Belo sur Mer BSI21d
ASMEL127-11	CASENT0426097-D15	10-Dec-2001	Belo sur Mer BSI21d
ASMEL128-11	CASENT0426468-D11	12-Mar-2002	Beroboka 80
ASMEL129-11	CASENT0426468-D12	12-Mar-2002	Beroboka 80
ASMEL130-11	CASENT0426468-D13	12-Mar-2002	Beroboka 80
ASMEL131-11	CASENT0426331-D11	12-Mar-2002	Beroboka 80
ASMEL132-11	CASENT0426331-D12	12-Mar-2002	Beroboka 80
ASMEL133-11	CASENT0426331-D13	12-Mar-2002	Beroboka 80
ASMEL134-11	CASENT0426362-D11	06-Mar-2002	Tsifota 70
ASMEL135-11	CASENT0426362-D12	06-Mar-2002	Tsifota 70
ASMEL136-11	CASENT0426362-D13	06-Mar-2002	Tsifota 70
ASMEL137-11	CASENT0426273-D11	27-Feb-2002	Andranomite 75
ASMEL138-11	CASENT0426273-D12	27-Feb-2002	Andranomite 75
ASMEL139-11	CASENT0426273-D13	27-Feb-2002	Andranomite 75
ASMEL140-11	CASENT0426267-D11	27-Feb-2002	Andranomite 75
ASMEL141-11	CASENT0426267-D12	27-Feb-2002	Andranomite 75
ASMEL142-11	CASENT0426267-D13	27-Feb-2002	Andranomite 75
ASMEL143-11	CASENT0426251-D11	27-Feb-2002	Andranomite 75
ASMEL144-11	CASENT0426251-D12	27-Feb-2002	Andranomite 75
ASMEL145-11	CASENT0426228-D13	27-Feb-2002	Andranomite 75
ASMEL146-11	CASENT0426228-D14	27-Feb-2002	Andranomite 75

ASMEL147-11	CASENT0426228-D15	27-Feb-2002	Andranomite 75
ASMEL148-11	CASENT0426210-D13	17-Feb-2002	Soamanitra 150
ASMEL149-11	CASENT0426210-D14	17-Feb-2002	Soamanitra 150
ASMEL150-11	CASENT0426210-D15	17-Feb-2002	Soamanitra 150
ASMEL151-11	CASENT0057341-D13	16-Apr-2005	Malaza 40
ASMEL152-11	CASENT0057341-D14	16-Apr-2005	Malaza 40
ASMEL153-11	CASENT0057341-D15	16-Apr-2005	Malaza 40
ASMEL154-11	CASENT0107796-D13	15-Dec-2004	Ambohimanga 250
ASMEL155-11	CASENT0107796-D14	15-Dec-2004	Ambohimanga 250
ASMEL156-11	CASENT0107796-D15	15-Dec-2004	Ambohimanga 250
ASMEL157-11	CASENT0107669-D11	14-Dec-2004	Ambohimanga 250
ASMEL158-11	CASENT0107669-D12	14-Dec-2004	Ambohimanga 250
ASMEL159-11	CASENT0107669-D13	14-Dec-2004	Ambohimanga 250
ASMEL160-11	CASENT0107663-D13	09-Dec-2004	Ambato 150
ASMEL161-11	CASENT0107663-D14	09-Dec-2004	Ambato 150
ASMEL162-11	CASENT0107663-D15	09-Dec-2004	Ambato 150
ASMEL163-11	CASENT0110516-D11	06-Dec-2004	Analamerana-Bobakindro
ASMEL164-11	CASENT0110516-D12	06-Dec-2004	Analamerana-Bobakindro
ASMEL165-11	CASENT0110516-D13	06-Dec-2004	Analamerana-Bobakindro
ASMEL166-11	CASENT0107648-D13	01-Dec-2004	Ambondrobe 10
ASMEL167-11	CASENT0107648-D14	01-Dec-2004	Ambondrobe 10
ASMEL168-11	CASENT0107648-D15	01-Dec-2004	Ambondrobe 10
ASMEL169-11	CASENT0056403-D11	01-Dec-2004	Ambondrobe 10
ASMEL170-11	CASENT0056403-D12	01-Dec-2004	Ambondrobe 10
ASMEL171-11	CASENT0056403-D13	01-Dec-2004	Ambondrobe 10
ASMEL172-11	CASENT0491253-D11	10-Feb-2003	Sahanafa 500
ASMEL173-11	CASENT0491253-D12	10-Feb-2003	Sahanafa 500
ASMEL174-11	CASENT0491253-D13	10-Feb-2003	Sahanafa 500
ASMEL175-11	CASENT0493489-D11	05-Feb-2003	Zombitse 770
ASMEL176-11	CASENT0493489-D12	05-Feb-2003	Zombitse 770
ASMEL177-11	CASENT0493489-D13	05-Feb-2003	Zombitse 770
ASMEL178-11	CASENT0492483-D11	01-Dec-2002	Saririaky 20
ASMEL179-11	CASENT0492483-D12	01-Dec-2002	Saririaky 20
ASMEL180-11	CASENT0492483-D13	01-Dec-2002	Saririaky 20

ASMEL181-11	CASENT0156974-D11	31-Oct-2009	Beanka VII 160
ASMEL182-11	CASENT0156974-D12	31-Oct-2009	Beanka VII 160
ASMEL183-11	CASENT0156974-D13	31-Oct-2009	Beanka VII 160
ASMEL184-11	CASENT0156621-D11	21-Oct-2009	Beanka II 250
ASMEL185-11	CASENT0156621-D12	21-Oct-2009	Beanka II 250
ASMEL186-11	CASENT0156621-D13	21-Oct-2009	Beanka II 250
ASMEL187-11	CASENT0156657-D11	21-Oct-2009	Beanka II 250
ASMEL188-11	CASENT0156657-D12	21-Oct-2009	Beanka II 250
ASMEL189-11	CASENT0156657-D13	21-Oct-2009	Beanka II 250
ASMEL190-11	CASENT0155912-D11	19-Oct-2009	Beanka II 250
ASMEL191-11	CASENT0155912-D12	19-Oct-2009	Beanka II 250
ASMEL287-11	CASENT0155912-D13	19-Oct-2009	Beanka II 250
ASMEL288-11	CASENT0121504-D11	09-Dec-2006	Behara Spiny Bush
ASMEL289-11	CASENT0121504-D12	09-Dec-2006	Behara Spiny Bush
ASMEL290-11	CASENT0121504-D13	09-Dec-2006	Behara Spiny Bush
ASMEL291-11	CASENT0122764-D11	09-Dec-2006	Behara Spiny Bush
ASMEL292-11	CASENT0122764-D12	09-Dec-2006	Behara Spiny Bush
ASMEL293-11	CASENT0122764-D13	09-Dec-2006	Behara Spiny Bush
ASMEL294-11	CASENT0121592-D11	08-Dec-2006	Vohidava 850
ASMEL295-11	CASENT0121592-D12	08-Dec-2006	Vohidava 850
ASMEL296-11	CASENT0121592-D13	08-Dec-2006	Vohidava 850
ASMEL297-11	CASENT0120367-D11	01-Dec-2006	Libanona 2006
ASMEL298-11	CASENT0120367-D12	01-Dec-2006	Libanona 2006
ASMEL299-11	CASENT0120367-D13	01-Dec-2006	Libanona 2006
ASMEL300-11	CASENT0120368-D11	01-Dec-2006	Libanona 2006
ASMEL301-11	CASENT0120368-D12	01-Dec-2006	Libanona 2006
ASMEL302-11	CASENT0120368-D13	01-Dec-2006	Libanona 2006
ASMEL303-11	CASENT0070978-D11	19-Apr-2006	Mahabo 20
ASMEL304-11	CASENT0070978-D12	19-Apr-2006	Mahabo 20
ASMEL305-11	CASENT0070978-D13	19-Apr-2006	Mahabo 20
ASMEL306-11	CASENT0070980-D11	19-Apr-2006	Mahabo 20
ASMEL307-11	CASENT0070980-D12	19-Apr-2006	Mahabo 20
ASMEL308-11	CASENT0070980-D13	19-Apr-2006	Mahabo 20
ASMEL309-11	CASENT0066757-D11	23-Nov-2005	Ambohidena 20

ASMEL310-11	CASENT0066757-D12	23-Nov-2005	Ambohidena 20
ASMEL311-11	CASENT0066757-D13	23-Nov-2005	Ambohidena 20
ASMEL312-11	CASENT0067296-D11	13-Nov-2005	Mandrisy
ASMEL313-11	CASENT0067067-D11	13-Nov-2005	Mandrisy
ASMEL314-11	CASENT0067067-D12	13-Nov-2005	Mandrisy
ASMEL315-11	CASENT0067067-D13	13-Nov-2005	Mandrisy
ASMEL316-11	CASENT0426159-D11	10-Jan-2001	Libanona
ASMEL317-11	CASENT0426159-D12	10-Jan-2001	Libanona
ASMEL318-11	CASENT0426159-D13	10-Jan-2001	Libanona
ASMEL319-11	CASENT0426062-D11	07-Jan-2001	Toliara airport
ASMEL320-11	CASENT0426062-D12	07-Jan-2001	Toliara airport
ASMEL321-11	CASENT0426062-D13	07-Jan-2001	Toliara airport

Appendix 2: Primers used to amplify *Melissotarsus insularis* and Diaspididae DNA

Primer Name	***Primer Sequence***	***Gene Region***	***Source***
LepF1	ATTCAACCAATCA TAAAGATATTGG	CO1	Hebert et al., 2004
LepR1	TAAACTTCTGGAT GTCCAAAAAATCA	CO1	Hajibabaei et al., 2005
C_ANTMR1D-AMR1deg_R	CAWCCWGTWCC KRMNCCWKCAT	CO1	Smith and Fisher, 2009
RonMWASPdeg_t1	TGTAAAACGACG GCCAGTGGWTC WCCWGATATAKC WTTTCC	CO1	Modified by M. Alex Smith from Pfunder et al., 2004
12Sai	AAACTAGGATTA GATACCCTATTAT	12S	Simon et al., 1994
12Sbi-f	GAAAATGACGGG CAATTTGT	12S	Modified from Simon *et al.*, 1994
H3F	ATGGCTCGTACC AAGCAGACVGC	Histone H3	Colgan *et al.*, 1998
H3R	ATATCCTTRGGC ATRATRGTGAC	Histone H3	Colgan *et al.*, 1998
Pcof	CCTTCAACTAATC ATAAAAATATYAG	CO1	Doo-Sang Park (unpublished)
D2B	GTCGGGTTGCTT GAGAGTGC	28S	Saux *et al.*, 2004
D3Ar	TCCGTGTTTCAA GACGGGTC	28S	Saux *et al.*, 2004

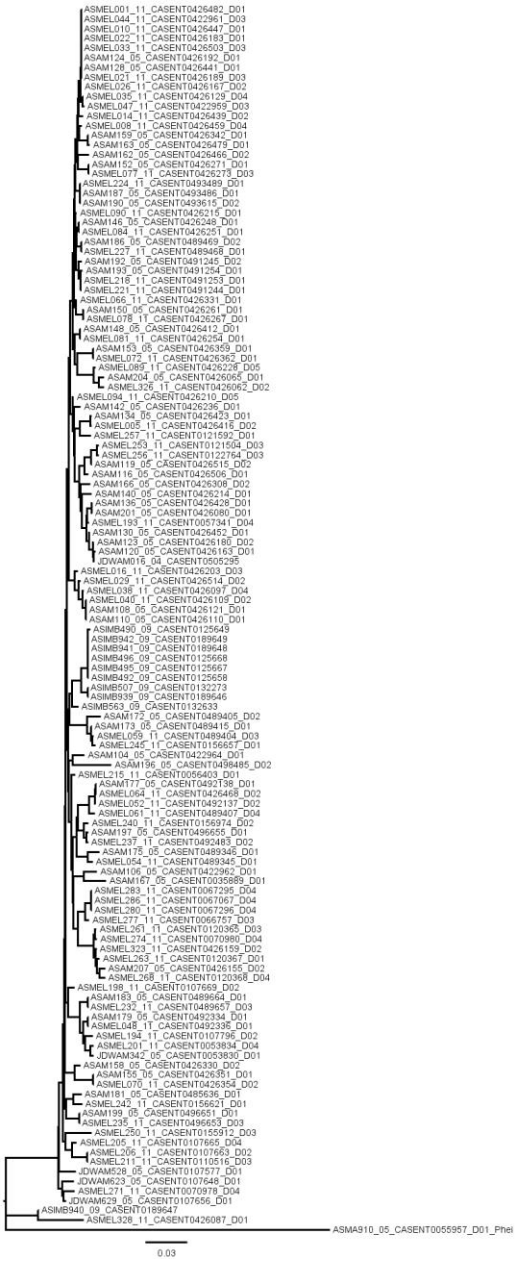
Appendix 3: PCR Temperature Profiles

Primer Set	***PCR Temperature Profile***
LepF1/ LepR1	94°C at 1:00 minutes, 5 cycles of 94°C at 0:40 seconds, 45°C at 0:40 seconds and 72°C at 1:00 minutes, 40 cycles of 94°C at 0:40 seconds, 51°C at 0:40 seconds and 72°C at 1:00 minutes. Final elongation at 72°C for 5:00 minutes. Stores at 4°C.
LepF1/C_ANTMR1D-AMR1deg_R	94°C at 1:00 minutes, 5 cycles of 94°C at 0:40 seconds, 45°C at 0:40 seconds and 72°C at 1:00 minutes, 40 cycles of 94°C at 0:40 seconds, 51°C at 0:40 seconds and 72°C at 1:00 minutes. Final elongation at 72°C for 5:00 minutes. Stores at 4°C.
RonMWASPdeg_t1/LepR1	94°C at 1:00 minutes, 5 cycles of 94°C at 0:40 seconds, 45°C at 0:40 seconds and 72°C at 1:00 minutes, 40 cycles of 94°C at 0:40 seconds, 51°C at 0:40 seconds and 72°C at 1:00 minutes. Final elongation at 72°C for 5:00 minutes. Stores at 4°C.
12Sai/12Sbi-f	94°C at 2:00 minutes, 35 cycles of 94°C at 1:00 minute, 50°C at 1:00 minute and 72°C at 2:00 minutes. Final elongation at 72°C for 5:00 minutes. Stores at 4°C. (Moreau, 2008)
H3F/ H3R	94°C at 2:00 minutes, 35 cycles of 94°C at 1:00 minute, 57°C at 1:00 minute and 72°C at 2:00 minutes. Final elongation at 72°C for 5:00 minutes. Stores at 4°C.
Pcof	94°C at 1:00 minutes, 5 cycles of 94°C at 0:40 seconds, 45°C at 0:40 seconds and 72°C at 1:00 minutes, 40 cycles of 94°C at 0:40 seconds, 51°C at 0:40 seconds and 72°C at 1:00 minutes. Final elongation at 72°C for 5:00 minutes. Stores at 4°C.
D2B/ D3Ar	94°C at 1:00 minutes, 5 cycles of 94°C at 0:40 seconds, 45°C at 0:40 seconds and 72°C at 1:00 minutes, 40 cycles of 94°C at 0:40 seconds, 51°C at 0:40 seconds and 72°C at 1:00 minutes. Final elongation at 72°C for 5:00 minutes. Stores at 4°C.

Appendix 4: Neighbor-joining, maximum parsimony and Bayesian phylogenetic trees for *Melissotarsus insularis*.

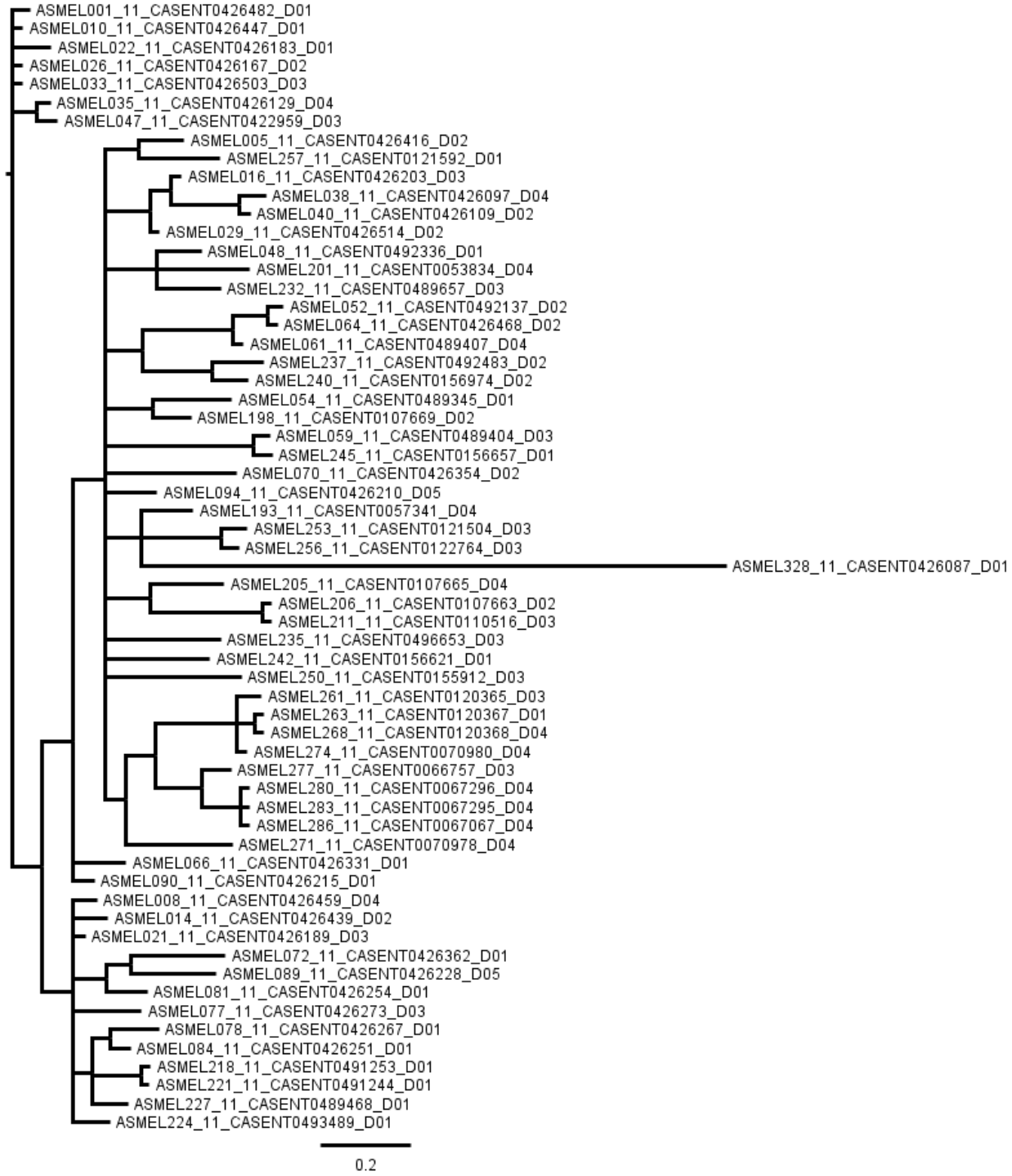
COI

Neighbor-joining



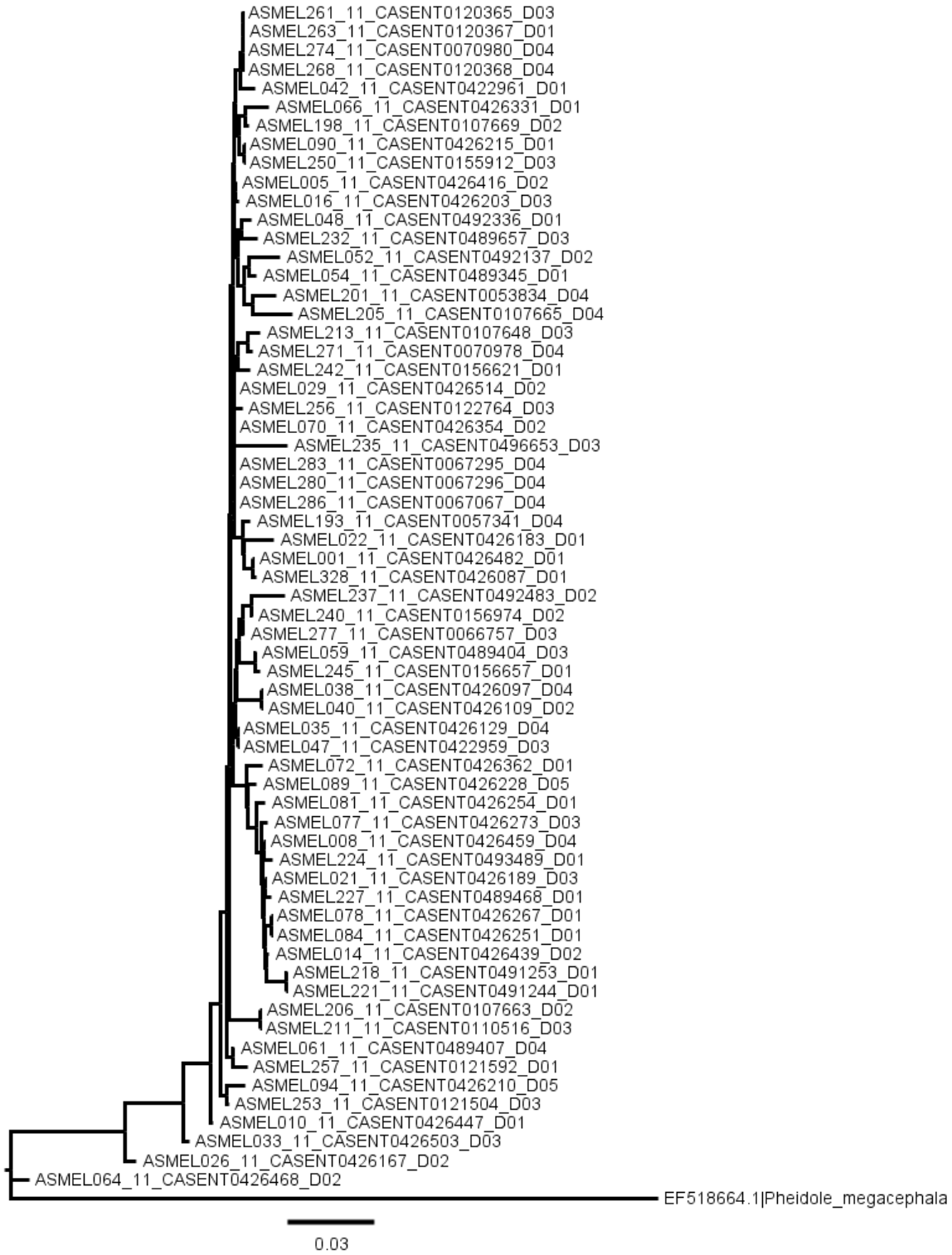
Maximum Parsimony

Partition

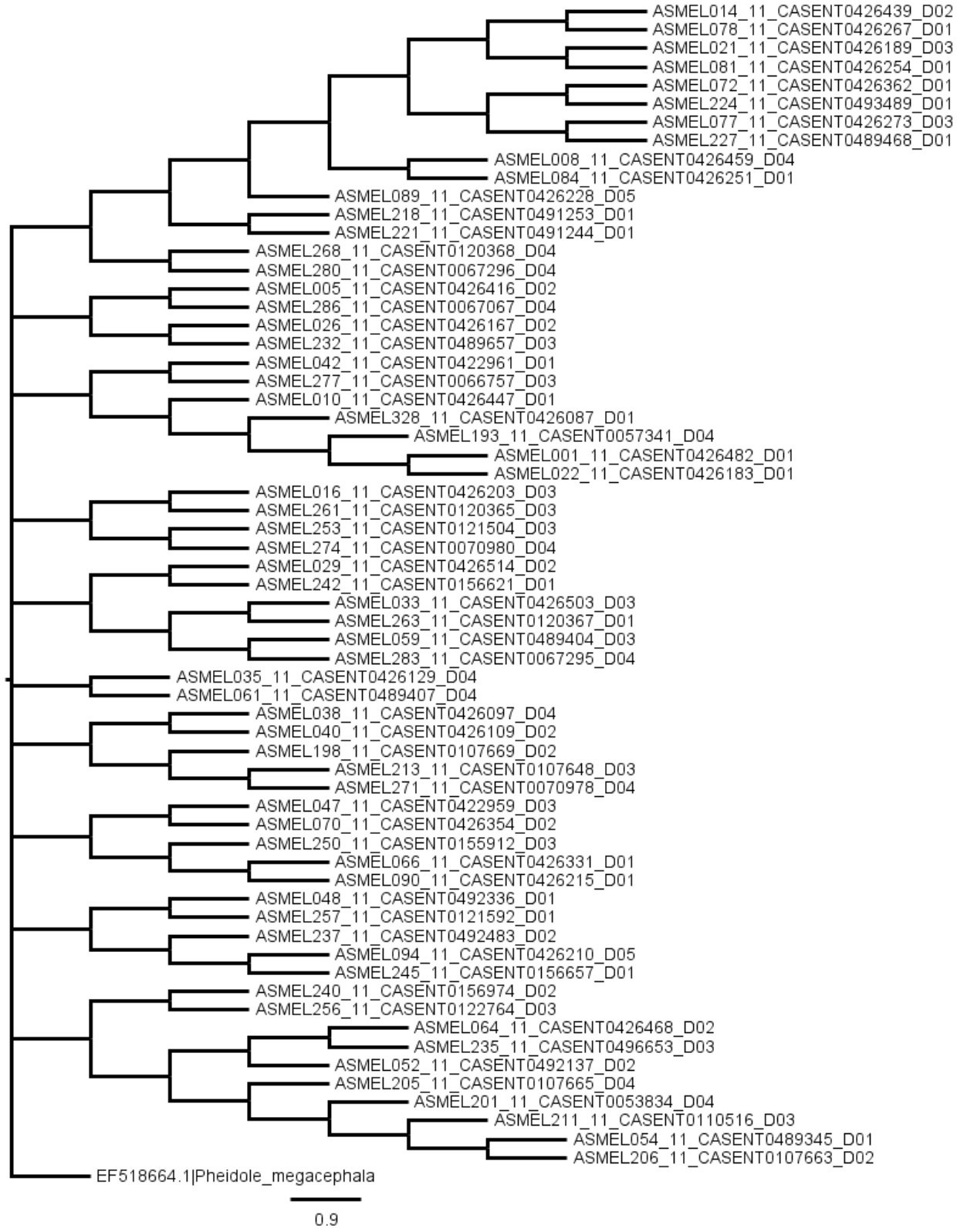

ASMEL001_11_CASENT0426482_D01
ASMEL010_11_CASENT0426447_D01
ASMEL022_11_CASENT0426183_D01
ASMEL026_11_CASENT0426167_D02
ASMEL033_11_CASENT0426503_D03
ASMEL035_11_CASENT0426129_D04
ASMEL047_11_CASENT0422959_D03
ASMEL005_11_CASENT0426416_D02
ASMEL257_11_CASENT0121592_D01
ASMEL016_11_CASENT0426203_D03
ASMEL038_11_CASENT0426097_D04
ASMEL040_11_CASENT0426109_D02
ASMEL029_11_CASENT0426514_D02
ASMEL048_11_CASENT0492336_D01
ASMEL201_11_CASENT0053834_D04
ASMEL232_11_CASENT0489657_D03
ASMEL052_11_CASENT0492137_D02
ASMEL064_11_CASENT0426468_D02
ASMEL061_11_CASENT0489407_D04
ASMEL237_11_CASENT0492483_D02
ASMEL240_11_CASENT0156974_D02
ASMEL054_11_CASENT0489345_D01
ASMEL198_11_CASENT0107669_D02
ASMEL059_11_CASENT0489404_D03
ASMEL245_11_CASENT0156657_D01
ASMEL070_11_CASENT0426354_D02
ASMEL094_11_CASENT0426210_D05
ASMEL193_11_CASENT0057341_D04
ASMEL253_11_CASENT0121504_D03
ASMEL256_11_CASENT0122764_D03
ASMEL328_11_CASENT0426087_D01
ASMEL205_11_CASENT0107665_D04
ASMEL206_11_CASENT0107663_D02
ASMEL211_11_CASENT0110516_D03
ASMEL235_11_CASENT0496653_D03
ASMEL242_11_CASENT0156621_D01
ASMEL250_11_CASENT0155912_D03
ASMEL261_11_CASENT0120365_D03
ASMEL263_11_CASENT0120367_D01
ASMEL268_11_CASENT0120368_D04
ASMEL274_11_CASENT0070980_D04
ASMEL277_11_CASENT0066757_D03
ASMEL280_11_CASENT0067296_D04
ASMEL283_11_CASENT0067295_D04
ASMEL286_11_CASENT0067067_D04
ASMEL271_11_CASENT0070978_D04
ASMEL066_11_CASENT0426331_D01
ASMEL090_11_CASENT0426215_D01
ASMEL008_11_CASENT0426459_D04
ASMEL014_11_CASENT0426439_D02
ASMEL021_11_CASENT0426189_D03
ASMEL072_11_CASENT0426362_D01
ASMEL089_11_CASENT0426228_D05
ASMEL081_11_CASENT0426254_D01
ASMEL077_11_CASENT0426273_D03
ASMEL078_11_CASENT0426267_D01
ASMEL084_11_CASENT0426251_D01
ASMEL218_11_CASENT0491253_D01
ASMEL221_11_CASENT0491244_D01
ASMEL227_11_CASENT0489468_D01
ASMEL224_11_CASENT0493489_D01
0.2

12S

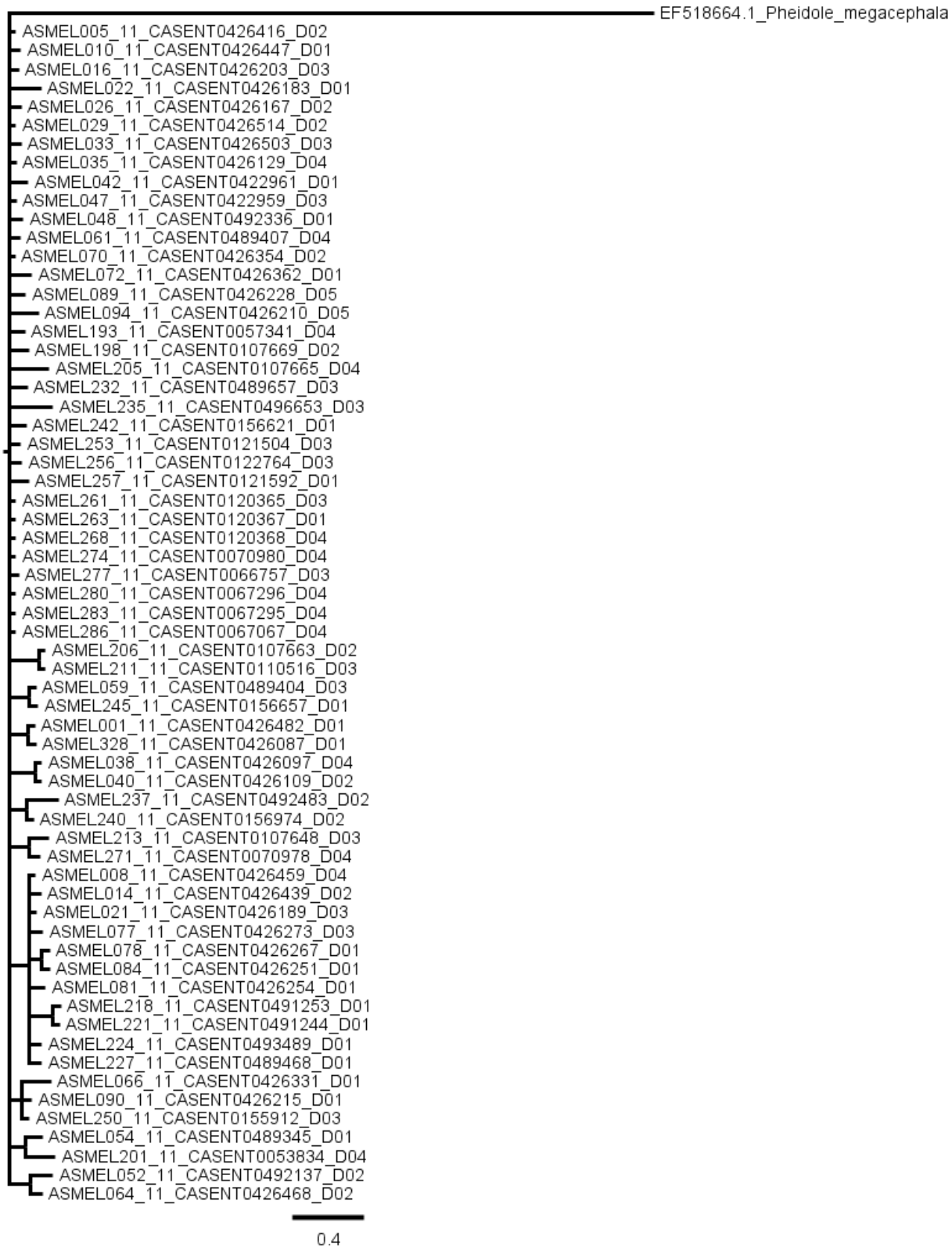
Neighbor-joining



Maximum Parsimony



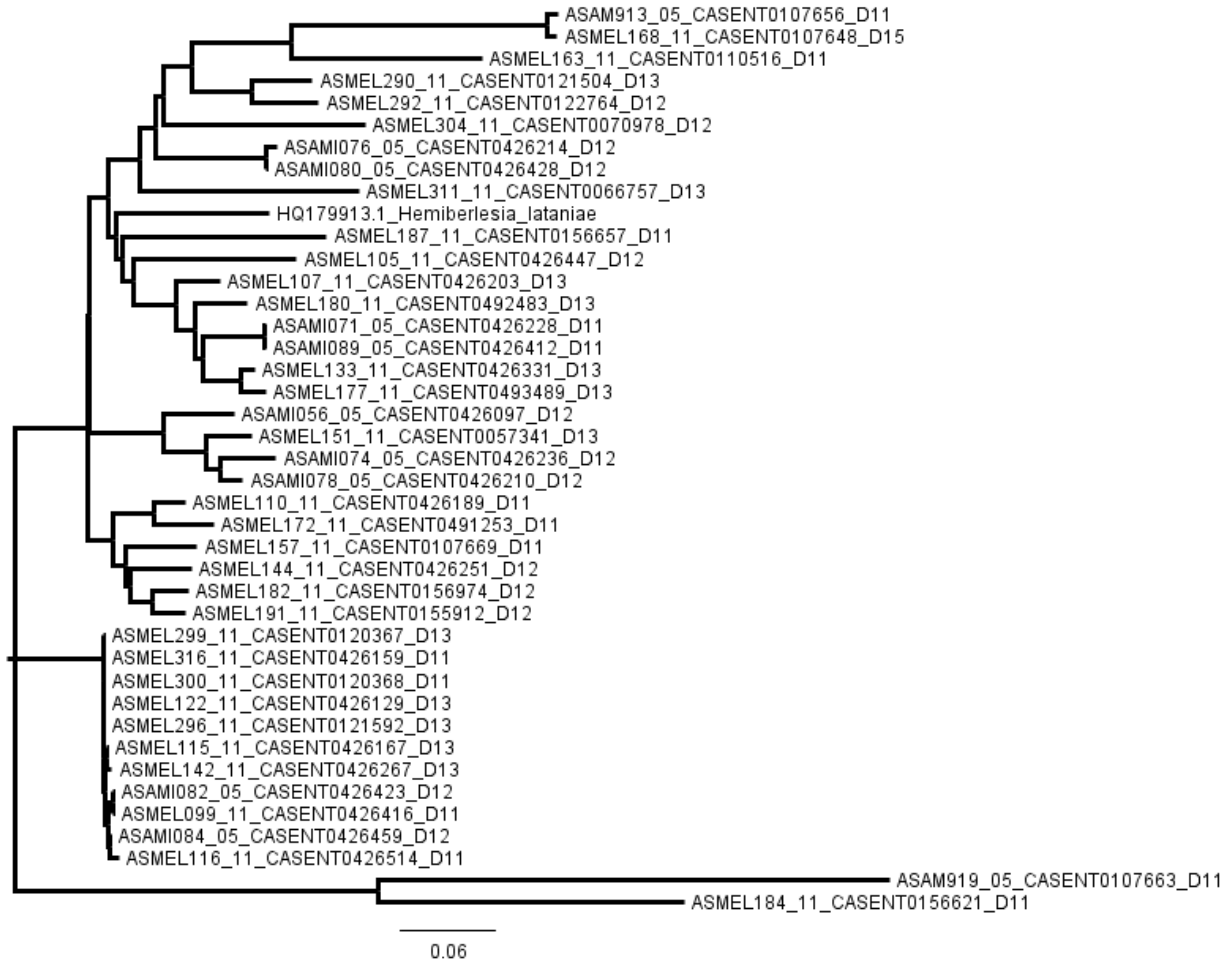
Bayesian


EF518664.1_Pheidole_megacephala
ASMEL005_11_CASENT0426416_D02
ASMEL010_11_CASENT0426447_D01
ASMEL016_11_CASENT0426203_D03
ASMEL022_11_CASENT0426183_D01
ASMEL026_11_CASENT0426167_D02
ASMEL029_11_CASENT0426514_D02
ASMEL033_11_CASENT0426503_D03
ASMEL035_11_CASENT0426129_D04
ASMEL042_11_CASENT0422961_D01
ASMEL047_11_CASENT0422959_D03
ASMEL048_11_CASENT0492336_D01
ASMEL061_11_CASENT0489407_D04
ASMEL070_11_CASENT0426354_D02
ASMEL072_11_CASENT0426362_D01
ASMEL089_11_CASENT0426228_D05
ASMEL094_11_CASENT0426210_D05
ASMEL193_11_CASENT0057341_D04
ASMEL198_11_CASENT0107669_D02
ASMEL205_11_CASENT0107665_D04
ASMEL232_11_CASENT0489657_D03
ASMEL235_11_CASENT0496653_D03
ASMEL242_11_CASENT0156621_D01
ASMEL253_11_CASENT0121504_D03
ASMEL256_11_CASENT0122764_D03
ASMEL257_11_CASENT0121592_D01
ASMEL261_11_CASENT0120365_D03
ASMEL263_11_CASENT0120367_D01
ASMEL268_11_CASENT0120368_D04
ASMEL274_11_CASENT0070980_D04
ASMEL277_11_CASENT0066757_D03
ASMEL280_11_CASENT0067296_D04
ASMEL283_11_CASENT0067295_D04
ASMEL286_11_CASENT0067067_D04
ASMEL206_11_CASENT0107663_D02
ASMEL211_11_CASENT0110516_D03
ASMEL059_11_CASENT0489404_D03
ASMEL245_11_CASENT0156657_D01
ASMEL001_11_CASENT0426482_D01
ASMEL328_11_CASENT0426087_D01
ASMEL038_11_CASENT0426097_D04
ASMEL040_11_CASENT0426109_D02
ASMEL237_11_CASENT0492483_D02
ASMEL240_11_CASENT0156974_D02
ASMEL213_11_CASENT0107648_D03
ASMEL271_11_CASENT0070978_D04
ASMEL008_11_CASENT0426459_D04
ASMEL014_11_CASENT0426439_D02
ASMEL021_11_CASENT0426189_D03
ASMEL077_11_CASENT0426273_D03
ASMEL078_11_CASENT0426267_D01
ASMEL084_11_CASENT0426251_D01
ASMEL081_11_CASENT0426254_D01
ASMEL218_11_CASENT0491253_D01
ASMEL221_11_CASENT0491244_D01
ASMEL224_11_CASENT0493489_D01
ASMEL227_11_CASENT0489468_D01
ASMEL066_11_CASENT0426331_D01
ASMEL090_11_CASENT0426215_D01
ASMEL250_11_CASENT0155912_D03
ASMEL054_11_CASENT0489345_D01
ASMEL201_11_CASENT0053834_D04
ASMEL052_11_CASENT0492137_D02
ASMEL064_11_CASENT0426468_D02
0.4

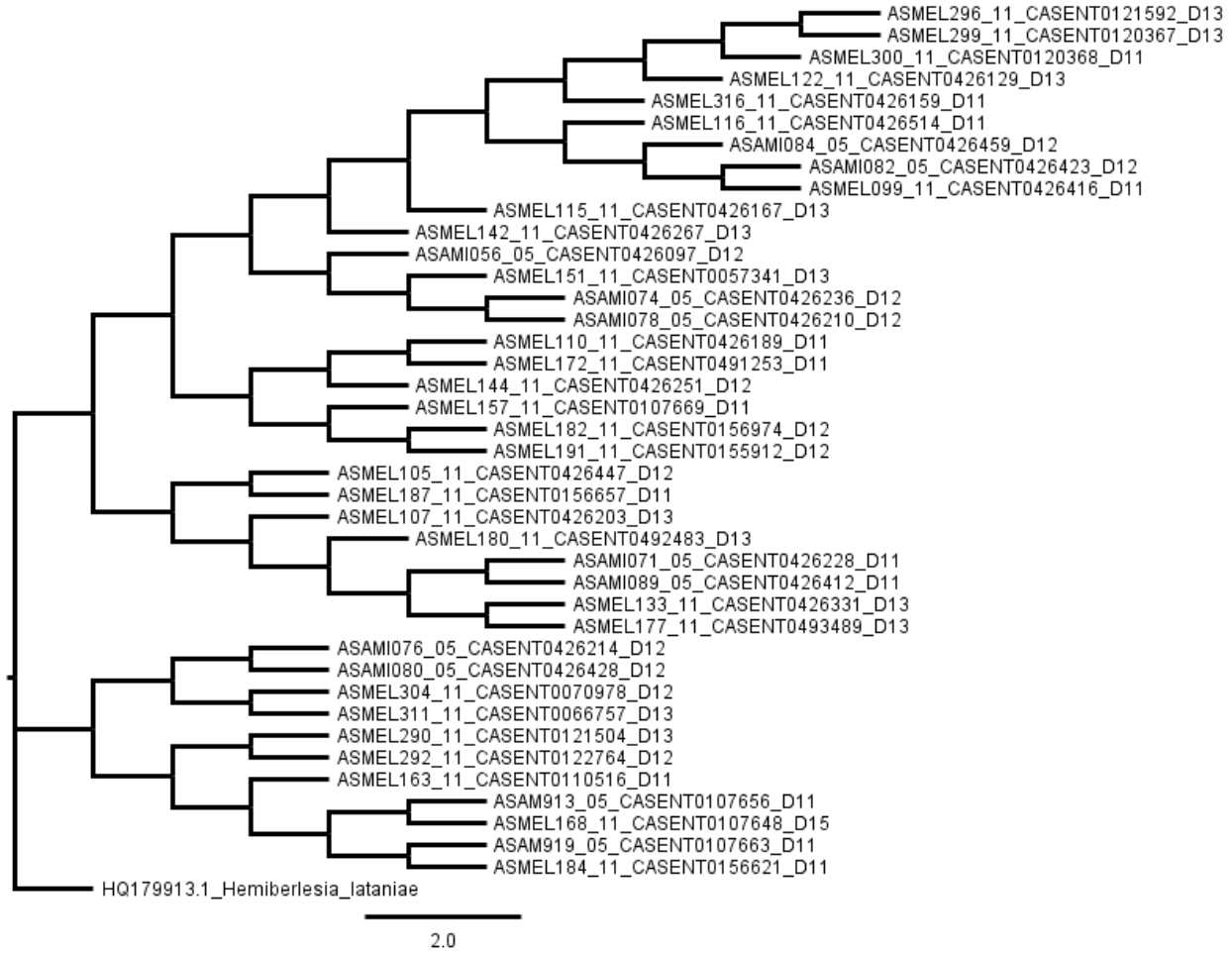
Appendix 5: Neighbor-joining, maximum parsimony and Bayesian phylogenetic trees for Diaspididae.

COI

Neighbor-joining

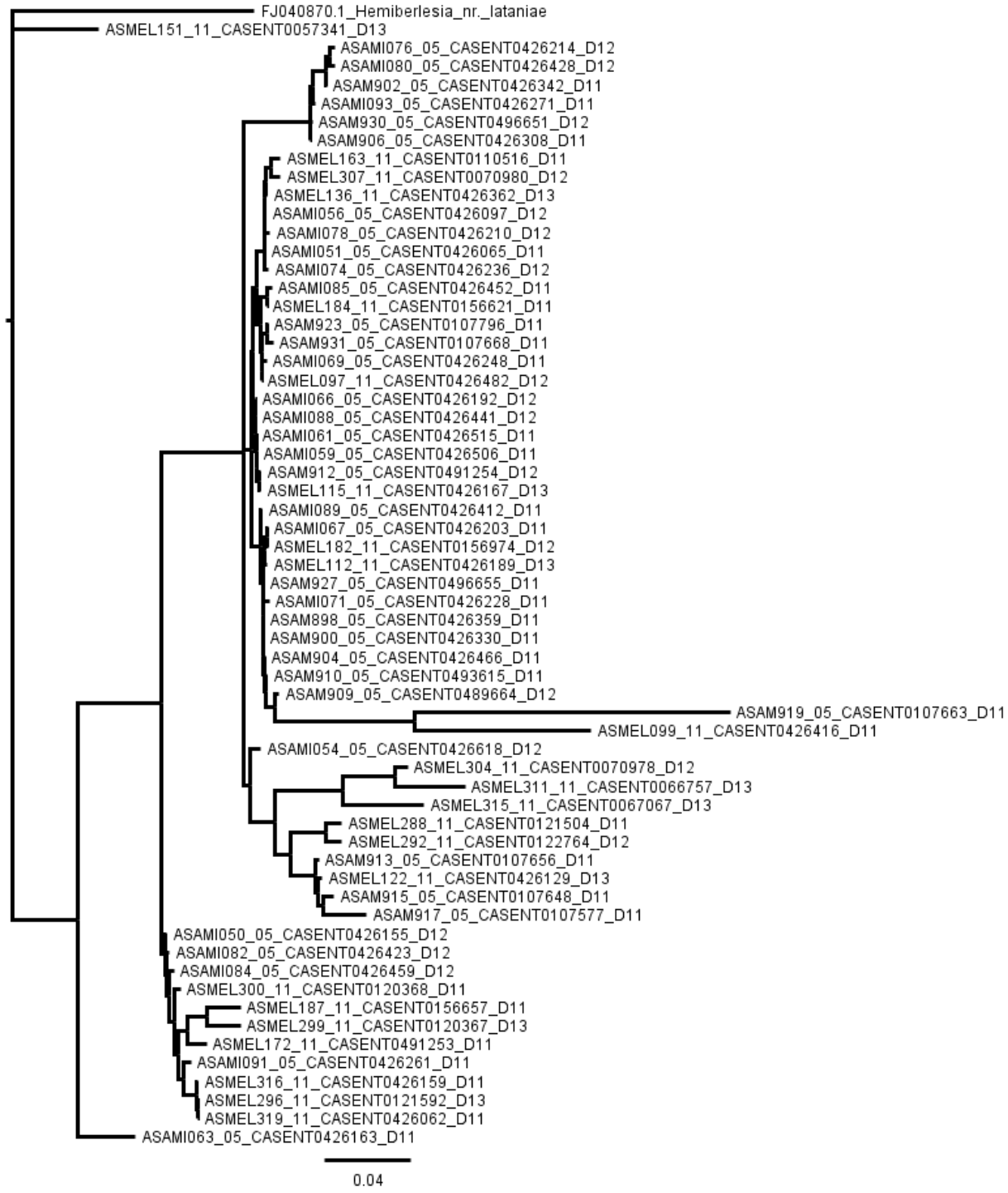


Maximum Parsimony

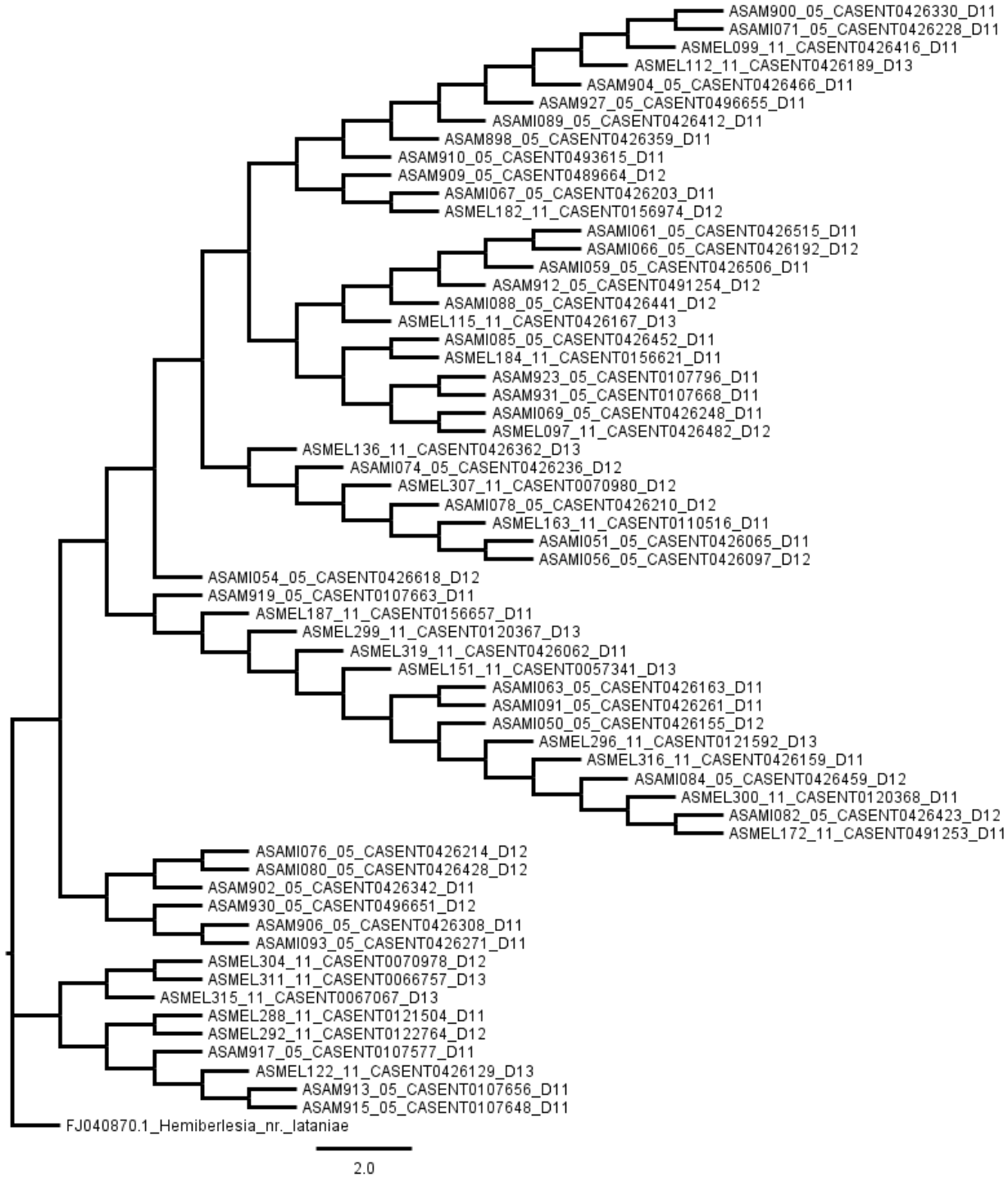


28S

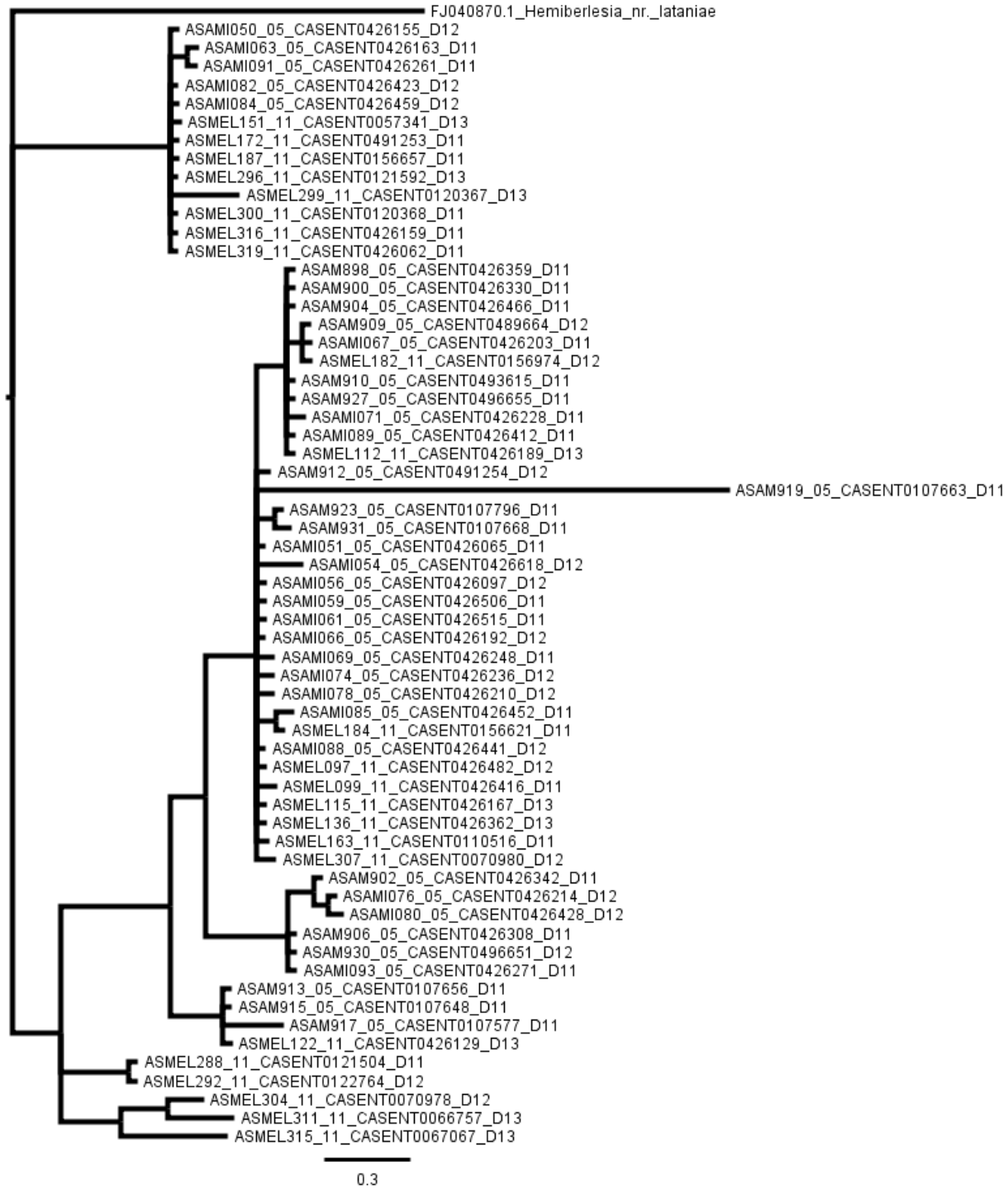
Neighbor-joining



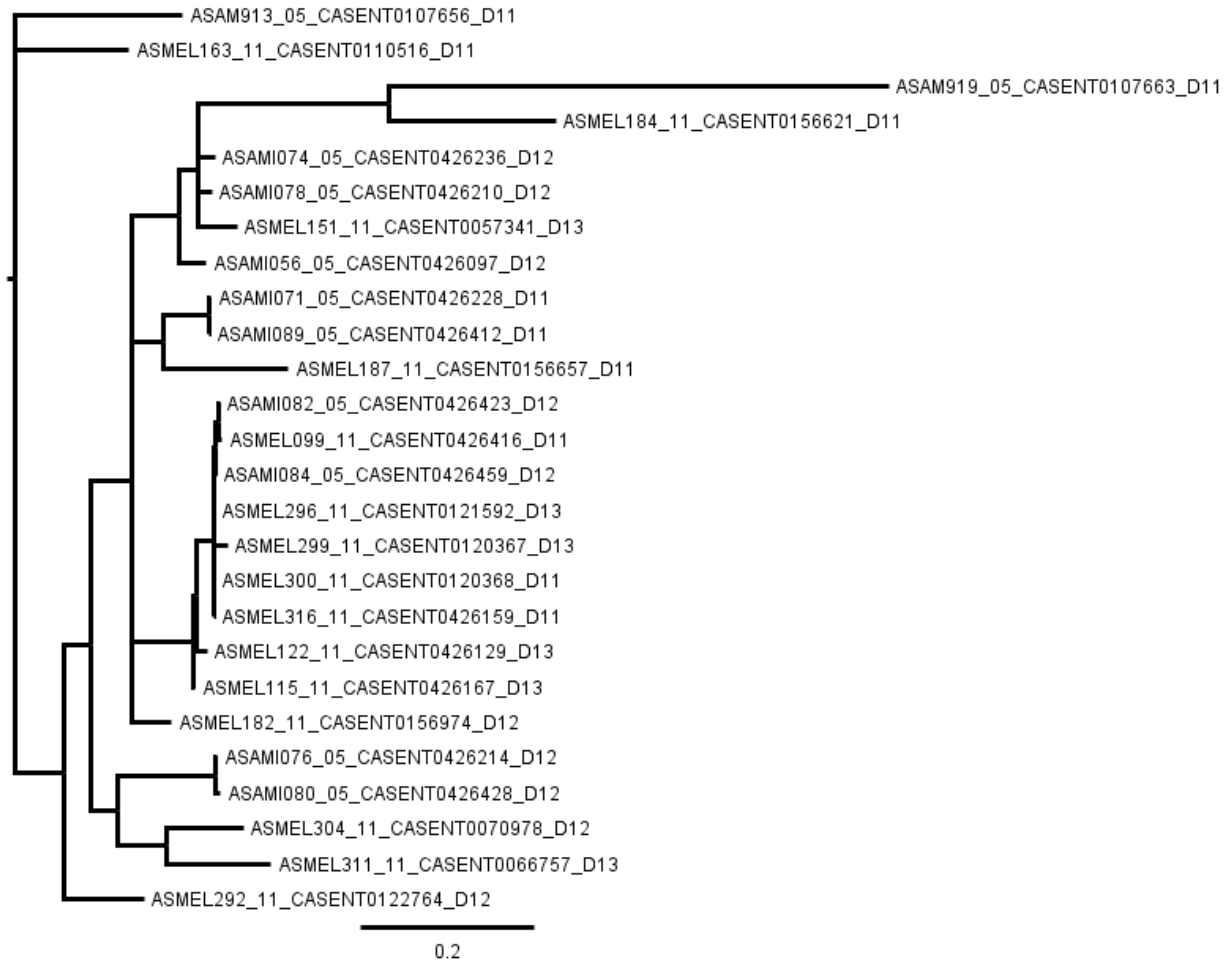
Maximum Parsimony



Bayesian

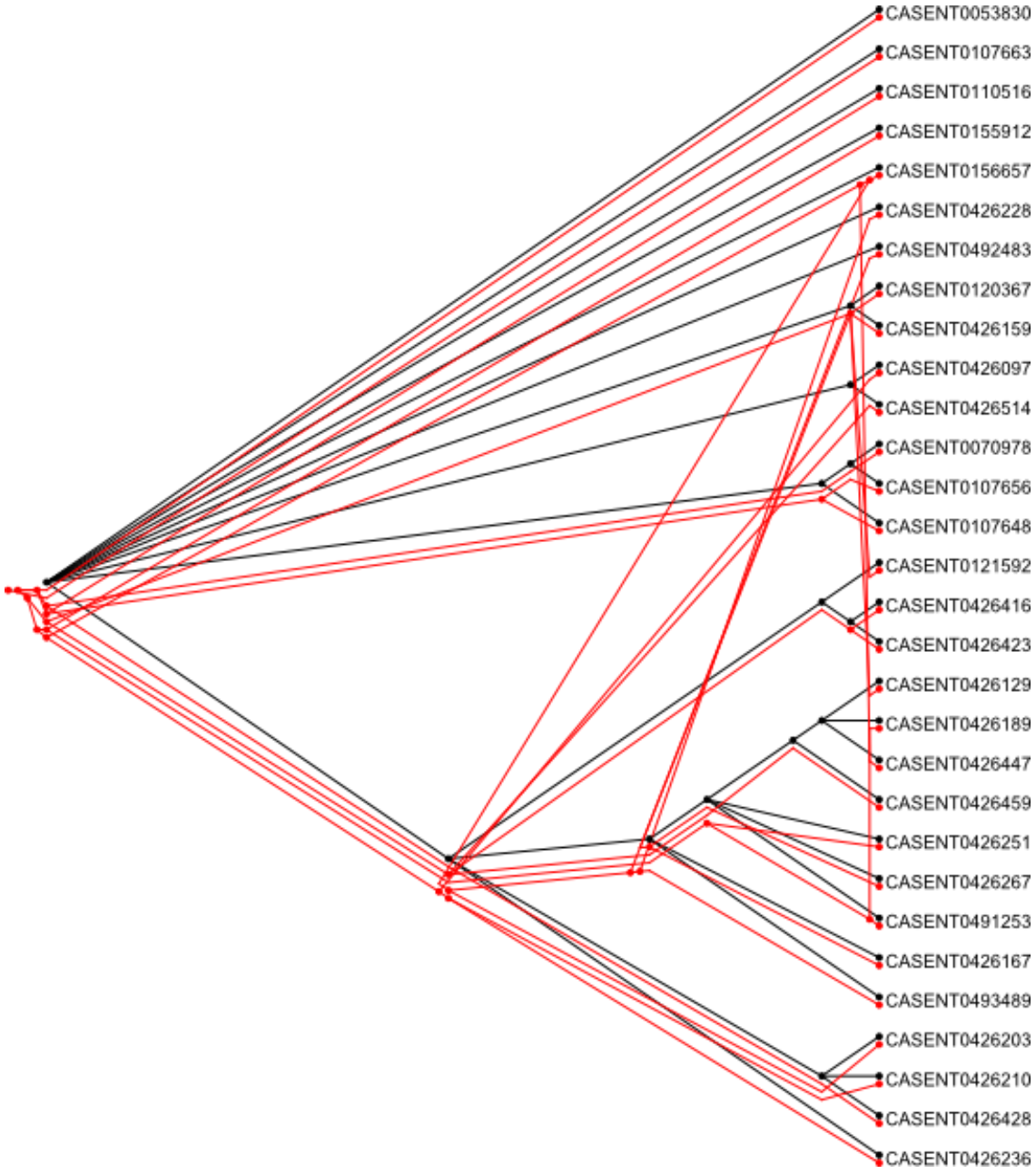

FJ040870.1_Hemiberlesia_nr._lataniae
ASAMI050_05_CASENT0426155_D12
ASAMI063_05_CASENT0426163_D11
ASAMI091_05_CASENT0426261_D11
ASAMI082_05_CASENT0426423_D12
ASAMI084_05_CASENT0426459_D12
ASMEL151_11_CASENT0057341_D13
ASMEL172_11_CASENT0491253_D11
ASMEL187_11_CASENT0156657_D11
ASMEL296_11_CASENT0121592_D13
ASMEL299_11_CASENT0120367_D13
ASMEL300_11_CASENT0120368_D11
ASMEL316_11_CASENT0426159_D11
ASMEL319_11_CASENT0426062_D11
ASAM898_05_CASENT0426359_D11
ASAM900_05_CASENT0426330_D11
ASAM904_05_CASENT0426466_D11
ASAM909_05_CASENT0489664_D12
ASAMI067_05_CASENT0426203_D11
ASMEL182_11_CASENT0156974_D12
ASAM910_05_CASENT0493615_D11
ASAM927_05_CASENT0496655_D11
ASAMI071_05_CASENT0426228_D11
ASAMI089_05_CASENT0426412_D11
ASMEL112_11_CASENT0426189_D13
ASAM912_05_CASENT0491254_D12
ASAM919_05_CASENT0107663_D11
ASAM923_05_CASENT0107796_D11
ASAM931_05_CASENT0107668_D11
ASAMI051_05_CASENT0426065_D11
ASAMI054_05_CASENT0426618_D12
ASAMI056_05_CASENT0426097_D12
ASAMI059_05_CASENT0426506_D11
ASAMI061_05_CASENT0426515_D11
ASAMI066_05_CASENT0426192_D12
ASAMI069_05_CASENT0426248_D11
ASAMI074_05_CASENT0426236_D12
ASAMI078_05_CASENT0426210_D12
ASAMI085_05_CASENT0426452_D11
ASMEL184_11_CASENT0156621_D11
ASAMI088_05_CASENT0426441_D12
ASMEL097_11_CASENT0426482_D12
ASMEL099_11_CASENT0426416_D11
ASMEL115_11_CASENT0426167_D13
ASMEL136_11_CASENT0426362_D13
ASMEL163_11_CASENT0110516_D11
ASMEL307_11_CASENT0070980_D12
ASAM902_05_CASENT0426342_D11
ASAMI076_05_CASENT0426214_D12
ASAMI080_05_CASENT0426428_D12
ASAM906_05_CASENT0426308_D11
ASAM930_05_CASENT0496651_D12
ASAMI093_05_CASENT0426271_D11
ASAM913_05_CASENT0107656_D11
ASAM915_05_CASENT0107648_D11
ASAM917_05_CASENT0107577_D11
ASMEL122_11_CASENT0426129_D13
ASMEL288_11_CASENT0121504_D11
ASMEL292_11_CASENT0122764_D12
ASMEL304_11_CASENT0070978_D12
ASMEL311_11_CASENT0066757_D13
ASMEL315_11_CASENT0067067_D13
0.3

Partition

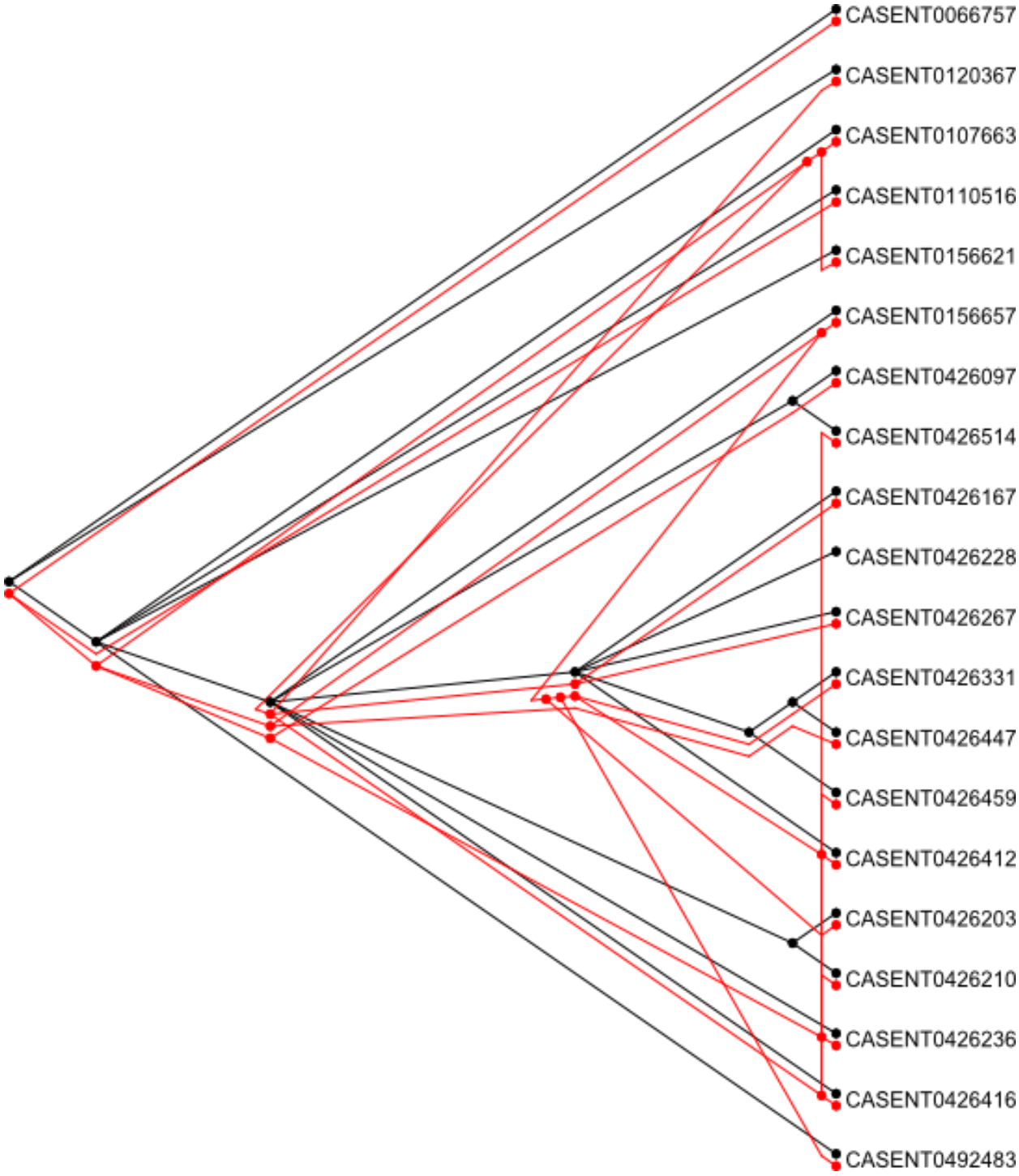

ASAM913_05_CASENT0107656_D11
ASMEL163_11_CASENT0110516_D11
ASAM919_05_CASENT0107663_D11
ASMEL184_11_CASENT0156621_D11
ASAMI074_05_CASENT0426236_D12
ASAMI078_05_CASENT0426210_D12
ASMEL151_11_CASENT0057341_D13
ASAMI056_05_CASENT0426097_D12
ASAMI071_05_CASENT0426228_D11
ASAMI089_05_CASENT0426412_D11
ASMEL187_11_CASENT0156657_D11
ASAMI082_05_CASENT0426423_D12
ASMEL099_11_CASENT0426416_D11
ASAMI084_05_CASENT0426459_D12
ASMEL296_11_CASENT0121592_D13
ASMEL299_11_CASENT0120367_D13
ASMEL300_11_CASENT0120368_D11
ASMEL316_11_CASENT0426159_D11
ASMEL122_11_CASENT0426129_D13
ASMEL115_11_CASENT0426167_D13
ASMEL182_11_CASENT0156974_D12
ASAMI076_05_CASENT0426214_D12
ASAMI080_05_CASENT0426428_D12
ASMEL304_11_CASENT0070978_D12
ASMEL311_11_CASENT0066757_D13
ASMEL292_11_CASENT0122764_D12
0.2

Appendix 6: Cophylogeny reconstruction of 25%, 50% and 75% reduced COI datasets for *Melissotarsus insularis* and Diaspididae.

25% Reduced Dataset



50% Reduced Dataset



75% Reduced Dataset