DOCTOR OF PHILOSOPHY

Advancing mitogenomics : a case study in the Araneae

Briscoe, Andrew

Award date:
2013

Awarding institution:
Bangor University

PRIFYSGOL
BANGOR
UNIVERSITY

Advancing mitogenomics: a case study in the Araneae

A thesis submitted to Bangor University for the degree of Doctor of Philosophy

Andrew G. Briscoe
April 2013
Molecular Ecology & Fisheries Genetics Laboratory
Bangor University
School of Biological Sciences
Environment Centre Wales
BANGOR
Gwynedd, LL57 2UW

Declaration and Consent

Details of the Work

I hereby agree to deposit the following item in the digital repository maintained by Bangor University and/or in any other repository authorized for use by Bangor University.

Author Name: ……..

Title: ……

Supervisor/Department:
……

Funding body (if any):
……

Qualification/Degree obtained: ……

This item is a product of my own research endeavours and is covered by the agreement below in which the item is referred to as "the Work". It is identical in content to that deposited in the Library, subject to point 4 below.

Non-exclusive Rights

Rights granted to the digital repository through this agreement are entirely non-exclusive. I am free to publish the Work in its present version or future versions elsewhere.

I agree that Bangor University may electronically store, copy or translate the Work to any approved medium or format for the purpose of future preservation and accessibility. Bangor University is not under any obligation to reproduce or display the Work in the same formats or resolutions in which it was originally deposited.

Bangor University Digital Repository

I understand that work deposited in the digital repository will be accessible to a wide variety of people and institutions, including automated agents and search engines via the World Wide Web.

I understand that once the Work is deposited, the item and its metadata may be incorporated into public access catalogues or services, national databases of electronic theses and dissertations such as the British Library's EThOS or any service provided by the National Library of Wales.

I understand that the Work may be made available via the National Library of Wales Online Electronic Theses Service under the declared terms and conditions of use (http://www.llgc.org.uk/index.php?id=4676). I agree that as part of this service the National Library of Wales may electronically store, copy or convert the Work to any approved medium or format for the purpose of future preservation and accessibility. The National Library of Wales is not under any obligation to reproduce or display the Work in the same formats or resolutions in which it was originally deposited.

Statement 1:

This work has not previously been accepted in substance for any degree and is not being concurrently submitted in candidature for any degree unless as agreed by the University for approved dual awards.

Signed …….. (candidate)

Date ……..

Statement 2:

This thesis is the result of my own investigations, except where otherwise stated. Where correction services have been used, the extent and nature of the correction is clearly marked in a footnote(s).

All other sources are acknowledged by footnotes and/or a bibliography.

Signed ……. (candidate)

Date …….

Statement 3:

I hereby give consent for my thesis, if accepted, to be available for photocopying, for inter-library loan and for electronic storage (subject to any constraints as defined in statement 4), and for the title and summary to be made available to outside organisations.

Signed ……. (candidate)

Date …….

NB: Candidates on whose behalf a bar on access has been approved by the Academic Registry should use the following version of **Statement 3:**

Statement 3 (bar):

I hereby give consent for my thesis, if accepted, to be available for photocopying, for inter-library loans and for electronic storage (subject to any constraints as defined in statement 4), after expiry of a bar on access.

Signed …… (candidate)

Date ……

Statement 4:

Choose **one** of the following options

a)	I agree to deposit an electronic copy of my thesis (the Work) in the Bangor University (BU) Institutional Digital Repository, the British Library ETHOS system, and/or in any other repository authorized for use by Bangor University and where necessary have gained the required permissions for the use of third party material.	
b)	I agree to deposit an electronic copy of my thesis (the Work) in the Bangor University (BU) Institutional Digital Repository, the British Library ETHOS system, and/or in any other repository authorized for use by Bangor University when the approved **bar on access** has been lifted.	
c)	I agree to submit my thesis (the Work) electronically via Bangor University's e-submission system, however I **opt-out** of the electronic deposit to the Bangor University (BU) Institutional Digital Repository, the British Library ETHOS system, and/or in any other repository authorized for use by Bangor University, due to lack of permissions for use of third party material.	

Options B should only be used if a bar on access has been approved by the University.

In addition to the above I also agree to the following:

1. That I am the author or have the authority of the author(s) to make this agreement and do hereby give Bangor University the right to make available the Work in the way described above.
2. That the electronic copy of the Work deposited in the digital repository and covered by this agreement, is identical in content to the paper copy of the Work deposited in the Bangor University Library, subject to point 4 below.
3. That I have exercised reasonable care to ensure that the Work is original and, to the best of my knowledge, does not breach any laws – including those relating to defamation, libel and copyright.
4. That I have, in instances where the intellectual property of other authors or copyright holders is included in the Work, and where appropriate, gained explicit permission for the inclusion of that material in the Work, and in the electronic form of the Work as accessed through the open access digital repository, *or* that I have identified and removed that material for which adequate and appropriate permission has not been obtained and which will be inaccessible via the digital repository.
5. That Bangor University does not hold any obligation to take legal action on behalf of the Depositor, or other rights holders, in the event of a breach of intellectual property rights, or any other right, in the material deposited.
6. That I will indemnify and keep indemnified Bangor University and the National Library of Wales from and against any loss, liability, claim or damage, including without limitation any related legal fees and court costs (on a full indemnity bases), related to any breach by myself of any term of this agreement.

Signature: …… Date : …….

To a wonderful person and an even better mother, without whom none of this would be possible.....

..... and to those who have provided their support along the way,

Thank you!

Abstract

Advances in sequencing technologies over the last decade have led to the rapid production of large volumes of sequencing data for phylogenetic analyses. True phylogenomic studies are now possible, although remain constrained by both labour and cost. However, the availability of mitochondrial genome sequences is growing as a result of increased access to Next-Generation Sequencing platforms. The phylogenetic analysis of complete mitochondrial genomes is becoming the marker of choice within some taxa and the comparative analysis of entire mitogenome sequences has revealed new informative characters. Despite these technological improvements, assembling longer sequences, such as that of entire mitochondrial genomes, has not been straightforward. Very few large-scale studies exist, and have been primarily limited to nominally intra-specific datasets.

This study explores the potential of amplifying, sequencing and assembling mitochondrial genomes from a diverse order; Araneae (spiders). To this end I have amplified spider mitogenomes via more traditional polymerase chain reaction (PCR) strategies and whole genome amplification (WGA) steps, in order to assess their utility in resolving phylogenetic in-congruence in hitherto understudied taxa.

Whilst complete mitochondrial genomes were successfully sequenced, this research highlights the obstacles to the rapid amplification of large divergent mitogenomic datasets. However, the data generated from the study has been critical in elucidating the evolutionary history of the Araneae, the generation of robust phylogenetic hypotheses, and in dating the major radiations of extant spiders. To my knowledge this study represents the most complete comparative mitogenomic analysis within the Araneae to date, and provides a framework from which subsequent analyses can further delineate the evolutionary history of the order.

Table of contents

List of figures

List of tables

List of abbreviations

A	Adenine
A-T	Adenine and thymine
BP	Base pairs
C	Cytosine
COI	Cytochrome oxidase subunit I
DMSO	Dimethyl sulphoxide
DNA	Dioxyribonucleic acid
dNTP	Deoxynucleotide triphosphate
EDTA	Ethylenediaminetetraacetic acid
emPCR	Emulsion PCR
EST	Expressed sequence tag
G	Guanine
G-C	Guanine and cytosine
Kb	Kilobases
MgCl2	Magnesium chloride
MID	Molecular identifier tags
MPS	Massively parallel sequencing
NaCl	Sodium chloride
NaOH	Sodium hydroxide
NGS	Next-Generation Sequencing
PCR	Polymerase chain reaction
RCA	Rolling circle amplification
RED	Restriction enzyme digest
RNA	Ribonucleic acid

rRNA	Ribosomal RNA
tRNA	Transfer RNA

List of publications, posters and communications

Publications:

Briscoe, A.G., Goodacre, S., Masta, S., Taylor, M.I., Arnedo, M.A., Penney, D., Kenny, J., Creer, S., 2013. Can long-range PCR be used to amplify genetically divergent mitochondrial genomes for comparative phylogenetics? A case study within spiders (Arthropoda: Araneae). PLOS ONE (doi: 10.1371/journal.pone.0062404).

Posters:

(2011) Systematics Association 11th Biennial Meeting, Belfast, UK. Poster presentation: Towards Large-scale Second-generation Sequencing of Full Mitochondrial Genomes in the Araneae.

(2009) CoSyst Review, Linnean Society of London, UK. Poster Presentation: Using Mitogenomics to Elucidate a Higher Level Spider Phylogeny.

(2008) 10th Young Systematist's Forum, Natural History Museum, London, UK.

Poster presentation: Using Mitogenomics to Elucidate a Higher Level Spider Phylogeny

Communications:

(2011) 13th Young Systematist's Forum, Natural History Museum, London, UK.

Oral presentation: Promises and pitfalls of the amplification of whole mitochondrial genomes: A case study within the Araneae.

(2009) RSF Beyond Boundaries Conference, Bangor University, UK.

Oral Presentation: Using Mitogenomics to Elucidate a Higher Level Spider Phylogeny.

(2007) Conference on Biomedical Egyptology, Cairo, Egypt. Oral Presentation: A Phylogenetic Study of the Sacred Scarab (Scarabaeus sacer).

1. General introduction

1.1. Arthropod diversity

The Araneae are among the oldest and most diverse groups of terrestrial organisms with the earliest fossils (and most Arachnid allies) dating from 380 million year old Devonian deposits (Shear et al., 1989). Current diversity is comprised of over 43,000 described species placed in more than 3,800 genera spanning 112 families (Platnick, 2012). Spiders are an unequivocally ecologically important guild, being the dominant predators of insects in natural and managed ecosystems. Furthermore, spiders are model organisms used in biochemical, behavioural, ecological, comparative development and speciation research (Blackledge et al., 2009; Foelix, 1996; Oxford and Gunnarsson, 2006; Scott et al., 2006).

Amongst other similarly sized orders, only the five largest insect groups (Coleoptera, Lepidoptera, Hymenoptera, Diptera, and Heteroptera) and the mites (Acari) are more speciose than spiders (Araneae). Recent estimates of the world's total spider diversity range from 76,000 (Platnick, 1999) to 170,000 (Coddington and Levi, 1991) indicating that between 20-50% of the world's potential total spider species have already been described and classified. This contrasts favourably with other non-vertebrate taxa; the 12,000 known species of millipedes, for example, are thought to represent approximately 10% of the actual total diversity (Hoffman et al., 2002; Sierwald and Bond, 2007), and the figure for mites is believed to be much lower (Maraun et al., 2007; Walter and Proctor, 1999).

1.2. Araneae systematics

The phylogenetic study of spiders has so far been limited and, prior to the turn of the century, phylogenetic analyses have been based almost entirely on morphological characters supplemented with behavioural information

(Coddington and Levi, 1991; Griswold et al., 1998; Griswold et al., 1999). The comparatively insignificant amount of genomic work to date on spiders has been uncoordinated and of little utility for broad-scale phylogenetic investigation. However, the advent of high-throughput DNA sequencing has made it feasible to examine substantial parts of the genome across a dense sampling of taxa (Philippe and Telford, 2006).

Prior to renewed efforts led by the American Museum of Natural History in resolving the phylogeny of the Araneae, evolutionary biologists have referred to a benchmark order-level phylogenetic framework, derived from morphological data proposed by the compilation of data from many individual studies (Coddington et al., 2004) (Figure. 1.1.). This framework has subsequently been exposed as inaccurate, as new data from more recent molecular studies has led to its continued alteration (Blackledge et al., 2009; Bond, 2012) (Figure. 1.2.) There is therefore a clear need for a comprehensive study incorporating enough distinguishing characters to provide a definitive and robust phylogeny (Jocque and Dippenaar-Schoeman, 2006).

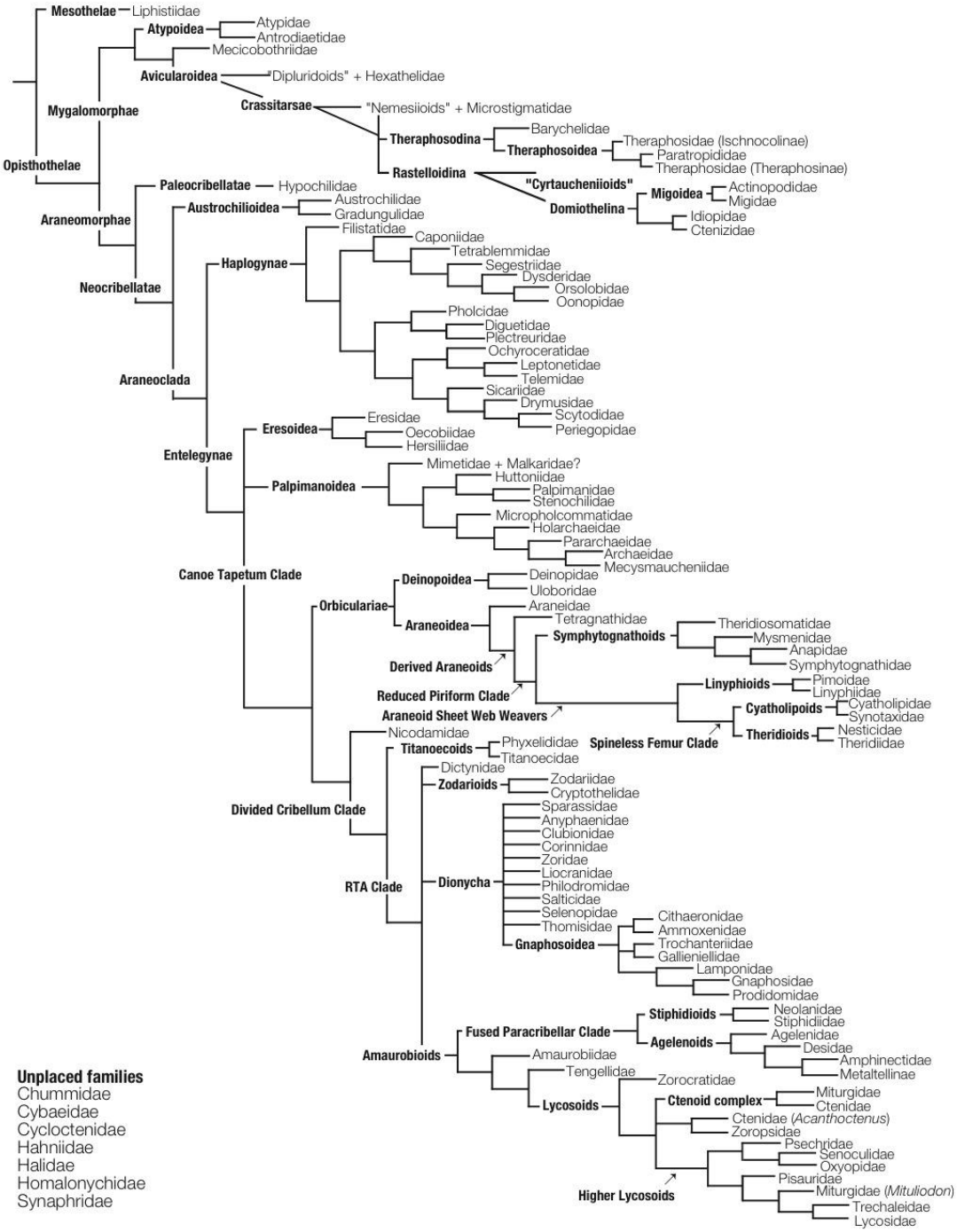


Figure 2.1. Cladistic hypotheses for all Araneomorph spider families using behavioural and morphological characters. From Coddington, J. A. (2004).

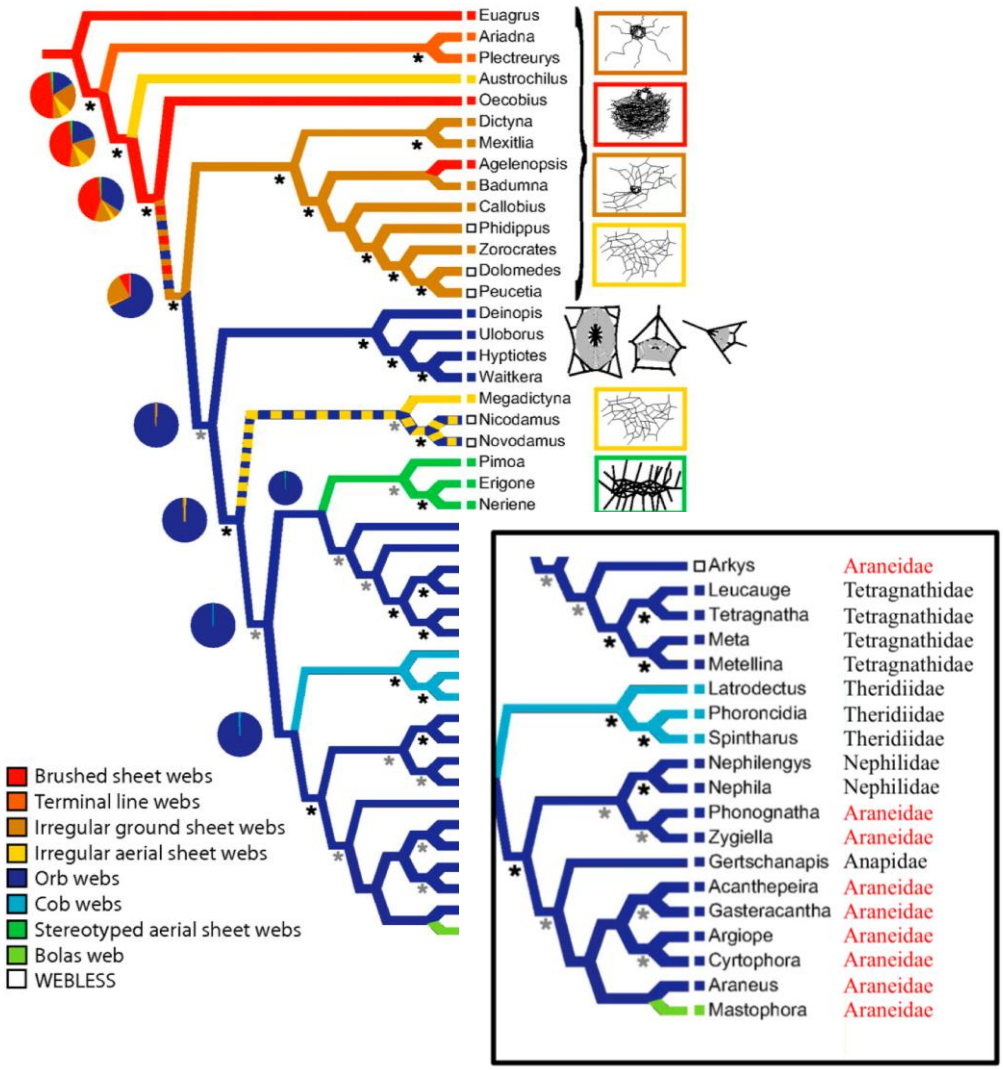


Figure 1.2. Phylogenetic framework for web building spiders from behavioural and molecular characters from Blackledge et al. (2009). The magnified section of the tree (black box) is annotated to show the families each genus is currently placed in. The family Araneidae (Orb-weavers) are in red text. The interspersal of four other families (Anapidae, Nephilidae, Tetragnathidae and Theridiidae) amongst the Araneidae highlights the issues with current taxonomic placements.

In spite of the increased availability of high-throughput technologies, comparatively few studies in the Araneae have moved away from utilising morphological and behavioural characters for resolving phylogenies (Coddington

et al., 2012; Moradmand and Jaeger, 2012) and none are yet to take advantage of Next-Generation Sequencers. The majority of studies that have incorporated molecular data are primarily focused on the Cytochrome Oxidase I (COI) bar-coding region and single gene markers that are frequently used to differentiate cryptic species, and resolve phylogenies at the genus and familial levels (Arnedo et al., 2004; Astrin et al., 2006; Barrett and Hebert, 2005; Wood et al., 2012). Genes from both the mitochondrial and nuclear genomes, that have been used in studies to resolve phylogenetic placement within the Aranaeae include, mitochondrial (mt) Cytochrome Oxidase I (COI) (Arnedo et al., 2004; Blackledge et al., 2009; Bruvo-Madaric et al., 2005; Hedin and Maddison, 2001) Actin-5C (Vink et al., 2008), mt12S Ribosomal RNA (Bruvo-Madaric et al., 2005; Fang et al., 2000; Park et al., 2007), mt16S ribosomal RNA (Arnedo et al., 2004; Bruvo-Madaric et al., 2005; Fang et al., 2000; Hedin and Maddison, 2001), nuclear 18S ribosomal RNA (Arnedo et al., 2004; Blackledge et al., 2009), nuclear 28S ribosomal RNA (Arnedo et al., 2004; Blackledge et al., 2009; Bruvo-Madaric et al., 2005; Hedin and Maddison, 2001), NADH dehydrogenase 1 (ND1) (Hedin and Maddison, 2001) and Histone H3 (Arnedo et al., 2004; Blackledge et al., 2009). However, most of these studies have been narrow in taxon sampling and have focused at the family and sub-family levels. The resulting lack of available molecular data has meant that very little is known about the process and patterns of molecular evolution within the Araneae (Masta, 2000a).

1.3. Arachnid mitochondrial genomes

The arachnid mitochondrial genome, similar to most animal mitochondrial genomes, is a small extra-chromosomal genome comprising 37 genes including 22 coding for transfer RNAs (tRNAs), 13 coding for proteins and 2 coding for

ribosomal RNAs (rRNAs) (Boore, 1999). As well as the 37 coding genes mitochondrial genomes also contain a large non-coding sequence that has been named the control region, due to its role in controlling the transcription and replication of the mtDNA molecule (Mardulyn et al., 2003). Within arachnids and other invertebrates the control region is also known as the 'A + T' rich region as it exhibits a higher frequency of deoxyadenylate and deoxythymidylate residues than any other section within the mitochondrial DNA molecule. All animal mitochondrial genomes, including those of the Araneae are typically closed-circular molecules with the exception of three Cnidarian classes (Cubozoa, Scyphozoa & Hydrozoa) that have linear mtDNA chromosomes (Boore, 1999).

Prior to this study, there have been six complete Araneae mitochondrial genomes published within the National Centre for Biotechnology's (NCBI) GenBank database (GenBank accession numbers: AY309258, AY309259, AY452691, NC_005942, NC_010777, NC010780). The first whole mitochondrial genome of a spider to be sequenced was that of the jumping spider *Habronattus oregonensis* (family Salticidae), which provided novel insights on the structure and arrangement of genes within the mitochondrial DNA of spiders (Masta and Boore, 2004). As well as differing in gene order from the most basal member of the Chelicerata, the horseshoe crab *Limulus polyphemus* (GenBank accession number JX983598.), the secondary structure of tRNAs were found to lack the distinctive canonical cloverleaf shape typically associated with their function. Truncations of tRNA coding sequences were found to be the smallest sequences recorded, even smaller than previously recorded nematode tRNAs (Ohtsuki et al., 2002; Okimoto and Wolstenholme, 1990), and highlighted the lack of understanding of the evolution of the Araneae, especially when contrasted with other arachnid orders.

The complete spider mitochondrial genomes that have been sequenced so far range in size from 13,874 bp to 14,436 bp in length (Lee et al, unpublished (Qiu et al., 2005)). Much larger mitochondrial genomes have been found in other classes, but these typically contain large non-coding regions and/or duplications of large portions of the genes present and thus do not exhibit variation in gene content (Denovanwright and Lee, 1992; Lilly and Havey, 2001; Yokobori et al., 2004). The complete spider mitochondrial genomes that have been sequenced so far have highlighted the potential of complete mtDNA sequences in resolving complex intra-order relations by opening the door to new avenues of bioinformatic research from which a more accurate and robust phylogeny can be created.

The sequenced Araneae mitochondrial genomes have led to the realisation of two potential tools that can be used, in conjunction with traditional methods, for constructing more accurate phylogenies. Differences within the Araneae in both gene order and configuration of the secondary structure of tRNAs can be used to test any phylogenetic reconstruction undertaken. Mitochondrial gene order was considered to be well conserved amongst the Class Arachnida, however, studies have since shown that this not the case (Black and Roehrdanz, 1998). When compared to the American horseshoe crab (*Limulus polyphemus*), believed to be the most basal member of the Arachnida, seven translocations of tRNA genes are apparent (Masta, 2000b; Masta and Boore, 2004). However, comparisons between the currently sequenced spider genomes have shown that these translocations have occurred as spiders diversified, and those more basal genomes do not exhibit any translocations.

Several mechanisms have been proposed for the occurrence of gene translocations within the mitochondrial genome, predominantly focused around tandem duplication events (Macey et al., 1997). Through this mechanism gene

duplications occur as a result of slipped strand mispairing followed by the subsequent deletion of one of the genes (Moritz, Dowling and Brown, 1986; Moritz and Brown, 1987). Other instruments of gene order changes known to occur include inversion-based mechanisms (Griffiths et al., 2000), whereby the coding sequences of genes change strands, and transposition, involving the movement of defined elements to new locations with the genome (McClintock, 1950). However, neither of these two mechanisms are likely to be involved in the differences exhibited between spider mitochondrial genomes as there are no currently observed reversals of coding polarity and spider mtDNA sequences lack the terminal repeats characteristic of transposable elements. More commonly, tRNA genes can alter their position via translocation and remoulding (Cantatore et al., 1987; Littlewood & Webster, 2012; Wang & Lavrov, 2011) where tRNA genes are duplicated through the process of DNA replication.

1.4. Phylogenetics

The advent of DNA sequencing, in the latter part of the twentieth century, soon led to its use in the reconstruction of complex phylogenies that could not be resolved solely using morphological and behavioural characters (Fitch, 1982; Wake, 1981, 1982). However, prior to the turn of the century, phylogenetic analyses have often relied solely on single nucleotide substitutions and insertions or deletions as indicators of divergence or common ancestry (Felsenstein, 1988; Hillis et al., 1994). Many factors have since been found to cloud the phylogenetic signal from such characters, including convergence, non-independent substitutions, functional constraints and saturation of mutations at variable sites making them unreliable for such studies (Rokas, 2000).

Phylogenetics follows directly from the theory of evolution and is a prerequisite to any evolutionary study. Contemporary phylogenetics relies heavily on mathematical models to identify homologous characteristics shared between organisms and to infer relationships based on that data (Goodman et al., 1974; Huelsenbeck and Crandall, 1997; Huelsenbeck et al., 2001; Lewis, 1998). The accuracy of phylogenetic reconstructions is strongly influenced by the quality of the models used to illustrate/explain the evolution of such characters (Holder and Lewis, 2003). Reconstructing evolutionary history based on living organisms alone has proved difficult (Quental and Marshall, 2010). Until the 1970's when advances in molecular sciences first allowed genetic comparisons, phylogeny was almost solely based on morphological and behavioural characters. This has worked well in respect to some studies but has proved difficult to establish relationships between organisms with few reliable homologous characteristics (Van Niel, 1955).

Access to molecular data has given evolutionary biologists better provision to a greater number of reliable homologous characters and was key in identifying the Archaea as the third major domain of life, in addition to Prokaryotes and Eukaryotes (Woese and Fox, 1977). Previous approaches to phylogenetics, including analysis of morphological characters (Sereno, 2008), analysis of fossil records (again morphological) (Smith, 1994) and comparisons of partial ribosomal RNA sequences (Mindell and Honeycutt, 1990), had resulted in the production of several conflicting topologies and were hindered by alignment ambiguities, artefactual associations of rapidly evolving taxa and substitutions within the ribosomal sequences studied (Delsuc et al., 2005). As more and more genes were analysed, topological conflicts arose (due in part to homoplasy and convergence) and single genes were seen as not containing enough information to form statistically sound support (Gee, 2003; Soltis et al., 1999).

1.5. Mitogenomics

In order to overcome the bias associated with tree assembly, ever-increasing numbers of genetic characters have been used in phylogenetic analyses, and coupled with advances in DNA extraction and sequencing technology has led to the phylogenetic study of whole genomes in a field termed phylogenomics (Delsuc et al., 2005). Phylogenomics uses existing phylogenetic principles to make sense of genomic data (Philippe et al., 2009). This method is helping to resolve complex relationships by expanding the number of characters that can be used for phylogenetic analysis (Wiens et al., 2008). This research has changed the focus of phylogenetic study from searching for informative characters to the development of improved reconstruction methods for the use of genomic data (Sanchez et al., 2011; Tarraga et al., 2007). Models that existed prior to this had failed to fully take in to account molecular evolutionary processes, including the mechanisms of evolution and the time scales involved. Current methods of using whole genomes to infer phylogeny focus on gene content and gene order coupled with the use of more traditional sequence-based methods (Boore and Brown, 1998).

Due to the size and associated sequencing costs of the nuclear genome, it has been preferable to analyse the genomes of intracellular organelles instead of an organism's complete nuclear DNA (Curole and Kocher, 1999; Miya and Nishida, 2000; Shaw et al., 2007). The use of intracellular organelles, such as mitochondria, for phylogenetic analyses offers many advantages over nuclear DNA (Simon et al., 1994). Benefits include the comparative ease of extracting homologous sequences, maternal inheritance (in most taxa) and lack of recombination, making the mitochondrial genome an increasingly popular marker in genetic studies of animals (Harrison, 1989). Many studies have now been undertaken to use whole mitochondrial genomes to resolve phylogenetic relationships (Morin et al., 2010;

Prosdocimi et al., 2012; Timmermans et al., 2010). These studies have highlighted the importance and potential benefits of such work, however, the nature of evolution and the limitations of current phylogenetic methods mean that parts of the tree of life may be difficult to resolve with confidence (Cameron, 2007).

Arthropods provide the second most studied phylum for mitochondrial DNA (after chordates). They illustrate perfectly the use of mitogenomics for phylogenetic/evolutionary biology purposes. A single translocation (within the genome) of one gene coding for one of 22 tRNAs has been used to show the close relationship of insects and crustaceans to the exclusion of myriapods, chelicerates, tardigrades and onychophorans (Boore et al., 1998) (Figure 1.3.). Gene arrangements are highly conserved within the arthropods with most variation occurring at the areas of large non-coding regions or the tRNAs.

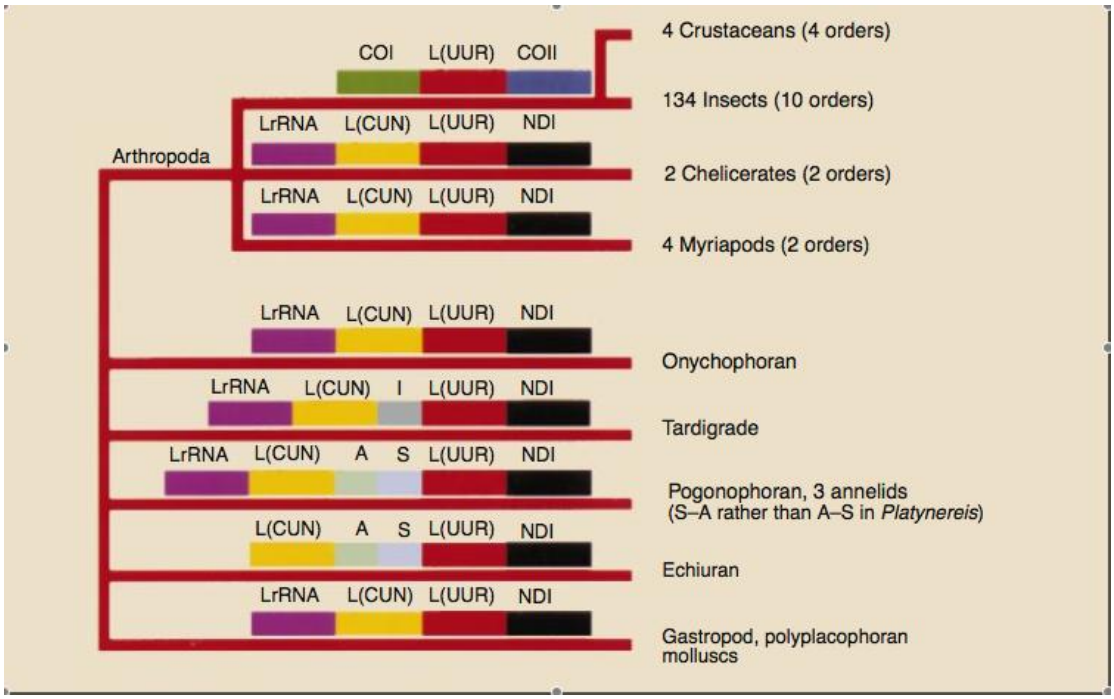


Figure 1.3. Phylogenetic reconstruction of Arthropoda and sister taxa at the sub-phylum level taken from Boore et al (1998). A single tRNA translocation

has been used to infer the close relationship of the phyla Crustacea and Insecta, to the exclusion of Chelicerata and Myriapoda.

The increase in whole mitochondrial genome datasets is providing new characters from which phylogenies can be better supported. These tools, known as rare genomic changes (RGCs) are becoming increasingly popular for resolving complex phylogenetic relationships where traditional methods have produced ambiguous, unresolved results (Rokas, 2000). RGCs occur much less frequently than base substitutions and are defined as large-scale mutational changes that have occurred within particular clades. Examples of rare genomic changes are intron indels, retroposon integrations, signature sequences and changes in organelle gene order, gene duplications and genetic code variants.

Gene arrangements and translocations have been found to be the most common and important RGCs in providing support to phylogenetic data (Bleidorn et al., 2006; Moret and Warnow, 2005). Statistically, it is highly unlikely that the occurrence of the 37 mitochondrial genes in the same order would appear independently across taxa. Shared-derived arrangements are most likely to be due to common ancestry and not convergent evolution or chance. The number of gene arrangements that have been found within whole mitochondrial genomes that have been published so far, suggest that gene arrangements are not selectively constrained, but those that do not lethally disrupt mitochondrial genome function are rare (Boore, 1995). Not all divergences will coincide with mtDNA rearrangements and as a result, the resolving power of strictly looking at gene order, within a genome, is weak (Luo, 2009; Boore, 1995). Subsequent independent changes in lineages may remove any trace of relatedness that was previously present, however, due to the low chance of sequence orders being independently repeated, relationships resolved this way are likely to be correct.

The secondary structure of tRNA molecules is another important feature that has been shown to vary among some metazoans (Dirheimer et al., 1995). In almost all Eukaryota and Prokaryota the secondary structure of tRNA is a highly conserved cloverleaf shape (with three arms), but truncated, non-cloverleaf configurations have been recorded in some animal phyla, including land snails, nematodes and now spiders (Masta, 2000a; Masta and Boore, 2004; Okimoto and Wolstenholme, 1990; Yokobori and Paabo, 1995). It has been hypothesized that such truncations were due to pressures on the mitochondrial genome to minimize its size (Yamazaki, 1997). In some metazoans, the T-arm has been lost completely and is often found in conjunction with the loss of the D-arm and its replacement with a new secondary structure (D-loop). Due to the widespread form of the cloverleaf secondary structure of the mitochondrial tRNAs it has been argued that truncations are actually deleterious and consequently, deleterious mutations may be accumulating in the tRNAs of the affected organisms. From a genomic perspective, the accumulation of deleterious mutations in large groups of metazoans appears counterintuitive (Masta and Boore, 2004, 2008). However, posttranscriptional modification has been suggested as a mechanism whereby deleterious mutations are corrected via RNA editing (Borner et al., 1997).

1.6. Second generation sequencing

The advent of Next-Generation Sequencing (NGS) has dramatically increased the volume of sequencing conducted by the scientific community and expanded the range of problems that can be addressed by the direct readouts of DNA sequence (Rothberg and Leamon, 2008). Key breakthroughs in the development of NGS sequencing platforms included higher throughput, simplified *in vitro* sample preparation and the miniaturization of sequencing chemistries. It has been this

miniaturization that has enabled massively parallel sequencing reactions to be carried out at a scale and cost that has not been previously possible (von Bubnoff, 2008). Since its conception, NGS sequencing has led to a better understanding of the structure of the human genome, led to the first non-Sanger sequence of an individual human and opened up new approaches to identify small RNAs (Rothberg and Leamon, 2008; Wheeler et al., 2008).

There has been a desire to increase the throughput of DNA sequencing ever since advent of chain termination or fragmentation approaches coupled with electrophoretic size separation (Sanger, 1981; Sanger and Coulson, 1978; Sanger et al., 1992). To date, the demand for sequencing throughput has mirrored that for computing power, with increases in sequencing capacity being quickly absorbed by ever-larger experiments. Several novel approaches were explored to replace Sanger as the dominant provider of sequencing technologies, including mass spectrometry resolution (Tang et al., 1994), direct imaging of DNA sequence by atomic force microscopy (Allison et al., 1996), several large efforts to apply microfluidics to sequencing (Paegel et al., 2003), a number of different approaches for sequencing by synthesis (Seo et al., 2005) and sequencing by hybridization (Gresham et al., 2008). However, three more improvements allowed sequencing to meet the challenges of the first human genome project: the use of fluorescent tags instead of radioactive labels to detect the terminated ladders; the use of capillary electrophoresis in place of slab gels; and the development of paired-end sequencing protocols incorporating hierarchical template sizes to provide sequence context and orientation beyond the constraints of the actual sequence read-length (Rothberg and Leamon, 2008; Shendure and Ji, 2008).

So far NGS sequencing platforms have been used to solve problems in many biological fields including human genetics, metagenomics, ecology, evolution and

paleobiology. Roche's 454 sequencing was the first technology other than Sanger's to sequence and assemble bacterial genomes *de novo* (Margulies et al., 2006) and the first non-Sanger technology to sequence an individual human (Wheeler et al., 2008). Many other studies have relied on the sequencing power of NGS technology to make scientific breakthroughs in areas as diverse as uncovering the potential cause of the disappearance of the honeybee (Cox-Foster et al., 2007), revealing the complexity of rearrangements between individual human genomes (Korbel et al., 2007), providing new approaches to understand infectious diseases (Palacios et al., 2008) and sequencing the whole mitochondrial genome of a Neanderthal (Green et al., 2008).

The concept of sequencing by synthesis, although unsuccessful during the early stages of development, provided the basis for miniaturization and scaling. Two general approaches had been proposed. On the one hand there were cyclic reversible termination technologies, based on sequencing by sequential addition of a labeled base, followed by fluorescent detection and cleavage of that fluorescent base (Metzker, 2005). On the other were technologies that sequenced by detecting pyrophosphate release with an enzymatic cascade ending in luciferase and detection of the emitted light (Ansorge, 2009). The latter approach was chosen for most NGS platforms because direct incorporation of natural nucleotides seemed more efficient than repeated cycles of incorporation, detection and cleavage. For example, 454 Roche sequencing was based on miniaturizing a pyrosequencing reaction and moving both the template preparation step and the pyrosequencing chemistry to the solid phase (Margulies et al., 2006).

Pyrosequencing has been available to the scientific community since the mid-1990s as a genotyping tool (Ronaghi et al., 1996), but was not considered powerful enough for standard sequencing needs because of the short read-lengths it

generated, relegating it to the role of single-nucleotide polymorphism (SNP)-based genotyping (Fakhrai-Rad et al., 2002). However, because pyrosequencing is based on the detection of light produced whenever a nucleotide is incorporated, it does not require a physical separation process like electrophoresis to resolve the next base in the DNA strand. This means that unlike electrophoresis, where the physical length required for achieving precise resolution of distinct fragments limits miniaturization, pyrosequencing could be reduced to any reaction volume that generates detectable levels of light, also enabling sequencing to be performed in parallel (Meyer et al., 2008). Short-read pyrosequencing reactions had the potential to operate in parallel at high densities, and so provided that the sequencing template and required enzymes could be immobilized on solid supports instead of occurring free in solution, the reactions could be miniaturized, parallelized and placed on a single suitable substrate (Meyer et al., 2008; Meyer et al., 2007). Additionally, the physical separation obtained by the discrete wells of a microtiter plate could be replaced by tightly packed reactions on a single solid support, where the reaction rates of strand polymerization and light generation can be carefully balanced against the diffusion of both reagents and products, allowing for a higher sequencing throughput per run (Ansorge, 2009; Metzker, 2010).

1.7. Summary

The aim of this thesis was to utilise massively parallel sequencing platforms to address both the paucity of divergent mitogenomic datasets and assess the utility of complete mitochondrial DNA sequences in resolving topological incongruence. Advances in sequencing technology has allowed ever increasing numbers of genes to be sequenced but true phylogenomic analysis, involving many

taxa, remain constrained for most taxonomic groups by the limited number of suitable nuclear markers currently analysed. This provides an opportunity to test the ability of smaller intracellular genomes, such as mitochondria, in reconstructing well-supported phylogenetic hypotheses within an understudied group such as the Araneae.

The overarching objectives of the thesis were;

1. Assess high throughput techniques by which taxonomically divergent complete mitochondrial genome datasets could be amplified;
2. Investigate the role of massively parallel sequencing platforms and associated bioinformatic processes in processing and building mitogenomic datasets.
3. Evaluate methods of phylogenetic analysis in order to best exploit the phylogenetically informative characters inherent in mitochondrial genomes.

1.8. Thesis organisation

Chapter 1 provides a general overview and background to the recent advances in sequencing technology and the emerging field of comparative mitogenomics whilst detailing the aims and objectives of the study.

In **Chapter 2** I assess the utility of long-range PCR amplification of taxonomically divergent complete mitochondrial genome datasets using 454 GS FLX sequencing to validate the amplification of mitochondrial DNA.

Chapter 3 describes the use of both long-range PCR and rolling circle amplification (RCA) in an effort to sequence the complete mitochondrial genomes of spiders. I also describe mitochondrial genome assembly and annotation following Illumina MiSeq sequencing.

Chapter 4 describes a comparative analysis of all available complete mitochondrial genomes, from Chapter 3 and previous studies, within the Araneae,

focusing on the examination of potential phylogenetic markers and describing novel characteristics.

In **Chapter 5** the utility of complete mitochondrial genomes is explored in relation to the creation of robust fossil calibrated phylogenetic hypotheses. The relationships and evolutionary history the Araneae is discussed.

Chapter 6 provides a general discussion with an overview of the results from this study and provides information and further direction for future work.

2. Can long-range PCR be used to amplify genetically divergent mitochondrial genomes for comparative phylogenetics? A case study within spiders (Arthropoda: Araneae).

2.1. Abstract

The development of second generation sequencing technology has resulted in the rapid production of large volumes of sequence data for relatively little cost, thereby substantially increasing in the quantity of data available for phylogenetic studies. Despite these technological advances, assembling longer sequences, such as that of entire mitochondrial genomes, has not been straightforward. Existing studies have been limited to using only incomplete or nominally intra-specific datasets resulting in a bottleneck between mitogenome amplification and downstream high-throughput sequencing. Here I assess the effectiveness of a wide range of targeted long-range PCR strategies, encapsulating single and dual fragment primer design approaches to provide full mitogenomic coverage within the Araneae (Spiders). Despite extensive rounds of optimisation, full mitochondrial genome PCR amplifications were stochastic in most taxa, although 454 Roche sequencing confirmed the successful amplification of 10 mitochondrial genomes out of the 33 trialled species. The low success rates of amplification using Long-Range PCR highlights the difficulties in consistently obtaining genomic amplifications using currently available DNA polymerases optimised for large genomic amplifications and suggests there may be opportunities for the use of alternative amplification methods.

2.2. Introduction

Mitochondrial DNA markers are the cornerstones of contemporary molecular systematics and contribute greatly to our understanding of organellar evolution (Boore, 1999; Moritz et al., 1987). However, robust phylogenetic reconstruction is often impeded by low sequence volume and number of comparative loci. Prior to the turn of

the century, single gene partitions were commonly used to infer phylogenetic histories, but poor nodal support and discordance between phylogenies derived from separate markers clearly revealed that additional data are needed for robust phylogenetic reconstruction (Delsuc et al., 2005).

As more sequence data becomes available for elucidating the tree of life, large-scale sequencing efforts and interrogation of expressed sequence tag (EST) libraries or sequenced transcriptomes (Murchison et al., 2010; Okazaki et al., 2002) have begun to yield large numbers of nuclear markers that can be used for phylogenetic reconstruction (Dellaporta et al., 2006). However, true phylogenomic analyses are still not practical for the *de novo* construction of phylogenetic hypotheses for taxa without sequenced transcriptomes (Dordel et al., 2010; Struck et al., 2011). A compromise between traditional phylogenetic methods and phylogenomics lies in the phylogenetic analysis of whole mitochondrial genomes (Morin et al., 2010; Timmermans et al., 2010; Zhang and Chen, 2011). However, the routine amplification of complete mitochondrial genomes from divergent taxa remains a significant hurdle to the widespread adoption of mitogenomic approaches.

The number of informative phylogenetic characters within the mitochondrial genome has been appreciated for some time (Boore, 1999) but most studies have limited themselves to only exploring parts of this information to resolve relationships at multiple levels (Rokas and Holland, 2000). The increase in whole mitochondrial genome datasets has provided new characters potentially allowing more robust phylogenies to be constructed. These markers, known as rare genomic changes (RGCs) have become increasingly popular for resolving complex phylogenetic relationships where traditional methods have produced ambiguous, unresolved results (Rokas and Holland, 2000).

RGCs, defined as large-scale mutational changes occur much less frequently than base substitutions and have long been used in phylogenetics as supporting data embedded in DNA sequences. Examples of RGCs include changes in organelle gene order, gene duplications and genetic code variants (Boore and Brown, 1998; Dermauw et al., 2010; Masta et al., 2010; Venkatesh et al., 1999).

Whilst the potential benefits of analysing whole mitochondrial genomes (including both coding regions and structural characters) in furthering our understanding of the mechanisms behind the evolution of organellar DNA are clear, all of the data produced to date have been of comparatively low volume. The lack of empirical data means that it is still not possible to accurately assess the utility of full mitogenomic sequences as a tool for resolving complex phylogenies within the Araneae (Willerslev, 2009). Advances in sequencing technology (e.g 454 Roche GSFLX series and Solexa/Illumina) are likely to resolve the sequencing throughput issue (Glenn, 2011). However, there are still clear limitations associated with the large-scale PCR amplification of divergent, interspecific, whole mitochondrial genomes. In spite of the recent revolution in sequencing technologies, current mitogenomic studies have been characterised by either analysing data from large, but incomplete mitochondrial genome fragments (Timmermans et al., 2010), low numbers of species (Prosdocimi et al., 2012) or by focusing on nominally intra-specific datasets (Morin et al., 2010). In order to investigate the resolution potential, and associated problems with the amplification of large, interspecific mitogenomic data sets I focus here on spiders (Arthropoda: Araneae).

The Araneae are among the oldest and most diverse groups of terrestrial organisms (Penney et al., 2012a; Penney and Selden, 2011), with a current diversity of more than 43,240 described species, placed in 111 families (Platnick, 2012). Spiders are

an unequivocally ecologically important guild, being the dominant predators of insects in natural and managed ecosystems (Sunderland et al., 1986). However, they have been relatively understudied from a higher-level molecular systematic perspective, and very little is known about inter-family relationships (Benjamin et al., 2008). More recent attempts to resolve the phylogeny of the Araneae, have revealed significant topological incongruence between morphological and multiple loci phylogenies (Blackledge et al., 2009). This makes the Araneae an ideal order to test the application of using mitochondrial phylogenomics to resolve complex relationships and better understand the evolutionary mechanisms underlying speciation and diversification. Furthermore, recent studies have shown repeated tRNA gene translocations (Masta and Boore, 2004, 2008) in combination with an extensive fossil record can be utilised in resolving high-level relationships, via calibrated gene trees in the Araneae (Ayoub et al., 2007; Bodner and Maddison, 2012; Dimitrov et al., 2012a; Dimitrov et al., 2012b). Here, I investigate the utility of direct long-range PCR amplification of whole mitochondrial spider genomes, using a large range of currently available long-range *Taq* polymerases. Thus, circumventing the need for large starting biomass associated with the direct pelleting of mitochondria. Whilst mitochondrial genomes can be readily amplified in large numbers of small fragments, this is both costly and labour intensive. To this end, I adopted both single and dual fragment amplification approaches across the phylogenetic breadth of the order to assess the feasibility of both conserved and directed approaches for expedient whole mitochondrial DNA amplification. Thus, circumventing the need for large starting biomass associated with the direct pelleting of mitochondria.

2.3. Materials & methods

2.3.1. Sample collection and DNA extraction

Spiders were obtained from across the United Kingdom and Gambia by the authors and members of the British Arachnological Society (BAS) and either ethanol preserved (70% -100%, stored at 4°C) or freshly frozen (stored at -80°C) directly from living individuals (Table 2.1.). Samples of *Selenops annulatus*, *Deinopis* sp. and *Cithaeron praedonius* had been stored in 70% ethanol and were incorporated in order to maximise taxonomic coverage across the order. No specific permits were required for the described field studies and no specific permissions were required for these locations or activities. No locations were privately owned or protected and the field studies did not involve protected or endangered species. All four legs were removed from the left side of the thorax prior to DNA extraction. The rest of the body was stored in 100% ethanol for vouchering and subsequent identification purposes. Whole genomic DNA was extracted from a single femur of each species using the Qiagen DNeasy Blood & Tissue Kits (Qiagen).

Table 2.1. Taxonomic information, storage conditions and GenBank accession numbers for specimens. Sample storage indicates the methods in which the specimens were preserved on collection; either freshly frozen at -80°C (frozen) or stored in 70-100% ethanol at 4°C (ethanol).

Family	Genus	Specific epithet	Sample ID	Locality	CoxI	16s	Sample storage
Agelenidae	*Malthonica*	*silvestris*	464_SC_AB	Kent, UK	JQ412460		Frozen
Amaurobiidae	*Amaurobius*	*similis*	483_SC_AB	Lancashire, UK		JQ406635	Frozen
Anyphaenidae	*Anyphaena*	*accentuata*	522_SC_AB	Kent, UK	JQ412439	JQ406633	Frozen
Araneidae	*Agalenatea*	*redii*	507_SC_AB	Kent, UK		JQ406637	Frozen
Araneidae	*Araneus*	*diadematus*	571_SC_AB	Dorset, UK	JQ412440	JQ406621	Frozen
Cithaeronidae	*Cithaeron*	*praedonius*	455_SC_AB	Kotu, Gambia	JQ412441		Ethanol
Clubionidae	*Clubiona*	*terrestris*	518_SC_AB	Kent, UK	JQ412442		Frozen
Corinnidae	*Messapus*	*martini*	453_SC_AB	Kerr Serign, Gambia		JQ406616	Ethanol
Corinnidae	*Phrurolithus*	*festivus*	503_SC_AB	Kent, UK		JQ406632	Ethanol
Ctenidae	*Anahita*	sp.	460_SC_AB	Kotu, Gambia		JQ406614	Ethanol
Cybaeidae	*Argyroneta*	*aquatica*	581_SC_AB	Dorset, UK		JQ406617	Frozen
Deinopidae	*Deinopis*	sp.	451_SC_AB	Gunjur, Gambia	JQ412443		Ethanol
Dictynidae	*Dictyna*	*latens*	499_SC_AB	Kent, UK	JQ412444	JQ406629	Ethanol
Dysderidae	*Dysdera*	*erythrina*	479_SC_AB	Kent, UK	JQ412445	JQ406627	Ethanol
Eresidae	*Stegodyphus*	sp.	454_SC_AB	Bijilo, Gambia	JQ412459		Ethanol
Gnaphosidae	*Zelotes*	*apricorum*	462_SC_AB	Kent, UK	JQ412463		Frozen
Idiopidae	*Gorgyrella*	sp.	448_SC_AB	UK Pet Trade	JQ412447		Frozen
Linyphiidae	Unknown	Unknown	559_SC_AB	Gwynedd, UK	JQ412448	JQ406636	Frozen
Lycosidae	*Pardosa*	*nigriceps*	477_SC_AB	Kent, UK	JQ412452	JQ406631	Ethanol
Philodromidae	*Philodromus*	*dispar*	517_SC_AB	Kent, UK	JQ412453	JQ406634	Frozen
Pisauridae	*Pisaura*	*mirabilis*	502_SC_AB	Kent, UK	JQ412454	JQ406630	Ethanol
Pholcidae	*Pholcus*	*phalangioides*	484_SC_AB	Lancashire, UK		JQ406625	Frozen
Salticidae	*Salticus*	*scenicus*	423_SC_AB	Gwynedd, UK	JQ412456	JQ406628	Ethanol
Segestriidae	*Segestria*	*senoculata*	583_SC_AB	Gwynedd, UK	JQ412457	JQ406615	Ethanol
Selenopidae	*Selenops*	*annulatus*	449_SC_AB	Kotu, Gambia	JQ412458		Ethanol
Sparassidae	*Micrommata*	*virescens*	461_SC_AB	Kent, UK	JQ412451	JQ406618	Frozen
Tetragnathidae	*Meta*	*menardi*	481_SC_AB	Gwynedd, UK	JQ412449	JQ406620	Frozen
Theraphosidae	*Eupalaestrus*	*campestratus*	446_SC_AB	UK Pet Trade	JQ412446	JQ406626	Frozen
Theraphosidae	*Grammostola*	*rosea*	445_SC_AB	UK Pet Trade		JQ406624	Frozen
Theraphosidae	*Psalmopoeus*	*cambridgei*	447_SC_AB	UK Pet Trade	JQ412455	JW406623	Frozen
Theridiosomatidae	*Theridiosoma*	*gemmosum*	489_SC_AB	Glamorgan, UK	JQ412461		Frozen
Thomisidae	*Xysticus*	*audax*	521_SC_AB	Kent, UK	JQ412462	JQ406622	Frozen
Uloboridae	*Miagrammopes*	sp.	452_SC_AB	Bijilo, Gambia	JQ412450	JQ406619	Ethanol

2.3.2. Primer anchoring strategy

In order to identify an efficient approach for the amplification of whole mitochondrial genomes I adopted two long-range PCR strategies, in order to amplify the complete mitogenome in one or two large fragments (Barnes, 1994). DNA from two anchoring regions, a c. 650 b.p. region of the Cytochrome Oxidase I (COI) and a 450 b.p. region spanning the large ribosomal subunit (16s rRNA), were amplified for 33 taxa. Both amplifications were performed in 25µl reactions using the primer combinations of CHELF1 (5′- TACTCTACTAATCATAAAGACATTGG) and CHELR2 (5′- GGATGGCCAAAAAATCAAAATAAATG) (*COI*) (Barrett and Hebert, 2005) and primers LR-N-13398 (5′- CGCCTGTTTAACAAAAACAT) and LR-J-12887 (5′- CCGGTCTGAACTCAGATCACGT) (16S) (Blackledge et al., 2009). PCR reactions comprised 1x PCR buffer, 1.5mM $MgCl_2$, 0.4 µM of each primer, 0.625 units ThermoPrime *Taq* DNA polymerase (Thermo Scientific) and 1µl template DNA and thermocycling was performed using a DNA engine (Tetrad 2) Peltier Thermal Cycler (BIORAD). Cycling conditions were 60 seconds at 94°C, 5 cycles of 60 seconds at 94°C, 90 seconds at 45°C and 90 seconds at 72°C followed by 35 cycles of 60 seconds at 94°C, 90 seconds at 50°C and 90 seconds at 72°C (Barrett and Hebert, 2005). Amplification success was checked using a 1% agarose gel stained with Ethidium Bromide (EtBr). Successful amplifications were cleaned with 1U shrimp alkaline phosphatase (Promega) to dephosphorylate residual deoxynucleotides and 0.5 U Exonuclease I (Promega) to degrade excess primers (Werle et al., 1994) and subsequently sequenced bidirectionally (Macrogen Inc, Seoul, Korea) using the same primers as for amplification.

2.3.3. Long-range primer design and PCR

Following the sanger sequencing of the CO1 and 16s anchoring regions, chromatographs were checked and sequences were manually edited where necessary, using CodonCode Aligner (v. 2.0.6, Codon Code Corporation), prior to alignment using Clustal W (Larkin, 2007). The initial aim of this process was to amplify regions of sufficient length, on opposite sides of the mitochondrial genome, from which 'universal' primers could be designed (Gai et al., 2008; Hwang et al., 2001; Masta and Boore, 2004). However, no conserved regions were found that would facilitate the design of long-range primers that could be used to amplify homologous loci from multiple families. Long-range primers were subsequently designed for individual taxa in order to amplify the entire mitochondrial genome in one or two large fragments that overlapped with the conserved COI and/or 16s ribosomal subunit using the Primer 3 software (Rozen and Skaletsky, 2000). Default values were used with the exception of length (22-30bp), primer T_m (57.0 – 70.0 ^oC) and GC content (40-60%), which followed consensus recommendations from the Taq manufacturers (Table 2.2.). For the single fragment protocol, primers were designed, within the COI sequences, with the light strand primer situated downstream of the heavy strand primer, thus taking advantage of the circular nature of the genome. For the dual fragment protocols two sets of primers were designed for each taxon, to bridge the gaps between the COI and 16S regions (Figure 2.1.).

The PCRs were performed initially in 50µl volumes using multiple polymerases recommended for long range PCR amplification, including: Clontech Advantage 2, Clontech Advantage Titanium, Clontech Advantage HD (Takara Biotech), NEB LongAmp (NEB), GeneAmp XL (Applied Biosystems) & Expand Long Range PCR (Roche), using the

protocols and cycling conditions recommended by the manufacturers. All DNA polymerases were initially tested on five phylogenetically disparate samples prior to using the most successful ones on the remaining samples. Successfully amplified PCR products were electrophoresed and subsequently purified using Qiagen QiaQuick Gel extraction kits (Qiagen) or Origene Rapid PCR Purification System (Origene), dependent on fragment size. Following long-range PCR amplification, primers were assigned to groups based on amplification success (good, stochastic or no amplification) and subsequently tested for significant differences in primer length, GC content and T_m using a Mann-Whitney U test in the statistics package SPSS (Corp., Released 2011).

Table 2.2. Taxon information and long-range primer sequences for specimens. All Primer sequences are shown in 5' – 3' orientation.

Genus	Specific Epithet	Single Fragment (COI)		Dual Fragment (COI – 16s)		Dual Fragment 16s – COI)	
		Forward	Reverse	Forward	Reverse	Forward	Reverse
Malthonica	*silvestris*	CTG CTG GTG GTG GTG ATC CAA TTT TA	CCA ACT ATA GCA GAT CAT GCC CCA AA				
Amaurobius	*similis*			CAC CAA CAA GCC CAG CAA ATA TGG	AAG GCC CCA ACA AAG TCG GAA TAT CT	GAT ATT CCG ACT TTG TTG GGG CCT TT	ATG GTG GCG GAG ATC CAG TGT TAT TT
Anyphaena	*accentuata*	TCC AGT TCC TAC TAT AGC AGC TCA	TCC TGC TGG AGG TGG AGA TCC TAT TT	GGA GGA TAT ACT GTT CAT CCA GCT CCT	AGC CGC AAT TTT TGT GCA AAG G	CAA CAT CGA GGT CGC AAT CTT T	TCC TGC TGG AGG TGG AGA TCC TAT TT
Araneus	*diadematus*	ACT TTC TTT ACC CGT TTT GGC AGG TG	TCT CCA ATA AAT CTC CCA GGC TGA CC				
Cithaeron	*praedonius*	CCA ATA ATC TTC CAG GCT GAC CCA AT	TGG GGG TGG GGA TCC TAT TTT ATT TC				
Clubiona	*terrestris*	TCT TTG ATC CTG CTG GAG GAG GAG AT	TCA GCT GCC CCT AAT ATC ATT GGA AC	TCT TTG ATC CTG CTG GAG GAG GAG AT	TGG CTT ACA CCG GTC TGA ACT CAA AT	CGG TGT AAG CCA GGT CGG TTT CT	TCA GCT GCC CCT AAT ATC ATT GGA AC
Deinopis	sp.	ATC TCC CAG GCT GAC CTA ATT CCG TA	GCC TGT TTT AGC TGG GGC GAT TAC TA				
Dictyna	*latens*	TGT TCC TAC TAT AGC CGA TCA AGC TCC A	TCT CTT CCT GTT TTA GCA GGG GCT AT	TGT TCC TAC TAT AGC CGA TCA AGC TCC A	AGC CGC AAT TTA CTG TGC TAA GGT AG	CTA CCT TAG CAC AGT AAA TTG CGG CT	TTT TTG ATC CTG CTG GAG GAG GTG AT
Dysdera	*erythrina*			GCC AAT TAC CAA AGC CCC CAA TTA TC	ATA AAG TTC TGC GGC TGT TTG TGT GC	AAG CTG GGG CGG CCT AAA TAA AAA	CTG CTG GTG GAG GGG ATC CTA TCT TA
Zelotes	*apricorum*	GCA TTT AGC GGG TGC TTC GTC TAT TA	TCC TAC TAT AGC AGC CCA TGA ACC AA				
Gorgyrella	sp.	TCA TCC CCT AAA AGT CTC CCA ACC TG	GGG GGA TCC GGT TTT GTT TCA A				
Unknown Linyphiidae	sp.			CCC CAA CAT TAA AGG CAC TAG CCA AT	TGC GTC TTA GAG TAA TTT TAA TTC AAC ATC GAG	CGC AAC TGT ATT AGA AGT AAG TCT GTT CGA C	GGG GGC TCC TGA TAT AGC TTT TCC TC
Pardosa	*nigriceps*	CAT AGC TGA CCA AAC TCC AAA C	AGG GAT GAC TAT GGA AAA GGT TCC TC	TCC CTA TCA TAG CTG ACC AAA CTC CA	GCT ACC TTA GCA CAA AAG CGG CCA TT	AGA TTG CGA CCT CGA TGT TGA T	GCC GGG GGA GGA GAT CCT ATT TTA TT
Philodromus	*dispar*	TCC AAC TAT AGC TGC TCA TGC ACC AA	CTG CAG GGG GAG GAG ATC CTA TTT TA	CTG CAG GGG GAG GAG ATC CTA TTT TA	TGA TCA AGT TAC CAT AGG GAT AAC AGC GTA	TGC TAC CTT AGC ACA TAA GCG GCT ATT TA	TAT TGC GGT TCC AAC TAT AGC TGC TC
Phrurolithus	*festivus*	TGG CGG AGA TCC TGT GTT ATT TCA AC	CTC CCA TGA GGG CTG GCA TTA TC	TGG CGG AGA TCC TGT GTT ATT TCA AC	TTC AAT AGG GTC TTG TCG TCT ACC TTT	ATC AAG CTA CCA TAG GGA TAA CAG CA	TGA CGT GCA TTA ACT GAG TAC GAA TGA
Pisaura	*mirabilis*			AAA CTG TTC AAC CAG CAC CAA CTC CT	TGC TAC CTT AGC ACA AAT GCG GCT AT	AGA TTG CGA CCT CGA TGT TGA T	ATC CAG CTG GTG GAG GTG ATC CTA TT
Salticus	*scenicus*			TAT ACC GTT CAA CCT GCC CCT ACT CC	TGG CAG ACA ATT GCA TTA GAA TTA GA	TCT AAT TCT AAT GCA ATT GTC TGC CA	TTT TTG ATC CTG CTG GAG GAG GAG AT
Meta	*menardi*			GGA GTA GGG AAG GCG GTA ACA ATC AA	GGG GAG ATT ATG GGT ACG AAA GAG TT	TTT CGT ACC CAT AAT CTC CCC TAC TG	TTG ATT GTT ACC GCC TTC CCT ACT CC
Eupalaestrus	*campestratus*	GTT CCA ACC ATC GAA GAC CAA ATC C	ATA GAG CGG GTA CCG TTG TTT GTG TG	CGG TGT TGG CTG GGG CAA TTA CTA T	ATT AAG CTA CCG CAG GGA TAA CAG CA	GTC GTC CCT CTT AGC AAA TGA AAG TC	CCC AAC AAA CTT CCA ACC TGA CCT AA
Psalmopoeus	*cambridgei*			TGC AGG AGG AGG AGA CCC TGT TTT AT	AAT CAA GCT ACC GCA GGG ATA ACA GC	GGT CTT GTC GTC CCT CTT GCT ATG AA	ATC CCA GAC CAC ACC CCA AAT AGA AA
Theridiosoma	*gemmosum*	ACC AGT ATT AGC AGG AGC AAT CAC	CCC AGG TTG CCC TAA TTC TCT ACG AA				
Xysticus	*audax*			TCC CAA CTA TAG CAG ATC AAG CAC CA	GAC GAC AAG ACC CTA TAG AAC TTT ACT TCA	TGA TTA TGC TAC CTT AGC ACG ATT GC	TCT TTA CAT TTA GCG GGT GCT TCT TC

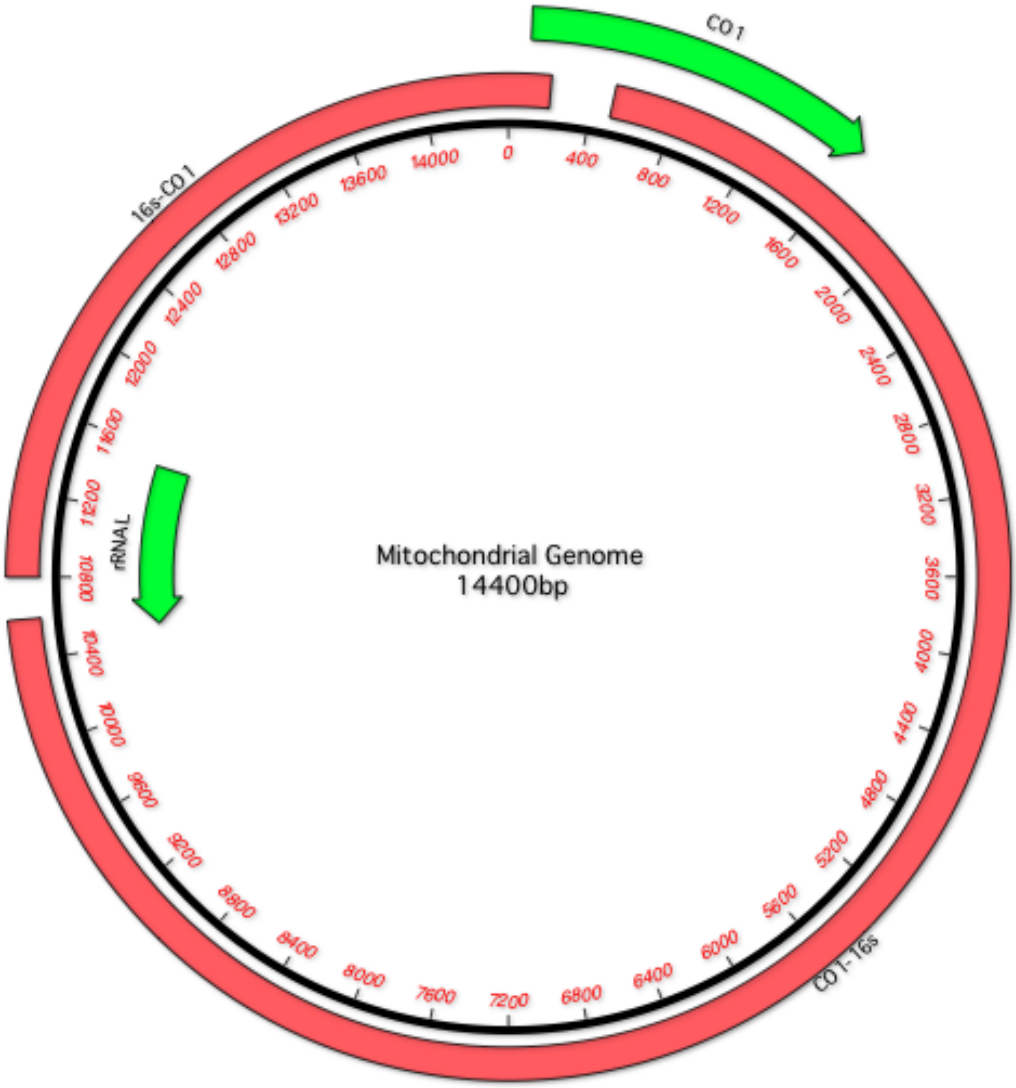


Figure 2.1. Araneae mitochondrial genome displaying anchoring regions. Example spider mitochondrial genome highlighting the genes from which the two fragment long-range PCR strategy was designed (green arrows – direction indicates gene sequence 5' – 3'). Solid red bars show the two long-range PCR products and approximate length.

2.3.4. Amplicon shearing, library construction and 454 Roche sequencing

Amplicons from eleven purified samples (Table 2.3.) were fragmented using a Covaris DNA shearer (Covaris) at 10% duty cycle, intensity: 4 with 200 cycles per burst for 65 seconds. Following quantification using a Qubit 2.0 fluorometer (Invitrogen) MID barcoded adapted sequencing libraries were constructed using the NEBNext DNA sample prep master mix set in accordance with manufacturer's instructions (New England Biolabs), pooled in equimolar concentrations and sequenced at low putative coverage on $1/16^{th}$ of a 454 Roche GSFLX Titanium platform (Centre for Genome Research, University of Liverpool). Please note that the sequencing step here was intended as a quality control measure and not intended to sequence to a depth required for full mitochondrial genome assembly (Figure 2.2.). The 454 Roche sequences were assembled using GS *De Novo* Assembler software (Roche). Following trials with Roche's GS *De Novo* software, MIRA and CLC Genomics Workbench, no substantial differences in number of contigs or length were present using any of the approaches. Therefore, in accordance with the proven functionality of the 454 Roche software, all sequences were assembled using the GS *De Novo* program. Following assembly, the resulting contigs were compared to existing Araneae mitochondrial DNA sequences within GenBank via BLAST (Altschul et al., 1990) in order to investigate sequence homology (GenBank accession numbers: AY309258, AY309259, AY452691, NC_005942, NC_010777, NC010780).

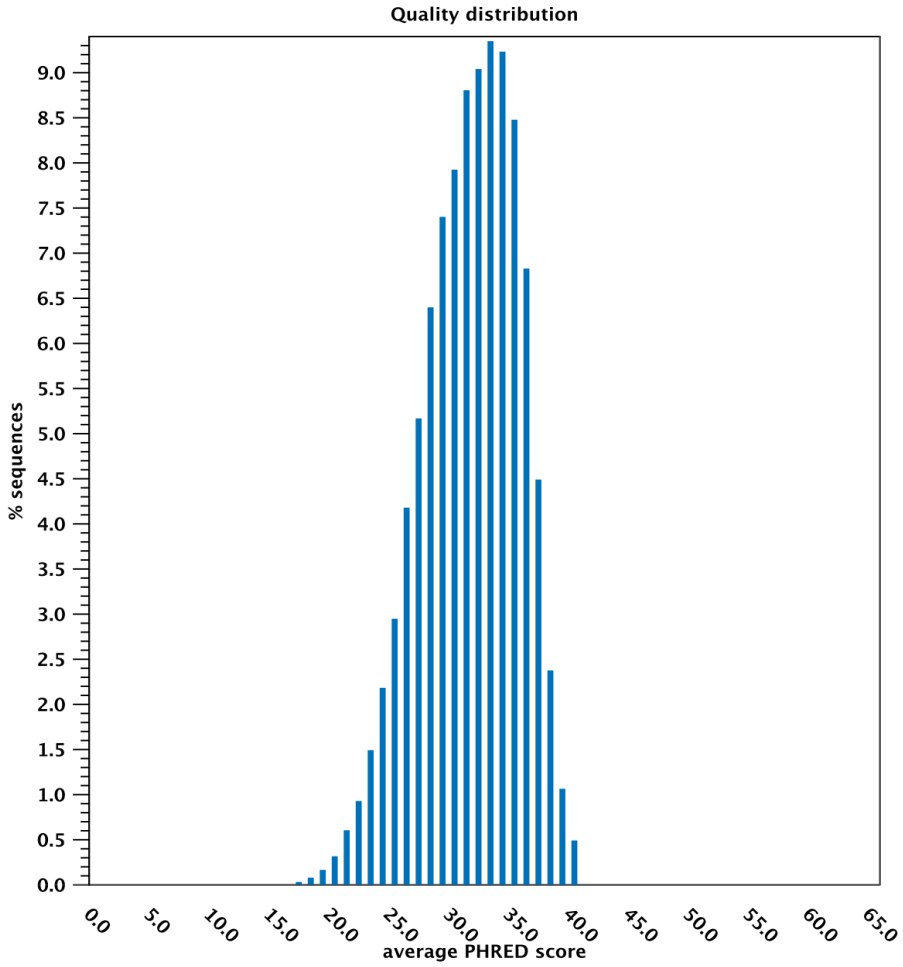


Figure 2.2. Quality scores from the total number of reads obtained from 454 sequencing.

2.4. Results

2.4.1. Anchoring regions

Of the 33 species used, 27 and 24 taxa were successfully amplified for the COI and the 16S regions respectively, yielding anchoring points for a total of 26 families (Table 2.1.) (GenBank accession numbers: COI - JQ412439 - JQ412463 and 16s - JQ406614-JQ406637).

2.4.2. Mitochondrial genome PCR amplifications

Initially, five specimens, covering a broad taxon range, were used to test the performance of the *Taq* polymerases for single fragment mitochondrial genome amplification (Table 2.3.). Following repetition of the *Taq* testing when opting for a dual fragment approach, it was found that NEB LongAmp was equally as effective as Clontech Advantage 2 *Taq* polymerase. Subsequently, complete mitochondrial genomes were successfully amplified in c. 7-9 kb or c. 15kb fragments using Clontech Advantage 2 (single fragment) or NEB LongAmp (dual fragment) *Taq* polymerases. The probability of obtaining successful amplifications of complete mitochondrial genomes utilising either the 1 or 2 fragment amplification protocol was not demonstrably different (based upon presence/absence of bands), suggesting that the limiting factor in the amplification of the genomes was not necessarily the size of the target fragment. In some two-fragment amplifications (Samples *Pardosa nigriceps*, *Dysdera erythrina*, *Pisaura mirabilis*, *Phrurolithus festivus* & *Anyphaena accentuata*) the COI to 16s region amplified consistently more often than the reverse, highlighting potential inhibition to amplification within the target sequence. However, of the initial 33 samples trialled, from which a subset of 22 (based on sequence availability and prior PCR amplification success) were used for long-range PCR, only 11 mitochondrial genomes could be amplified robustly and with a guaranteed level of reproducibility (*Meta menardi, Xysticus audax, Psalmopoeus cambridgei,* a *Gorgyrella* sp*., Eupalaestrus campestratus,* a Linyphiidae sp*., Zelotes apricorum, Malthonica silvestris, Araneus diadematus, Pisaura mirabilis* & *Dysdera erythrina*). The NEB LongAmp polymerase (NEB) amplified genomes in two fragments with greater reproducibility whilst Clontech Advantage 2 (Takara Biotech) amplified single fragment genomes with greater consistency. Statistical

analysis of the primer properties (length, GC content and T_m) yielded no significant differences between any combination of primers that amplified repeatedly, intermittently, or those that consistently failed to amplify mitochondrial genome fragments.

Table 2.3. Enzymatic information for successful long-range amplifications. Clontech Advantage Titanium and Roche GeneAmplification XL are not shown but failed to amplify any samples. * Indicates good amplification from a dual fragment PCR approach. The term Good -indicates robust and consistent amplification. The term Stochastic indicates a non-consistently reproducible amplification.

Sample reference	**Clontech**		**NEB LongAmp**	**Expand Long Range**
	Advantage 2	**Advantage LA**		
448_SC_AB	Stochastic	No amplification	No amplification	Good
462_SC_AB	Good	No amplification	Weak	No amplification
464_SC_AB	Good	No amplification	No amplification	No amplification
446_SC_AB	No amplification *	No amplification	No amplification *	No amplification
451_SC_AB	Stochastic	Stochastic	No amplification	Stochastic

2.4.3. 454 Roche sequencing

The sequencing of the 11 amplicon libraries provided a total of 27,892 tagged reads with an average length of approximately 278 bases (Table 2.4.). Reads from sample 464_SC_AB (*Malthonica silvestris*) were unable to be recovered by the MID identifying software. *De novo* read assembly was unable to construct complete mitochondrial genomes but provided an average of an estimated 65% sequence assembly (assuming a ca. 15kb mitogenome target; contigs and sequences available on DRYAD entry doi:10.5061/dryad.8dd3n).

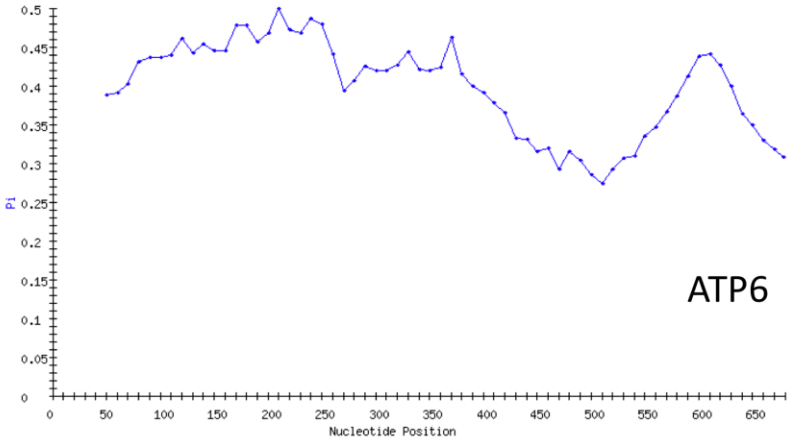
In light of the likelihood of tRNA gene rearrangements and because only 6 mitochondrial genomes were available on Genbank at the time of analysis, a highly dissimilar BLAST search was performed on the nucleotide collection database. Of the 10 samples, all contigs created through short-read assembly had maximum identification values ranging from 78-100% Max ID for spider mitochondrial DNA, either through an unrestricted nucleotide blast search or with searches focused on Araneae accessions. Nucleotide sequences for all reads were deposited in the GenBank short-read archive (SRA051390.1).

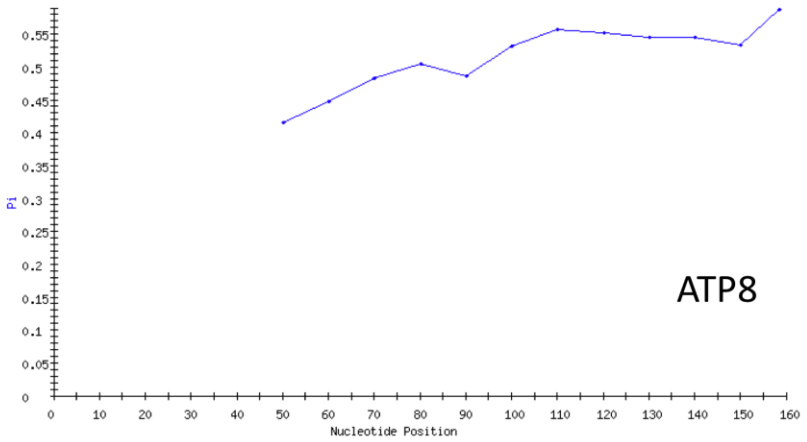
Table 2.4. Sequencing information following 454 GS FLX run and subsequent GS *De Novo* Assembler contig assembly. Number of reads per sample represents only those from which the indicator MID sequence was recovered. Number of contigs represents only those formed with a length greater than 100 bases. Average contig length refers only to contigs over 100 bases in length. Longest contig length shows the largest single contig assembled.

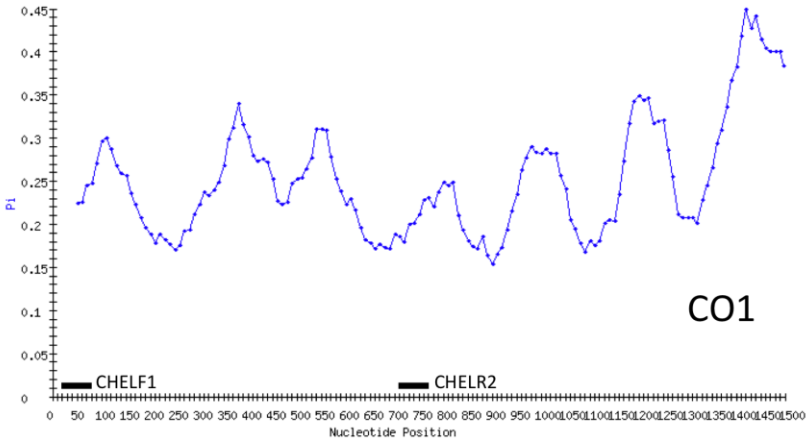
Species Name	Sample Reference	Number of Reads	Number of Contigs	Average Contig Length	Roche MID Identifier	Longest Contig Length
Araneus diadematus	571_SC_AB	1,313	20	340	1	1,058
Psalmopoeus cambridgei	447_SC_AB	1,689	17	944	2	4,483
Eupalaestrus campestratus	446_SC_AB	380	5	849	3	1,916
Gorgyrella sp.	448_SC_SB	6,203	49	665	4	2,849
Xysticus audax	521_SC_AB	4,887	35	423	5	2,257
Pisaura mirabilis	502_SC_AB	1,192	14	505	6	872
Dysdera erythrina	479_SC_AB	1,554	16	733	7	1,523
Linyphiidae sp.	559_SC_AB	503	12	540	8	1,310
Zelotes apricorum	462_SC_AB	9,081	26	619	9	6,434
Meta menardi	481_SC_AB	1,090	18	532	11	1,452

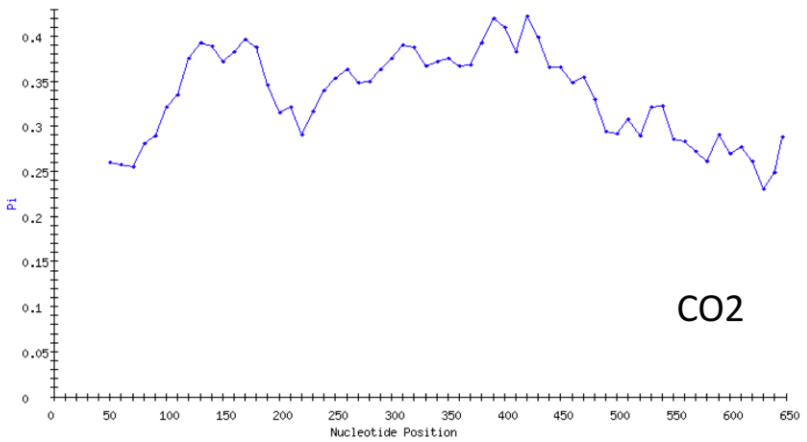
2.5. Discussion

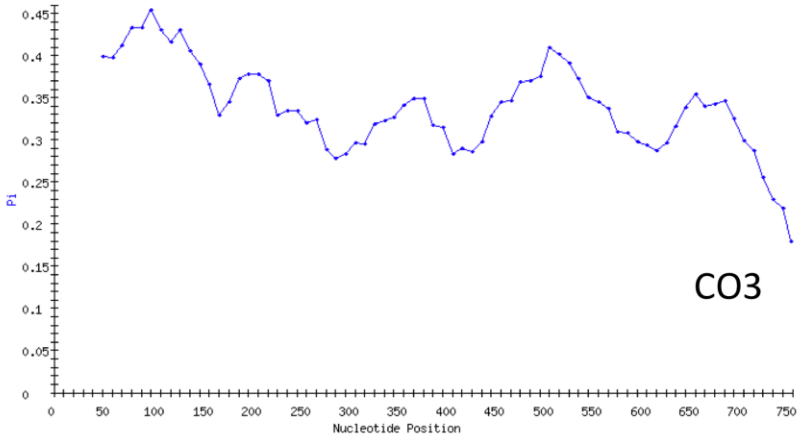
I was able to amplify most of the target genes (COI and 16s) across our taxon range, but was unable to identify or develop degenerate primers for any other potential anchoring region within the spider mitochondrial genome following sliding window analyses of all regions most commonly used in spider phylogenetics. Sliding window analysis was performed on all 13 protein coding genes and both ribosomal RNAs (rRNAs) of the six previously published spider mitochondrial genomes using the *Drosophila* Polymorphism database, SNP Graphics (http://dpdb.uab.es/dpdb/diversity.asp) (Figure 2.3.). This most likely highlights the high level of mitochondrial genetic diversity that can be found within the spiders along with the absence of universal primers available for no more than a handful of protein coding genes, resulting in few reference sequences available for comparison for large portions of the mitogenome (Hosseini et al., 2007).

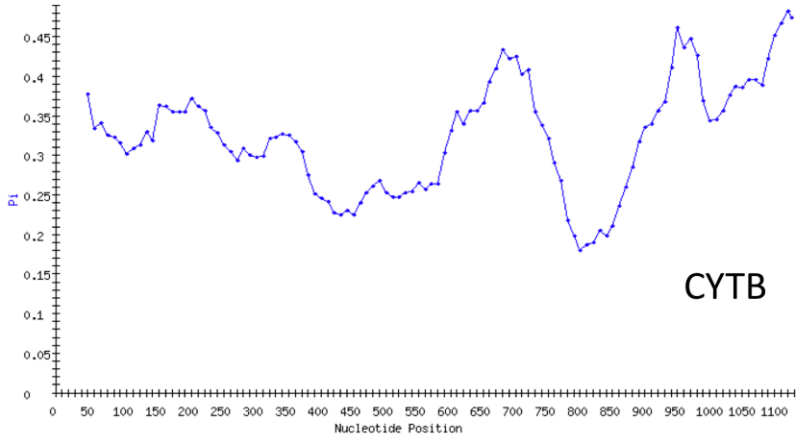

Pi
Nucleotide Position
ATP6

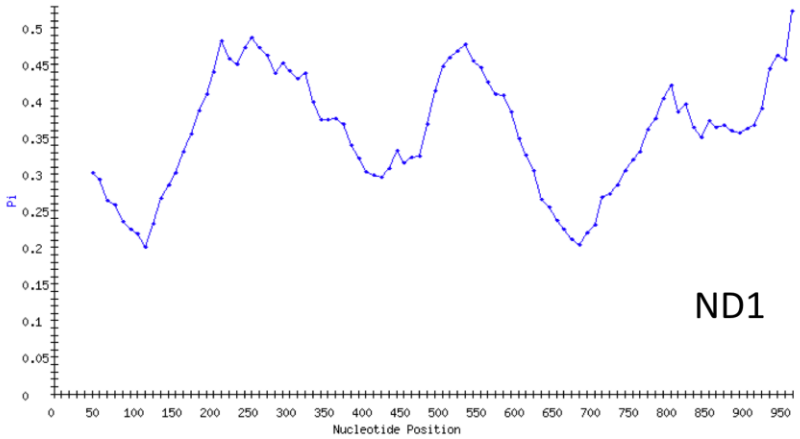

Pi
Nucleotide Position
ATP8

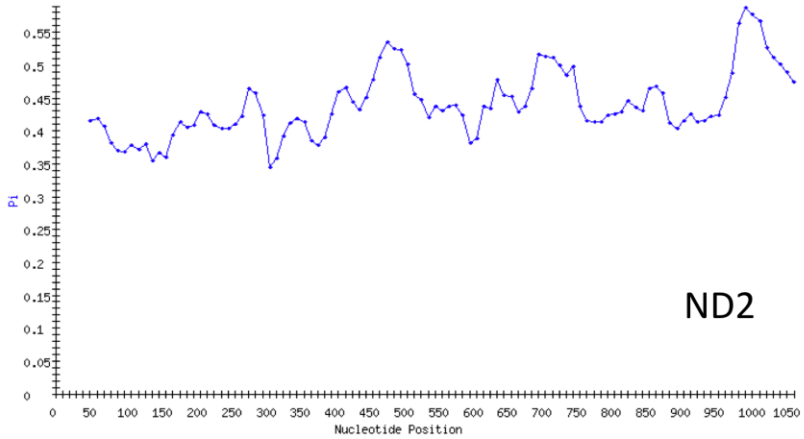

Pi
CHELF1
CHELR2
Nucleotide Position
CO1

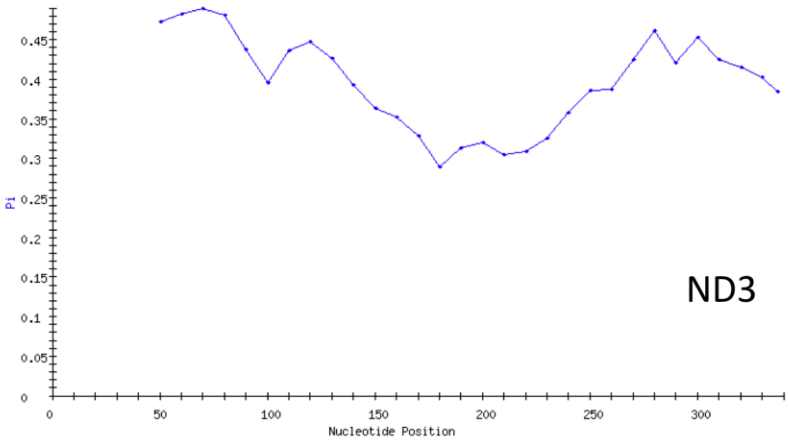

Pi
Nucleotide Position
CO2

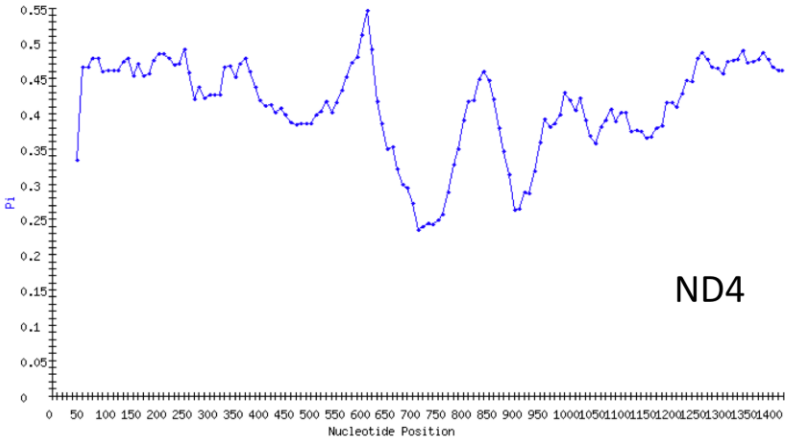

Pi
Nucleotide Position
CO3

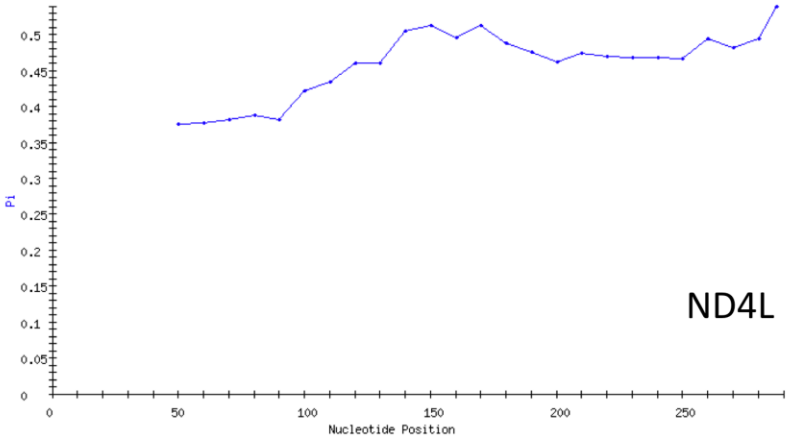

Pi
Nucleotide Position
CYTB

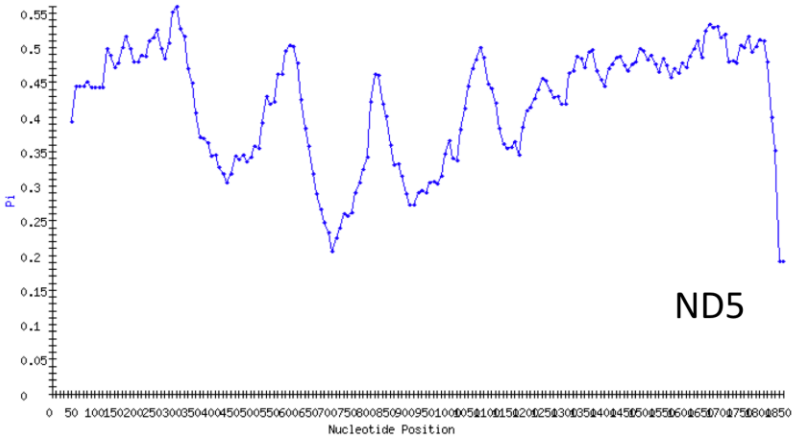

Pi
Nucleotide Position
ND1


Pi
Nucleotide Position
ND2


ND3
Pi
Nucleotide Position


ND4
Pi
Nucleotide Position


ND4L
Pi
Nucleotide Position


ND5
Pi
Nucleotide Position

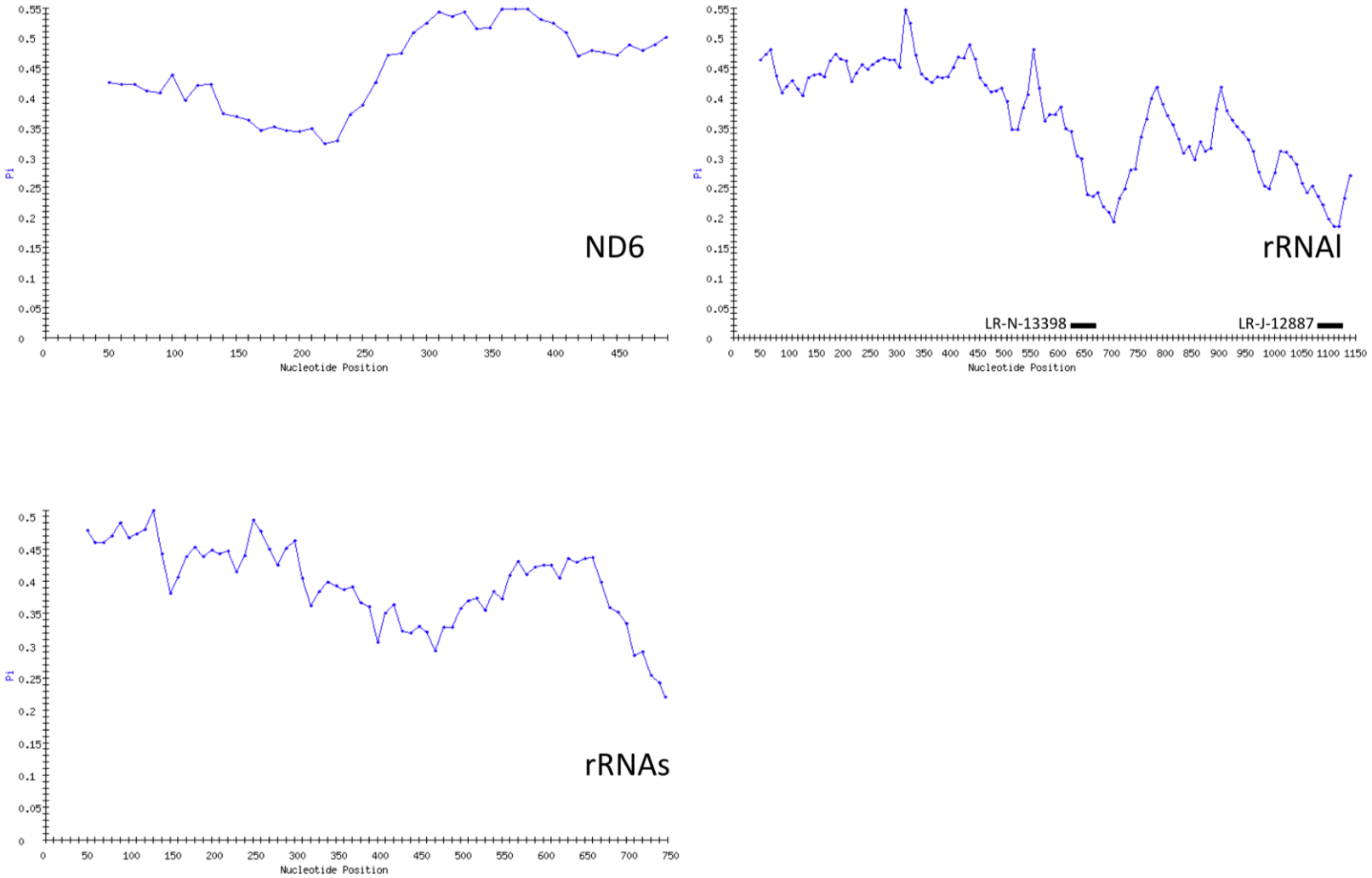


Figure 2.3. Sliding window plots of nucleotide diversity (pi) across all 13 mitochondrial protein coding genes and both ribosomal RNA genes of six Araneae taxa. The generepresented is indicated in bold. The analyses were performed on Clustal alignments of *Habronattus oregonensis , Nephila clavata, Ornithoctuns huwena, Heptathela hangzhouensis, Calisoga longitarsus* and *Hypochilus thorelli* gene sequences using a window size of 99 and step size of 10. Approximate positions of priming sites for anchoring regions are indicated by black bars along side primer name.

Predictably, the 454 Roche low coverage sequencing step resulted in a wide variation of read coverage per amplicon pool, most likely due to differences in MID primer tag design and read recovery. Nevertheless, even highly covered mitochondrial genomes did not result in a full assembly, suggesting further optimisation may be required in either the shearing or bioinformatic steps of mitogenome assembly (Alkan et al., 2011). The BLAST search of the assembled contigs showed that spider mitochondrial genomes had indeed been amplified and that I had not inadvertently amplified nuclear DNA or DNA from a contaminating source such as the *Wolbachia* bacterium (Goodacre et al., 2006). The need to use searches focused on Araneae accessions, in order to get positive matches, most likely highlights the genetic divergence between the trialled taxa and the complete mitogenomic data currently available for 6 out of the 112 described families of spiders.

Nevertheless, even following an extensive campaign of PCR strategies and optimization, using a variety of long range *Taqs*, I could not consistently amplify approximately two thirds of the species, a substantial proportion of our sample taxa. There is no reason to attribute PCR failure to degraded DNA resulting from sample storage conditions, since all specimens were preserved directly from living organisms using tried and tested preservation media. Moreover, of the three samples stored in 70% EtOH, although sample *Selenops annulatus* did not amplify, samples *Deinopis sp*. and *Cithaeron praedonius* could be amplified, albeit inconsistently, suggesting that 70% EtOH preservation over ca. three years may be sufficient to preserve mitogenome integrity for PCR amplification. However, I could not rule out the possibility that sub-optimal storage conditions can adversely affect the availability of suitable templates for long-range PCR. It is also unlikely that the primers affected long-range PCR success rate, since comparative analysis

between the primer sets revealed no discernable statistical differences in physical or chemical properties (i.e. GC content, length, T_m). The arachnid mitochondrial genome, as for most animal mitochondrial genomes, is a small extra-chromosomal genome comprising 37 genes including 22 coding for transfer RNAs (tRNAs), 13 coding for proteins and 2 coding for ribosomal RNAs (rRNAs) (Boore, 1999). As well as the 37 coding genes, mitochondrial genomes also contain a small, approximately 1-2 kb, non-coding control region, named due to its perceived role in controlling the transcription and replication of the mtDNA molecule (Mardulyn et al., 2003). However, the protein coding gene arrangement of spiders is highly conserved and shared amongst many other chelicerates and so is not likely to cause differences in amplification success. Previously published Araneae mitochondrial genomes are considered to be very A/T rich (64-76%), reflecting the low G/C content of the mitochondrial genomes of other arthropod orders (Beard et al., 1993). In spite of the low G/C content of the DNA no definitive reason is forthcoming for the failure to obtain more consistent long PCR amplifications from our target mitochondrial genomes. Whilst the larger region (COI-16s) of the dual fragment approach amplified successfully more often than the shorter 16s-COI counterpart, the G/C content of both fragments is comparable (approximate difference of 2.6% averaged across all currently published Araneae mitochondrial genomes) with no significant differences between the highest and lowest values of the comprising genes within each fragment. However the lesser-amplified small fragment (16s-COI) does contain the control region, also known as the A+T rich region due to its low G/C content, which has been known to inhibit DNA polymerases in insects and platyhelminths (Hu et al., 2007). However, enzyme inhibition is more likely to be caused by tandem repeats (Zhang, 1997) present in

and around the control region as comparative analysis of previously published arachnid mitochondrial genomes show other genes to have lower G/C content.

Our results provide evidence for the successful amplification of several whole mitochondrial genomes, in one or two long fragments overlapping the COI and/or 16s regions, using a variety of commercially available DNA polymerases optimised for large genomic amplifications. However, I have also highlighted the difficulty in obtaining amplifications with a guaranteed level of reproducibility, using multiple currently available polymerases and thus highlighted potential limitations to the feasibility of mitogenomic studies featuring large numbers of independently amplified taxa using single, or dual fragment approaches. Recent studies utilising mitogenomics to resolve phylogenetic incongruence have avoided this issue by focusing on the production of data sets comprising of either incomplete genome sequences (Timmermans et al., 2010) or that are nominally intra-specific (Morin et al., 2010). Whilst these data still provide a large number of informative phylogenetic characters, the lack of the whole mitochondrial genome precludes the acquisition of maximum phylogenetic resolution, including RGCs, from the mtDNA genome (Ernsting et al., 2009; Quinn and Wilson, 1993; Wenink et al., 1994). However, the importance of recovering complete mitogenomic sequences in order to more accurately reconstruct phylogenetic relationships has long been understood and so remains an important avenue of exploration in contemporary phylogenetics (Cummings et al., 1995). Although I present data on a single order (Araneae – Spiders) I believe that the results highlight the limitations to the feasibility of generating diverse, interspecific complete mitochondrial genome data sets from limited long range PCR amplifications, in terms of both cost and efficiency. Whilst many studies have successfully used a multitude of PCR amplifications in order to generate mitochondrial genomes (Ernsting et al., 2009;

Jiang et al., 2009; Sorenson et al., 1999), this is both labour and time intensive. This highlights the potential need to utilise alternative, non-PCR based methods that are able to amplify complete mitochondrial genomes both quickly and cost effectively. Direct recovery of organellar genomes and mitochondrial gene partitions from whole shotgun genome sequencing and EST libraries is possible (Groenenberg et al., 2012; Kane et al., 2012), but expedient mechanisms to isolate only the mtDNA locus are needed to optimise organellar coverage for large numbers of taxa. While methods such as Rolling Circle Amplification (RCA) (Tang and Hyman, 2005) have been trialled successfully on a limited number of species, further investigations across a range of taxa will be desirable to investigate their full potential to create divergent, multi-taxon datasets for comparative mitogenomics. Such methods hold advantages over PCR-based strategies due to the non-specific nature of the amplification process. By using random hexamers, as opposed to synthesising bespoke taxon-specific oligonucleotides, it is possible to avoid the amplification failure shown by this study whilst allowing for the creation of large, diverse mitochondrial genome datasets from low amounts of starting material.

3. Rolling circle amplification & paired-end sequencing of complete Araneae (true spider) mitochondrial genomes

3.1. Abstract

Despite the potential that the advent of Next-Generation sequencing allows for the expanding field of mitogenomics, limitations in the PCR-based amplification of whole mitochondrial genomes has slowed the accumulation of information and limited its resolution potential by restricting analysis to the use of incomplete or nominally intra-specific datasets. Whole-genome amplification methods such as Rolling Circle Amplification (RCA), that utilise multiple displacement amplifications, have provided an insight into the potential for the rapid acquisition of taxonomically diverse datasets that have been previously unobtainable through other amplification strategies. Our results show that following an alkaline-lysis DNA extraction spider mitochondrial genomes have been successfully amplified in large enough concentrations to allow for the preparation and subsequent Illumina MiSeq sequencing and assembly of complete mitochondrial genomes, independently verified by PCR amplification. Nevertheless, the *de novo* recovery of full mitochondrial genomes was not possible and is hindered by low coverage and the lack of closely related spider reference sequences. This method provides the basis from which a viable option for the rapid and inexpensive construction of large divergent mitochondrial data sets can be designed.

3.2. Introduction

Phylogenetic reconstructions utilising mitochondrial DNA have traditionally relied on the use of several key, universally amplifiable loci (COI, 16S, CYTB) (Simon et al., 1994; Zardoya and Meyer, 1996). However, in order to better resolve the tree of life, more informative characters must be used to redress conflicting topologies and form more robust phylogenetic hypotheses (Jabbar et al., 2013; Rokas and Carroll, 2005; Spinks et al., 2009; Wortley et al., 2005). The rapid progression of DNA sequencing technologies has opened up the possibility of acquiring complete mitochondrial genomic data sets (Morin et al., 2010; Prosdocimi et al., 2012; Zhang et al., 2013). The acquisition of complete mitogenomes avoids the constraints of analysing only highly conserved protein coding sequences, whilst allowing for more robust phylogenetic analyses (Delsuc et al., 2005; Philippe et al., 2009; Pick et al., 2011; Struck et al., 2011) through the incorporation of multiple loci and rare genomic changes (RCA) (Rokas and Holland, 2000).

Recent studies utilising long-range PCR methods (Cheng et al., 1994; Cooper et al., 2001; Hu et al., 2007) in order to amplify complete mitochondrial genomes in multiple overlapping fragments, have been limited to either analysing data from large, but incomplete mitochondrial genome fragments (Timmermans et al., 2010), low numbers of species (Bi et al., 2012; Prosdocimi et al., 2012) or by focusing on nominally intra-specific datasets (Morin et al., 2010). Whilst the ability of mitochondrial DNA alone in resolving complex phylogenetic relationships is a highly debated topic (Aguileta et al., 2008; Mueller, 2006; Zardoya and Meyer, 1996), in order to fully investigate the potential of complete mitogenomic DNA sequences the bottleneck associated with their amplification must be removed.

One way in which this has been achieved is through the direct shotgun sequencing of both mitochondria (Gilbert et al., 2007; Ogihara et al., 2005) and chloroplasts (Dempewolf et al., 2010). However, this proves difficult for taxa where starting biomass is a limiting factor, and not cost effective for the recovery of mitochondrial genomes from genomic DNA preparations. More recently, Whole Genome Amplification (WGA) methods, i.e. alternatives to long-range PCR, have been used to amplify both mitochondrial and nuclear DNA (Martino et al., 2012; Rodrigue et al., 2009; Stokes et al., 2011).

Rolling Circle Amplification (RCA) is one such method of WGA that offers the potential to overcome the limitations of long-range PCR amplification (Figure 3.1).

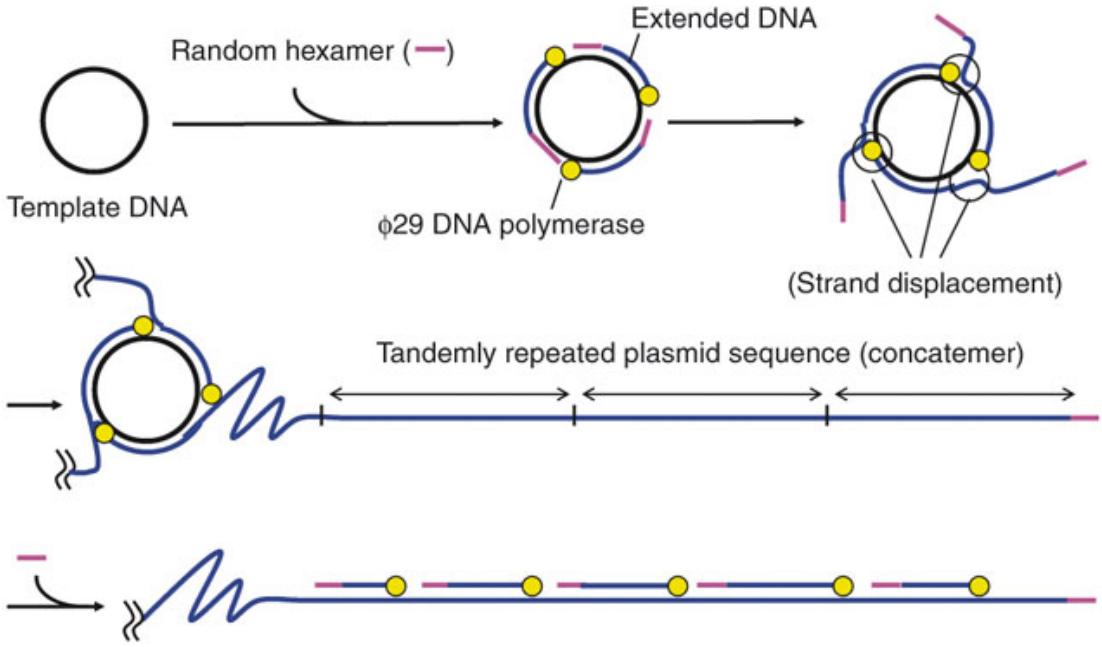


Figure 3.1. Rolling Circle Amplification (RCA). Random hexamers are used in place of taxon specific oligonucleotides to anneal to the template DNA in the presence of the strand displacing enzyme Phi 29, preferentially amplifying circular molecules.

Whilst originally designed for the rapid amplification of plasmid and phage DNA for the creation of high-throughput sequencing libraries (Dean et al., 2001; Fire and Xu, 1995; Lizardi et al., 1998), RCA has been successfully trialled for complete mitochondrial genome amplifications (Tang and Hyman, 2005). Whilst taking advantage of both the high copy number and circular nature of the mitochondrial genome, RCA uses a strand displacing polymerase, phi29, in the presence of random hexamers in order to amplify the complete mitochondrial genome from relatively low concentration starting material (Fire and Xu, 1995).

Primarily limited to nematodes for the amplification of mitochondrial DNA (Hyman et al., 2011; Tang and Hyman, 2005; Tang and Hyman, 2007), RCA has already been widely used in environmental meta-genomic studies looking at viral and bacterial DNA (Bexfield and Kellam, 2011; Kim and Bae, 2011) and biosecurity (Goransson et al., 2012). The inherent difficulties associated with using RCA to amplify mitochondrial sequences, as opposed to bacterial genomes, has been nuclear and extra cellular contamination, thus highlighting the need to purify samples of extraneous DNA (Simison et al., 2006). Once these constraints can be removed, RCA would allow for the rapid construction of mitogenomic libraries for second-generation sequencing platforms, facilitating the subsequent analysis of broad taxonomic data sets to test the utility of mitochondrial genomes in resolving complex phylogenies (Pacheco et al., 2011; Zhang et al., 2013).

Here I explored the utility of Rolling Circle Amplification in order to create divergent mitogenomic data sets using the Araneae (Spiders) as a test order. Whilst PCR-based methodologies have been found to be hindered, in terms of both labour and financial limitations, whole genome amplification (WGA) methods offer viable alternatives to the generation of intra-specific mitogenomic data sets in order to create robust phylogenetic hypotheses. To this end, I provide evidence of

the successful rolling circle amplification of complete mitochondrial genomes, from low starting biomass, corroborated by long-range PCR amplifications, whilst highlighting the difficulties that must be overcome in order to make the construction of divergent WGA datasets a reality.

3.3. Materials & Methods

3.3.1 Sample Collection and DNA extraction

Spiders were obtained from across the United Kingdom by the authors and members of the British Arachnological Society (BAS) and either ethanol preserved (100%, stored at 4°C) or freshly frozen (stored at -80°C) directly from living individuals (Table 3.1.).

Table 3.1. Taxonomic information, storage conditions and GenBank accession numbers for specimens. Sample storage indicates the methods in which the specimens were preserved on collection; either freshly frozen at -80°C (frozen) or stored in 70-100% ethanol at 4°C (ethanol).

Family	Genus	Specific epithet	Sample ID	Locality	CoxI	16s	Sample storage
Agelenidae	*Malthonica*	*silvestris*	464_SC_AB	Kent, UK	JQ412460		Frozen
Araneidae	*Araneus*	*diadematus*	571_SC_AB	Dorset, UK	JQ412440	JQ406621	Frozen
Dysderidae	*Dysdera*	*erythrina*	479_SC_AB	Kent, UK	JQ412445	JQ406627	Ethanol
Gnaphosidae	*Zelotes*	*apricorum*	462_SC_AB	Kent, UK	JQ412463		Frozen
Idiopidae	*Gorgyrella*	sp.	448_SC_AB	UK Pet Trade	JQ412447		Frozen
Linyphiidae	Unknown	Unknown	559_SC_AB	Gwynedd, UK	JQ412448	JQ406636	Frozen
Pisauridae	*Pisaura*	*mirabilis*	502_SC_AB	Kent, UK	JQ412454	JQ406630	Ethanol
Tetragnathidae	*Meta*	*menardi*	481_SC_AB	Gwynedd, UK	JQ412449	JQ406620	Frozen
Theraphosidae	*Psalmopoeus*	*cambridgei*	447_SC_AB	UK Pet Trade	JQ412455	JW406623	Frozen
Thomisidae	*Xysticus*	*audax*	521_SC_AB	Kent, UK	JQ412462	JQ406622	Frozen

All four legs were removed from the left side of the thorax prior to DNA extraction. The rest of the body was stored in 100% ethanol for vouchering and subsequent identification purposes. Companion PCR amplifications were used to validate the sequences obtained from the RCA amplifications, whole genomic DNA was extracted from a single femur of each of four species using (*Meta menardi*, *Pisaura mirabilis*, *Psalmopoeus cambridgei* & *Zelotes apricorum*) the Qiagen DNeasy Blood & Tissue Kits (Qiagen). For samples undergoing rolling circle amplification (*Araneus diadematus*, *Dysdera erythrina*, *Gorgyrella sp*., *Linyphiidae sp*., *Malthonica silvestrus*, *Meta menardi*, *Pisaura mirabilis*, *Psalmopoeus cambridgei, Xysticus audax* & *Zelotes apricorum*) the above extraction was preceded by an alkaline lysis step (Raimond et al., 1999). Tissues were dissected in 40µl (1 volume) of extraction buffer (0.1 M NaOH 10 mM Tris, 1 mM EDTA). A total of 80µl extraction buffer was added (2 volumes) on ice for 5 minutes followed by 60µl 3 M potassium acetate (1.5 volumes) on ice for 10 minutes. This allowed for the enrichment of mitochondrial DNA in order to provide a suitable template for the RCA (Rolling Circle Amplification) reaction by reducing the amount of nuclear DNA in the extraction (Raimond et al., 1999; Tang and Hyman, 2005).

3.3.2. Long-range primer design and PCR

Long-range primers, for individual taxa, had been previously designed in this study in order to amplify the entire mitochondrial genome in one or two large fragments that overlapped with the conserved COI and/or 16S ribosomal subunit using the Primer 3 software (Rozen and Skaletsky, 2000) (See Chapter 2, Table 2.2.)

The companion PCR reactions were performed in 50µl volumes using either Clontech Advantage 2 or NEB LongAmp (NEB) long-range polymerases, using the

protocols and cycling conditions recommended by the manufacturers. Successfully amplified PCR products were electrophoresed and subsequently purified using Qiagen QiaQuick Gel extraction kits (Qiagen) or Origene Rapid PCR Purification System (Origene), dependent on fragment size. Purified samples were subsequently quantified using a Qubit 2.0 fluorometer (Invitrogen).

3.3.3. Rolling circle amplification

The rolling Circle amplification reactions were conducted using the Illustra Templiphi amplification kit (GE Healthcare) according to the manufacturer's instructions and incubated at 30˚C for 18 hours. Reactions were performed in twice the recommended volumes due to the low concentrations of the starting material (less than 5ng/µl). Amplification success was checked using a 1% agarose gel stained with Ethidium Bromide (EtBr).

3.3.4. Restriction enzyme digest & purification

Successful RCA amplifications were subjected to restriction enzyme digests, carried out with the EcoR1 (Promega) restriction enzyme using recommended procedures in 30µl reactions for 4 hours. Following digestion, samples were electrophoresed and subsequently purified using Qiagen QiaexII suspension gel extraction kits (Qiagen, Germany). Purified samples were quantified using a Qubit 2.0 fluorometer (Invitrogen).

3.3.5. Amplicon shearing, library construction and Illumina sequencing

Amplicons from thirteen purified samples (Table 3.2.) were fragmented using a Covaris DNA shearer (Covaris) at 10% duty cycle, intensity: 4 with 200 cycles per burst for 65 seconds, followed by a second round at 10% duty cycle,

intensity: 5 with 200 cycles per burst for 30 seconds. Following quantification using a Qubit 2.0 fluorometer (Invitrogen), index adapted sequencing libraries were constructed using the TruSeq DNA sample prep master mix set in accordance with manufacturer's instructions (Illumina), pooled in equimolar concentrations and sequenced on an Illumina MiSeq flow cell (Centre for Genome Research, University of Liverpool).

Illumina paired reads were separated according to their indexes at source, following the completion of the MisSeq run. Contigs were assembled using CLC Genomics Workbench version 4.7.0 (CLC Bio). Following trials of low coverage assemblies with CLC Genomics Workbench and MIRA, no substantial differences in number of contigs or length were present using any of the approaches. Therefore, in order to take advantage of the inbuilt sequence manipulation and visualisation functionalities, paired-reads were assembled using the CLC Genomics Workbench program. The resulting contigs were compared to existing Araneae mitochondrial DNA sequences within GenBank via a restricted nucleotide BLAST search (Altschul et al., 1990) in order to investigate sequence homology with existing Araneae mitochondrial DNA sequences (GenBank accession numbers: AY309258, AY309259, AY452691, NC_005942, NC_010777, NC010780).

Table 3.2. Taxonomic information and amplification strategy for specimens. A tick in the long-range PCR or RCA columns indicates whether or not a sample was amplified using that method. The sequenced column denotes which successful amplifications were processed for Illumina sequencing. * The RCA amplified sample from *Zelotes apricorum* failed to pass pre-sequencing quality checks and so was excluded from the Illumina Miseq run.

Family	**Taxon name**	**Sample ID**	**Long-range PCR**	**RCA**	**Sequenced**
Agelenidae	*Malthonica silvestris*	464_SC_AB		✓	RCA
Araneidae	*Araneus diadematus*	571_SC_AB		✓	RCA
Dysderidae	*Dysdera erythrina*	479_SC_AB		✓	RCA
Gnaphosidae	*Zelotes apricorum*	462_SC_AB	✓	✓	PCR*
Idiopidae	*Gorgyrella* sp.	448_SC_AB		✓	RCA
Linyphiidae	Linyphiidae sp.	559_SC_AB		✓	RCA
Pisauridae	*Pisaura mirabilis*	502_SC_AB	✓	✓	Both
Tetragnathidae	*Meta menardi*	481_SC_AB	✓	✓	Both
Theraphosidae	*Psalmopoeus cambridgei*	447_SC_AB	✓	✓	Both
Thomisidae	*Xysticus audax*	521_SC_AB		✓	RCA

3.4. Results

3.4.1. Mitochondrial genome PCR amplifications

All four of the mitochondrial genomes were successfully PCR-amplified in one or two fragment approaches (Figure 3.2.).

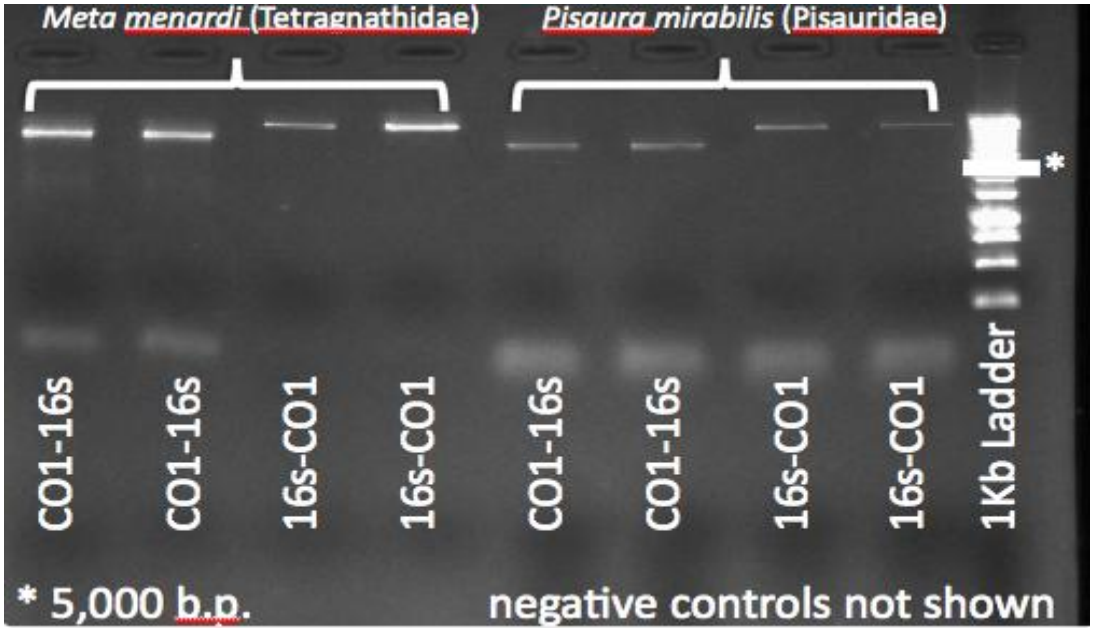


Figure 3.2. Successful long-range PCR amplification of dual fragment approach. Agarose gel containing duplicates of two dual fragment long PCR amplified products. Taxon name and family is shown (top of image) as well as amplified fragment orientation (below bands) (See chapter 1 Figure 2.1.). Approximate size of 5kb (highest rung = 10kb) is indicated by white bar in lane 9 (1kb DNA ladder).

3.4.2 Rolling circle amplification & restriction enzyme digest

All ten rolling circle amplifications were successful and yielded large amounts of product for restriction enzyme digest. The restriction enzymes EcoRI and XbaI had been chosen following in depth tests of the six previously published

Araneae mitochondrial genomes using the RestrictionMapper software (http://www.restrictionmapper.org/) (Table 3.3.).

Table 3.3. Taxonomic information, Genbank accession numbers and restiction site information for six previously published spider mitochondrial genomes. Restriction Sites shows the number of restriction sites present in the circular DNA molecule for each given restriction enzyme.

Taxon Name	**Family**	**Genbank Accession**	**Restriction Sites**				
			EcoRI	**XbaI**	**SacI**	**HindIII**	**XhoI**
Heptathela hangzhouensis	Liphistiidae	AY309258	4	5	2	3	0
Hypochilus thorelli	Hypochilidae	NC010777	2	10	1	4	0
Nephila clavata	Nephilidae	AY452691	4	3	2	6	0
Calisoga longitarsus	Nemesiidae	NC010780	2	3	1	9	1
Habronattus oregonensis	Salticidae	NC_005942	9	8	0	1	1
Haplopelma schmidti	Theraphosidae	AY309259	3	9	5	0	1

Following 4 hour restriction enzyme digests, no discernable bands were present below 10,000 bp in length, using either restriction enzyme, that would indicate the presence of successfully amplified mitochondrial genomes (Figure 3.3.) (Tang and Hyman, 2005). However, in both instances, a band was present above the range of a 1kb ladder (Promega) that was approximate in length to the product of a single fragment long-PCR amplification (Figure 3.4.).

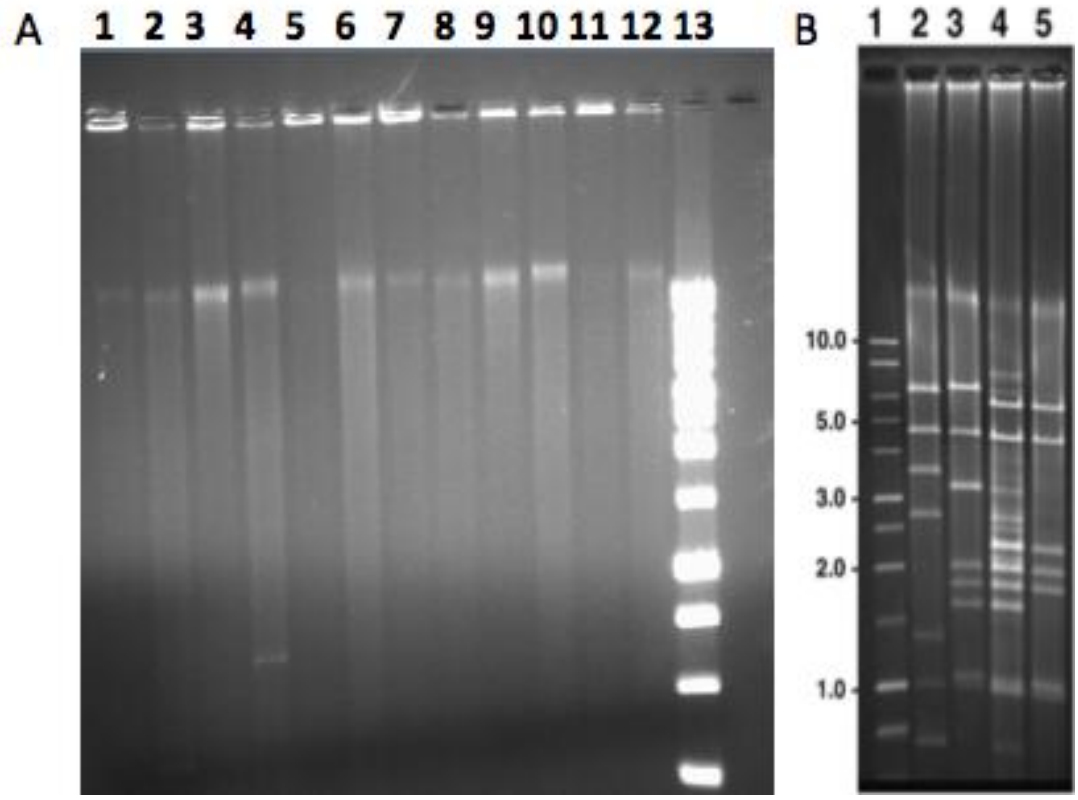


Figure 3.3. Visualisation of a Rolling Circle Amplification following restriction enzyme digest. (A) Twelve RCA amplification products (lanes 1-12), following 4 hour restriction enzyme digest with XbaI (lanes 1-6) and EcoRI (lanes 7-12) with 1kb DNA ladder (lane 13). The only visible product of the restriction enzyme digest is a prominent band in lane 4. (B) Four RCA amplification products from the nematode *Thaumamermis cosgrovei* (lanes 2-5) (from Tang & Hyman, 2005), following BgIII restriction enzyme digest, and 1kb DNA ladder (lane 1).

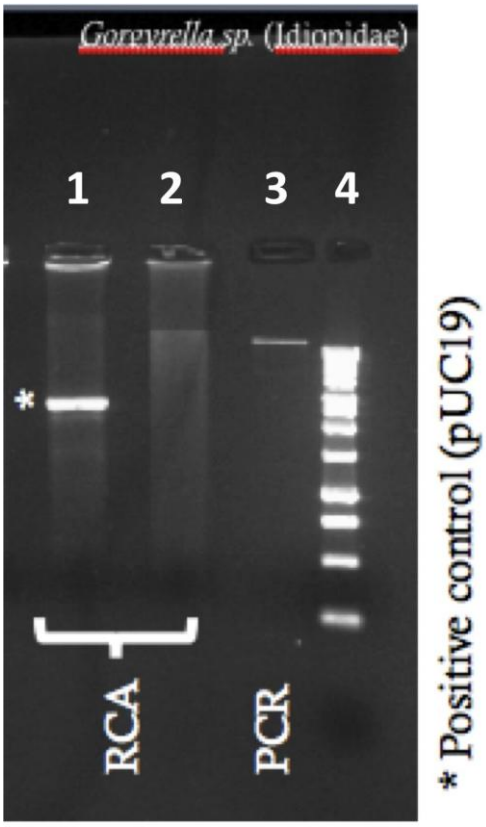


Figure 3.4. Restriction enzyme digested RCA product and long-range PCR amplification from a single taxon. Size comparison of rolling circle amplifications following 4 hour EcoRI digest (lanes 1 & 2) and single fragment long-range PCR product (lane 3) in the trap door spider genus *Gorgyrella.* Lane 5 contains 1kb DNA ladder.

3.4.3. Illumina MiSeq Sequencing

The sequencing of the 12 amplicon libraries provided a total of 11,974,246 high quality tagged paired-end reads with an average length of approximately 145 bases (Figure 3.5.). The RCA amplified sample 462_SC_AB (*Zelotes apricorum*) did not pass the necessary quality control steps and so was removed from the run. The entire mitochondrial genome sequence of each of the four specimens trialled for long-range PCR was amplified in one (*Zelotes apricorum*) or two (*Meta menardi*,

Pisaura mirabilis, *Psalmopoeus cambridgei*) overlapping fragments, resulting in mitochondrial genomes of 14,222, 14,399, 14,013 and 13,767 base pairs in length respectively (Table 3.4.). *De novo* read assembly was able to reconstruct complete mitochondrial genomes for each of the four PCR amplified samples, with over 1000 times average coverage (using default settings), but was unable to assemble mitochondrial genome fragments from the rolling circle amplified samples (Table 3.5.), in spite of repeated optimisation steps involving the rigorous testing of stringency settings. However, a large number of non-spider contigs up to 5,386 bp in length were recovered from the data and identified by comparison to sequences contained within the Genbank database via BLAST (Table 3.6.). Due to the substantial number of assembled sequences generated from *de novo* assembly (Table 3.5.) it was not possible to elucidate the identity of all contigs. Nevertheless, three of the rolling circle amplified sequences were successfully assembled using their PCR amplified counterpart as a reference sequence with less than 1% sequence divergence in the resulting assembly. In order to assess the feasibility of using non-identical reference sequences from which to assemble spider mitochondrial genomes, the unassembled reads from each RCA amplification product were mapped to the closest related complete Araneae mitochondrial genome based on current topologies. Default parameters were used in CLC Genomics Workbench with the exception of percentage identity, which was reduced from 80% to 70% based on previous analysis of nucleotide diversity (See Chapter 2 Figure 2.3) (Table 3.7.).

The identities of all putative Araneae mitochondrial genomes were confirmed through an unrestricted BLAST search. Following assembly it was evident that there was an uneven coverage across the newly assembled mitochondrial genomes. In order to assess if the observed bias generated through

the sequencing run was attributable to compositional extremes within the amplicons, as previous studies have noted (Aird et al., 2011), sliding window analysis plots were generated to display G/C content across each of the four complete mitochondrial genome sequences (Figure 3.6.).

Table 3.4. Taxonomic information with sequencing data and assembly results for complete mitochondrial genomes. Number of reads indicates how many paired-end reads were recovered for each indexed taxon. Genome length and contig coverage show the final length of the fully assembled mitochondrial genome and average total coverage respectively. The bracketed figure for contig coverage shows the highest average coverage of a recovered contig.

Family	**Genus**	**Specific epithet**	**Sample ID**	**No. of reads**	**Genome length**	**Contig coverage**
Gnaphosidae	*Zelotes*	*apricorum*	462_SC_AB	850,404	14,222	7,875 (8,375)
Pisauridae	*Pisaura*	*mirabilis*	502_SC_AB	761,576	14,013	6,412 (25,181)
Tetragnathidae	*Meta*	*menardi*	481_SC_AB	967,416	14,399	9,657 (9,657)
Therasphosidae	*Psalmopoeus*	*cambridgei*	447_SC_AB	1,276,916	13,767	11,434 (32,787)

Table 3.5. Sequencing information following Illumina Miseq run and subsequent CLC Genomics Workbench contig assembly. Number of reads per sample represents those from which the indexing sequence was recovered. Number of contigs represents all contigs created through *de novo* assembly. Longest contig length shows the largest single contig assembled.

Taxon Name	Sample Reference	Amplification strategy	Number of Reads	Number of Contigs	Illumina Index	Longest Contig Length
Psalmopoeus cambridgei	447_SC_AB	PCR	1,276,916	43	1	10,037
Zelotes apricorum	462_SC_AB	PCR	850,404	136	2	13,023
Pisaura mirabilis	502_SC_AB	RCA	1,093,640	5,194	3	5,386
Pisaura mirabilis	502_SC_SB	PCR	761,576	43	4	10,161
Gorgyrella sp.	448_SC_AB	RCA	944,612	4,889	5	5,386
Xysticus audax	521_SC_AB	RCA	1,369,646	224	7	1,653
Linyphiidae sp.	559_SC_AB	RCA	857,504,	4,923	8	5,386
Malthonica silvestris	464_SC_AB	RCA	687,244	4,588	9	5,386
Psalmopoeus cambridgei	464_SC_AB	RCA	746,818	1,349	10	5,386
Meta menardi	481_SC_AB	PCR	967,416	45	11	14,399
Meta menardi	481_SC_AB	RCA	871,414	1,780	12	5,386
Araneus diadematus	571_SC_AB	RCA	1,547,056	1,036	13	5,386

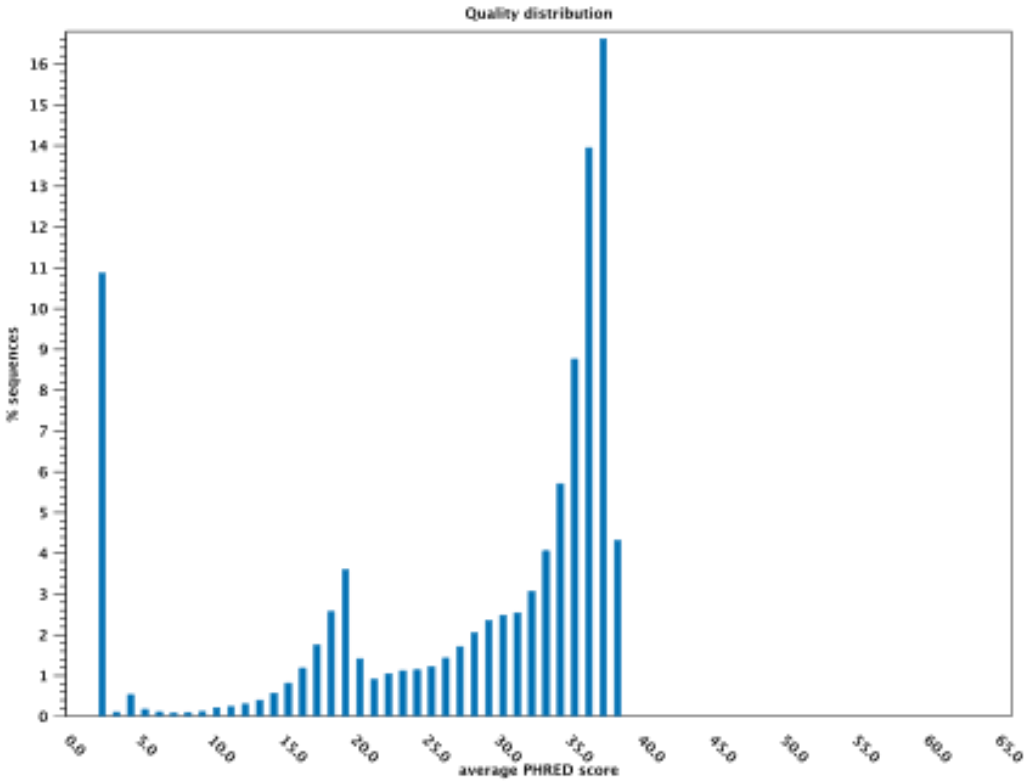


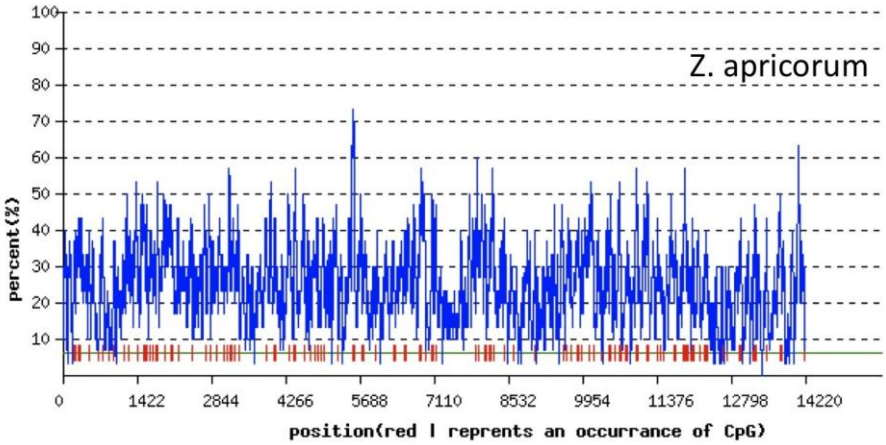
Figure 3.5. Frequency plot of phred quality scores from the total number of reads obtained from Illumina MiSeq sequencing.

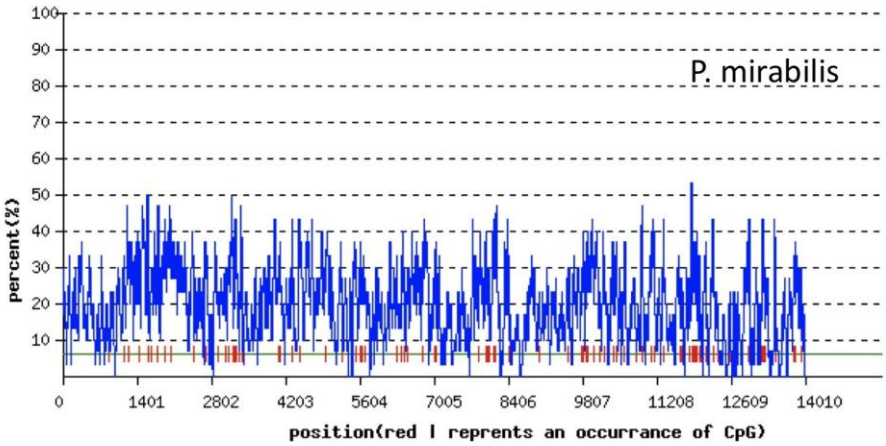
Table 3.6. BLAST results of the five largest unique, non-Araneae contigs created through *de novo* assembly of Illumina sequenced rolling circle amplifications. Contig length represents the length in bases. BLAST match shows the top hit (highest score) of an unrestricted BLAST search with Max ID (% match to query sequence), Coverage (% overlap), E-value (number of matches expected by chance) and the Genbank accession number. Sample denotes which amplifications the contig was found in following *de novo* assembly.* Contaminating sequence was also found in PCR amplified samples.

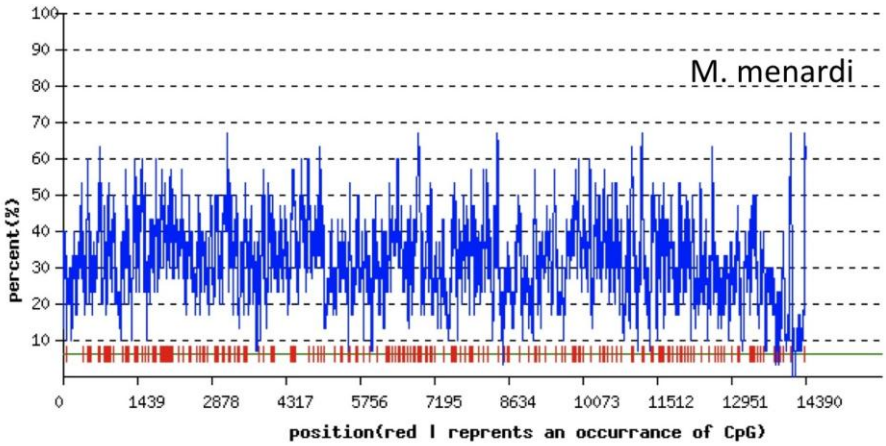
Contig length	**Sample**	**BLAST match**	**Max ID**	**Coverage**	**E-value**	**Genbank accession**
5,386	All*	Bacteriophage S13	99	100	0.0	M14428.1
3,424	464_SC_AB	Human DNA	99	94	0.0	AY308744.1
3,272	571_SC_AB	Cloning vector	100	53	0.0	JQ354897.1
1,653	521_SC_AB	*Wolbachia*	79	42	4e-54	HE660029.1
1,650	521_SC_AB	*Priopioni* bacteria	99	97	0.0	CP003084.1

Table 3.7. Assembly results from mapping unassembled paired-reads from the sequencing of RCA amplifications to the most closely related complete Araneae mitochondrial genome. Number of reads denotes the total number of reads left following *de novo* assembly. Match length shows the total length of the reference genome. Contig length represents the total length in nucleotides of the mapped reads.

Taxon name	**Sample ID**	**Number of reads**	**Closest match**	**Match length**	**Contig length**
Gorgyrella sp.	448_SC_AB	808,556	*Calisoga longitarsis*	14,070	3,713
Xysticus audax	521_SC_AB	1,365,742	*habronattus oregonensis*	14,381	6,567
Linyphiidae sp.	559_SC_AB	787,898	*Nephila clavata*	14,436	4,904
Malthonica silvestris	464_SC_AB	622,254	*Nephila clavata*	14,436	4,897
Araneus diadematus	571_SC_AB	1,539,770	*Nephila clavata*	14,436	7,588


Z. apricorum
percent(%)
100
90
80
70
60
50
40
30
20
10
0
1422
2844
4266
5688
7110
8532
9954
11376
12798
14220
position(red | reprents an occurrance of CpG)


P. mirabilis
percent(%)
100
90
80
70
60
50
40
30
20
10
0
1401
2802
4203
5604
7005
8406
9807
11208
12609
14010
position(red | reprents an occurrance of CpG)


M. menardi
percent(%)
100
90
80
70
60
50
40
30
20
10
0
1439
2878
4317
5756
7195
8634
10073
11512
12951
14390
position(red | reprents an occurrance of CpG)

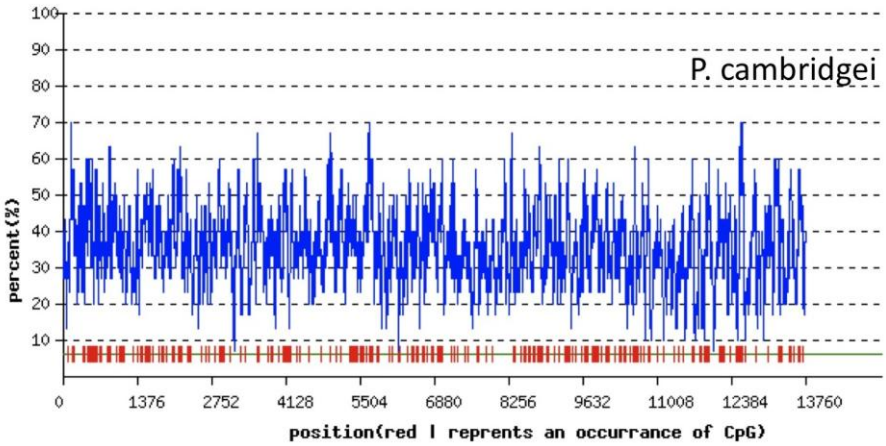

P. cambridgei
percent(%)
100
90
80
70
60
50
40
30
20
10
0
1376
2752
4128
5504
6880
8256
9632
11008
12384
13760
position(red | reprents an occurrance of CpG)

Figure 3.6. Sliding window plots of G/C content (percent %) across the complete mitochondrial genome of four Araneae taxa. The taxon name is indicated in the top right corner of each plot. The analyses were performed using the entire genome sequence with a window size of 30 and step size of 30 (http://www.bitgene/cgi/gene_analysis.cgi)

3.4.4 Mitochondrial genome annotation

The complete mitochondrial genomes of *Meta menardi*, *Pisaura mirabilis*, *Psalmopoeus cambridgei* & *Zelotes apricorum* were initially annotated using the online pipeline MITOS (http://mitos.bioinf.uni-leipzig.de). After subsequent reorientation (to best ensure gene order homology of linear sequences), gene boundaries were amended via BLAST alignment with previously published Araneae mitochondrial genomes (GenBank accession numbers: AY309258, AY309259, AY452691, NC_005942, NC_010777, NC010780). All 37 expected genes were found to be present in each of the new Araneae mitochondrial genomes (Boore, 1999; Masta and Boore, 2004), but a novel position of the tRNA Ser2 gene was observed in the mitochondrial gene order of *Meta menardi* (Figure 3.7.), as well as a major truncation of the 5' end of the region coding for NADH dehydrogenase 1.

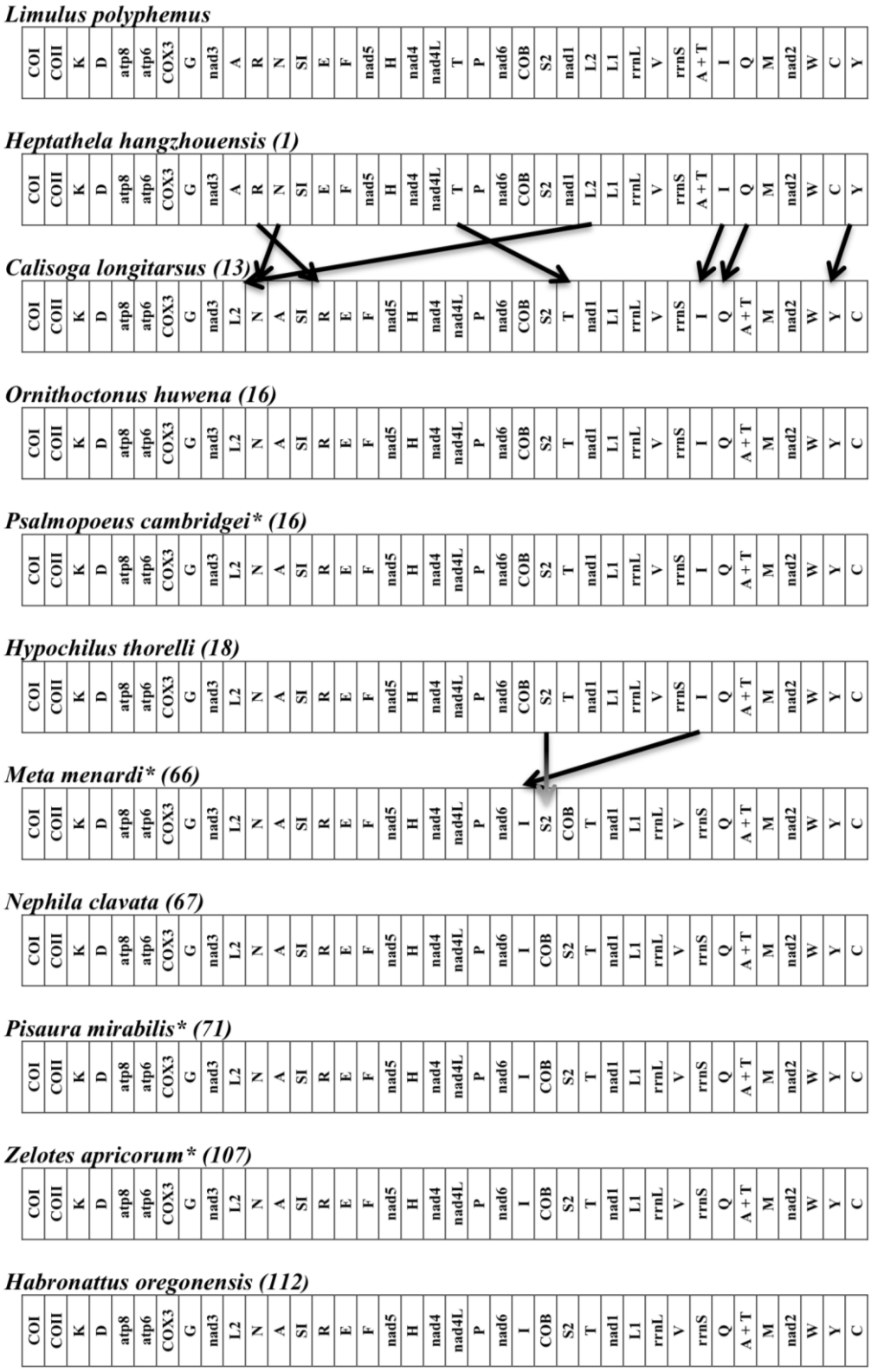


Figure 3.7. Gene order changes in sequenced Araneae mitochondrial genomes. Comparison of complete mitochondrial gene orders of all sequenced spider mitochondrial genomes to date, as well as the most basal member of the chelicerates (*Limulus polyphemus* – American Horseshoe crab). Black arrows

represent gene translocations between taxa with the graduated arrow representing a gene order change found only in *Meta menardi* (Tetragnathidae). Numbers in brackets represent taxonomic order of family according to the World spider catalogue (Platnick, 2013). Taxa marked with * have been sequenced as part of this study

3.5. Discussion

The Illumina MiSeq sequencing step resulted in high coverage of complete mitochondrial genomes assembling in one or two fragments. The high levels of coverage suggest that a large number of individual genomes can be sequenced on a single MiSeq run. Although all rolling circle reactions performed were successful (presence of band over 10kb in length could be visualised) it is impossible to tell, prior to sequencing, exactly how much of the RCA product is mitochondrial in origin (Simison et al., 2006). In published RCA studies, the use of restriction enzymes are used to isolate the mitochondrial DNA, potentially from nuclear or extracellular nucleic acids. In the trialled spiders however, the enzyme restrictions failed to produce any visible fragments smaller than the complete mitochondrial genome. Further restriction mapping analysis of the four mitochondrial genomes sequenced for this study has shown that the site for EcoR1 is seemingly pervasive throughout Araneae mitochondrial genomes. As the initial tests suggested, EcoRI restriction sites were found in each of the newly sequenced mitochondrial genomes. With the exception of *Pisaura mirabilis,* which contained only one, three restriction sites were found in each mitochondrial genome. The reason for the failure of the restriction enzyme digest to produce visible bands in the presence of mitochondrial DNA is unclear, however, background contamination and low total concentrations of fragmented mitochondrial DNA within the rolling circle

amplification product are most likely contributing factors. It has previously been suggested that background contamination, responsible for the visualised smear through gel electrophoresis, only contains a minimal amount of chromosomal DNA (Tang and Hyman, 2005). Evidence from this study corroborates such findings as the majority of the assembled contigs were of bacterial and viral origin (Table 3.5.). The reference assembled RCA amplified sampled (Table 3.7.) were identified as being spider mitochondrial DNA following comparison to the NCBI database via a BLAST search. Whilst this is unsurprising considering the fragments were matched to spider mitochondrial DNA through the assembly process, high percentage identities (>88%) to taxa that were not the reference sequence supports the notion that the sequences are not artificial. However, they are highly fragmented and of little utility for further analysis without independent corroboration that they are not chimeras.

Following *de novo* assembly, of PCR amplified mitochondrial genomes, large differences in average total coverage were observed between contigs (Table 3.4.), despite equimolar pooling prior to sequencing and the generation of high quality data from the MiSeq run (Figure 3.5.). Whilst the high coverage achieved for each mitochondrial genome should negate the incorporation of sequencing error (Dohm et al., 2008), Illumina sequencing platforms are susceptible to under-representation of regions of extreme base composition (A/T or G/C rich amplicons) throughout a run, due in part to the standardised reaction conditions necessary for massively high-throughput sequencing (Aird et al., 2011). Although no substantial variation was found that would explain the occurrence of c. 2000 bp contigs exhibiting high coverage (Figure 3.6.).

The four assembled mitochondrial genomes were 14,222, 14,399, 14,013 and 13,767 base pairs in length (*Zelotes apricorum*, *Meta menardi*, *Pisaura mirabilis*

and *Psalmopoeus cambridgei* respectively) **.** Whilst these figures are in alignment with the length of previously published Araneae mitochondrial genomes (13,874 – 14,456 bp), the 13,767 bp *Psalmopoeus cambridgei* genome is the shortest spider mitochondrial sequence currently known (de Meo et al., 2012). This represents a difference of over 100 bases in length when compared to the mitochondrial genome of *Ornithoctonus huwena*, a member of the same family (*Theraphosidae*). Whilst this difference is smaller than those found within single families of other chelicerates such as the Acari (mites and ticks) (Gissi et al., 2008), it still highlights the extreme diversity of spider mitochondrial genomes.

Despite the successful reconstruction of the PCR amplified mitochondrial genomes, mitogenomic sequences could not be recovered *de novo* from the sequenced RCA products. However, complete mitochondrial genomes were successfully assembled, for three taxa (*Meta menardi, Pisaura mirabilis* & *Psalmopoeus cambridgei*), using their PCR amplified counterpart as a reference mapping template, thus highlighting the current limitations of assembly algorithms in the *de novo* reconstruction of short-read data (Alkan et al., 2011).

One confounding factor known to affect the assembly of RCA amplified material is the formation of chimeric sequences, not only through the Illumina sequencing process (Kircher et al., 2011) but as an artefact of multiple displacement amplification reactions (Lasken and Stockwell, 2007). Low total coverage of the mitochondrial genome coupled with the formation and presence of chimeric sequences creates conflicting reconstructions that would be filtered out of the assembly, thus highlighting the need for improved mitochondrial isolation techniques to allow for an enriched template (Lang and Burger, 2007).

Whilst the MITOS annotation tool was able to approximately locate most of the protein coding genes for each of the four fully assembled mitochondrial

genomes, less than half of the 22 tRNAs had been found using the MITFI algorithm for the creation of structure annotated covariance models (Juehling et al., 2012a). However, this was in part caused by the inability of the software to allow the incorporation of highly truncated and overlapping tRNA sequences that have been reported in some Metazoan mitochondrial genomes (Masta, 2000; Yamazaki, 1997). Other popular tRNA predicting algorithms including; ARWEN (Laslett and Canback, 2007), tRNAscan-SE (Lowe and Eddy, 1997) and DOGMA (Wyman et al., 2004) were also used in order to locate tRNA secondary structures within the complete mitochondrial genome sequences. However, despite repeated attempts at optimisation via the manipulation of cove scores, as has proved successful with truncated nematode tRNAs, none of these methods were as successful as the MITFI algorithm. Until these limitations are addressed, the bioinformatic processes essential to understanding the output of ever improving Next-Generation Sequencing Platforms, could create a bottleneck in the production of large divergent mitogenomic datasets (Shendure and Ji, 2008).

My results provide evidence of the successful amplification and sequencing of several complete mitochondrial genomes by both long-range PCR, in one and two fragment approaches, and rolling circle amplification. I have highlighted the potential ability of multiple displacement amplification reactions, such as RCA, in creating divergent mitochondrial genome datasets with low concentrations of starting material from somatic tissues; the first time this has been achieved for Metazoans outside of the phylum Nematoda (Simison et al., 2006). Whilst RCA offers an inexpensive and expedient solution to the problems associated with utilising Long-range PCR for mitogenomics, it is clear that significant obstacles still remain in the recovery of low coverage sequences without adequate references from which reads can be mapped (Alkan et al., 2011).

4. Araneae mitochondrial genomes; A comparative analysis

4.1. Abstract

Animal mitochondrial genomes are small, circular DNA molecules with length ranging from ca. 10,000 bp to over 20,000 bp usually encoding 37 genes (13 protein-coding, 22 transfer RNA, and 2 ribosomal RNA genes). The number of complete mitochondrial genomes from a range of taxa is steadily rising following the widespread introduction of Next-Generation Sequencing technologies. The increasing availability of complete mitochondrial genome data invites comparative study and provides a large amount of nucleotide data that is useful for both deep and shallow phylogenetic studies. In addition to sequence data mitochondrial genomes possess a number of evolutionarily interesting features such as length variation, non-canonical transfer RNA secondary structures, atypical start codons, base compositional bias and transfer RNA gene rearrangement. Studies have shown these features appear to be lineage specific, but this insight can only be obtained from comparative analysis at various taxonomic levels. The Araneae are an exceptionally diverse order with over 43,000 described species. However, very little is known of their evolutionary history. Here I present a comparative analysis of the complete mitochondrial genomes of ten spiders (featuring four sequenced as part of this study) in order to assess their utility in providing characters for phylogenetic analysis and elucidating relationships within the Araneae.

4.2. Introduction

Mitochondria are semiautonomous organelles, dependent on the cell to replicate, with the universally recognised function of producing cellular adenosine triphosphate (ATP) by the process of oxidative phosphorylation (Handa, 2003). This role is well conserved throughout eukaryotic cells; animal, fungi and plant, however, substantial structural variation exists between taxa and some organisms have lost their mitochondrial genome entirely in favour of a non-respiring, mostly anaerobic lifestyle (Williams and Keeling, 2003). Mitochondrial genomes vary significantly in size ranging from approximately 10kb in Ctenophores (Pett et al., 2011) to over 2000 kb in length in some seed bearing plants (Alverson et al., 2010). Whilst remaining functionally constrained, genes have been both gained and lost throughout evolutionary history, with the number of protein coding genes varying from 3 to 67, while tRNA gene content ranges from 0 to 27 (Adams and Palmer, 2003). Mitochondrial genes can be classified into three distinct groups; protein coding genes, transfer RNAs (tRNAs) and ribosomal RNAs (rRNAs) (Wolstenholme, 1992). Although the protein coding and tRNA gene content varies substantially across lineages, both the large and small ribosomal subunits are universally present in all sequenced mitochondrial genomes, varying only in size and complexity, with some studies revealing a fragmented gene sequence that is post-transcriptionally assembled in certain taxa (Milbury et al., 2010). It is the differences in apparent structure and order that make mitochondrial genes evolutionary interesting features and potential tools for delineating evolutionary relationships (Cameron, 2007; Cooper, 2001; Cerasale, 2012).

The Arthropoda are one of the most species rich Metazoan phyla (Black and Roehrdanz, 1998) and to date, over 500 complete mitochondrial genome sequences have been published from Arthropods (de Meo et al., 2012), representing most of the major extant sub-phyla and orders. Whilst mitochondrial gene order has been found to be mostly well conserved across arthropods as a whole, some orders exhibit repeated, widespread rearrangements of tRNA genes (Boore, 1995).

Including the four sequences generated from this study, there are now 10 available complete mitochondrial genomes from the Araneae, representing a total of 9 separate families under current taxonomic placements (Platnick, 2012). Prior to this study, the only analysis of the features encoded in the mitochondrial genomes of spiders has taken place at the level of tRNA secondary structure (Masta, 2000a; Masta and Boore, 2004, 2008). It is well documented that some metazoan lineages, including the Araneae, lack the typical tRNA cloverleaf shape associated with the function of the molecule (Yamazaki, 1997; Lavrov, 2000; Wolstenholme, 1987; Montiel, 2006; Keddie, 1998). However, these 22 genes only comprise a small amount of the approximately 14,000 base pairs that comprise the complete spider mitochondrial genome.

Here I take a comparative approach to the analysis of the structures and rare genomic changes that are present in the currently available data. The direct comparison of mitogenomic data from multiple families provides unique insights into the evolution of the mitochondrial genome. All ten of the available spider mitochondrial genomes are used to emphasize unique features and shared characteristics and to highlight particular parts of the mitochondrial DNA that

has caused problems for downstream bioinformatic processes such as annotation.

4.3. Mitogenome organisation and gene content

The mitochondrial genomes of spiders range in size from 13,787 bp in length (this study) to 14,436 (*Nephila clavata*, *Lee et al.* unpublished) (Table 4.1.). All ten genomes are circular and code for 13 protein-coding genes, 22 tRNAs, 2 rRNAs and a large non-coding region that are found in all arachnid mitochondrial genomes (Figures 4.1.-4.4.). Despite pronounced variation in overall size of the mitochondrial genomes there is very little difference between the lengths of individual protein coding genes and the total length of all protein coding sequences. This intra-familial conservation in comparative gene size, suggests that although Araneae protein coding genes are substantially smaller than those found in more basal members of the Chelicerata, gene size is well conserved throughout spider lineages.

The gene order of the most basal of the Araneae, *Heptathela hangzhoensis* (sub-order Mesothelae) is identical to that of the Horseshoe crab (*Limulus polyphemus*) (Staton et al., 1997) as well as whip spiders (Solifugae) (Fahrein, 2009) and soft-bodied ticks of the order Acari (Shao et al., 2004) and so is considered to be the ancestral state of chelicerates (Masta et al., 2008). The currently sequenced mitochondrial genomes of spiders show that there have been at least nine mitochondrial gene translocations since the divergence of the Mesothelae and Opisthothelae over 300 million years ago (see Chapter 3; Figure 3.7.).

Table 4.1. Taxon information and length of protein coding genes (bases) in all ten complete mitochondrial genomes. PRC shows the total number of bases in each mitochondrial genome that code for proteins. Total genome represents the entire length of the fully sequenced mitochondrial genome. Grey shaded taxa were sequenced as part of this study.

Taxon name	**Gene length**														
	CO1	**CO2**	**CO3**	**Cytb**	**ATP6**	**ATP8**	**ND1**	**ND2**	**ND3**	**ND4**	**ND4L**	**ND5**	**ND6**	**Total PRC**	**Total genome**
Heptathela hangzhouensis	1,533	667	774	1,122	661	150	915	963	330	1,300	282	1,656	432	10,785	14,215
Ornithoctonus huwena	1,536	669	784	1,135	669	153	912	921	322	1,278	277	1,639	429	10,724	13,874
Psalmopoeus cambridgei	1,536	664	783	1,135	669	150	918	931	325	1,291	259	1,645	418	10,724	13,767
Calisoga longitarsis	1,536	666	784	1,135	666	150	915	934	336	1,288	259	1,642	427	10,738	14,070
Hypochilus thorelli	1,537	654	786	1,131	666	153	918	960	345	1,287	259	1,633	424	10,753	13,991
Meta menardi	1,533	669	786	1,134	669	156	885	960	340	1,270	271	1,641	438	10,752	14,399
Nephila clavata	1,536	663	784	1,131	663	159	906	948	333	1,275	270	1,635	426	10,729	14,436
Pisaura mirabilis	1,533	669	780	1,104	660	159	912	964	358	1,283	268	1,639	425	10,754	14,013
Zelotes apricorum	1,512	663	791	1,105	660	172	924	966	331	1,279	271	1,639	427	10,740	14,222
Habronattus oregonensis	1,542	666	786	1,105	660	153	921	960	350	1,288	268	1,636	429	10,764	14,381

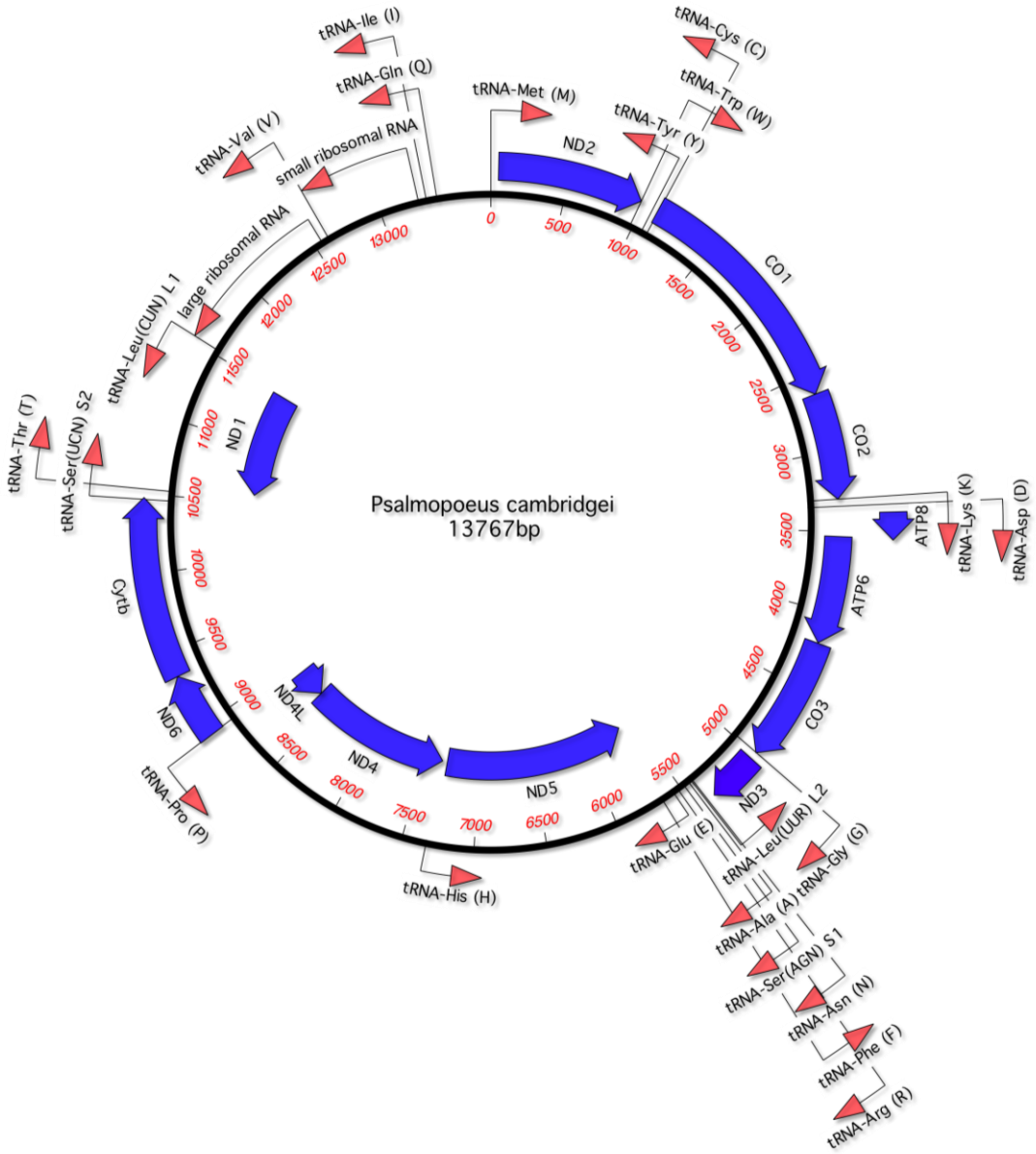


Figure 4.1. Araneae mitochondrial genomes displaying annotations. Spider mitochondrial genome of *Psalmopoeus cambridgei* displaying all gene annotations. Solid arrows represent protein coding genes, tRNAs and rRNAs appear as line arrows. In all instances the direction of the arrow indicates gene sequence 5' – 3' orientation.

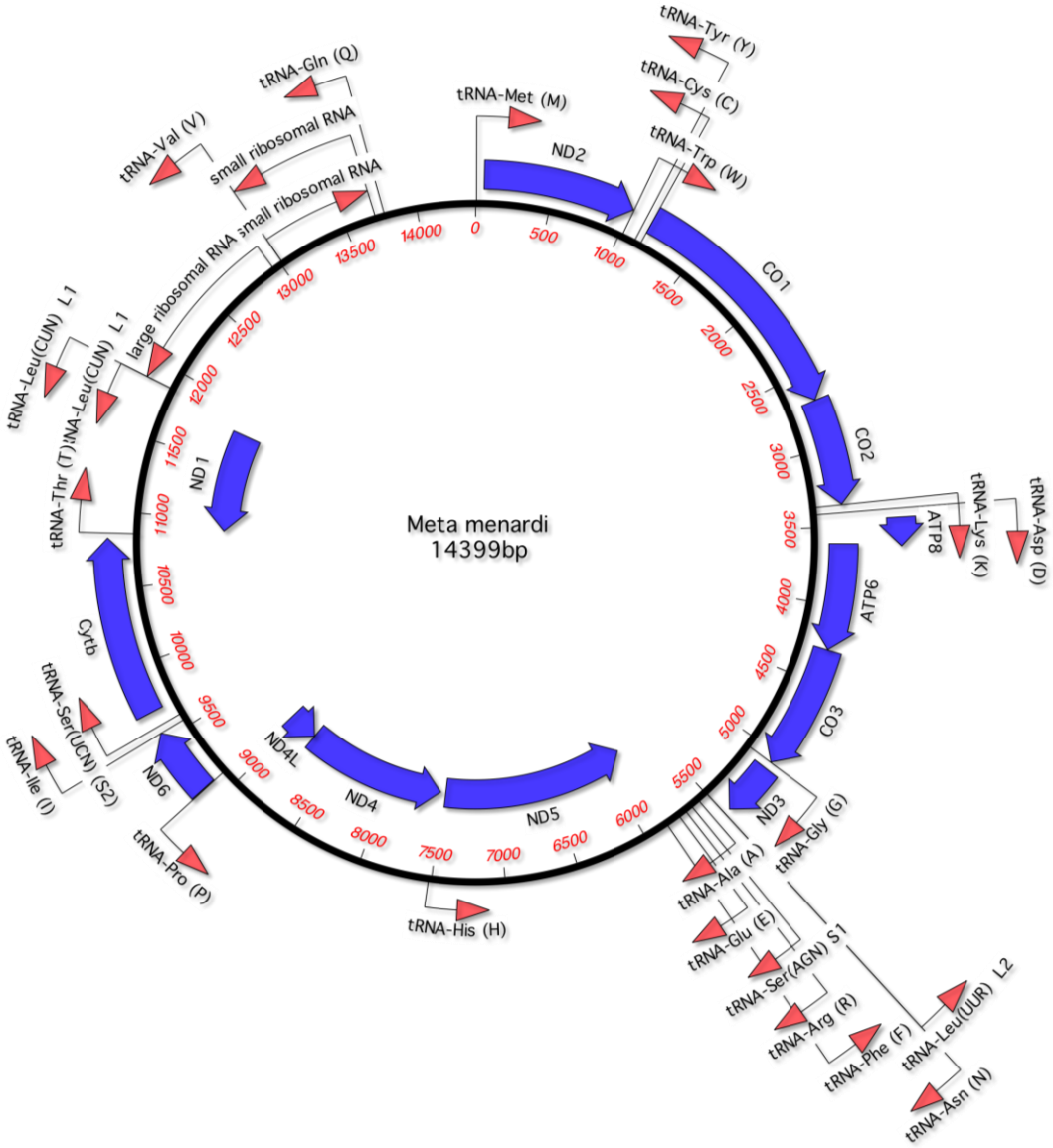


Figure 4.2. Araneae mitochondrial genomes displaying annotations. Spider mitochondrial genome of *Meta menardi* displaying all gene annotations. Solid arrows represent protein coding genes, tRNAs and rRNAs appear as line arrows. In all instances the direction of the arrow indicates gene sequence 5' – 3' orientation.

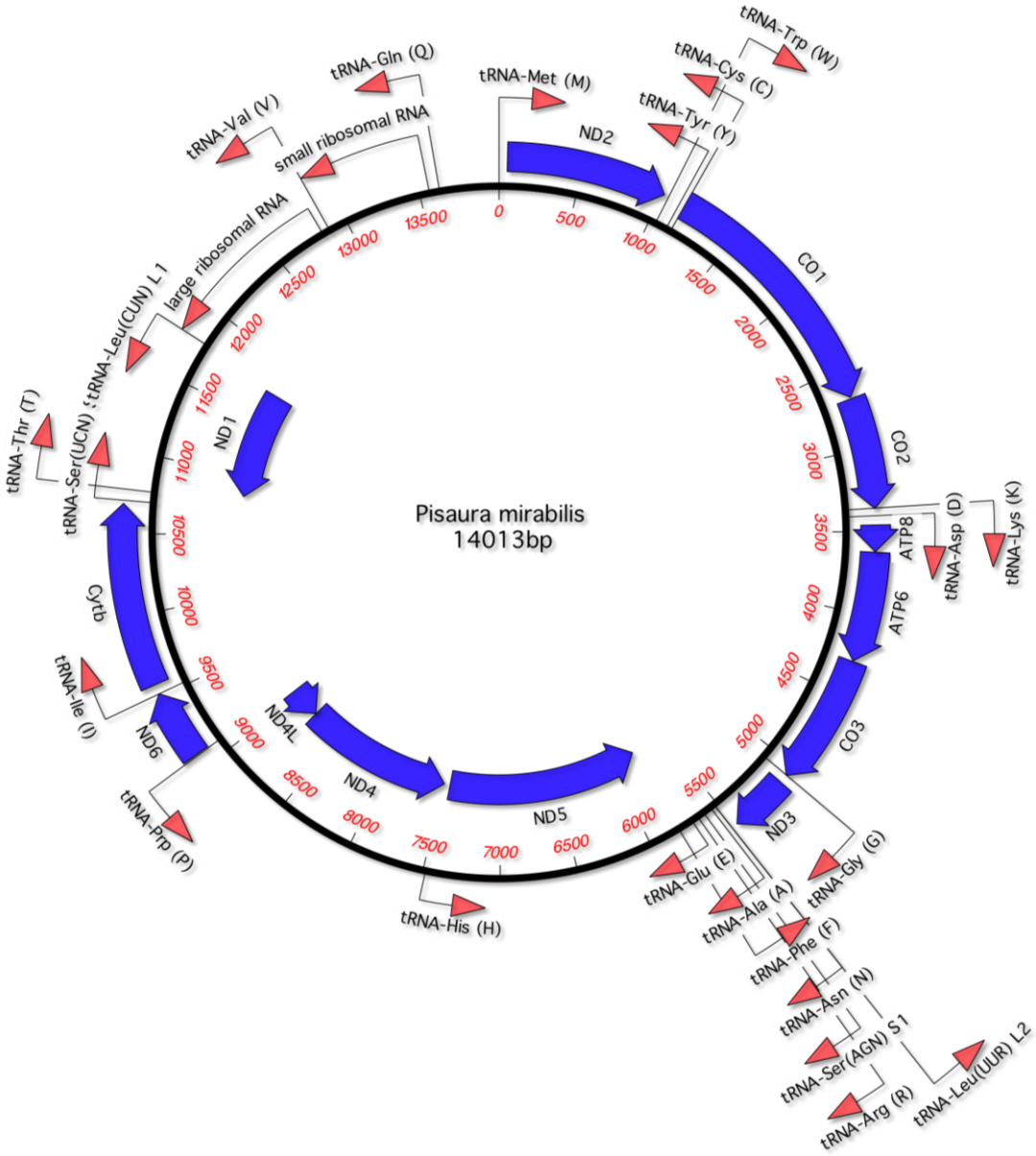


Figure 4.3. Araneae mitochondrial genomes displaying annotations. Spider mitochondrial genome of *Pisaura mirabilis* displaying all gene annotations. Solid arrows represent protein coding genes, tRNAs and rRNAs appear as line arrows. In all instances the direction of the arrow indicates gene sequence 5' – 3' orientation.

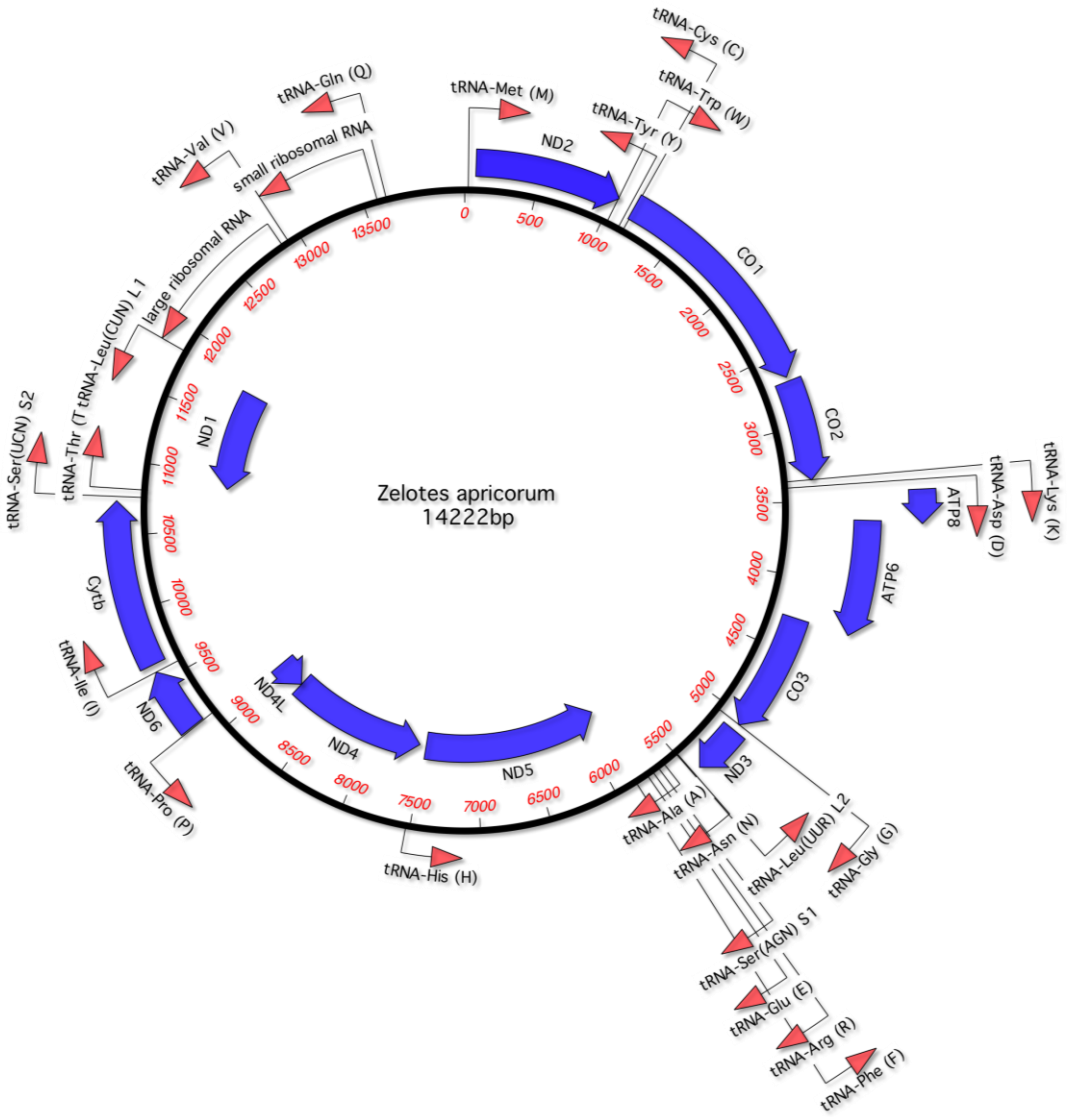


Figure 4.4. Araneae mitochondrial genomes displaying annotations. Spider mitochondrial genomes of *Zelotes apricorum* displaying all gene annotations. Solid arrows represent protein coding genes, tRNAs and rRNAs appear as line arrows. In all instances the direction of the arrow indicates gene sequence 5' – 3' orientation.

Changes in gene order have been confined strictly to transfer RNAs (tRNA) genes and none have involved changes from one strand of mitochondrial DNA (light or heavy) to the other (strand transposition). Unlike the larger mitochondrial genomes found in some metazoans and higher plants (Alverson et al., 2011; Alverson et al., 2010) there are only very small non-coding regions within the spider mitochondrial genomes (Table 4.2.). The number of non-coding regions present in the spider mitochondrial genomes varies from 3 to 11 (including the putative control region) and account for most of the variation in size between genomes. Most of these spacers consist of 2 to 10 nucleotides between adjacent genes, however, each spider mitochondrial genome has one non-coding region substantially larger than the inter-genic spacers, at a position in the genome where the A-T rich control region occurs in the Horseshoe crab, *Limulus polyphemus*. The large non-coding sections are not substantially more A-T rich, in comparison to the rest of the mitochondrial genome, as they have been found to be in the Insecta (Zhang, 1995, Zhang, 1997). However, due to it's relative positioning within the mitochondrial genome the large non-coding region is thought to be analogous to the putative insect 'Control Region' despite the apparent size reduction exhibited in spiders (Masta and Boore, 2004).

Table 4.2. Occurrence and size of inter-genic sequences (non-coding regions) in sequenced Araneae mitochondrial genomes. Non-coding regions shows the number of instances of non-coding regions within each mitochondrial genome. The length of the largest spacer is shown as its length in bases along with A-T content (%).

Taxon name	**Family**	**Non-coding regions**	**Largest Non-coding region**	**A-T Content**
Heptathela hanzhouensis	Liphistiidae	11	340	80.6
Ornithoctonus huwena	Theraphosidae	3	396	68.2
Psalmopoeus cambridgei	Theraphosidae	3	371	66.6
Calisoga longitarsis	Nemesiidae	8	478	66.3
Hypochilus thorelli	Hypochilidae	4	503	77.3
Meta menardi	Tetragnathidae	11	629	80.4
Nephila clavata	Nephilidae	8	670	81.2
Pisaura mirabilis	Pisauridae	8	385	82.7
Zelotes apricorum	Gnaphosidae	6	584	75.8
Habronattus oregonensis	Salticidae	9	716	77.1

4.4. Gene content

In order to compare the changes in nucleotide content across Araneae mitochondrial genomes the relative G-C and A-T compositions were calculated, along with G-C and A-T skew (differences in ratios between the light and heavy strands) (Table 4.3.). Arthropod mitochondrial genomes are known to be particularly A-T rich (Beckenbach, 2011; Hu et al., 2010; Stewart and Beckenbach, 2005; Wilson et al., 2000), although this is not consistently found to be the case within the mitogenomes of the Araneae. The sequenced mygalamorph genomes (*Ornithoctonus huwena, Psalmopoeus cambridgei,*

Calisoga longitarsis) have G-C contents above 30%, however, the more basal representative of the Mesothelae (*Heptathela hangzhouensis*) is substantially more AT rich, as are the more recently diverged spiders (<74%). This therefore suggests that a lower mitochondrial G-C content is an ancestral trait and that a reduction in genome size is linked with the loss of A and T nucleotides from the non-coding regions. The only exception to this would appear to be *Meta menardi*, which has the largest recorded spider mitochondrial genome (14,399 bp) yet exhibits a relatively high G-C content of 30%. The G-C content exhibited across spider mitochondrial genomes varied more within genes than it did between the genes of each taxa. The asymmetric usage of the four base pairs between the major coding (heavy) and minor coding strand (light), termed G-C skew, was found to be similar in almost all spider mitochondrial genomes. The G-C skew was positive (approx 0.25), showing guanine to be preferentially located on the heavy strand and the A-T skew was found to be negative (approx -0.1) indicating more adenine in the light strand, with *Heptathela hangzhouensis* exhibiting a heavy strand bias towards cytosine.

Metazoan mitochondrial genomes typically possess an overall Cytosine and Adenine (C-A) nucleotide bias on their heavy strand (Saccone et al., 1999). This C-A bias is believed to result from the asymmetrical way mitochondrial genomes replicate, during which the two strands of DNA are single-stranded for differing periods of time, leaving one strand more prone to mutations. This is known to create a mutational bias in the occurrence of deaminations of cytosine and adenine (Faith and Pollock, 2003; Reyes et al., 1998). Arthropods, however, exhibit a reverse nucleotide bias to most metazoans resulting from an inversion of the origin of replication (Hassanin, 2006).

Table 4.3. Taxon information and G-C content of protein coding genes (percentage) in all ten complete mitochondrial genomes. PRC shows the G-C content in each mitochondrial genome for the entire protein-coding region. Total DNA represents the G-C content of the fully sequenced mitochondrial genome. Skew represents strand bias of nucleotides expressed as (G-C)/(G+T) or (A-T)/(A+T).

Taxon name	**G-C content**															**GC skew**	**AT Skew**
	CO1	**CO2**	**CO3**	**Cytb**	**ATP6**	**ATP8**	**ND1**	**ND2**	**ND3**	**ND4**	**ND4 L**	**ND5**	**ND6**	**Total PRC**	**Total DNA**		
Heptathela hangzhouensis	32.5	32.1	32.3	31.3	26.6	25.5	26.0	25.4	22.4	27.2	25.9	28.5	21.3	28.5	27.8	-0.23	-0.02
Ornithoctonus huwena	33.6	31.5	35.5	30.8	28.8	37.3	29.5	27.8	30.1	28.8	28.9	29.1	24.9	30.4	30.2	0.34	-0.08
Psalmopoeus cambridgei	39.1	39.0	36.4	35.7	35.7	31.3	36.8	36.0	36.6	36.0	34.0	37.8	35.2	36.8	35.9	0.36	-0.09
Calisoga longitarsis	40.0	39.8	39.0	35.8	34.8	26.0	33.7	36.6	33.3	35.3	38.2	37.2	35.8	36.7	36.0	0.36	-0.14
Hypochilus thorelli	36.3	33.5	33.7	31.4	29.9	22.9	30.3	29.5	29.3	28.7	32.0	29.9	22.9	31.0	29.7	0.27	-0.14
Meta menardi	36.4	36.0	38.4	36.2	32.1	33.3	34.2	31.8	28.2	31.6	28.8	32.5	34.3	33.9	32.0	0.22	0.00
Nephila clavata	30.3	27.2	29.0	26.4	27.1	14.5	24.0	22.2	21.0	22.0	19.6	23.4	23.7	25.1	24.0	0.24	-0.05
Pisaura mirabilis	27.3	23.5	25.0	22.6	21.4	13.2	18.3	16.7	14.8	18.3	17.5	19.0	16.2	20.7	20.0	0.17	-0.08
Zelotes apricorum	30.0	29.7	28.2	28.1	24.1	20.4	26.0	22.8	23.3	24.9	21.8	24.3	22.5	26.0	25.3	0.30	-0.14
Habronattus oregonensis	30.0	29.0	27.7	27.2	26.4	21.6	24.9	26.1	19.3	24.2	19.8	26.0	21.7	26.1	25.6	0.30	-0.11

The codon usage exhibited across spider mitochondrial genomes has been reported before (Masta and Boore, 2004) with an ATN (where N represents A,T,G or C) initiation codon favoured the most, with a TTG start codon proposed for some genes which would avoid coding sequence overlap. However, as has been in found in Coleoptera, without an explicit RNA expression study it is impossible to determine exactly where *each protein-coding gene* starts (Sheffield et al., 2008). The most frequent termination codons used by spider mitochondrial genomes were TAG and TAA, which was again characteristic of arthropod mitochondrial genomes. Several genes, nad4, nad4L, nad5 and Cytb, did not contain these stop codons for all taxa and it has been previously inferred that these genes code for truncated stop codons (T), polyadenylated after transcription to form TAA stop codons (Masta and Boore, 2004; Ojala et al., 1981).

4.5. Transfer RNA secondary structure

Mitochondrial transfer RNA (tRNA) gene sequences have been the focus of many recent studies due to translocation throughout the mitochondrial genome and the atypical secondary structure that is found in some metazoan lineages (Juehling et al., 2012a; Juehling et al., 2012b; Meer et al., 2010; Segovia et al., 2011). Both these features are found in the mitochondrial tRNAs of spiders and 4 types of secondary structure have been inferred from arachnid mitochondrial genomes (Masta and Boore, 2008). The typical configuration of metazoan tRNAs are to form a cloverleaf shape, however, tRNAs have also been found which lack a T-arm, lack a D-arm, and those that lack sequence that are able to encode a paired acceptor stem and a T-arm (Figure 4.5.).

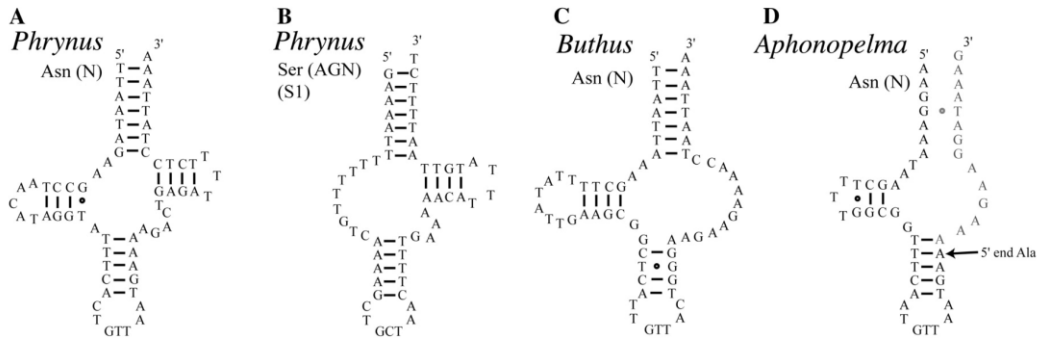


Figure 4.5. Examples of the 4 types of inferred tRNA secondary structures found in arachnids. (A) Cloverleaf structure, as found in tRNAAsn in the amblypygid *Phrynus*. (B) Structure lacking a D-arm, as found in tRNASer(AGN) in the amblypygid *Phrynus*. (C) Structure lacking a T-arm, as found in tRNAAsn in the scorpion *Buthus*. (D) Structure lacking both a T-arm and a well-paired aminoacyl acceptor stem, as found in tRNAAsn in the spider *Calisoga*. The sequence in grey in (D) (AAAGAAGGATAAAGT) encodes the downstream gene.

The 4 described secondary structures of mitochondrial tRNAs were found in all of the ten spider mitochondrial genomes and no new structures were discovered (Table 4.4.). Similarly all secondary structures were noted on both strands (heavy and light), thus, it does not appear that the strand on which a tRNA is encoded influences whether it maintains a T arm and 3′ acceptor stem sequence.

Table 4.4. Inferred secondary structure of mitochondrially enconded tRNAs in the Araneae. Numbers 1-3 represent the configuration of secondary structure; 1 – cloverleaf, 2 – structure lacking D-arm and 3 – structure lacking T-arm.

Taxon name	**Transfer RNA genes**																				
	A	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**K**	**L1**	**L2**	**M**	**N**	**P**	**Q**	**R**	**S1**	**S2**	**T**	**V**	**W**	**Y**
Heptathela hangzhouensis	3	3	3	1	3	3	3	1	1	1	1	1	1	1	1	1	2	3	1	1	3	1
Ornithoctonus huwena	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	2	2	1	3	3	3
Psalmopoeus cambridgei	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	2	2	1	3	3	3
Calisoga longitarsis	2	3	3	1	3	3	3	3	3	3	3	3	3	1	1	3	2	2	1	3	3	3
Hypochilus thorelli	3	3	3	3	3	3	3	3	3	3	3	3	3	1	1	3	2	1	1	3	3	3
Meta menardi	1	3	3	1	3	3	3	3	3	3	3	3	3	3	1	3	2	2	1	1	3	1
Nephila clavata	1	3	3	1	3	3	1	1	3	3	3	3	3	3	1	3	2	2	1	1	3	1
Pisaura mirabilis	1	3	3	1	3	3	1	1	3	3	3	3	3	3	1	3	2	2	1	1	3	1
Zelotes apricorum	1	3	3	1	3	3	1	1	3	3	1	3	3	3	1	3	2	2	1	1	3	1
Habronattus oregonensis	1	3	3	1	3	3	1	1	3	3	1	3	1	3	1	3	2	1	1	1	3	1

The number and identity of mitochondrially encoded tRNAs that have become truncated varies substantially throughout the Araneae. However, the variation of losses of one or more structural features can be mapped onto a phylogeny to show that the occurrence of a reduction in tRNA secondary structure is not seen to be reversible, although some reductions may have occurred independently through evolutionary time in different taxa, and that the trait can define synapomorphic clades (Figure 4.6.). This highlights the potential utility of transfer RNA secondary structures in phylogenetic reconstruction. From these data it can be inferred that the common ancestor of extant spiders most likely lost sequences to encode canonical secondary structures for multiple different mitochondrial tRNAs (Cys, Phe, Asp, Gly, Trp,) and that mygalamorph families (*Calisoga longitarsis, Ornithoctonus huwena* and *Psalmopoeus cambridgei*) have undergone further structural changes since the Mygalamorphae and Araneomorphae diverged (See Chapter 1 Figure 1.1.).

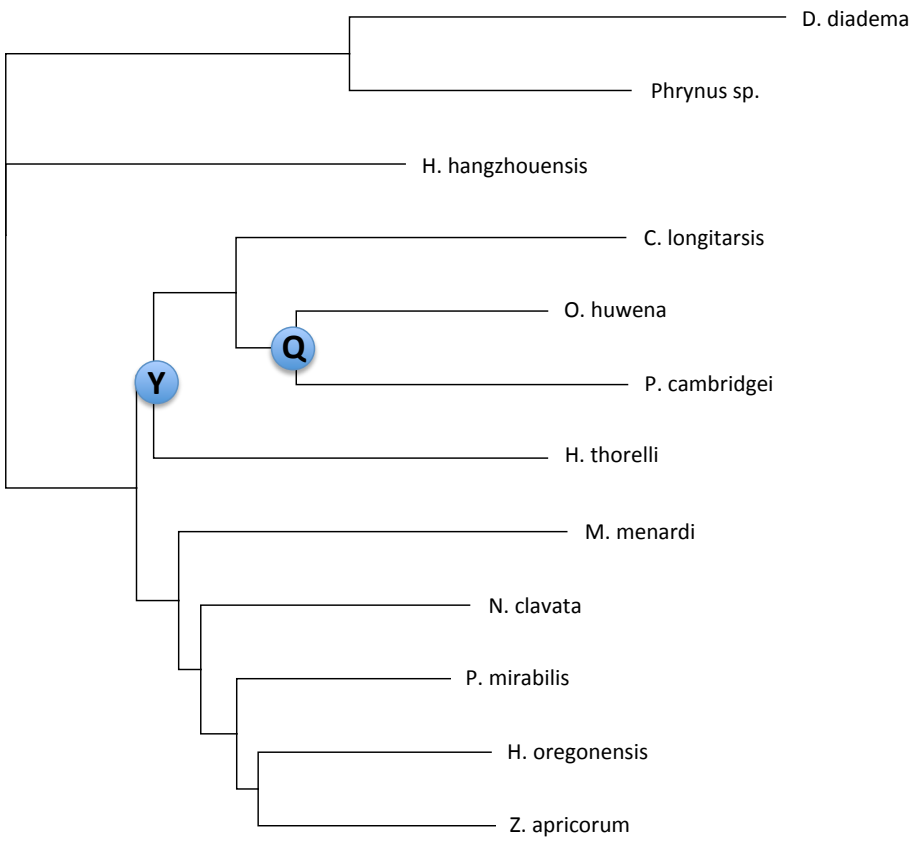


Figure 4.6. tRNA secondary structure changes mapped on to Araneae phylogeny. Preliminary neighbour-joining tree constructed from concatenated complete spider mitochondrial gene sequences (*Heptathela hangzhouensis, Calisoga longitarsis, Ornithoctonus huwena, Psalmopoeus cambridgei, Hypochilus thorelli, Meta menardi, Nephila clavata, Pisaura mirabilis, Habronattus oregonensis* and *Zelotes apricorum*) with two whip spiders (Ambypygi; *Phrynus* sp. and *Damon diadema*) as outgroups (Bootstrap values not shown). Circles over nodes represent change in secondary structure of tRNA through the loss of the T arm and letters show the identity of the gene.

4.6. Ribosomal RNAs

Ribosomal RNA sequences lengths were found to be similar throughout the spider mitochondrial genomes (Table 4.5.) Whilst G-C content varied substantially between taxa, the differences between the two subunits were very small within taxa (approximately 1-8% difference between the rrnL genes and 1-5% between the rrnS genes when compared to the total genome. Substantially less variation in G-C content was exhibited between rRNA encoding genes than between the protein coding genes analysed. Both the large and small ribosomal subunits are approximately 100bp shorter than that of the most basal chelicerate (*Limulus polyphemus*) (Lavrov et al., 2000a) but similar in size to the figures provided for mites (Black and Roehrdanz, 1998). Such a reduction in size of more recently diverged chelicerates would suggest an evolutionary mechanism that selected towards minimising the length of mitochondrial genes.

Table 4.5. Taxonomic information, length and G-C content of ribosomal RNA genes in all ten complete mitochondrial genomes. Length is shown as number of nucleotides in the coding sequence and G-C represents the residue total of Guanine and Cytosine as a percentage.

Taxon name	Family	Large ribosomal subunit		Small ribosomal subunit	
		Length	G-C	Length	G-C
Heptathela hangzhouensis	Liphistiidae	1119	24.4	698	27.8
Ornithoctonus huwena	Theraphosidae	1048	30.2	666	28.8
Psalmopoeus cambridgei	Theraphosidae	1021	30.8	662	32.9
Calisoga longitarsis	Nemesiidae	1062	32.7	691	34.0
Hypochilus thorelli	Hypochilidae	1041	24.6	688	24.9
Meta menardi	Tetragnathidae	1070	29.5	700	28.4
Nephila clavata	Nephilidae	1046	20.9	695	21.5
Pisaura mirabilis	Pisauridae	1043	16.8	688	16.3
Zelotes apricorum	Gnaphosidae	1020	20.9	690	21.0
Habronattus oregonensis	Salticidae	1018	21.3	691	22.6

4.7. Synthesis and overview

The mitochondrial genomes found within the Araneae represent some of the smallest reported from any chelicerate, with only mites (Acari; Acaraiformes) of the sub-order Prostigmata being shorter in length (Van Leeuwen et al., 2008; Yuan et al., 2010). The mitochondrial genomes of chelicerates are generally smaller than those recorded in other arthropods (Jeyaprakash and Hoy, 2007; Lavrov et al., 2004; Lewis et al., 1995; Masta and Boore, 2004) achieved by a reduction in size via the truncation of gene coding sequences. Spider mitochondrial genomes in

particular have achieved a compacted state via reduction in gene size and notably, the ribosomal RNA (rRNA) and transfer RNA (tRNA) genes. The tRNA genes seen in spider mitochondrial genomes are acutely reduced at their 3′ ends so that many no longer encode a T arm (Figure 4.5.) and it is plausible that most do not encode a 3′ acceptor stem (Masta and Boore, 2008). Thus, the tRNA genes found in spiders are even smaller than the truncated nematode mitochondrial tRNA genes, once described as the most minimal tRNA structure (Ohtsuki et al., 2002). The possibility that tRNAs overlap with adjacent genes seems unlikely and has already been refuted in the mitochondrial genome of *Habronattus oregonensis* (family Salticidae) (Masta and Boore, 2004). The most important evidence to strengthen this argument comes from the inability to find the necessary stable base pairs from which to form the 5′ acceptor stem in the surrounding sequences. The coding regions surrounding tRNA genes have been found to exhibit very little, or no potential, to form any of the required base pairs in those tRNAs that lack the stem. While there is not enough evidence to completely rule out the possibility that the truncated mitochondrial tRNAs could be non-functional, and that the mitochondrion instead utilizes nuclear coded tRNAs imported from the cytoplasm, this does not seem a plausible explanation. Although the mitochondrial importation of nuclear derived tRNAs has been described in some taxa (Schneider, 2011), it is not common for all 22 tRNA sequences to remain present in the mitochondrial genome. Many taxa that have evolved to rely on tRNAs imported in to the mitochondria have seen complete loss of tRNA coding sequences within the mitochondrial genome (Gray et al., 1998).

A plausible explanation for the widespread loss of the tRNA acceptor stem coding sequences found in spider mitochondrial genomes is the existence of a post-transcriptional RNA editing mechanism (Hiesel et al., 1989; Maier et al., 1996;

Schuster et al., 1993). Whilst other metazoan mitochondrial genomes, such as those in nematodes are also characterised by the loss of the T arm in some tRNAs, they still retain well-paired acceptor stem sequences (Watanabe et al., 1994; Wolstenholme et al., 1987). Both gastropods and centipedes have been found to exhibit acceptor stem coding sequence loss, to varying degrees, within their mitochondrial genomes and centipedes have been found to possess a novel mechanism for post-transcriptional editing (Lavrov et al., 2000b). This would suggest that such mechanisms have arisen independently. However, confirmation of the existence of an editing mechanism in spiders will require the sequencing and analysis of the processed tRNA transcripts.

The relationships amongst the Araneae are poorly resolved and very little work has been conducted at the family level and above (Blackledge et al., 2009; Coddington et al., 2004; Coddington and Levi, 1991). The comparative analysis of complete mitochondrial genomes, spanning more taxa than previous studies, has uncovered potential new markers from which to resolve contentious relationships, previously assessed using behavioural and morphological traits (Arnedo et al., 2009; Miller et al., 2010; Platnick et al., 2012). Analyses of the mitochondrial genomes of spiders have shown trends both in changes of gene order and alterations to the secondary structure of tRNAs. More taxa would be required to delineate relationships between the Araneae from gene order changes and structural characteristics alone. Although, the suggestions from other studies (Masta and Boore, 2004; Rokas and Holland, 2000) that such rare genomic changes, having occurred, are unlikely to return to the ancestral state, makes them intriguing characters for resolving higher level topological incongruence. The information provided from this study would suggest that the family Hypochilidae (*Hypochilus thorelli*) is more closely related to the mygalamorph spiders (*Calisoga*

longitarsis, Ornithoctonus huwena and *Psalmopoeus cambridgei*), as opposed to being the sister taxon of the Araneomorphae (*Meta menardi, Nephila clavata, Pisaura mirabilis, Habronattus oregonensis* and *Zelotes apricorum*), as previously believed (Coddington et al., 2004). Thus, providing further support for the molecular phylogenetic analyses, which whilst only preliminary here, will be discussed in more detail in Chapter 5.

The reduction in size of tRNA genes, ribosomal genes, and at least half of the protein-coding genes in the mitochondrial genome of the Araneae suggests there has been an overall trend toward minimization of mitochondrial genome size in spiders. The Araneae possesses some of the smallest known tRNA genes, and the reduction in size and number of non-coding regions within the two Tarantula mitochondrial genomes (*Ornithoctonus huwena* & *Psalmopoeus cambridgei*) also lends support to minimising overall mitochondrial genome size. Deleterious mutations have been found to be accumulating at a faster rate in metazoan mitochondrial tRNAs than in nuclear encoded tRNAs (Lynch, 1996). Whilst this is a contributing factor to the truncation of tRNA sequences and structures this could lead to issues with fitness if the proposed RNA editing mechanisms cannot co-evolve as rapidly and maintain function.

This study represents the first comprehensive comparative analysis across a broad range of complete spider mitochondrial genomes. I find that Araneae mitochondria deviate substantially from the proposed ancestral chelicerate arrangement, following the divergence of the Mesothelae and Opisthothelae, with no apparent gene deletions or duplications. There are several shared features that many spider families have in common, such as short intergenic spacers, truncated 'AT rich' control regions and atypical secondary structure of several tRNAs. I have

also demonstrated the importance of comparative analysis in understanding the evolution of the mitochondrial genome using the Araneae as a model order.

5. Mitogenomic phylogenetics and evolution in a fossil rich arthropod group: A case study within the Araneae (Spiders).

5.1. Abstract

The Araneae are one of the most diverse arthropod orders with fossils dating back over 300 million years. However phylogenetic relationships amongst families are primarily based on morphological distinctions and have long remained unclear. Recent molecular studies have presented alternate hypotheses and identified the misplacement of taxa within currently recognised families. Here I present a strongly supported phylogeny for spiders based on the complete mitochondrial genome sequences of ten extant spiders and two whip spiders (Amblypygi). The concatenated mitogenomic datasets provide a well-resolved topology for the Araneae, with the Mesothelae at the basal position within the order with the Mygalamorphae and Araneomorphae forming sister clades. This study shows strong support for the family Hypochilidae (Paleocribellate) to be considered more closely related to the mygalamorph spiders than the Araneomorph spiders, as previously thought. Molecular dating of the mtDNA divergence times suggests a rapid radiation of spiders between the end of the Jurassic and the start of the Cretaceous periods. This coincides with major radiations described in other taxa such as birds, frogs and the spread of flowering plants.

5.2. Introduction

The Araneae (spiders) are an exceptionally diverse order, with over 43,000 currently described species placed in 112 families (Platnick, 2012). However, very little is known about the origins of their diversification (Arnedo and Gillespie, 2006; Blackledge et al., 2009; Dimitrov et al., 2008), and inter-family relationships remain poorly resolved (Ayoub et al., 2007; Davila, 2003). Despite rapid advances associated with sequencing technologies (Shendure and Ji, 2008) and the phylogenetic analysis of large datasets (Dutilh et al., 2007; Lerner and Fleischer, 2010), higher-level systematics of spiders is still predominantly focused on the use of morphological and behavioural data (Coddington et al., 2012; Coddington and Levi, 1991; Griswold et al., 2005; Lopardo and Hormiga, 2008; Platnick et al., 2012). Whilst recent systematic studies below the genus level have incorporated more molecular characters (Bond et al., 2012a; Bond et al., 2012b; Murphy et al., 2006; Rix et al., 2010), spiders (and most arachnids) are still poorly characterised at the genomic level (Ayoub et al., 2007). The sequencing of complete spider mitochondrial genomes (Masta and Boore, 2004, 2008; Qiu et al., 2005) has shown a lack of conservation in mitochondrial gene order and in doing so have provided new characters for resolving high-level phylogenetic incongruence (Dong et al., 2012; Masta et al., 2010). Relatively few studies have utilised the nuclear genome of spiders resulting in only a small number of conserved nuclear markers currently available for systematics (Ayoub et al., 2007). Advances in sequencing technologies have allowed for the generation of expressed sequence tag (EST) libraries, however, few additional phylogenetic markers have been gained (Ayoub et al., 2007; Mattila et al., 2012).

The advent of Next-Generation Sequencing (NGS) platforms have allowed increasing numbers of characters to be incorporated into systematics without

limiting the number of taxa analysed (Straub et al., 2012). Despite increases in the numbers of available sequenced transcriptomes (Chiari et al., 2012; Hedin et al., 2012; McKain et al., 2012), the realisation of large, divergent phylogenomic analyses currently remains constrained by labour and cost despite recent developments in sequencing methodologies (Baird et al., 2008; Kakioka et al., 2013). Mitogenomics provides a potential alternative, where information from the complete mitochondrial genome can be used to create a robust phylogenetic framework (Shamblin et al., 2012). The metazoan mitochondrial genome typically comprises of 37 genes including 22 coding for transfer RNAs (tRNAs), 13 coding for proteins and 2 coding for ribosomal RNAs (rRNAs) (Boore, 1999). It is the smallest extant genome (Cameron, 2007), with typical sizes in spiders ranging from 13, 787 – 14,399 bp, and so is tractable to sequencing in its entirety whilst being an order of magnitude larger than most of the single genes used in current arachnid systematics (Arnedo et al., 2009; Bodner and Maddison, 2012). In spite of difficulties associated with the PCR-based amplification of divergent complete mitochondrial genomic datasets (Briscoe *et al.* 2013), recent studies have showcased their utility in resolving complex phylogenetic relationships (Bjork et al., 2011; Cameron et al., 2012; Morin et al., 2010). Whilst nuclear markers are often preferred over mitochondrial loci due to the effects of incomplete lineage sorting and hybridization of mitogenomes (Tang et al., 2012), mitochondrial partitions have been important in resolving evolutionary relationships in many evolutionary relationships that are prone to low branch support (Duchene et al., 2011). Mitogenomic phylogenies therefore have the potential to resolve hitherto poorly supported polytomies, whilst additionally providing strongly supported date estimates (Duchene et al., 2012; Wang and Yang, 2011); important attributes for any molecular phylogenetic locus.

Dating divergences and major radiations has played an increasingly important role in evolutionary biology, providing more accurate timing whilst offering insights in to many evolutionary phenomena (Ronquist et al., 2012). The precision of dating divergence times estimated from sequence data is dependent on accurate fossil records in order to correctly calibrate phylogenetic trees (Beilstein et al., 2010) and complete mitochondrial genomes of fossil rich taxa have been utilised to resolve discordant phylogenetic hypotheses (Duchene et al., 2012; Fulton and Strobeck, 2010).

The Araneae (spiders) present an ideal opportunity in which to explore the potential of utilising complete mitochondrial genomes in dating major divergences within the order using both sequence data and rare genomic changes. Spiders have the best documented fossil record of all arachnids (Dunlop and Penney, 2012) (Figure 5.1.) but due to the fragility of the morphological characters from which spiders can be identified, their occurrence in the fossil record is mostly limited to sites with exceptional preservation of fossil deposits (Konservat-Lagerstatten) such as amber, the highly polymerised form of tree resin (Saupe and Selden, 2011). However, spider fossils in sedimentary deposits date back as early as the Carboniferous (c. 358 mya) but the majority of fossilised samples recorded have come from Baltic and Dominican amber ranging from the Cretaceous (c. 145 mya) to the Neogene (c. 23 mya) periods (Penney et al., 2012a). These records cover approximately 1150 described species (Penney et al., 2012a) and provide potential calibration points from which to date molecular phylogenies.

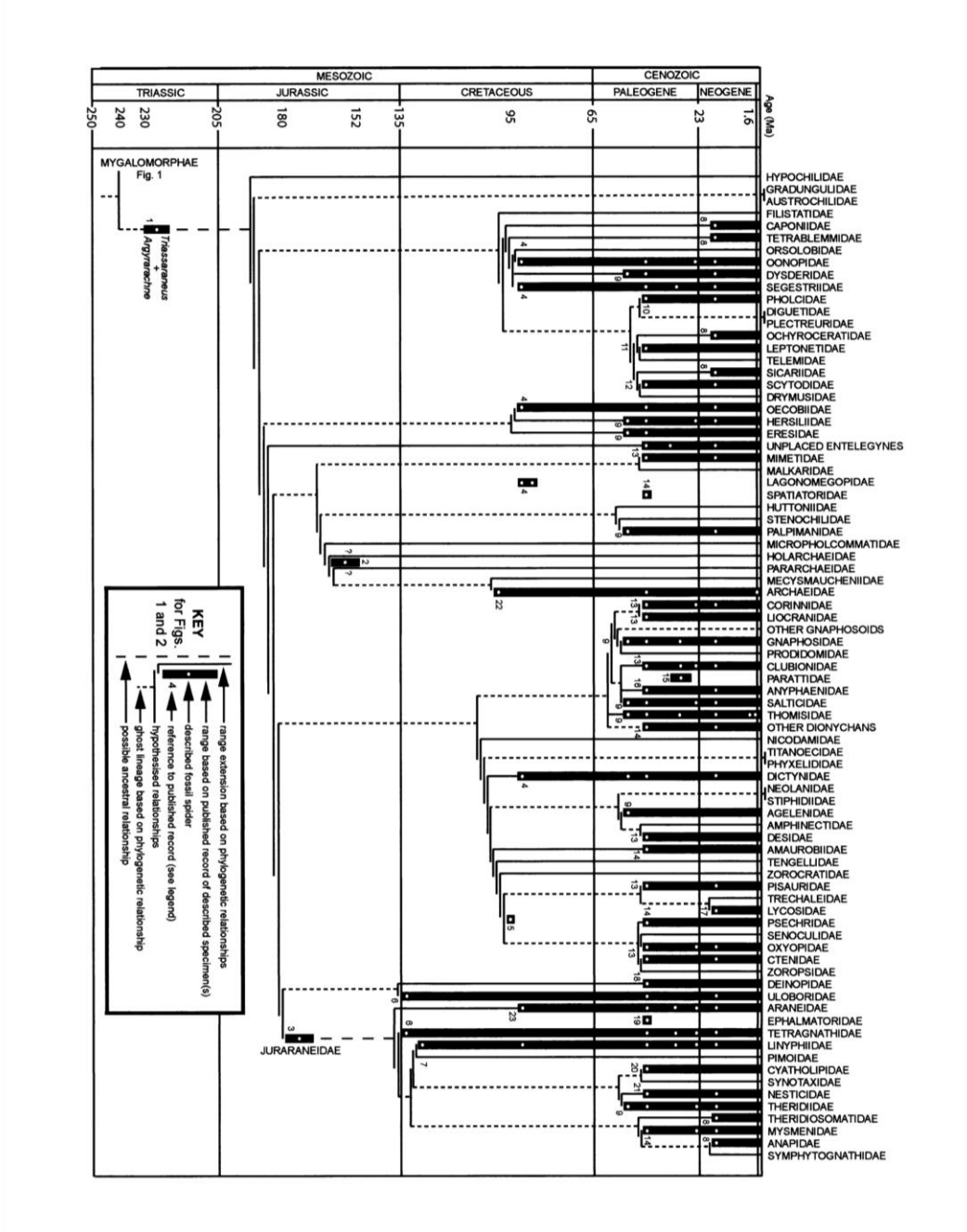


Figure 5.1. Evolutionary tree of spiders calibrated with described Fossils. Cladistic phylogeny superimposed onto geological time scale from Penney *et al.* 2003.

This study investigates the potential of utilising the information encoded within complete mitochondrial genomes of extant spiders using the available extensive fossil record in order to calibrate the phylogeny and provide an insight

into the timing of the major radiations that have led to the present day diversity. Taking advantage of the growing availability of complete mitochondrial genome sequences within the order, Bayesian analysis is used for its computational efficiency and ability to incorporate complex models (Huelsenbeck et al., 2001; Lakner et al., 2008; Nylander et al., 2004; Posada and Buckley, 2004), in showcasing the potential of mitogenomic studies in delineating inter-familial relationships in an understudied order.

5.3. Materials & methods

5.3.1. Taxon sampling & alignments

The complete mitochondrial genomes of ten spiders representing nine different families (Figure 5.2.) were used for this study along with the mitochondrial genomes of two whip spiders (Amblypygi; *Phrynus sp.* and *Damon diadema*) (Table 5.1.), previously determined to be the most closely related arachnid ally of the Araneae (Wheeler and Hayashi, 1998). In order to assess the effect of gene inclusion/exclusion, in light of repeated transfer RNA (tRNA) gene rearrangements, two approaches were used. Complete mitochondrial genome sequences were taken from this study or downloaded from GenBank and datasets were created by aligning either the concatenated sequences of all genes (protein coding, tRNA and rRNA) or concatenated sequences of protein coding genes, using Clustal Omega (Sievers et al., 2011). Default settings were used for each of the Clustal alignments. The decision to concatenate data into single partitions was taken following the generation of bootstrapped (1000 replicates) Neighbour-joining trees from alignments of single genes using Geneious. As found in other studies (Gadagkar et al., 2005; Nery et al., 2012; Phillips et al., 2004; Rokas et al., 2003; Zhou et al., 2012; Zhou et al., 2011), the use of highly divergent taxa with

single gene partitions creates poorly supported topologies with little concordance between genes (O'Huigin et al., 2002; Talavera and Vila, 2011). Thus, concatenated sequences were deemed a more appropriate choice for the mitochondrial genome datasets.

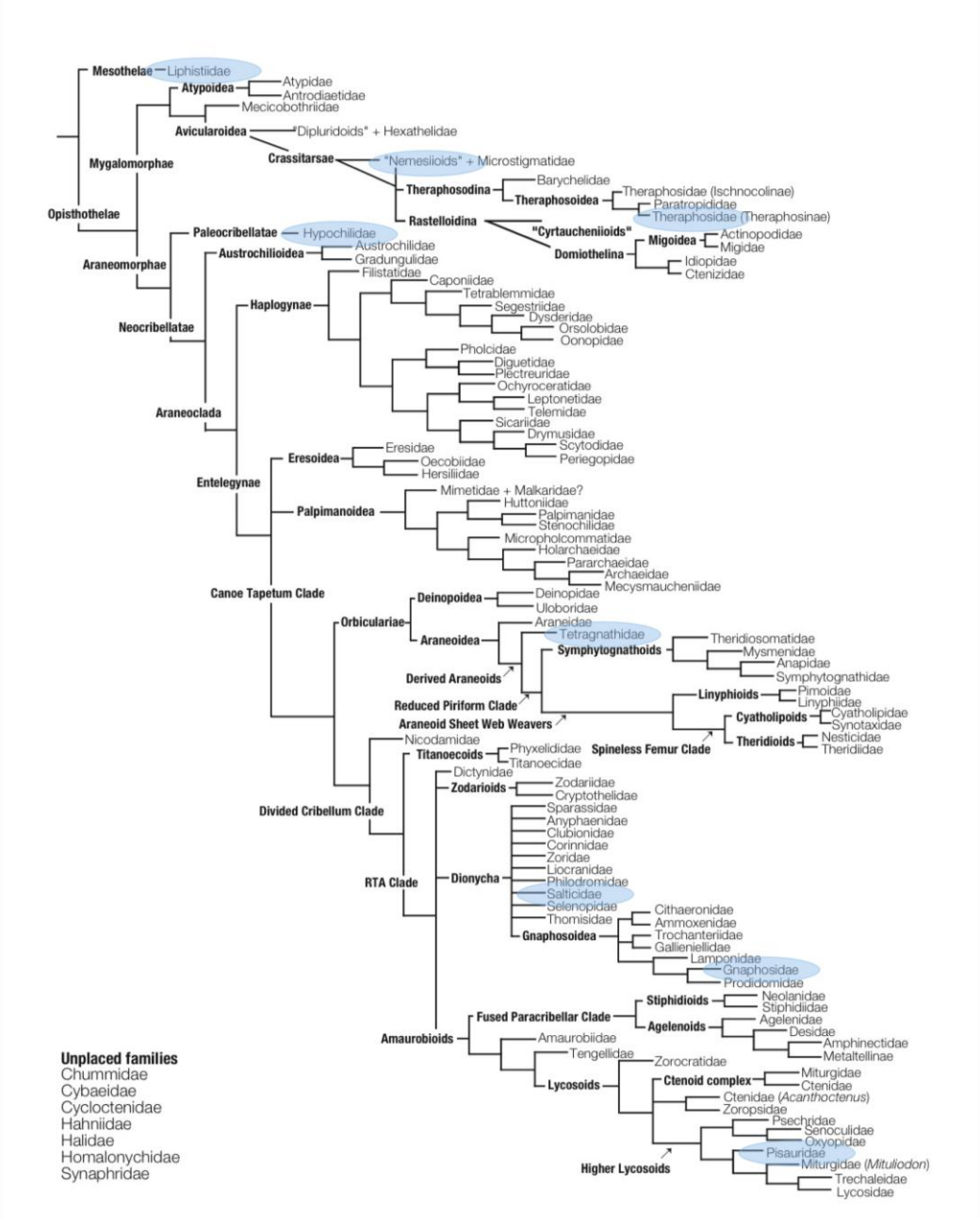


Figure 5.2. Cladistic hypotheses for all Araneomorph spider families using behavioural and morphological characters. From Coddington, J. A. (2004).

Families used in this study are highlighted; Nephilidae (*Nephila clavata*) is not placed within the phylogeny.

Table 5.1. Taxonomic information and GenBank accession numbers for specimens. Length represents the total nucleotide sequence length in base pairs of the complete mitochondrial genome.

Order	Family	Genus	Specific epithet	Length	Genbank accession	Reference
Araneae	*Gnaphosidae*	*Zelotes*	*apricorum*	14,222		This study
Araneae	*Hypochilidae*	*Hypochilus*	*thorelli*	13,991	NC_010777	Masta & Boore 2008
Araneae	*Liphistiidae*	*Heptathela*	*hangzhouensis*	14,215	AY309258	Qiu et al. 2005
Araneae	*Nemisiidae*	*Calisoga*	*longitarsis*	14,070	NC010780	Masta & Boore 2008
Araneae	*Nephilidae*	*Nephila*	*clavata*	14,436	AY452691	Lee et al unpublished
Araneae	*Pisauridae*	*Pisaura*	*mirabilis*	14,013		This study
Araneae	*Salticidae*	*Habronattus*	*oregonensis*	14,381	NC_005942	Masta & Boore 2003
Araneae	*Tetragnathidae*	*Meta*	*menardi*	14,399		This study
Araneae	*Theraphosidae*	*Ornithoctonus*	*huwena*	13,874	AY309259	Qiu et al. 2005
Araneae	*Theraphosidae*	*Psalmopoeus*	*cambridgei*	13,767		This study
Amblypygi	*Phrynichidae*	*Damon*	*diadema*	14,786	FJ204233	Fahrein et al. 2009
Amblypygi	*Phrynidae*	*Phrynus*	sp.	14,764	EU520641	Masta & Boore 2008

5.3.2. Phylogenetic analyses

The substitution model for each of the aligned datasets was selected using the FindModel web software (available at http://www.hiv.lanl.gov). The program Weighbour (Bruno et al., 2000) was used to generate a weighted neighbour-joining tree based on Jukes-Cantor distances and Akaike's Information Criterion scores were calculated using MODELTEST (Posada and Crandall, 1998). The Generalised time-reversible model (GTR) plus Gamma was found to be the most appropriate model and was used in all subsequent analyses.

All mitochondrial genome phylogenies for both of the datasets were reconstructed using the Bayesian phylogenetic software BEAST v1.7.5 (Drummond et al., 2012) and run for 40,000,000 MCMC generations, logging every 4000 steps

with the first 4,000,000 generations discarded as burn in and convergence was assessed graphically (Drummond et al., 2002).

5.3.3. Estimation of divergence times

Dating of the major radiations and divergences within the Araneae was performed using a relaxed molecular clock approach (Lemey et al., 2010) and two fossil calibration points. The minimum age used for the emergence of the spiders was set as 300 million years based on the age of the oldest recorded spider fossil from the late Cretaceous (Selden, 1996). As a second calibration point the genus *Nephila* (family Nephilidae) has fossils dating back to the middle Jurassic and a calibration point was set with an estimated age of 135 million years (Selden et al., 2011). Finally fossil information was available for the family Theraphosidae, which according to the fossil record, is one of the most recently diverged spider families with the oldest fossil estimated to be 16.5 million years (Dunlop et al., 2007). However, previous phylogenetic assessment indicates the lineage could be much older and so this fossil reference was not used for calibration.

Nodal ages were estimated using Bayesian approaches implemented in BEAST, again using the GTR + Gamma substitution model. Aside from the calibration points (normal distribution, standard deviation = 1 and lognormal relaxed clock model with default rate of 0.1) default priors were used for analysis with the exception of; Yule speciation on a previously attained tree and ucld.mean (initial = 0.01, gamma prior of 1, 0.03). Forty million Markov chain Monte Carlo (MCMC) steps were sampled every 4,000 generations. Convergence was assessed by comparing multiple runs in Tracer v1.5 after discarding the first 4 million states as burn-in. All effective sample size (ESS) values exceeded accepted values (>200),

showing the run length was sufficient. All analyses were performed on the CIPRES web portal using XSEDE and run across 8 processors (Miller *et al*. 2010).

5.4. Results

5.4.1. Phylogenetic analysis

Ten mitochondrial genomes were aligned using the two whip spider (Amblypygi) mitochondrial genomes for each of the concatenated sequences and single gene datasets. Phylogenetic trees were reconstructed using Bayesian inference with the BEAST program (Drummond et al., 2012). A similar topology was recovered for each of the datasets varying only in the relationships between *Nephila clavata* and *Meta menardi* (Figure 5.3.). The results confirm most of the current familial placements and suspected relationships, which have been previously assessed with morphological and behavioural data, with strong statistical support (Coddington et al., 2004; Coddington and Levi, 1991). However, both topologies place the Paleocribellate Hypochilidae (*Hypochilus thorelli*) as a sister taxa to the Mygalamorphae rather than the Araneomorphae as previously described (2005; Catley, 1994; Forster et al., 1987; Song et al., 1999). The results suggest the Mesothelae is likely to be the most basal of the Araneae and provides strong statistical support for all nodes.

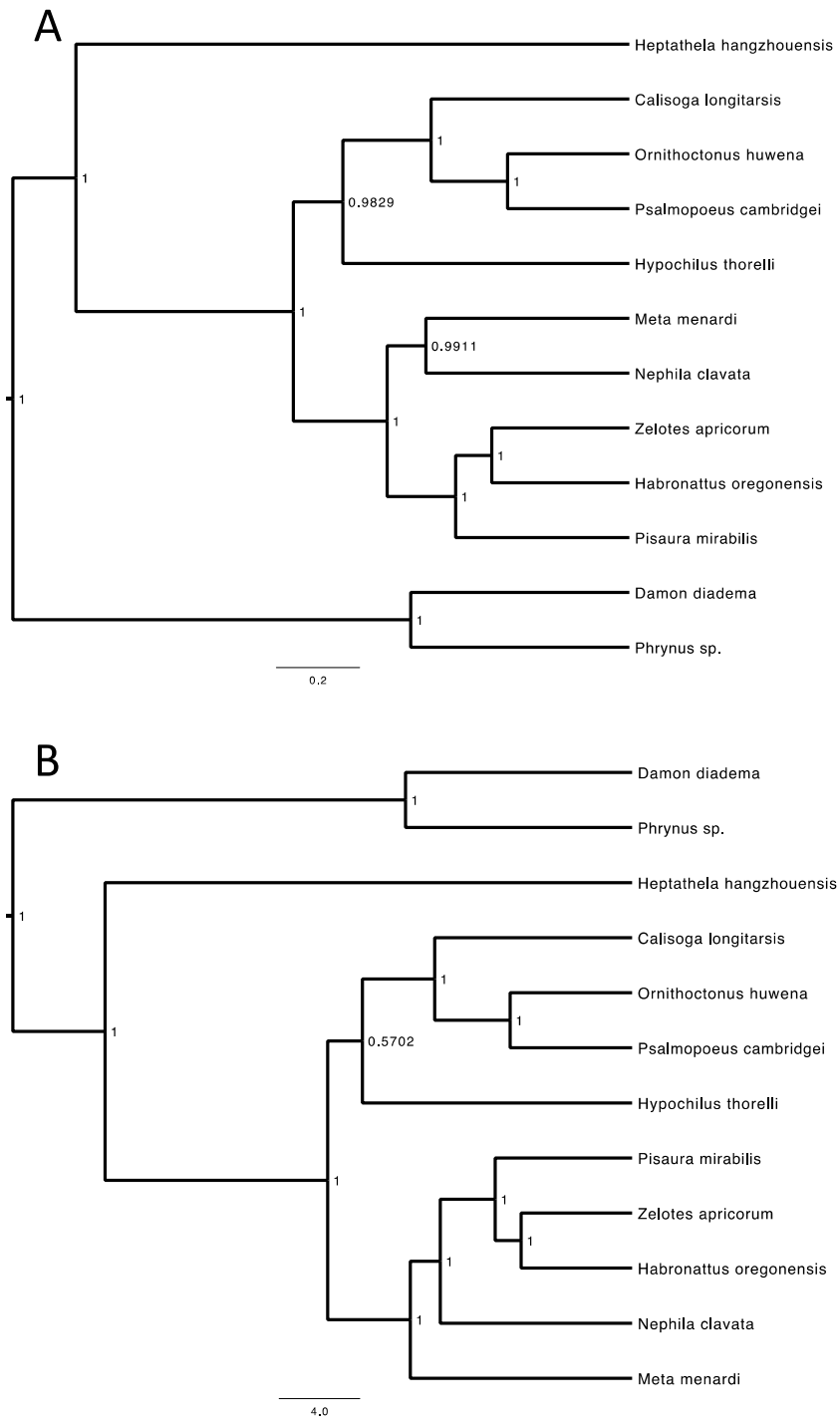
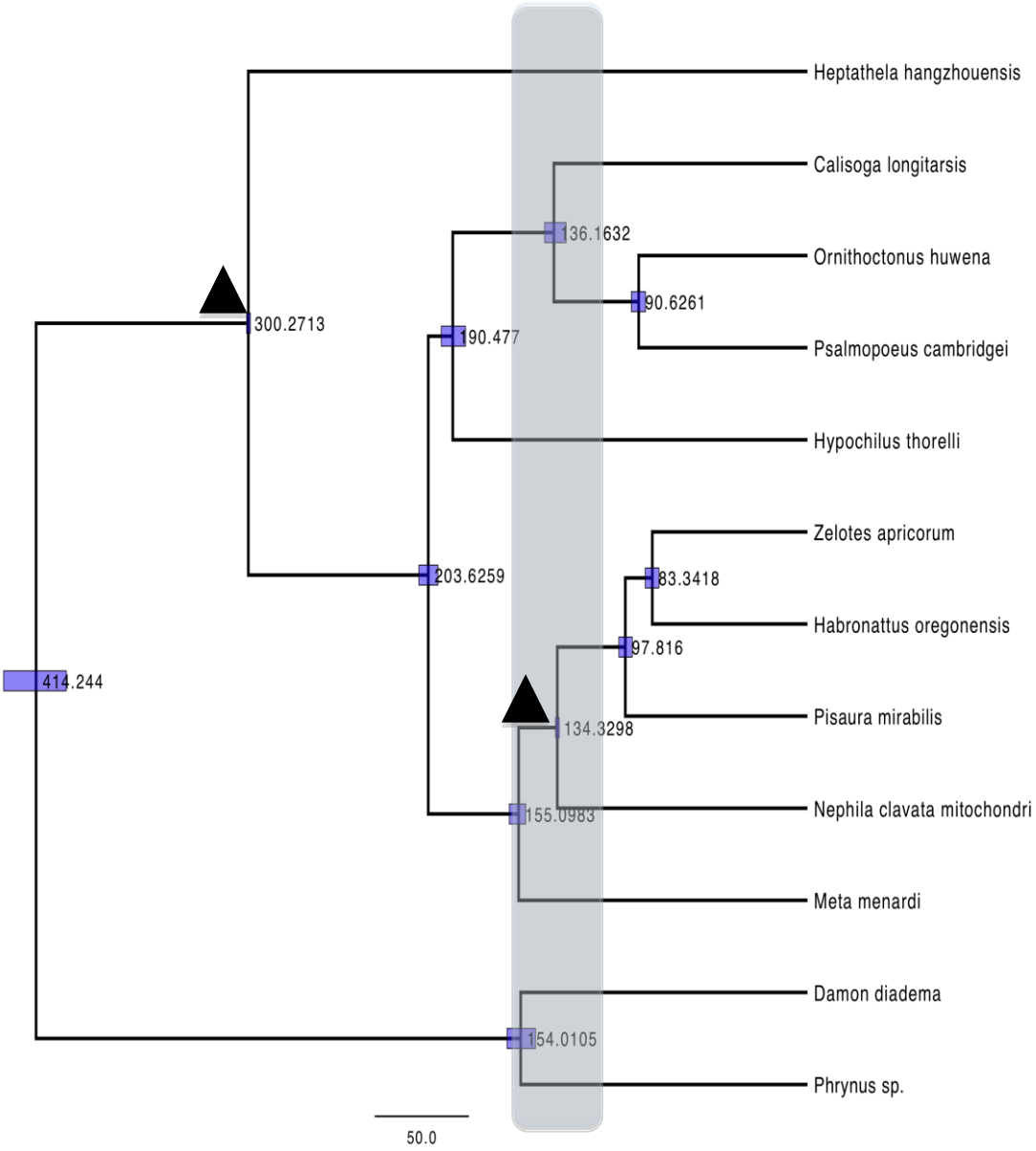


Figure 5.3. Bayesian phylogenetic trees recovered from concatenated datasets. All lineages evolved according to a relaxed clock and the GTR+Γ_4 substitution model. Numbers above the nodes indicate the posterior probabilities for both of the alignments; A) Concatenated sequence of all protein coding genes and B) Concatenated sequence of all thirty-seven protein coding and RNA genes.

5.4.2. Estimation of divergence times

Two fossil calibration points were used to estimate the mtDNA divergence dates within the Araneae (see Materials and Methods) (Figure 5.4.). The posterior mean of the divergence time between members of the Araneomorphae was estimated at approximately 155 million years ago, which is concordant with previous estimates based on the available fossil record (Penney, 2010). The initial divergence within the Arachnida that separated the Araneae and Amblypygi was estimated to have occurred between 398 and 431 million years ago. The next divergence following the split of the Mesothelae and Opisthothelae 300 mya, which separated the mygalamorph and araneomorph spiders is dated at 198 to 208 mya. Whilst most dates provided by the calibrated tree seem plausible, fossils from both lineages have been recorded from a slightly earlier time (approx 230 mya) (Dunlop and Penney, 2012). The calibrated phylogeny also suggests a major radiation between 160-129 million years ago, which is similar to diversification recorded in other taxa. Finally the suggested age of the family Theraphosidae (*Ornithoctonus huwe*na and *Psalmopoeus cambridgei*) is dated as being 87-94 million years. Whilst this is far earlier than the fossil record would suggest (16.5 mya) it is concordant with current phylogenetic hypotheses (Penney et al., 2003).



D	C	P	T	J	K	Pg	N

Figure 5.4. Bayesian phylogenetic tree displayed as a chronogram from the BEAST analysis of the unpartitioned mitochondrial genome alignment. All lineages evolved according to the GTR+Γ_4 substitution model. Posterior probabilities not shown but all nodes were supported in at least 99% of topologies. Node bars illustrate the width of the 95% highest posterior density (age range). Numbers indicate the mean estimates of divergence times. Grey bar shows period of divergence. Black triangles denote calibration points.

5.5. Discussion

This study represents the most comprehensive sampling of complete mitochondrial genomes within the Araneae, covering 11 extant families and all three of the major spider lineages; Mesothelae, Mygalamorphae and Araneomorphae. Whilst there are large numbers of mitochondrial genomes available for other Arthropods, few studies (Timmermans, 2010; Cameron, 2007; Sheffield, 2009; Fenn, 2008) have focused on their potential to confidently resolve phylogenetic discordance. This analysis has allowed the phylogenetic topology of spiders to be resolved with high support values, the first time this has been achieved solely based on molecular characters. Differing from previously reported phylogenies (Coddington et al., 2004; Coddington and Levi, 1991), all phylogenetic reconstructions have shown the Paleocribellate spider family Hypochilidae (*Hypochilus thorelli*) to have diverged from the common ancestor of the Mygalamorphae spiders and not the Araneomorphae. Hypochilidae is the only family within the Paleocribellate grouping and shares morphological characteristics of both Opisthothelae lineages. Similar to mygalamorph spiders the family Hypochilidae is characterised by having only two pairs of book lungs, as opposed to the four found in all other Araneomorph taxa (Jocque and Dippenaar-Schoeman, 2006). However, as with all other Araneomorph spiders, members of the family Hypochilidae possess intersecting chelicerae and like some families have formed a cribellum, a silk spinning organ that is a functional homologue of the anterior median spinnerets of spiders belonging to the Mesothelae and Mygalamorphae (Opell, 2002) (Figure 5.5.).

Figure 5.5. Electron micrograph image of a spider cribellum. The small tubes on the cribellar gland release strands of silk of varying diameters.

Although the presence of a cribellum was initially used to differentiate between cribellate (crebellum) and ecribellate (no cribellum) Araneomorph families, changes to taxonomic classifications have placed spiders with both morphologies in to the same family (Lehtinen 1967). Thus, leading to the conclusion that the ancestor of all Araneomorphae was cribellate and this feature has been lost secondarily in some taxa (Coddington and Levi, 1991). Whilst many ecribellate Araneomorph spiders feature a colulus (reduced cribellum) some taxa appear to have evolved independently without cribellate precursors (Foelix, 1996), making it more plausible that the ancestors of all Opisthothelae spiders were cribellate and this feature has been lost in present mygalamorphs. Whilst there are no fossils that can provide evidence for or against this, the inclusion of the Hypochilidae as a sister taxon to the Mygalamorphae is further supported in the secondary structure of tRNAs (See Chapter 4 Table 4.4.). *Hypochilus thorelli* not only shares the same mitochondrial gene order of other mygalamorph spiders but also exhibits greater similarities in the losses of tRNA T arms.

The phylogenetic reconstruction also highlights the reasons for earlier difficulties in obtaining strong support for individual gene trees, as most of the internal branches in the concatenated sequence phylogenetic trees are very short. Short internal branch lengths have been known to create conflicting topologies in both nuclear (O'HUigin et al., 2002) and mitochondrial DNA (Rogaev et al., 2006; Rohland et al., 2007).

Using complete mitochondrial genome sequences from nine families of spiders I found evidence for a rapid radiation that occurred approximately 160-129 million years ago at the boundary of the Jurassic and Cretaceous periods in line with the suggested hypotheses based on morphological, behavioural characters calibrated with fossil data (Penney et al., 2003). As flowering plants (Angiosperms) were also undergoing rapid diversification (Bremer et al., 2004) at this time I suggest that a change in floral landscapes may have spurred diversification within the Araneae by providing new ecological niches from which to diversify web structures (Dunlop et al., 2007). The diversification of flowering plants was also linked to the radiation of insects (Grimaldi, 1999), particularly those involved in pollination. An increase in prey in association with habitat and niche expansion may explain the high levels of speciation seen within spider families that have abandoned prey capture webs altogether, such as the mega-diverse Salticidae. My results clearly demonstrate the power of mitogenomic analyses for resolving complicated phylogenetic relationships with strong statistical support using fossils of extant lineages from which to calibrate divergence estimates.

6. General discussion

6.1. Overview

Advances in sequencing technology and in particular the development of second generation sequencing platforms have resulted in the rapid dissemination of large volumes of sequence data for relatively little cost (Ansorge, 2009; Metzker, 2005; Meyer et al., 2007; Rothberg and Leamon, 2008). This has substantially increased the quantity of data available for comparative evolutionary analysis and allowed for ever increasing numbers of taxa and characters to be incorporated into resolving the branches of the tree of life (Lerner and Fleischer, 2010; Metzker, 2010). The relatively small size of mitochondrial genomes makes them tractable to sequencing in their entirety and they have the potential to be utilised for the design and optimisation of phylogenomic analyses (Dordel et al., 2010; Dutilh et al., 2007; Hedin et al., 2012; McKain et al., 2012; Philippe et al., 2009).

Despite these technological advances, the amplification and subsequent recovery of longer sequences, such as that of complete mitochondrial genomes, has not been straightforward. In some taxa a bottleneck exists between mitogenome amplification and downstream high-throughput sequencing and is restricting the creation of divergent datasets from which we can resolve higher-level topological incongruence (Kocot et al., 2011; Philippe et al., 2009; Wiens et al., 2010).

The aim of this study was to examine these issues using the Araneae as a model order to explore the phylogenetic utility of divergent mitogenomic data sets. Spiders, in spite of their status as a well-described and speciose order, have been relatively understudied from a molecular perspective, especially in comparison with other orders of similar magnitude. In this thesis I have assessed multiple amplification techniques and processes from which to create divergent complete mitochondrial genome data sets. This has been achieved through the robust trialling and subsequent analysis of data from both PCR and WGA methodologies

across a range of disparate taxa. Through this endeavour I have highlighted both the successes and hindrances associated with each field and provided an insight in to the phylogenetic potential of the complete mitochondrial genome.

6.2. Spider mitochondrial genomes

This study represents the first comparative analysis of divergent mitochondrial genomes within the Araneae and has highlighted the high levels of variability therein. The comparative mitogenomic analysis of spider mitochondrial genomes sheds light on the architecture and evolution of the organelle across the order Araneae. Whilst rare genomic changes have long been thought to allow mitochondrial genomes to be phylogenetically informative at higher-levels (Rokas, 2000; Boore, 1995), recent studies have found that both nucleotide diversity and mitogenomic traits ca be informative at subfamily level (Li et al., 2012). Whilst the inclusion of a broader range of taxa representing more taxonomic levels will be critical to fully understanding the mitogenomic evolution of spiders it is encouraging to have been able to identify potential markers for delineating relationships within the structure and content of mitochondrial genomes.

6.3. Advances in fossil morphological identification

Whilst this study serves to highlight the potential of complete mitochondrial genomes in resolving topological incongruence with high statistical support, many studies have shown the utility of morphological characters in combination with molecular data (Carden, 2012; Hills, 2012). Recent advances in imaging technology (CT scanning) have allowed the 3D reconstruction of fossils from both rock and amber. This technique has been important in the identification of fossil spiders through the visualisation of morphological features that traditional

microscopy could not view and has been key in the correct identification to the species level (Dunlop et al., 2012; Penney et al., 2012b) . Whilst allowing for the correct identification of fossil samples to ensure correct familial placements, based on current understandings enhanced fossil visualisation also provides new and more accurate calibration points from which to date molecular phylogenies. In this study I have shown how fossil calibrations across large geological timescales can be incorporated with molecular data from extant species in order to accurately date the divergence and major adaptive radiations within an order whilst highlighting potential gaps in the fossil record. As more and more fossils are recovered the ability to accurately use this information in determining infra-order relationships will be invaluable for correctly hypothesising evolutionary relationships between disparate taxa.

6.4. Bioinformatic issues

Next-generation sequencing technologies have allowed the generation of ever increasing numbers of molecular characters and enabled us to perform large-scale analyses in ways never before realised (Ansorge, 2009). While novel tools have been developed specifically for such massively parallel high-throughput sequencers, the increasing complexity of samples has presented difficult challenges and exposed a number of analytical bottlenecks in light of problems inherent with assembly or read-based analysis (Alkan et al., 2011).

Based on the already apparent issues with down stream bioinformatics associated with NGS technologies it is expected that sequencing will soon cease to be the limiting factor in genomic studies. This study highlighted the difficulties in recovering mitochondrial DNA from low sequence coverage and in the presence of

contaminant sequences. Without improved algorithms and advances in bioinformatics it could soon become the main bottleneck for genomic projects.

6.5. Advances in sequencing approaches

This study has provided support for the potential of mitochondrial genomes to be utilised in creating robust phylogenetic frameworks. However, even if the hindrance associated the genomic amplification of divergent taxa is overcome, there still remains significant difficulties associated with assembly and annotation (Salzberg et al., 2012; Vezzi et al., 2012). Some new approaches have sought to minimise these issues by working in reverse (Johnson et al., 2013). Targeted Restricted Assembly Method (TRAM), known protein sequences from genes of interest in closely related taxa are used to identify reads from a shotgun NGS sequencing run. Reads can then be pooled and assembled into contigs before being aligned to reference cDNA in order to remove non-coding regions. EST, or expressed sequence tags have also been used with good success to analyse large parts of the transcriptome and in doing so provide large amounts of microsatellite data that can be used for population genetic studies (Mercati et al., 2013; Mohanty et al., 2013). However this is limited by the need for exceptional specimen preservation in order to assure the availability of mRNA and can be unsuitable or more difficult for some taxa. RAD tag sequencing has recently been used to exploit the massively high throughput capabilities of Illumina sequencing platforms (Scaglione et al., 2012). Restriction enzymes are used to digest sites within the nuclear genome and sequncing adapters and indexes are simultaneously annealed to the resulting fragments. Once sequenced this is used to identify single nucleotide polymorphisms (SNPs) for use in population genetics. With each new study and technique it is evident that advances in sequencing technologies have

opened up more avenues from which genetic data can be retrieved to allow greater and more robust comparative analyses.

6.6. Third generation sequencing

In spite of NGS, or second generation sequencing technologies, still being relatively in their infancy, the next-generation of DNA sequencing technologies is already being developed (Rusk, 2009; Schadt et al., 2011). Unlike NGS platforms that rely on sequencing by synthesis (SBS) methods where DNA is PCR amplified in clusters before being bound to a solid surface for imaging, third generation sequencing involves the interrogation of only a single molecule of DNA (Eid et al., 2009; Luan et al., 2010). Ultimately this technology centres on the ability to sequence a single molecule of DNA at a time without the need to pause between reads and thus greatly reducing the time the associated with sequencing, allowing even higher-throughput than the systems used currently (Meyer et al., 2007). Several methods have been developed by which to achieve this goal, but currently they are either still restricted to PCR based sequencing such as Ion Torrent's semi conductor sequencer (Merriman et al., 2012; Rothberg et al., 2011) or have not yet become truly feasible at a commercial scale. Once third-generation sequencing methods become established they will have the potential to eliminate all bias associated with the amplification of divergent datasets and the only step needed for sequencing preparation will revolve around the isolation of the required genome be it nuclear or organellar in origin.

6.7. Summary of thesis

The aim of this thesis, as set out in the introduction, was to harness the potential of next-generation sequencing platforms to address both the lack of molecular data available for the Araneae, and to address the issues associated with the construction of divergent mitogenomic datasets. Whilst the goal of creating large datasets from which I could attempt to resolve the conflicting relationships within the Araneae (Blackledge et al., 2009) was not realised, the mitochondrial genome data generated by this study was instrumental in allowing a true comparative analysis of mitochondrial features that has helped elucidate the evolutionary history of spiders. Moreover, the comparative analysis of complete mitochondrial genome sequences has been able to assess the variation of mitochondrial traits, such as tRNA secondary structure, across all currently recognized major lineages of spiders. Whilst the phylogenetic analysis of mitogenomic sequences used within the study were able to robustly reassess current taxonomic classification within the Araneae, the incorporation of rare genomic changes and other gene features will be critical in providing support for phylogenetic hypotheses, particularly where few taxa have been sampled.

In relation to the initial aims of the thesis, I feel that this study has provided a framework from which future analyses can be designed and offered potential new insights in to the utilisation of non-PCR based amplification strategies for the construction of divergent mitogenomic datasets.

7. References

Adams, K.L., Palmer, J.D., 2003. Evolution of mitochondrial gene content: gene loss and transfer to the nucleus. Molecular Phylogenetics and Evolution 29, 380-395.

Aguileta, G., Marthey, S., Chiapello, H., Lebrun, M.H., Rodolphe, F., Fournier, E., Gendrault-Jacquemard, A., Giraud, T., 2008. Assessing the performance of single-copy genes for recovering robust phylogenies. Systematic Biology 57, 613-627.

Aird, D., Ross, M.G., Chen, W.-S., Danielsson, M., Fennell, T., Russ, C., Jaffe, D.B., Nusbaum, C., Gnirke, A., 2011. Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. Genome Biology 12. doi: 10.1186/gb-2011-12-2-r18.

Alkan, C., Sajjadian, S., Eichler, E.E., 2011. Limitations of next-generation genome sequence assembly. Nature Methods 8.

Allison, D.P., Kerper, P.S., Doktycz, M.J., Spain, J.A., Modrich, P., Larimer, F.W., Thundat, T., Warmack, R.J., 1996. Direct atomic force microscope imaging of EcoRI endonuclease site specifically bound to plasmid DNA molecules. Proceedings of the National Academy of Sciences of the United States of America 93, 8826-8829.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic Local Alignment Search Tool. Journal of Molecular Biology 215, 403-410.

Alverson, A.J., Rice, D.W., Dickinson, S., Barry, K., Palmer, J.D., 2011. Origins and recombination of the bacterial-sized multichromosomal mitochondrial genome of cucumber. Plant Cell 23, 2499-2513.

Alverson, A.J., Wei, X., Rice, D.W., Stern, D.B., Barry, K., Palmer, J.D., 2010. Insights into the evolution of mitochondrial genome size from complete sequences of

Citrullus lanatus and *Cucurbita pepo* (Cucurbitaceae). Molecular Biology and Evolution 27, 1436-1448.

Ansorge, W.J., 2009. Next-generation DNA sequencing techniques. New Biotechnology 25, 195-203.

Arnedo, M.A., Coddington, J., Agnarsson, I., Gillespie, R.G., 2004. From a comb to a tree: phylogenetic relationships of the comb-footed spiders (Araneae, Theridiidae) inferred from nuclear and mitochondrial genes. Molecular Phylogenetics and Evolution 31, 225-245.

Arnedo, M.A., Gillespie, R.G., 2006. Species diversification patterns in the Polynesian jumping spider genus *Havaika* Proszynski, 2001 (Araneae, Salticidae). Molecular Phylogenetics and Evolution 41, 472-495.

Arnedo, M.A., Hormiga, G., Scharff, N., 2009. Higher-level phylogenetics of linyphiid spiders (Araneae, Linyphiidae) based on morphological and molecular evidence. Cladistics 25, 231-262.

Astrin, J.J., Huber, B.A., Bernhard, M., Klutsch, C.F.C., 2006. Molecular taxonomy in pholcid spiders (Pholcidae, Araneae): evaluation of species identification methods using CO1 and 16S rRNA. Zoologica Scripta 35, 441-457.

Ayoub, N.A., Garb, J.E., Hedin, M., Hayashi, C.Y., 2007. Utility of the nuclear protein-coding gene, elongation factor-1 gamma (EF-1 gamma), for spider systematics, emphasizing family level relationships of tarantulas and their kin (Araneae : Mygalomorphae). Molecular Phylogenetics and Evolution 42.

Baird, N.A., Etter, P.D., Atwood, T.S., Currey, M.C., Shiver, A.L., Lewis, Z.A., Selker, E.U., Cresko, W.A., Johnson, E.A., 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS One 3(10): e3376.

Barnes, W.M., 1994. PCR amplification of up to 35-KB DNA with high-fidelity and high-yield from lambda-bacteriophage templates. Proceedings of the National Academy of Sciences of the United States of America 91, 2216-2220.

Barrett, R.D.H., Hebert, P.D.N., 2005. Identifying spiders through DNA barcodes. Canadian Journal of Zoology 83, 481-491.

Beard, C.B., Mills Hamm, D., Collins, F.H., 1993. The mitochondrial genome of the mosquito *Anopheles gambiae*: DNA sequence, genome organization, and comparisons with mitochondrial sequences of other insects. Insect Molecular Biology 2, 103-124.

Beckenbach, A.T., 2011. Mitochondrial genome sequences of representatives of three families of scorpionflies (Order Mecoptera) and evolution in a major duplication of coding sequence. Genome 54, 368-376.

Beilstein, M.A., Nagalingum, N.S., Clements, M.D., Manchester, S.R., Mathews, S., 2010. Dated molecular phylogenies indicate a Miocene origin for *Arabidopsis thaliana*. Proceedings of the National Academy of Sciences of the United States of America 107, 18724-18728.

Benjamin, S.P., Dimitrov, D., Gillespie, R.G., Hormiga, G., 2008. Family ties: molecular phylogeny of crab spiders (Araneae : Thomisidae). Cladistics 24, 708-722.

Bexfield, N., Kellam, P., 2011. Metagenomics and the molecular identification of novel viruses. Veterinary Journal 190, 191-198.

Bi, X., Huang, L., Jing, M., Zhang, L., Feng, P., Wang, A., 2012. The complete mitochondrial genome sequence of the black-capped capuchin (*Cebus apella*). Genetics and Molecular Biology 35, 545-552.

Bjork, A., Liu, W., Wertheim, J.O., Hahn, B.H., Worobey, M., 2011. Evolutionary history of chimpanzees inferred from complete mitochondrial genomes. Molecular Biology and Evolution 28, 615-623.

Black, W.C., Roehrdanz, R.L., 1998. Mitochondrial gene order is not conserved in arthropods: Prostriate and metastriate tick mitochondrial genomes. Molecular Biology and Evolution 15, 1772-1785.

Blackledge, T.A., Scharff, N., Coddington, J.A., Szuts, T., Wenzel, J.W., Hayashi, C.Y., Agnarsson, I., 2009. Reconstructing web evolution and spider diversification in the molecular era. Proceedings of the National Academy of Sciences of the United States of America 106, 5229-5234.

Bleidorn, C., Podsiadlowski, L., Bartolomaeus, T., 2006. The complete mitochondrial genome of the orbiniid polychaete *Orbinia latreillii* (Annelida, Orbiniidae) - A novel gene order for Annelida and implications for annelid phylogeny. Gene 370, 96-103.

Bodner, M.R., Maddison, W.P., 2012. The biogeography and age of salticid spider radiations (Araneae: Salticidae). Molecular Phylogenetics and Evolution 65.

Bond, J.E., 2012. Phylogenetic treatment and taxonomic revision of the trapdoor spider genus *Aptostichus* Simon (Araneae, Mygalomorphae, Euctenizidae). Zookeys, 1-209.

Bond, J.E., Hamilton, C.A., Garrison, N.L., Ray, C.H., 2012a. Phylogenetic reconsideration of *Myrmekiaphila* systematics with a description of the new trapdoor spider species *Myrmekiaphila tigris* (Araneae, Mygalomorphae, Cyrtaucheniidae, Euctenizinae) from Auburn, Alabama. Zookeys, 95-109.

Bond, J.E., Hendrixson, B.E., Hamilton, C.A., Hedin, M., 2012b. A Reconsideration of the classification of the spider infraorder Mygalomorphae (Arachnida:

Araneae) based on three nuclear genes and morphology. PLoS One 7(6): e38753.

Boore, J.L., 1999. Animal mitochondrial genomes. Nucleic Acids Research 27, 1767-1780.

Boore, J.L., Brown, W.M., 1998. Big trees from little genomes: mitochondrial gene order as a phylogenetic tool. Current Opinion in Genetics & Development 8, 668-674.

Boore, J.L., Lavrov, D.V., Brown, W.M., 1998. Gene translocation links insects and crustaceans. Nature 392, 667-668.

Borner, G.V., Yokobori, S., Morl, M., Dorner, M., Paabo, S., 1997. RNA editing in metazoan mitochondria: Staying fit without sex. EBS Letters 409, 320-324.

Bremer, K., Friis, E.M., Bremer, B., 2004. Molecular phylogenetic dating of asterid flowering plants shows early Cretaceous diversification. Systematic Biology 53, 496-505.

Briscoe, A.G., Goodacre, S., Masta, S., Taylor, M.I., Arnedo, M.A., Penney, D., Kenny, J., Creer, S., 2013. Can long-range PCR be used to amplify genetically divergent mitochondrial genomes for comparative phylogenetics? A case study within spiders (Arthropoda: Araneae). PLoS One 8(5): e62404.

Bruno, W.J., Socci, N.D., Halpern, A.L., 2000. Weighted neighbor joining: A likelihood-based approach to distance-based phylogeny reconstruction. Molecular Biology and Evolution 17, 189-197.

Bruvo-Madaric, B., Huber, B.A., Steinacher, A., Pass, G., 2005. Phylogeny of pholcid spiders (Araneae: Pholcidae): Combined analysis using morphology and molecules. Molecular Phylogenetics and Evolution 37, 661-673.

Cameron, S.L., Lo, N., Bourguignon, T., Svenson, G.J., Evans, T.A., 2012. A mitochondrial genome phylogeny of termites (Blattodea: Termitoidae):

Robust support for interfamilial relationships and molecular synapomorphies define major clades. Molecular Phylogenetics and Evolution 65, 163-173.

Cantatore, P., Gadaleta, M.N., Roberti, M., Saccone, C., Wilson, A.C., 1987. Duplication and remolding of tRNA genes during the evolutionary rearrangement of mitochondrial genomes. Nature 329, 853-855.

Catley, K.M., 1994. Descriptions of new *Hypochilus* species from New Mexico and California with a cladistic analysis of the Hypochilidae (Araneae). American Museum Novitates 3088, 1-27.

Cheng, S., Chang, S.Y., Gravitt, P., Respess, R., 1994. LONG PCR. Nature 369, 684-685.

Chiari, Y., Cahais, V., Galtier, N., Delsuc, F., 2012. Phylogenomic analyses support the position of turtles as the sister group of birds and crocodiles (Archosauria). BMC Biology 10:65 doi: 10.1186/1741-7007-10-65.

Coddington, J.A., Giribet, G., Harvey, M.S., Prendini, L., Walter, D.E., 2004. Arachnida. Oxford Univ Press, New York.

Coddington, J.A., Kuntner, M., Opell, B.D., 2012. Systematics of the spider family Deinopidae with a revision of the genus *Menneus*. Smithsonian Contributions to Zoology 636, i-iv, 1-61.

Coddington, J.A., Levi, H.W., 1991. Systematics and evoloution of spiders (Araneae). Annual Review of Ecology and Systematics 22, 565-592.

Cooper, A., Lalueza-Fox, C., Anderson, S., Rambaut, A., Austin, J., Ward, R., 2001. Complete mitochondrial genome sequences of two extinct moas clarify ratite evolution. Nature 409, 704-707.

Corp., I., Released 2011. IBM SPSS Statistics for Windows, Version 20.0. IBM Corp., Armonk, NY.

Cox-Foster, D.L., Conlan, S., Holmes, E.C., Palacios, G., Evans, J.D., Moran, N.A., Quan, P.L., Briese, T., Hornig, M., Geiser, D.M., Martinson, V., vanEngelsdorp, D., Kalkstein, A.L., Drysdale, A., Hui, J., Zhai, J.H., Cui, L.W., Hutchison, S.K., Simons, J.F., Egholm, M., Pettis, J.S., Lipkin, W.I., 2007. A metagenomic survey of microbes in honey bee colony collapse disorder. Science 318, 283-287.

Cummings, M.P., Otto, S.P., Wakeley, J., 1995. Sampling properties of DNA-sequence data in phylogenetic analysis. Molecular Biology and Evolution 12(5): 814-822.

Curole, J.P., Kocher, T.D., 1999. Mitogenomics: digging deeper with complete mitochondrial genomes. Trends in Ecology & Evolution 14, 394-398.

Davila, D.S., 2003. Higher-level relationships of the spider family Ctenidae (Araneae: Ctenoidea). Bulletin of the American Museum of Natural History 274, 1-86.

de Meo, P.D.O., D'Antonio, M., Griggio, F., Lupi, R., Borsani, M., Pavesi, G., Castrignano, T., Pesole, G., Gissi, C., 2012. MitoZoa 2.0: a database resource and search tools for comparative and evolutionary analyses of mitochondrial genomes in Metazoa. Nucleic Acids Research 40, D1168-D1172.

Dean, F.B., Nelson, J.R., Giesler, T.L., Lasken, R.S., 2001. Rapid amplification of plasmid and phage DNA using *phi29* DNA polymerase and multiply-primed rolling circle amplification. Genome Research 11, 1095-1099.

Dellaporta, S.L., Xu, A., Sagasser, S., Jakob, W., Moreno, M.A., Buss, L.W., Schierwater, B., 2006. Mitochondrial genome of *Trichoplax adhaerens* supports Placozoa as the basal lower metazoan phylum. Proceedings of the National Academy of Sciences of the United States of America 103, 8751-8756.

Delsuc, F., Brinkmann, H., Philippe, H., 2005. Phylogenomics and the reconstruction of the tree of life. Nature Reviews Genetics 6, 361-375.

Dempewolf, H., Kane, N.C., Ostevik, K.L., Geleta, M., Barker, M.S., Lai, Z., Stewart, M.L., Bekele, E., Engels, J.M.M., Cronk, Q.C.B., Rieseberg, L.H., 2010. Establishing genomic tools and resources for *Guizotia abyssinica* (L.f.) Cass.-the development of a library of expressed sequence tags, microsatellite loci, and the sequencing of its chloroplast genome. Molecular Ecology Resources 10, 1048-1058.

Denovanwright, E.M., Lee, R.W., 1992. Comparative-analaysis of the mitochondrail genomes of *Chlamydomonas eugamentos* and *Chlamydomonas moewusii*. Current Genetics 21, 197-202.

Dermauw, W., Vanholme, B., Tirry, L., Van Leeuwen, T., 2010. Mitochondrial genome analysis of the predatory mite Phytoseiulus persimilis and a revisit of the Metaseiulus occidentalis mitochondrial genome. Genome 53, 285-301.

Dimitrov, D., Arnedo, M.A., Ribera, C., 2008. Colonization and diversification of the spider genus *Pholcus* Walckenaer, 1805 (Araneae, Pholcidae) in the Macaronesian archipelagos: Evidence for long-term occupancy yet rapid recent speciation. Molecular Phylogenetics and Evolution 48, 596-614.

Dimitrov, D., Astrin, J.J., Huber, B.A., 2012a. Pholcid spider molecular systematics revisited, with new insights into the biogeography and the evolution of the group. Cladistics 29(2), 132-146.

Dimitrov, D., Lopardo, L., Giribet, G., Arnedo, M.A., Alvarez-Padilla, F., Hormiga, G., 2012b. Tangled in a sparse spider web: single origin of orb weavers and their spinning work unravelled by denser taxonomic sampling. Proceedings of the Royal Society B-Biological Sciences 279, 1341-1350.

Dirheimer, G., Keith, G., Dumas, P., Westhof, E., 1995. Primary, secondary, and tertiary structures of tRNAs. American Society for Microbiology (ASM).

Dohm, J.C., Lottaz, C., Borodina, T., Himmelbauer, H., 2008. Substantial biases in ultra-short read data sets from high-throughput DNA sequencing. Nucleic Acids Research 36(16): e105.

Dong, Y., Xu, J.-J., Hao, S.-J., Sun, H.-Y., 2012. The complete mitochondrial genome of the giant pill millipede, *Sphaerotheriidae* sp. (Myriapoda: Diplopoda: Sphaerotheriida). Mitochondrial DNA 23, 333-335.

Dordel, J., Fisse, F., Purschke, G., Struck, T.H., 2010. Phylogenetic position of Sipuncula derived from multi-gene and phylogenomic data and its implication for the evolution of segmentation. Journal of Zoological Systematics and Evolutionary Research 48, 197-207.

Drummond, A.J., Nicholls, G.K., Rodrigo, A.G., Solomon, W., 2002. Estimating mutation parameters, population history and genealogy simultaneously from temporally spaced sequence data. Genetics 161, 1307-1320.

Drummond, A.J., Suchard, M.A., Xie, D., Rambaut, A., 2012. Bayesian Phylogenetics with BEAUti and the BEAST 1.7. Molecular Biology and Evolution 29, 1969-1973.

Duchene, S., Archer, F.I., Vilstrup, J., Caballero, S., Morin, P.A., 2011. Mitogenome phylogenetics: the impact of using single regions and partitioning schemes on topology, Substitution Rate and Divergence Time Estimation. PLoS One 6(11): e27138.

Duchene, S., Frey, A., Alfaro-Nunez, A., Dutton, P.H., Gilbert, M.T.P., Morin, P.A., 2012. Marine turtle mitogenome phylogenetics and evolution. Molecular Phylogenetics and Evolution 65, 241-250.

Dunlop, J.A., Harms, D., Penney, D., 2007. A fossil tarantula (Araneae: Theraphosidae) from Miocene Chiapas amber, Mexico. Revista Iberica de Aracnologia 15, 9-17.

Dunlop, J.A., Penney, D., 2012. Fossil arachnids. [Siri Scientific Press Monograph Series Volume 2.], 1-192.

Dunlop, J.A., Wirth, S., Penney, D., McNeil, A., Bradley, R.S., Withers, P.J., Preziosi, R.F., 2012. A minute fossil phoretic mite recovered by phase-contrast X-ray computed tomography. Biology Letters 8, 457-460.

Dutilh, B.E., van Noort, V., van der Heijden, R.T.J.M., Boekhout, T., Snel, B., Huynen, M.A., 2007. Assessment of phylogenomic and orthology approaches for phylogenetic inference. Bioinformatics 23, 815-824.

Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G., Peluso, P., Rank, D., Baybayan, P., Bettman, B., Bibillo, A., Bjornson, K., Chaudhuri, B., Christians, F., Cicero, R., Clark, S., Dalal, R., deWinter, A., Dixon, J., Foquet, M., Gaertner, A., Hardenbol, P., Heiner, C., Hester, K., Holden, D., Kearns, G., Kong, X., Kuse, R., Lacroix, Y., Lin, S., Lundquist, P., Ma, C., Marks, P., Maxham, M., Murphy, D., Park, I., Pham, T., Phillips, M., Roy, J., Sebra, R., Shen, G., Sorenson, J., Tomaney, A., Travers, K., Trulson, M., Vieceli, J., Wegener, J., Wu, D., Yang, A., Zaccarin, D., Zhao, P., Zhong, F., Korlach, J., Turner, S., 2009. Real-time DNA sequencing from single polymerase molecules. Science 323, 133-138.

Ernsting, B.R., Edwards, D.D., Aldred, K.J., Fites, J.S., Neff, C.R., 2009. Mitochondrial genome sequence of *Unionicola foili* (Acari: Unionicolidae): a unique gene order with implications for phylogenetic inference. Experimental and Applied Acarology 49, 305-316.

Faith, J.J., Pollock, D.D., 2003. Likelihood analysis of asymmetrical mutation bias gradients in vertebrate mitochondrial genomes. Genetics 165, 735-745.

Fakhrai-Rad, H., Pourmand, N., Ronaghi, M., 2002. Pyrosequencing (TM): An accurate detection platform for single nucleotide polymorphisms. Human Mutation 19, 479-485.

Fang, K., Yang, C.C., Lue, B.W., Chen, S.H., Lue, K.Y., 2000. Phylogenetic corroboration of superfamily Lycosoidae spiders (Araneae) as inferred from partial mitochondrial 12S and 16S ribosomal DNA sequences. Zoological Studies 39, 107-113.

Felsenstein, J., 1988. Phylogenies from molecular sequences – inference and reliability. Annual Review of Genetics 22, 521-565.

Fire, A., Xu, S.Q., 1995. Rolling replication of short DNA circles. Proceedings of the National Academy of Sciences of the United States of America 92, 4641-4645.

Fitch, W.M., 1982. The Challenges to Darwinism Since the Last Centennial and the Impact of Molecular Studies. Evolution 36, 1133-1143.

Foelix, R.F., 1996. Biology of spiders. Second edition. Oxford University Press, USA.

Forster, R.R., Platnick, N.I., Gray, M.R., 1987. A Review of the Spider Superfamilies Hypochiloidea and Austrochiloidea (Araneae, Araneomorphae). Bulletin of the American Museum of Natural History 185, 1-116.

Fulton, T.L., Strobeck, C., 2010. Multiple fossil calibrations, nuclear loci and mitochondrial genomes provide new insight into biogeography and divergence timing for true seals (Phocidae, Pinnipedia). Journal of Biogeography 37, 814-829.

Gadagkar, S.R., Rosenberg, M.S., Kumar, S., 2005. Inferring species phylogenies from multiple genes: Concatenated sequence tree versus consensus gene tree. Journal of Experimental Zoology Part B-Molecular and Developmental Evolution 304B, 64-74.

Gai, Y., Song, D., Sun, H., Yang, Q., Zhou, K., 2008. The complete mitochondrial genome of *Symphylella* sp. (Myriapoda: Symphyla): Extensive gene order rearrangement and evidence in favor of Progoneata. Molecular Phylogenetics and Evolution 49, 574-585.

Gee, H., 2003. Evolution - Ending incongruence. Nature 425, 782-782.

Gilbert, M.T.P., Tomsho, L.P., Rendulic, S., Packard, M., Drautz, D.I., Sher, A., Tikhonov, A., Dalen, L., Kuznetsova, T., Kosintsev, P., Campos, P.F., Higham, T., Collins, M.J., Wilson, A.S., Shidlovskiy, F., Buigues, B., Ericson, P.G.P., Germonpre, M., Goetherstroem, A., Iacumin, P., Nikolaev, V., Nowak-Kemp, M., Willerslev, E., Knight, J.R., Irzyk, G.P., Perbost, C.S., Fredrikson, K.M., Harkins, T.T., Sheridan, S., Miller, W., Schuster, S.C., 2007. Whole-genome shotgun sequencing of mitochondria from ancient hair shafts. Science 317, 1927-1930.

Gissi, C., Iannelli, F., Pesole, G., 2008. Evolution of the mitochondrial genome of Metazoa as exemplified by comparison of congeneric species. Heredity 101, 301-320.

Glenn, T.C., 2011. Field guide to next-generation DNA sequencers. Molecular Ecology Resources 11, 759-769.

Goodacre, S.L., Martin, O.Y., Thomas, C.F.G., Hewitt, G.M., 2006. Wolbachia and other endosymbiont infections in spiders. Molecular Ecology 15, 517-527.

Goodman, M., Moore, G.W., Barnabas, J., Matsuda, G., 1974. Phylogeny of human globin genes investigated by maximum parsimony method. Journal of Molecular Evolution 3, 1-48.

Goransson, J., Ke, R., Nong, R.Y., Howell, W.M., Karman, A., Grawe, J., Stenberg, J., Granberg, M., Elgh, M., Herthnek, D., Wikstrom, P., Jarvius, J., Nilsson, M.,

2012. Rapid identification of bio-molecules applied for detection of biosecurity agents using rolling circle amplification. PLoS One 7(2): e31068.

Gray, M.W., Lang, B.F., Cedergren, R., Golding, G.B., Lemieux, C., Sankoff, D., Turmel, M., Brossard, N., Delage, E., Littlejohn, T.G., Plante, I., Rioux, P., Saint-Louis, D., Zhu, Y., Burger, G., 1998. Genome structure and gene content in protist mitochondrial DNAs. Nucleic Acids Research 26, 865-878.

Green, R.E., Malaspinas, A.-S., Krause, J., Briggs, A.W., Johnson, P.L.F., Uhler, C., Meyer, M., Good, J.M., Maricic, T., Stenzel, U., Pruefer, K., Siebauer, M., Burbano, H.A., Ronan, M., Rothberg, J.M., Egholm, M., Rudan, P., Brajkovic, D., Kucan, Z., Gusic, I., Wikstrom, M., Laakkonen, L., Kelso, J., Slatkin, M., Paeaebo, S., 2008. A complete Neanderthal mitochondrial genome sequence determined by high-throughput sequencing. Cell 134, 416-426.

Gresham, D., Dunham, M.J., Botstein, D., 2008. Comparing whole genomes using DNA microarrays. Nature Reviews Genetics 9, 291-302.

Griffith, A.J.F., Miller, J.H., Suzuki, D.T., 2000. An introduction to genetic analysis 7th edition. New York' W.H. Freeman.

Grimaldi, D., 1999. The co-radiations of pollinating insects and angiosperms in the Cretaceous. Annals of the Missouri Botanical Garden 86, 373-406.

Griswold, C.E., Coddington, J.A., Hormiga, G., Scharff, N., 1998. Phylogeny of the orb-web building spiders (Araneae, Orbiculariae: Deinopoidea, Araneoidea). Zoological Journal of the Linnean Society 123, 1-99.

Griswold, C.E., Coddington, J.A., Platnick, N.I., Forster, R.R., 1999. Towards a phylogeny of entelegyne spiders (Araneae, Araneomorphae, Entelegynae). Journal of Arachnology 27, 53-63.

Griswold, C.E., Ramirez, M.J., Coddington, J.A., Platnick, N.I., 2005. Atlas of phylogenetic data for entelgyne spiders (Araneae: Araneomorphae:

Entelegynae) with comments on their phylogeny. Proceedings of the California Academy of Sciences 56, 1-324.

Groenenberg, D.S.J., Pirovano, W., Gittenberger, E., Schilthuizen, M., 2012. The complete mitogenome of *Cylindrus obtusus* (Helicidae, Ariantinae) using Illumina next generation sequencing. BMC Genomics 13:114 doi:10.1186/1471-2164-13-114.

Handa, H., 2003. The complete nucleotide sequence and RNA editing content of the mitochondrial genome of rapeseed (*Brassica napus L.*): comparative analysis of the mitochondrial genomes of rapeseed and *Arabidopsis thaliana*. Nucleic Acids Research 31, 5907-5916.

Harrison, R.G., 1989. Animal mitochondrial-DNA as a genetic-marker in population and evolotionary biology. Trends in Ecology & Evolution 4, 6-11.

Hassanin, A., 2006. Phylogeny of Arthropoda inferred from mitochondrial sequences: Strategies for limiting the misleading effects of multiple changes in pattern and rates of substitution. Molecular Phylogenetics and Evolution 38, 100-116.

Hedin, M., Starrett, J., Akhter, S., Schoenhofer, A.L., Shultz, J.W., 2012. Phylogenomic resolution of Paleozoic divergences in harvestmen (Arachnida, Opiliones) via analysis of next-generation transcriptome Data. PLoS One 7(8): e42888.

Hedin, M.C., Maddison, W.P., 2001. A combined molecular approach to phylogeny of the jumping spider subfamily Dendryphantinae (Araneae: Salticidae). Molecular Phylogenetics and Evolution 18, 386-403.

Hiesel, R., Wissinger, B., Schuster, W., Brennicke, A., 1989. RNA editing in plant-mitochondria. Science 246, 1632-1634.

Hillis, D.M., Huelsenbeck, J.P., Cunningham, C.W., 1994. Application and accuracy of molecular phylogenies. Science 264, 671-677.

Hoffman, R.L., Golovatch, S.I., Adis, J., de Morais, J.W., 2002. Diplopoda. Amazonian Arachnida and Myriapoda: identification keys to all classes, orders, families, some genera, and lists of known terrestrial species. [Pensoft Series Faunistica No 24.]. 505-533.

Holder, M., Lewis, P.O., 2003. Phylogeny estimation: traditional and Bayesian approaches. Nature Reviews Genetics 4, 275-284.

Hosseini, R., Keller, M.A., Schmidt, O., Framenau, V.W., 2007. Molecular identification of wolf spiders (Araneae: Lycosidae) by multiplex polymerase chain reaction. Biological Control 40, 128-135.

Hu, J., Zhang, D., Hao, J., Huang, D., Cameron, S., Zhu, C., 2010. The complete mitochondrial genome of the yellow coaster, *Acraea issoria* (Lepidoptera: Nymphalidae: Heliconiinae: Acraeini): sequence, gene organization and a unique tRNA translocation event. Molecular Biology Reports 37, 3431-3438.

Hu, M., Jex, A.R., Campbell, B.E., Gasser, R.B., 2007. Long PCR amplification of the entire mitochondrial genome from individual helminths for direct sequencing. Nature Protocols 2, 2339-2344.

Huelsenbeck, J.P., Crandall, K.A., 1997. Phylogeny estimation and hypothesis testing using maximum likelihood. Annual Review of Ecology and Systematics 28, 437-466.

Huelsenbeck, J.P., Ronquist, F., Nielsen, R., Bollback, J.P., 2001. Evolution - Bayesian inference of phylogeny and its impact on evolutionary biology. Science 294, 2310-2314.

Hwang, U.W., Park, C.J., Yong, T.S., Kim, W., 2001. One-step PCR amplification of complete arthropod mitochondrial genomes. Molecular Phylogenetics and Evolution 19, 345-352.

Hyman, B.C., Lewis, S.C., Tang, S., Wu, Z., 2011. Rampant gene rearrangement and haplotype hypervariation among nematode mitochondrial genomes. Genetica 139, 611-615.

Jabbar, A., Jex, A.R., Mohandas, N., Hall, R.S., Littlewood, D.T.J., Gasser, R.B., 2013. The mitochondrial genome of *Aelurostrongylus abstrusus*-diagnostic, epidemiological and systematic implications. Gene 516, 294-300.

Jeyaprakash, A., Hoy, M.A., 2007. The mitochondrial genome of the predatory mite Metaseiulus occidentalis (Arthropoda: Chelicerata: Acari: Phytoseiidae) is unexpectedly large and contains several novel features. Gene 391, 264-274.

Jiang, S.-T., Hong, G.-Y., Yu, M., Li, N., Yang, Y., Liu, Y.-Q., Wei, Z.-J., 2009. Characterization of the complete mitochondrial genome of the giant silkworm moth, *Eriogyna pyretorum* (Lepidoptera: Saturniidae). International Journal of Biological Sciences 5(4), 351-365.

Jocque, R., Dippenaar-Schoeman, A.S., 2006. Spider families of the world. Musee royal de l'Afrique centrale.

Johnson, K.P., Walden, K.K.O., Robertson, H.M., 2013. Next-generation phylogenomics using a target restricted assembly method. Molecular Phylogenetics and Evolution 66, 417-422.

Juehling, F., Puetz, J., Bernt, M., Donath, A., Middendorf, M., Florentz, C., Stadler, P.F., 2012a. Improved systematic tRNA gene annotation allows new insights into the evolution of mitochondrial tRNA structures and into the mechanisms of mitochondrial genome rearrangements. Nucleic Acids Research 40, 2833-2845.

Juehling, F., Puetz, J., Florentz, C., Stadler, P.F., 2012b. Armless mitochondrial tRNAs in Enoplea (Nematoda). RNA Biology 9, 1161-1166.

Kakioka, R., Kokita, T., Kumada, H., Watanabe, K., Okuda, N., 2013. A RAD-based linkage map and comparative genomics in the gudgeons (genus Gnathopogon, Cyprinidae). BMC Genomics 14, 32-32.

Kane, N., Sveinsson, S., Dempewolf, H., Yang, J.Y., Zhang, D., Engels, J.M.M., Cronk, Q., 2012. Ultra-barcoding in Cacao (*Theobroma* spp.; Malvaceae) using whole chloroplast genomes and nuclear ribosomal DNA. American Journal of Botany 99.

Kim, K.-H., Bae, J.-W., 2011. Amplification methods bias metagenomic libraries of uncultured single-stranded and double-stranded DNA viruses. Applied and Environmental Microbiology 77, 7663-7668.

Kircher, M., Heyn, P., Kelso, J., 2011. Addressing challenges in the production and analysis of Illumina sequencing data. BMC Genomics 12: 382 doi: 10.186/1471-2164-12-382.

Kocot, K.M., Cannon, J.T., Todt, C., Citarella, M.R., Kohn, A.B., Meyer, A., Santos, S.R., Schander, C., Moroz, L.L., Lieb, B., Halanych, K.M., 2011. Phylogenomics reveals deep molluscan relationships. Nature 477, 452.

Korbel, J.O., Urban, A.E., Affourtit, J.P., Godwin, B., Grubert, F., Simons, J.F., Kim, P.M., Palejev, D., Carriero, N.J., Du, L., Taillon, B.E., Chen, Z.T., Tanzer, A., Saunders, A.C.E., Chi, J.X., Yang, F.T., Carter, N.P., Hurles, M.E., Weissman, S.M., Harkins, T.T., Gerstein, M.B., Egholm, M., Snyder, M., 2007. Paired-end mapping reveals extensive structural variation in the human genome. Science 318, 420-426.

Lakner, C., Van Der Mark, P., Huelsenbeck, J.P., Larget, B., Ronquist, F., 2008. Efficiency of Markov chain Monte Carlo tree proposals in Bayesian phylogenetics. Systematic Biology 57, 86-103.

Lang, B.F., Burger, G., 2007. Purification of mitochondrial and plastid DNA. Nature Protocols 2, 652-660.

Lasken, R.S., Stockwell, T.B., 2007. Mechanism of chimera formation during the Multiple Displacement Amplification reaction. BMC Biotechnology 7, 19-19.

Laslett, D., Canback, D., 2008. ARWEN: a program to detect tRNA genes in metazoan mitochondrial nucleotide sequences. Bioinformatics 24, 172-175.

Lavrov, D.V., Boore, J.L., Brown, W.M., 2000a. The complete mitochondrial DNA sequence of the horseshoe crab *Limulus polyphemus*. Molecular Biology and Evolution 17, 813-824.

Lavrov, D.V., Brown, W.M., Boore, J.L., 2000b. A novel type of RNA editing occurs in the mitochondrial tRNAs of the centipede *Lithobius forficatus*. Proceedings of the National Academy of Sciences of the United States of America 97, 13738-13742.

Lavrov, D.V., Brown, W.M., Boore, J.L., 2004. Phylogenetic position of the Pentastomida and (pan)crustacean relationships. Proceedings of the Royal Society B-Biological Sciences 271, 537-544.

Lemey, P., Rambaut, A., Welch, J.J., Suchard, M.A., 2010. Phylogeography takes a relaxed random walk in continuous space and time. Molecular Biology and Evolution 27, 1877-1885.

Lerner, H.R.L., Fleischer, R.C., 2010. Prospects for the use of next-generation sequencing methods in ornithology. Auk 127, 4-15.

Lewis, D.L., Farr, C.L., Kaguni, L.S., 1995. Drosophila melanogaster mitochondrial DNA: Completion of the nucleotide sequence and evolutionary comparisons. Insect Molecular Biology 4, 263-278.

Lewis, P.O., 1998. A genetic algorithm for maximum-likelihood phylogeny inference using nucleotide sequence data. Molecular Biology and Evolution 15, 277-283.

Li, H., Liu, H., Song, F., Shi, A., Zhou, X., Cai, W., 2012. Comparative mitogenomic analysis of damsel bugs representing three tribes in the Family Nabidae (Insecta: Hemiptera). PLoS One 7(9): e45925.

Lilly, J.W., Havey, M.J., 2001. Small, repetitive DNAs contribute significantly to the expanded mitochondrial genome of cucumber. Genetics 159, 317-328.

Lizardi, P.M., Huang, X.H., Zhu, Z.R., Bray-Ward, P., Thomas, D.C., Ward, D.C., 1998. Mutation detection and single-molecule counting using isothermal rolling-circle amplification. Nature Genetics 19, 225-232.

Lopardo, L., Hormiga, G., 2008. Phylogenetic placement of the Tasmanian spider *Acrobleps hygrophilus* (Araneae, Anapidae) with comments on the evolution of the capture web in Araneoidea. Cladistics 24, 1-33.

Lowe, T.M., Eddy, S.R., 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res 25, 955-964.

Luan, B., Peng, H., Polonsky, S., Rossnagel, S., Stolovitzky, G., Martyna, G., 2010. Base-by-base ratcheting of single stranded DNA through a solid-state nanopore. Physical Review Letters 104, Issue 23.

Lynch, M., 1996. Mutation accumulation in transfer RNAs: Molecular evidence for Muller's ratchet in mitochondrial genomes. Molecular Biology and Evolution 13, 209-220.

Macey, J.R., Larson, A., Anajeva, N.B., Fang, Z., Papernfuss, T.J., 1997Two novel gene orders and the role of light-strand replication in rearrangement of the vertebrate mitochondrial genome. Mol. Biol. Evol. 14, 91-104.

Maier, R.M., Zeltz, P., Kossel, H., Bonnard, G., Gualberto, J.M., Grienenberger, J.M., 1996. RNA editing in plant mitochondria and chloroplasts. Plant Molecular Biology 32, 343-365.

Maraun, M., Schatz, H., Scheu, S., 2007. Awesome or ordinary? Global diversity patterns of oribatid mites. Ecography 30, 209-216.

Mardulyn, P., Termonia, A., Milinkovitch, M.C., 2003. Structure and evolution of the mitochondrial control region of leaf beetles (Coleoptera : Chrysomelidae): A hierarchical analysis of nucleotide sequence variation. Journal of Molecular Evolution 56, 38-45.

Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L.A., Berka, J., Braverman, M.S., Chen, Y.J., Chen, Z.T., Dewell, S.B., de Winter, A., Drake, J., Du, L., Fierro, J.M., Forte, R., Gomes, X.V., Godwin, B.C., He, W., Helgesen, S., Ho, C.H., Hutchison, S.K., Irzyk, G.P., Jando, S.C., Alenquer, M.L.I., Jarvie, T.P., Jirage, K.B., Kim, J.B., Knight, J.R., Lanza, J.R., Leamon, J.H., Lee, W.L., Lefkowitz, S.M., Lei, M., Li, J., Lohman, K.L., Lu, H., Makhijani, V.B., McDade, K.E., McKenna, M.P., Myers, E.W., Nickerson, E., Nobile, J.R., Plant, R., Puc, B.P., Reifler, M., Ronan, M.T., Roth, G.T., Sarkis, G.J., Simons, J.F., Simpson, J.W., Srinivasan, M., Tartaro, K.R., Tomasz, A., Vogt, K.A., Volkmer, G.A., Wang, S.H., Wang, Y., Weiner, M.P., Willoughby, D.A., Yu, P.G., Begley, R.F., Rothberg, J.M., 2006. Genome sequencing in microfabricated high-density picolitre reactors. Nature 441, 120-120.

Martino, A.J., Rhodes, M.E., Biddle, J.F., Brandt, L.D., Tomsho, L.P., House, C.H., 2012. Novel degenerate PCR method for whole-genome amplification applied to Peru Margin (ODP Leg 201) subsurface samples. Frontiers in microbiology 3, 17-17.

Masta, S.E., 2000a. Mitochondrial sequence evolution in spiders: Intraspecific variation in tRNAs lacking the T psi C arm. Molecular Biology and Evolution 17, 1091-1100.

Masta, S.E., 2000b. Phylogeography of the jumping spider *Habronattus pugillis* (Araneae : Salticidae): Recent vicariance of sky island populations? Evolution 54, 1699-1711.

Masta, S.E., Boore, J.L., 2004. The complete mitochondrial genome sequence of the spider *Habronattus oregonensis* reveals rearranged and extremely truncated tRNAs. Molecular Biology and Evolution 21, 893-902.

Masta, S.E., Boore, J.L., 2008. Parallel evolution of truncated transfer RNA genes in arachnid mitochondrial genomes. Molecular Biology and Evolution 25, 949-959.

Masta, S.E., Klann, A.E., Podsiadlowski, L., 2008. A comparison of the mitochondrial genomes from two families of Solifugae (Arthropoda: Chelicerata): Eremobatidae and Ammotrechidae. Gene 417, 35-42.

Masta, S.E., McCall, A., Longhorn, S.J., 2010. Rare genomic changes and mitochondrial sequences provide independent support for congruent relationships among the sea spiders (Arthropoda, Pycnogonida). Molecular Phylogenetics and Evolution 57, 59-70.

Mattila, T.M., Bechsgaard, J.S., Hansen, T.T., Schierup, M.H., Bilde, T., 2012. Orthologous genes identified by transcriptome sequencing in the spider genus *Stegodyphus*. BMC Genomics 13.

McClintlock, B., 1950. The origin and behaviour of mtatable loci in maize. Proc Natl Acad Sci USA 36, 344-355.

McKain, M.R., Wickett, N., Zhang, Y., Ayyampalayam, S., McCombie, W.R., Chase, M.W., Pires, J.C., dePamphilis, C.W., Leebens-Mack, J., 2012. Phylogenomic analysis of transcriptome data elucidates co-occurrence of a paleopolyploid event and the origin of bimodal karyotypes in Agavoideae (Asparagaceae). American Journal of Botany 99, 397-406.

Meer, M.V., Kondrashov, A.S., Artzy-Randrup, Y., Kondrashov, F.A., 2010. Compensatory evolution in mitochondrial tRNAs navigates valleys of low fitness. Nature 464, 279-U159.

Mercati, F., Riccardi, P., Leebens-Mack, J., Abenavoli, M.R., Falavigna, A., Sunseri, F., 2013. Single nucleotide polymorphism isolated from a novel EST dataset in garden asparagus (*Asparagus officinalis L.*). Plant Science 203-204, 115-123.

Merriman, B., Rothberg, J.M., Ion Torrent, R., Team, D., 2012. Progress in Ion Torrent semiconductor chip based sequencing. Electrophoresis 33, 3397-3417.

Metzker, M.L., 2005. Emerging technologies in DNA sequencing. Genome Research 15, 1767-1776.

Metzker, M.L., 2010. Applications of next-generation sequencing. Sequencing technologies - the next generation. Nature Reviews Genetics 11, 31-46.

Meyer, M., Stenzel, U., Hofreiter, M., 2008. Parallel tagged sequencing on the 454 platform. Nature Protocols 3, 267-278.

Meyer, M., Stenzel, U., Myles, S., Prufer, K., Hofreiter, M., 2007. Targeted high-throughput sequencing of tagged nucleic acid samples. Nucleic Acids Research 35(15): e97.

Milbury, C.A., Lee, J.C., Cannone, J.J., Gaffney, P.M., Gutell, R.R., 2010. Fragmentation of the large subunit ribosomal RNA gene in oyster mitochondrial genomes. BMC Genomics 11: doi: 10.1186/1471-2164-11-485.

Miller, J.A., Carmichael, A., Ramirez, M.J., Spagna, J.C., Haddad, C.R., Rezac, M., Johannesen, J., Kral, J., Wang, X.-P., Griswold, C.E., 2010. Phylogeny of entelegyne spiders: Affinities of the family Penestomidae (NEW RANK), generic phylogeny of Eresidae, and asymmetric rates of change in spinning

organ evolution (Araneae, Araneoidea, Entelegynae). Molecular Phylogenetics and Evolution 55, 786-804.

Mindell, D.P., Honeycutt, R.L., 1990. Ribosomal-RNA in Veretbrates – evolution and phylogenetic applications. Annual Review of Ecology and Systematics 21, 541-566.

Miya, M., Nishida, M., 2000. Use of mitogenomic information in teleostean molecular phylogenetics: A tree-based exploration under the maximum-parsimony optimality criterion. Molecular Phylogenetics and Evolution 17, 437-455.

Mohanty, P., Sahoo, L., Parida, K., Das, P., 2013. Development of polymorphic EST-SSR markers in *Macrobrachium rosenbergii* by data mining. Conservation Genetics Resources 5, 133-136.

Moradmand, M., Jaeger, P., 2012. Taxonomic revision of the huntsman spider genus *Eusparassus* Simon, 1903 (Araneae: Sparassidae) in Eurasia. Journal of Natural History 46, 2439-2496.

Moret, B.M.E., Warnow, T., 2005. Advances in phylogeny reconstruction from gene order and content data. Molecular Evolution: Producing the Biochemical Data, Part B 395, 673.

Morin, P.A., Archer, F.I., Foote, A.D., Vilstrup, J., Allen, E.E., Wade, P., Durban, J., Parsons, K., Pitman, R., Li, L., Bouffard, P., Nielsen, S.C.A., Rasmussen, M., Willerslev, E., Gilbert, M.T.P., Harkins, T., 2010. Complete mitochondrial genome phylogeographic analysis of killer whales (*Orcinus orca*) indicates multiple species. Genome Research 20, 908-916.

Moritz, C., Brown, W.M., 1986. Tandem duplications od D-loop and ribosomal RNA sequences in lizard mitochondrial DNA. Science 233, 1425-1427.

Moritz, C., Dowling, T.E., Brown, W.M., 1987. Evolution of animal mitochondrial-DNA – relevance for population biology and systematics. Annual Review of Ecology and Systematics 18, 269-292.

Mueller, R.L., 2006. Evolutionary rates, divergence dates, and the performance of mitochondrial genes in Bayesian phylogenetic analysis. Systematic Biology 55, 289-300.

Murchison, E.P., Tovar, C., Hsu, A., Bender, H.S., Kheradpour, P., Rebbeck, C.A., Obendorf, D., Conlan, C., Bahlo, M., Blizzard, C.A., Pyecroft, S., Kreiss, A., Kellis, M., Stark, A., Harkins, T.T., Graves, J.A.M., Woods, G.M., Hannon, G.J., Papenfuss, A.T., 2010. The Tasmanian devil transcriptome reveals Schwann cell origins of a clonally transmissible cancer. Science 327, 84-87.

Murphy, N.P., Framenau, V.W., Donnellan, S.C., Harvey, M.S., Park, Y.C., Austin, A.D., 2006. Phylogenetic reconstruction of the wolf spiders (Araneae : Lycosidae) using sequences from the 12S rRNA, 28S rRNA, and NADH1 genes: Implications for classification, biogeography, and the evolution of web building behavior. Molecular Phylogenetics and Evolution 38, 583-602.

Nery, M.F., Gonzalez, D.J., Hoffmann, F.G., Opazo, J.C., 2012. Resolution of the laurasiatherian phylogeny: Evidence from genomic data. Molecular Phylogenetics and Evolution 64, 685-689.

Nylander, J.A.A., Ronquist, F., Huelsenbeck, J.P., Nieves-Aldrey, J.L., 2004. Bayesian phylogenetic analysis of combined data. Systematic Biology 53, 47-67.

O'Huigin, C., Satta, Y., Takahata, N., Klein, J., 2002. Contribution of homoplasy and of ancestral polymorphism to the evolution of genes in anthropoid primates. Molecular Biology and Evolution 19, 1501-1513.

Ogihara, Y., Yamazaki, Y., Murai, K., Kanno, A., Terachi, T., Shiina, T., Miyashita, N., Nasuda, S., Nakamura, C., Mori, N., Takumi, S., Murata, M., Futo, S., Tsunewaki,

K., 2005. Structural dynamics of cereal mitochondrial genomes as revealed by complete nucleotide sequencing of the wheat mitochondrial genome. Nucleic Acids Research 33, 6235-6250.

Ohtsuki, T., Kawai, G., Watanabe, K., 2002. The minimal tRNA: unique structure of *Ascaris suum* mitochondrial tRNA(UCU)(Ser) having a short T arm and lacking the entire D arm. EBS Letters 514, 37-43.

Ojala, D., Montoya, J., Attardi, G., 1981. Transfer-RNA punctuation model of RNA processing in human mitochondria. Nature 290, 470-474.

Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Nikaido, I., Osato, N., Saito, R., Suzuki, H., Yamanaka, I., Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D.P., Bult, C., Hume, D.A., Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuda, H., Batalov, S., Beisel, K.W., Blake, J.A., Bradt, D., Brusic, V., Chothia, C., Corbani, L.E., Cousins, S., Dalla, E., Dragani, T.A., Fletcher, C.F., Forrest, A., Frazer, K.S., Gaasterland, T., Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S., Gustincich, S., Hirokawa, N., Jackson, I.J., Jarvis, E.D., Kanai, A., Kawaji, H., Kawasawa, Y., Kedzierski, R.M., King, B.L., Konagaya, A., Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons, P.A., Maglott, D.R., Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T., Numata, K., Okido, T., Pavan, W.J., Pertea, G., Pesole, G., Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ramachandran, S., Ravasi, T., Reed, J.C., Reed, D.J., Reid, J., Ring, B.Z., Ringwald, M., Sandelin, A., Schneider, C., Semple, C.A.M., Setou, M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale, R.D., Tomita, M., Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y., Wells, C., Wilming, L.G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P., Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M., Sakazume, N.,

Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K., Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K., Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E.S., Rogers, J., Birney, E., Hayashizaki, Y., Consurtium, F., Grp, R.G.E.R., 2002. Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. Nature 420, 563-573.

Okimoto, R., Wolstenholme, D.R., 1990. A set of transfer-RNAS that lack either the T-psi-C arm or the dihydrouridine arm – towards a minimal transfer-RNA adapter. Embo Journal 9, 3405-3411.

Opell, B.D., 2002. How spider anatomy and thread configuration shape the stickiness of cribellar prey capture threads. Journal of Arachnology 30, 10-19.

Oxford, G.S., Gunnarsson, B., 2006. Spatial variation in colour morph, spotting and allozyme frequencies in the candy-stripe spider, Enoplognatha ovata (Theridiidae) on two Swedish archipelagos. Genetica 128, 51-62.

Pacheco, M.A., Battistuzzi, F.U., Lentino, M., Aguilar, R.F., Kumar, S., Escalante, A.A., 2011. Evolution of modern birds revealed by mitogenomics: timing the radiation and origin of major orders. Molecular Biology and Evolution 28, 1927-1942.

Paegel, B.M., Blazej, R.G., Mathies, R.A., 2003. Microfluidic devices for DNA sequencing: sample preparation and electrophoretic analysis. Current Opinion in Biotechnology 14, 42-50.

Palacios, G., Druce, J., Du, L., Tran, T., Birch, C., Briese, T., Conlan, S., Quan, P., Hui, J., Marshall, J., Simons, J.F., Egholm, M., Paddock, C.D., Shieh, W., Goldsmith, C.S., Zaki, S.R., Catton, M., Lipkin, W.I., 2008. A new arenavirus in a cluster of fatal transplant-associated diseases. New England Journal of Medicine 358, 991-998.

Park, Y.C., Yoo, J.S., Schwarz, M.P., Murphy, N., Kim, J.P., 2007. Molecular phylogeny of east Asian wolf spiders (Araneae : Lycosidae) inferred from mitochondrial 12S ribosomal DNA. Annals of the Entomological Society of America 100, 1-8.

Penney, D., 2010. Dominican amber. Biodiversity of fossils in amber from the major world deposits., 22-41.

Penney, D., Dunlop, J.A., Marusik, Y.M., 2012a. Summary statistics for fossil spider species taxonomy. Zookeys.

Penney, D., Green, D.I., McNeil, A., Bradley, R.S., Marusik, Y.M., Withers, P.J., Preziosi, R.F., 2012b. A new species of Craspedisia (Araneae: Theridiidae) in Miocene Dominican amber, imaged using X-ray computed tomography. Paleontological Journal 46, 583-588.

Penney, D., Selden, P.A., 2011. Fossil spiders: the evolutionary history of a mega-diverse order. Fossil spiders: the evolutionary history of a mega-diverse order. Siri Scientific Press Monograph Series Volume 1.

Penney, D., Wheater, C.P., Selden, P.A., 2003. Resistance of spiders to Cretaceous-Tertiary extinction events. Evolution 57, 2599-2607.

Pett, W., Ryan, J.F., Pang, K., Mullikin, J.C., Martindale, M.Q., Baxevanis, A.D., Lavrov, D.V., 2011. Extreme mitochondrial evolution in the ctenophore *Mnemiopsis leidyi*: Insight from mtDNA and the nuclear genome. Mitochondrial DNA 22, 130-142.

Philippe, H., Derelle, R., Lopez, P., Pick, K., Borchiellini, C., Boury-Esnault, N., Vacelet, J., Renard, E., Houliston, E., Queinnec, E., Da Silva, C., Wincker, P., Le Guyader, H., Leys, S., Jackson, D.J., Schreiber, F., Erpenbeck, D., Morgenstern, B., Woerheide, G., Manuel, M., 2009. Phylogenomics revives traditional views on deep animal relationships. Current Biology 19, 706-712.

Philippe, H., Telford, M.J., 2006. Large-scale sequencing and the new animal phylogeny. Trends in Ecology & Evolution 21, 614-620.

Phillips, M.J., Delsuc, F., Penny, D., 2004. Genome-scale phylogeny and the detection of systematic biases. Molecular Biology and Evolution 21, 1455-1458.

Pick, C., Ebersberger, I., Spielmann, T., Bruchhaus, I., Burmester, T., 2011. Phylogenomic analyses of malaria parasites and evolution of their exported proteins. BMC Evolutionary Biology 11.

Platnick, N.I., 1999. Dimensions of biodiversity: targeting megadiverse groups. Columbia University Press.

Platnick, N.I., 2012. The world spider catalog, version 13.0. American Museum of Natural History. American Museum of Natural History, online at http://research.amnh.org/iz/spiders/catalog. DOI: 10.5531/db.iz.0001.

Platnick, N.I., Abrahim, N., Alvarez-Padilla, F., Andriamalala, D., Baehr, B.C., Baert, L., Bonaldo, A.B., Brescovit, A.D., Chousou-Polydouri, N., Duperre, N., Eichenberger, B., Fannes, W., Gaublomme, E., Gillespie, R.G., Grismado, C.J., Griswold, C.E., Harvey, M.S., Henrard, A., Hormiga, G., Izquierdo, M.A., Jocque, R., Kranz-Baltensperger, Y., Kropf, C., Ott, R., Ramirez, M.J., Raven, R.J., Rheims, C.A., Ruiz, G.R.S., Santos, A.J., Saucedo, A., Sierwald, P., Szuets, T., Ubick, D., Wang, X.-P., 2012. Tarsal organ morphology and the phylogeny of goblin spiders (Araneae, Oonopidae), with notes on basal genera. American Museum Novitates, 1-52.

Posada, D., Buckley, T.R., 2004. Model selection and model averaging in phylogenetics: Advantages of akaike information criterion and Bayesian approaches over likelihood ratio tests. Systematic Biology 53, 793-808.

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

Prosdocimi, F., de Carvalho, D.C., de Almeida, R.N., Beheregaray, L.B., 2012. The complete mitochondrial genome of two recently derived species of the fish genus *Nannoperca* (Perciformes, Percichthyidae). Molecular Biology Reports 39, 2767-2772.

Qiu, Y., Song, D.X., Zhou, K.Y., Sun, H.Y., 2005. The mitochondrial sequences of *Heptathela hangzhouensis* and *Ornithoctonus huwena* reveal unique gene arrangements and atypical tRNAs. Journal of Molecular Evolution 60, 57-71.

Quental, T.B., Marshall, C.R., 2010. Diversity dynamics: molecular phylogenies need the fossil record. Trends in Ecology & Evolution 25, 434-441.

Quinn, T.W., Wilson, A.C., 1993. Sequence evolution in and around the mitochondrial control region in birds. Journal of Molecular Evolution 37, 417-425.

Raimond, R., Marcade, I., Bouchon, D., Rigaud, T., Bossy, J.P., Souty-Grosset, C., 1999. Organization of the large mitochondrial genome in the isopod *Armadillidium vulgare*. Genetics 151, 203-210.

Reyes, A., Gissi, C., Pesole, G., Saccone, C., 1998. Asymmetrical directional mutation pressure in the mitochondrial genome of mammals. Molecular Biology and Evolution 15, 957-966.

Rix, M.G., Harvey, M.S., Roberts, J.D., 2010. A revision of the textricellin spider genus Raveniella (Araneae : Araneoidea : Micropholcommatidae): exploring patterns of phylogeny and biogeography in an Australian biodiversity hotspot. Invertebrate Systematics 24, 209-237.

Rodrigue, S., Malmstrom, R.R., Berlin, A.M., Birren, B.W., Henn, M.R., Chisholm, S.W., 2009. Whole genome amplification and *de novo* assembly of single bacterial cells. PLoS One 4.

Rogaev, E.I., Moliaka, Y.K., Malyarchuk, B.A., Kondrashov, F.A., Derenko, M.V., Chumakov, I., Grigorenko, A.P., 2006. Complete mitochondrial genome and phylogeny of Pleistocene mammoth Mammuthus primigenius. PLoS Biology 4, 403-410.

Rohland, N., Malaspinas, A.S., Pollack, J.L., Slatkin, M., Matheus, P., Hofreiter, M., 2007. Proboscidean mitogenomics: Chronology and mode of elephant evolution using mastodon as outgroup. PLoS Biology 5, 1663-1671.

Rokas, A., Carroll, S.B., 2005. More genes or more taxa? The relative contribution of gene number and taxon number to phylogenetic accuracy. Molecular Biology and Evolution 22, 1337-1344.

Rokas, A., Holland, P.W.H., 2000. Rare genomic changes as a tool for phylogenetics. Trends in Ecology & Evolution 15, 454-459.

Rokas, A., Williams, B.L., King, N., Carroll, S.B., 2003. Genome-scale approaches to resolving incongruence in molecular phylogenies. Nature 425, 798-804.

Ronaghi, M., Karamohamed, S., Pettersson, B., Uhlen, M., Nyren, P., 1996. Real-time DNA sequencing using detection of pyrophosphate release. Analytical Biochemistry 242, 84-89.

Ronquist, F., Klopfstein, S., Vilhelmsen, L., Schulmeister, S., Murray, D.L., Rasnitsyn, A.P., 2012. A total-evidence approach to dating with fossils, applied to the early radiation of the Hymenoptera. Systematic Biology 61, 973-999.

Rothberg, J.M., Hinz, W., Rearick, T.M., Schultz, J., Mileski, W., Davey, M., Leamon, J.H., Johnson, K., Milgrew, M.J., Edwards, M., Hoon, J., Simons, J.F., Marran, D., Myers, J.W., Davidson, J.F., Branting, A., Nobile, J.R., Puc, B.P., Light, D., Clark, T.A., Huber, M., Branciforte, J.T., Stoner, I.B., Cawley, S.E., Lyons, M., Fu, Y., Homer, N., Sedova, M., Miao, X., Reed, B., Sabina, J., Feierstein, E., Schorn, M., Alanjary, M., Dimalanta, E., Dressman, D., Kasinskas, R., Sokolsky, T., Fidanza,

J.A., Namsaraev, E., McKernan, K.J., Williams, A., Roth, G.T., Bustillo, J., 2011. An integrated semiconductor device enabling non-optical genome sequencing. Nature 475, 348-352.

Rothberg, J.M., Leamon, J.H., 2008. The development and impact of 454 sequencing. Nature Biotechnology 26, 1117-1124.

Rozen, S., Skaletsky, H., 2000. Primer3 on the WWW for general users and for biologist programmers. Methods in Molecular Biology (Clifton, N.J.) 132.

Rusk, N., 2009. Cheap third-generation sequencing. Nature Methods 6, 244-245.

Saccone, C., De Giorgi, C., Gissi, C., Pesole, G., Reyes, A., 1999. Evolutionary genomics in Metazoa: the mitochondrial DNA as a model system. Gene 238, 195-209.

Salzberg, S.L., Phillippy, A.M., Zimin, A., Puiu, D., Magoc, T., Koren, S., Treangen, T.J., Schatz, M.C., Delcher, A.L., Roberts, M., Marcais, G., Pop, M., Yorke, J.A., 2012. GAGE: A critical evaluation of genome assemblies and assembly algorithms. Genome Research 22, 557-567.

Sanchez, R., Serra, F., Tarraga, J., Medina, I., Carbonell, J., Pulido, L., de Maria, A., Capella-Gutierrez, S., Huerta-Cepas, J., Gabaldon, T., Dopazo, J., Dopazo, H., 2011. Phylemon 2.0: a suite of web-tools for molecular evolution, phylogenetics, phylogenomics and hypotheses testing. Nucleic Acids Research 39, W470-W474.

Sanger, F., 1981. Determination of nucleotide-sequences in DNA. Bioscience Reports 1, 3-18.

Sanger, F., Coulson, A.R., 1978. Use of thin acrylamide gels for DNA sequencing. EBS Letters 87, 107-110.

Sanger, F., Nicklen, S., Coulson, A.R., 1992. DNA sequencing with chain-terminating inhibitors. 1977. Biotechnology 24, 104-108.

Saupe, E.E., Selden, P.A., 2011. The study of fossil spider species. Comptes Rendus Palevol 10, 181-188.

Scaglione, D., Acquadro, A., Portis, E., Tirone, M., Knapp, S.J., Lanteri, S., 2012. RAD tag sequencing as a source of SNP markers in *Cynara cardunculus L*. BMC Genomics 13:3 doi: 10.1186/1471-2164-13-3.

Schadt, E.E., Turner, S., Kasarskis, A., 2011. A window into third generation sequencing. Human Molecular Genetics 20, 853-853.

Schneider, A., 2011. Mitochondrial tRNA Import and Its Consequences for Mitochondrial Translation. Annual Review of Biochemistry, Vol 80, 1033-1053.

Schuster, W., Hiesel, R., Brennicke, A., 1993. RNA editing in plant mitochondria. Seminars in Cell Biology 4, 279-284.

Scott, A.G., Oxford, G.S., Selden, P.A., 2006. Epigeic spiders as ecological indicators of conservation value for peat. Biological Conservation 127, 420-428.

Segovia, R., Pett, W., Trewick, S., Lavrov, D.V., 2011. Extensive and evolutionarily persistent mitochondrial tRNA editing in velvet worms (Phylum Onychophora). Molecular Biology and Evolution 28, 2873-2881.

Selden, P.A., 1996. First fossil mesothele spider, from the Carboniferous of France. Revue Suisse de Zoologie Volume Hors Serie 2, 585-596.

Selden, P.A., Shih, C., Ren, D., 2011. A golden orb-weaver spider (Araneae: Nephilidae: *Nephila*) from the Middle Jurassic of China. Biology Letters 7, 775-778.

Seo, T.S., Bai, X.P., Kim, D.H., Meng, Q.L., Shi, S.D., Ruparelt, H., Li, Z.M., Turro, N.J., Ju, J.Y., 2005. Four-color DNA sequencing by synthesis on a chip using photocleavable fluorescent nucleotides. Proceedings of the National Academy of Sciences of the United States of America 102, 5926-5931.

Sereno, P.C., 2008. Logical basis for morphological characters in phylogenetics. Cladistics 24, 396.

Shamblin, B.M., Bjorndal, K.A., Bolten, A.B., Hillis-Starr, Z.M., Lundgren, I., Naro-Maciel, E., Nairn, C.J., 2012. Mitogenomic sequences better resolve stock structure of southern Greater Caribbean green turtle rookeries. Molecular Ecology 21, 2330-2340.

Shao, R., Aoki, Y., Mitani, H., Tabuchi, N., Barker, S.C., Fukunaga, M., 2004. The mitochondrial genomes of soft ticks have an arrangement of genes that has remained unchanged for over 400 million years. Insect Molecular Biology 13, 219-224.

Shaw, J., Lickey, E.B., Schilling, E.E., Small, R.L., 2007. Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: The tortoise and the hare III. American Journal of Botany 94, 275-288.

Shear, W.A., Palmer, J.M., Coddington, J.A., Bonamo, P.M., 1989. A Devonian spinneret - early evidence of spiders and silk use. Science 246, 479-481.

Sheffield, N.C., Song, H., Cameron, L., Whiting, M.F., 2008. A comparative analysis of mitochondrial genomes in Coleoptera (Arthropoda: Insecta) and genome descriptions of six new beetles. Molecular Biology and Evolution 25, 2499-2509.

Shendure, J., Ji, H., 2008. Next-generation DNA sequencing. Nature Biotechnology 26, 1135-1145.

Sierwald, P., Bond, J.E., 2007. Current status of the myriapod class Diplopoda (Millipedes): Taxonomic diversity and phylogeny. Annual Review of Entomology 52, 401-420.

Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Soeding, J., Thompson, J.D., Higgins, D.G., 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Molecular Systems Biology 7:539 doi:10.1038/msb.2011.75.

Simison, W.B., Lindberg, D.R., Boore, J.L., 2006. Rolling circle amplification of metazoan mitochondrial genomes. Molecular Phylogenetics and Evolution 39, 562-567.

Simon, C., Frati, F., Beckenbach, A., Crespi, B., Liu, H., Flook, P., 1994. Evolution, weighting, and phylogenetic utility of mitochondrial gene-sequences and a compilation of conserved polymerase chain-reaction primers. Annals of the Entomological Society of America 87, 651-701.

Smith, A.B., 1994. Systematics and the fossil record: documenting evolutionary patterns., Oxford, Blackwell Science, i-viii, 1-223.

Soltis, P.S., Soltis, D.E., Chase, M.W., 1999. Angiosperm phylogeny inferred from multiple genes as a tool for comparative biology. Nature 402, 402-404.

Song, D., Zhu, M., Chen, J., 1999. The spiders of China.. Hebai, Science and Tecnology Publishing House, 1-640.

Sorenson, M.D., Ast, J.C., Dimcheff, D.E., Yuri, T., Mindell, D.P., 1999. Primers for a PCR-based approach to mitochondrial genome sequencing in birds and other vertebrates. Molecular Phylogenetics and Evolution 12(2), 105-114.

Spinks, P.Q., Thomson, R.C., Lovely, G.A., Shaffer, H.B., 2009. Assessing what is needed to resolve a molecular phylogeny: simulations and empirical data from emydid turtles. BMC Evolutionary Biology 9(56): doi: 10.1186/1471-2148-9-56.

Staton, J.L., Daehler, L.L., Brown, W.M., 1997. Mitochondrial gene arrangement of the horseshoe crab *Limulus polyphemus L*: Conservation of major features among arthropod classes. Molecular Biology and Evolution 14, 867-874.

Stewart, J.B., Beckenbach, A.T., 2005. Insect mitochondrial genomics: the complete mitochondrial genome sequence of the meadow spittlebug *Philaenus spumarius* (Hemiptera : Auchenorrhyncha : Cercopoidae). Genome 48, 46-54.

Stokes, A., Drozdov, I., Guerra, E., Ouzounis, C.A., Warnakulasuriya, S., Gleeson, M.J., McGurk, M., Tavassoli, M., Odell, E.W., 2011. Copy number and loss of heterozygosity detected by SNP array of formalin-fixed tissues using whole-genome amplification. PLoS One 6(9): e24503.

Straub, S.C.K., Parks, M., Weitemier, K., Fishbein, M., Cronn, R.C., Liston, A., 2012. Navigating the tip of the genomic iceberg: next-generation sequencing for plant systematics. American Journal of Botany 99, 349-364.

Struck, T.H., Paul, C., Hill, N., Hartmann, S., Hoesel, C., Kube, M., Lieb, B., Meyer, A., Tiedemann, R., Purschke, G., Bleidorn, C., 2011. Phylogenomic analyses unravel annelid evolution. Nature 471, 95-113.

Sunderland, K.D., Fraser, A.M., Dixon, A.F.G., 1986. Field and laboratory studies on money spiders (Linyphiidae) as predators of cereal aphids. Journal of Applied Ecology 23, 433-447.

Talavera, G., Vila, R., 2011. What is the phylogenetic signal limit from mitogenomes? The reconciliation between mitochondrial and nuclear data in the Insecta class phylogeny. BMC Evolutionary Biology 11(315): doi:10.1186/1471-2148-11-315.

Tang, K., Taranenko, N.I., Allman, S.L., Chang, L.Y., Chen, C.H., 1994. Detection of 500-Nucleotide DNA by laser-desorption mass-spectrometry. Rapid Communications in Mass Spectrometry 8, 727-730.

Tang, Q.-Y., Liu, S.-Q., Yu, D., Liu, H.-Z., Danley, P.D., 2012. Mitochondrial capture and incomplete lineage sorting in the diversification of balitorine loaches (Cypriniformes, Balitoridae) revealed by mitochondrial and nuclear genes. Zoologica Scripta 41, 233-247.

Tang, S., Hyman, B.C., 2005. Rolling circle amplification of complete nematode mitochondrial genomes. Journal of Nematology 37, 236-241.

Tang, S., Hyman, B.C., 2007. Mitochondrial genome haplotype hypervariation within the isopod parasitic nematode *Thaumamermis cosgrovei*. Genetics 176, 1139-1150.

Tarraga, J., Medina, I., Arbiza, L., Huerta-Cepas, J., Gabaldon, T., Dopazo, J., Dopazo, H., 2007. Phylemon: a suite of web tools for molecular evolution, phylogenetics and phylogenomics. Nucleic Acids Research 35, W38-W42.

Timmermans, M.J.T.N., Dodsworth, S., Culverwell, C.L., Bocak, L., Ahrens, D., Littlewood, D.T.J., Pons, J., Vogler, A.P., 2010. Why barcode? High-throughput multiplex sequencing of mitochondrial genomes for molecular systematics. Nucleic Acids Research 38: doi: 10.1093/nar/gkq807.

Van Leeuwen, T., Vanholme, B., Van Pottelberge, S., Van Nieuwenhuyse, P., Nauen, R., Tirry, L., Denholm, I., 2008. Mitochondrial heteroplasmy and the evolution of insecticide resistance: Non-Mendelian inheritance in action. Proceedings of the National Academy of Sciences of the United States of America 105, 5980-5985.

Venkatesh, B., Ning, Y., Brenner, S., 1999. Late changes in spliceosomal introns define clades in vertebrate evolution. Proceedings of the National Academy of Sciences of the United States of America 96, 10267-10271.

Vezzi, F., Narzisi, G., Mishra, B., 2012. Feature-by-feature - evaluating *de novo* sequence sssembly. PLoS One 7(2): e31002.

Vink, C.J., Hedin, M., Bodner, M.R., Maddison, W.P., Hayashi, C.Y., Garb, J.E., 2008. Actin 5C, a promising nuclear gene for spider phylogenetics. Molecular Phylogenetics and Evolution 48, 377-382.

von Bubnoff, A., 2008. Next-generation sequencing: The race is on. Cell 132, 721-723.

Wake, D.B., 1981. The application of allozyme evidence to problems in the evolution of morphology. Evolution today. Proceedings of the second International Congress of Systematic and Evolutionary Biology., 257-270.

Wake, D.B., 1982. Functional and Evolutionary Morphology. Perspectives in Biology and Medicine 25, 603-620.

Walter, D.E., Proctor, H.C., 1999. Mites: ecology, evolution and behaviour. Mites: ecology, evolution and behaviour., i-x, 1-322.

Wang, X., Lavrov, D.V., 2011. Gene recruitment – A common mechanism in the evolution of transfer RNA gene families. Gene 475, 22-29.

Wang, J., Yang, G., 2011. The complete mitogenome of the snakehead *Channa argus* (Perciformes: Channoidei): Genome characterization and phylogenetic implications. Mitochondrial DNA 22, 120-129.

Watanabe, Y.I., Tsurui, H., Ueda, T., Furushima, R., Takamiya, S., Kita, K., Nishikawa, K., Watanabe, K., 1994. Primary-order and high-order structures of nematode (*Ascaris suum*) mitochondrial transfer-RNAs lacking either the T-stem or D-stem. Journal of Biological Chemistry 269, 22902-22906.

Webster, B., Littlewood, D.T.J., 2012. Mitochondrial gene order change in *Schistosoma* (Platyhelminthes: Digenea: Schistosomatidae). International Journal for Parasitology 42, 313-321.

Wenink, P.W., Baker, A.J., Tilanus, M.G.J., 1994. Mitochondrial control-region in 2 shorbird species, the Turnstone and the Dunlin, and their utility and population genetic-studies. Molecular Biology and Evolution 11, 22-31.

Werle, E., Schneider, C., Renner, M., Volker, M., Fiehn, W., 1994. Convenient single step, one tube purification of PCR products for direct sequencing. Nucleic Acids Research 22, 4354-4355.

Wheeler, D.A., Srinivasan, M., Egholm, M., Shen, Y., Chen, L., McGuire, A., He, W., Chen, Y.J., Makhijani, V., Roth, G.T., Gomes, X., Tartaro, K., Niazi, F., Turcotte, C.L., Irzyk, G.P., Lupski, J.R., Chinault, C., Song, X.Z., Liu, Y., Yuan, Y., Nazareth, L., Qin, X., Muzny, D.M., Margulies, M., Weinstock, G.M., Gibbs, R.A., Rothberg, J.M., 2008. The complete genome of an individual by massively parallel DNA sequencing. Nature 452, 872-875.

Wheeler, W.C., Hayashi, C.Y., 1998. The phylogeny of the extant chelicerate orders. Cladistics 14, 173-192.

Wiens, J.J., Kuczynski, C.A., Smith, S.A., Mulcahy, D.G., Sites, J.W., Jr., Townsend, T.M., Reeder, T.W., 2008. Branch lengths, support, and congruence: Testing the phylogenomic approach with 20 nuclear loci in snakes. Systematic Biology 57, 420-431.

Wiens, J.J., Kuczynski, C.A., Townsend, T., Reeder, T.W., Mulcahy, D.G., Sites, J.W., Jr., 2010. Combining phylogenomics and fossils in higher-level squamate reptile phylogeny: molecular data change the placement of fossil taxa. Systematic Biology 59, 674-688.

Williams, B.A.P., Keeling, P.J., 2003. Cryptic organelles in parasitic protists and fungi. Advances in Parasitology 54, 9-68.

Wilson, K., Cahill, V., Ballment, E., Benzie, J., 2000. The complete sequence of the mitochondrial genome of the crustacean Penaeus monodon: Are

malacostracan crustaceans more closely related to insects than to branchiopods? Molecular Biology and Evolution 17, 863-874.

Woese, C.R., Fox, G.E., 1977. Phylogenetic structures of prokaryotic domain – primary Kingdoms. Proceedings of the National Academy of Sciences of the United States of America 74, 5088-5090.

Wolstenholme, D.R., 1992. Animal mitochondrial-DNA – structure and evolution. International Review of Cytology 141, 173-216.

Wolstenholme, D.R., Macfarlane, J.L., Okimoto, R., Clary, D.O., Wahleithner, J.A., 1987. Bizarre Transfer-RNAs Inferred from DNA-Sequences of Mitochondrial Genomes of Nematode Worms. Proceedings of the National Academy of Sciences of the United States of America 84, 1324-1328.

Wood, H.M., Griswold, C.E., Gillespie, R.G., 2012. Phylogenetic placement of pelican spiders (Archaeidae, Araneae), with insight into evolution of the "neck" and predatory behaviours of the superfamily Palpimanoidea. Cladistics 28, 598-626.

Wortley, A.H., Rudall, P.J., Harris, D.J., Scotland, R.W., 2005. How much data are needed to resolve a difficult phylogeny? Case study in Lamiales. Systematic Biology 54, 697-709.

Wyman, S.K., Jansen, R.K., Boore, J.L., 2005. Automatic annotation of organellar genomes with DOGMA. Bioinformatics 20, 3252-3255.

Yokobori, S., Fukuda, N., Nakamura, M., Aoyama, T., 2004. Long-term conservation of six duplicated structural genes in cephalopod mitochondrial genomes. Molecular Biology and Evolution 21, 2034-2046.

Yokobori, S., Paabo, S., 1995. Transfer-RNA editing in land snail mitochondria. Proceedings of the National Academy of Sciences of the United States of America 92, 10432-10435.

Yuan, M.-L., Wei, D.-D., Wang, B.-J., Dou, W., Wang, J.-J., 2010. The complete mitochondrial genome of the citrus red mite *Panonychus citri* (Acari: Tetranychidae): high genome rearrangement and extremely truncated tRNAs. BMC Genomics 11(597): doi: 10.1186/1471-2164-11-597.

Zardoya, R., Meyer, A., 1996. Phylogenetic performance of mitochondrial protein-coding genes in resolving relationships among vertebrates. Molecular Biology and Evolution 13, 933-942.

Zhang, H., Chen, L., 2011. The complete mitochondrial genome of dhole *Cuon alpinus*: phylogenetic analysis and dating evolutionary divergence within canidae. Molecular Biology Reports 38, 1651-1660.

Zhang, H.-L., Zeng, H.-H., Huang, Y., Zheng, Z.-M., 2013. The complete mitochondrial genomes of three grasshoppers, *Asiotmethis zacharjini*, *Filchnerella helanshanensis* and *Pseudotmethis rubimarginis* (Orthoptera: Pamphagidae). Gene 517, 89-98.

Zhou, X., Xu, S., Xu, J., Chen, B., Zhou, K., Yang, G., 2012. Phylogenomic analysis resolves the interordinal relationships and rapid diversification of the Laurasiatherian mammals. Systematic Biology 61, 150-164.

Zhou, X., Xu, S., Yang, Y., Zhou, K., Yang, G., 2011. Phylogenomic analyses and improved resolution of Cetartiodactyla. Molecular Phylogenetics and Evolution 61, 255-264.

8.Appendix

OPEN ACCESS Freely available online

PLOS ONE

Can Long-Range PCR Be Used to Amplify Genetically Divergent Mitochondrial Genomes for Comparative Phylogenetics? A Case Study within Spiders (Arthropoda: Araneae)

Andrew G. Briscoe[1], Sara Goodacre[2], Susan E. Masta[3], Martin I. Taylor[1], Miquel A. Arnedo[4], David Penney[5], John Kenny[6], Simon Creer[1]*

1 Environment Centre Wales Building, Molecular Ecology and Fisheries Genetics Laboratory, School of Biological Sciences, College of Natural Sciences, Bangor University, Bangor, United Kingdom, **2** Institute of Genetics, Queens Medical Centre, University of Nottingham, Nottingham, United Kingdom, **3** Department of Biology, Portland State University, Portland, Oregon, United States of America, **4** Departament Biologia Animal, Universitat de Barcelona, Barcelona, Spain, **5** Faculty of Life Sciences, The University of Manchester, Manchester, United Kingdom, **6** Centre for Genomic Research, School of Biological Sciences, University of Liverpool, Liverpool, United Kingdom

Abstract

The development of second generation sequencing technology has resulted in the rapid production of large volumes of sequence data for relatively little cost, thereby substantially increasing the quantity of data available for phylogenetic studies. Despite these technological advances, assembling longer sequences, such as that of entire mitochondrial genomes, has not been straightforward. Existing studies have been limited to using only incomplete or nominally intra-specific datasets resulting in a bottleneck between mitogenome amplification and downstream high-throughput sequencing. Here we assess the effectiveness of a wide range of targeted long-range PCR strategies, encapsulating single and dual fragment primer design approaches to provide full mitogenomic coverage within the Araneae (Spiders). Despite extensive rounds of optimisation, full mitochondrial genome PCR amplifications were stochastic in most taxa, although 454 Roche sequencing confirmed the successful amplification of 10 mitochondrial genomes out of the 33 trialled species. The low success rates of amplification using long-Range PCR highlights the difficulties in consistently obtaining genomic amplifications using currently available DNA polymerases optimised for large genomic amplifications and suggests that there may be opportunities for the use of alternative amplification methods.

Citation: Briscoe AG, Goodacre S, Masta SE, Taylor MI, Arnedo MA, et al. (2013) Can Long-Range PCR Be Used to Amplify Genetically Divergent Mitochondrial Genomes for Comparative Phylogenetics? A Case Study within Spiders (Arthropoda: Araneae). PLoS ONE 8(5): e62404. doi:10.1371/journal.pone.0062404

Editor: Diego Fontaneto, Consiglio Nazionale delle Ricerche (CNR), Italy

Received January 3, 2013; **Accepted** March 21, 2013; **Published** May 8, 2013

Funding: This work was supported by a CoSyst Grant awarded to SC and a Bangor University funded PhD studentship to AGB. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: s.creer@bangor.ac.uk

Introduction

Mitochondrial DNA markers are the cornerstones of contemporary molecular systematics and contribute greatly to our understanding of organellar evolution [1,2]. However, robust phylogenetic reconstruction is often impeded by low sequence volume and number of comparative loci. Prior to the turn of the century, single gene partitions were commonly used to infer phylogenetic histories, but poor nodal support and discordance between phylogenies derived from separate markers clearly revealed that additional data are needed for robust phylogenetic reconstruction [3].

As more sequence data becomes available for elucidating the tree of life, large-scale sequencing efforts and interrogation of expressed sequence tag (EST) libraries or sequenced transcriptomes [4,5] have begun to yield large numbers of nuclear markers that can be used for phylogenetic reconstruction [6]. However, true phylogenomic analyses are still not practical for the *de novo* construction of phylogenetic hypotheses for taxa without sequenced transcriptomes [7,8]. A compromise between traditional phylogenetic methods and phylogenomics lies in the phylogenetic analysis of whole mitochondrial genomes [9–11]. However, the routine amplification of complete mitochondrial genomes from divergent taxa remains a significant hurdle to the widespread adoption of mitogenomic approaches.

The number of informative phylogenetic characters within the mitochondrial genome has been appreciated for some time [1] but most studies have limited themselves to only exploring parts of this information to resolve relationships at multiple levels [12]. The increase in whole mitochondrial genome datasets has provided new characters potentially allowing more robust phylogenies to be constructed. These markers, known as rare genomic changes (RGCs) have become increasingly popular for resolving complex phylogenetic relationships where traditional methods have produced ambiguous, unresolved results [12]. RGCs, defined as large-scale mutational changes occur much less frequently than base substitutions and have long been used in phylogenetics as supporting data embedded in DNA sequences. Examples of RGCs include changes in organelle gene order, gene duplications and genetic code variants [13–16].

Whilst the potential benefits of analysing whole mitochondrial genomes (including both coding regions and structural characters) in furthering our understanding of the mechanisms behind the evolution of organellar DNA are clear, all of the data produced to date have been of comparatively low volume. The lack of empirical data means that we still cannot accurately assess the utility of full mitogenomic sequences as a tool for resolving complex phylogenies in many taxa [17]. Advances in sequencing technology (e.g 454 Roche GSFLX series and Solexa/Illumina) are likely to resolve the sequencing throughput issue [18]. However, there are still clear limitations associated with the large-scale PCR amplification of divergent, interspecific, whole mitochondrial genomes. In spite of the recent revolution in sequencing technologies, current mitogenomic studies have been characterised by either analysing data from large, but incomplete mitochondrial genome fragments [10], low numbers of species [19] or by focusing on nominally intra-specific datasets [9]. In order to investigate the resolution potential, and associated problems with the amplification of large, interspecific mitogenomic data sets we focus here on spiders (Arthropoda: Araneae).

The Araneae are among the oldest and most diverse group of terrestrial organisms [20,21], with a current diversity of more than 43,240 described species, placed in 111 families [22]. Spiders are an unequivocally ecologically important guild, being the dominant predators of insects in natural and managed ecosystems [23]. However, they have been relatively understudied from a higher-level molecular systematic perspective, and very little is known about inter-family relationships [24]. More recent attempts to resolve the phylogeny of the Araneae, have revealed significant topological incongruence between morphological and multiple loci phylogenies [25]. This makes the Araneae an ideal order to test the application of using mitochondrial phylogenomics to resolve complex relationships and better understand the evolutionary mechanisms underlying speciation and diversification. Furthermore, recent studies have shown repeated tRNA gene translocations [26,27] in combination with an extensive fossil record can be utilised in resolving high-level relationships, via calibrated gene trees in the Araneae [28–31]. Here, we investigate the utility of direct long-range PCR amplification of whole mitochondrial spider genomes, using a large range of currently available long-range *Taq* polymerases. The PCR approach was chosen as it circumvents the need for (often unavailable) large starting biomass associated with the direct pelleting of mitochondria. Whilst mitochondrial genomes can be readily amplified in large numbers of small fragments, this is both costly and labour intensive. To this end, we adopted both single and dual fragment amplification approaches across the phylogenetic breadth of the order to assess the feasibility of both conserved and directed approaches for expedient whole mitochondrial DNA amplification.

Materials and Methods

Sample Collection and DNA Extraction

Spiders were obtained from across the United Kingdom and Gambia by the authors and members of the British Arachnological Society (BAS) and either ethanol preserved (70%–100%, stored at 4°C) or freshly frozen (stored at −80°C) directly from living individuals (Table 1). Samples of *Selenops annulatus*, *Deinopis* sp. and *Cithaeron praedonius* had been stored in 70% ethanol and were incorporated in order to maximise taxonomic coverage across the order. No specific permits were required for the described field studies and no specific permissions were required for these locations or activities. No locations were privately owned or protected and the field studies did not involve protected or endangered species. All four legs were removed from the left side of the thorax prior to DNA extraction. The rest of the body was stored in 100% ethanol for vouchering and subsequent identification purposes. Whole genomic DNA was extracted from a single femur of each species using the Qiagen DNeasy Blood & Tissue Kits (Qiagen).

Primer Anchoring Strategy

In order to identify an efficient approach for the amplification of whole mitochondrial genomes we adopted two long-range PCR strategies to amplify the complete mitogenome in one or two large fragments [32]. DNA from two anchoring regions, a ca. 650 b.p. region of the Cytochrome Oxidase I (COI) and a 450 b.p. region spanning the large ribosomal subunit (16 s rRNA), were amplified for 33 taxa. Both amplifications were performed in 25 µl reactions using the primer combinations of CHELF1 (5′- TACTCTAC-TAATCATAAAGACATTGG) and CHELR2 (5′-GGATGGC-CAAAAAATCAAAATAAATG) (COI) [33] and primers LR-N-13398 (5′- CGCCTGTTTAACAAAAACAT) and LR-J-12887 (5′- CCGGTCTGAACTCAGATCACGT) (16 s) [25]. PCR reactions comprised 1× PCR buffer, 1.5 mM $MgCl_2$, 0.4 µM of each primer, 0.625 units ThermoPrime *Taq* DNA polymerase (Thermo Scientific) and 1 µl template DNA and thermocycling was performed using a DNA engine (Tetrad 2) Peltier Thermal Cycler (BIORAD). Cycling conditions were 60 seconds at 94°C, 5 cycles of 60 seconds at 94°C, 90 seconds at 45°C and 90 seconds at 72°C followed by 35 cycles of 60 seconds at 94°C, 90 seconds at 50°C and 90 seconds at 72°C [33]. Amplification success was checked using a 1% agarose gel stained with Ethidium Bromide (EtBr). Successful amplifications were cleaned with 1 U shrimp alkaline phosphatase (Promega) to dephosphorylate residual deoxynucleotides and 0.5 U Exonuclease I (Promega) to degrade excess primers [34] and subsequently sequenced bidirectionally (Macrogen Inc, Seoul, Korea) using the same primers as for amplification.

Long-Range Primer Design and PCR

Following the Sanger sequencing of the COI and 16 s anchoring regions, chromatographs were checked and sequences were manually edited where necessary, using CodonCode Aligner (v. 2.0.6, Codon Code Corporation), prior to alignment using Clustal W [35]. The initial aim of this process was to amplify regions of sufficient length, on opposite sides of the mitochondrial genome, from which 'universal' primers could be designed [26,36,37]. However, no conserved regions were found that would facilitate the design of long-range primers that could be used to amplify homologous loci from multiple families. long-range primers were subsequently designed for individual taxa in order to amplify the entire mitochondrial genome in one or two large fragments that overlapped with the conserved COI and/or 16 s ribosomal subunit using the Primer 3 software [38]. Default values were used with the exception of length (22–30 bp), primer T_m (57.0–70.0°C) and GC content (40–60%) which followed consensus recommendations from the *Taq* manufacturers (Primer sequences available on DRYAD entry doi:10.5061/dryad.8dd3n). For the single fragment protocol, primers were designed within the COI sequences, with the light strand primer situated downstream of the heavy strand primer, thus taking advantage of the circular nature of the genome. For the dual fragment protocols, two sets of primers were designed for each taxon, to bridge the gaps between the COI and 16 s regions (Figure 1).

The PCRs were performed initially in 50 µl volumes using multiple polymerases recommended for long-range PCR amplification, including: Clontech Advantage 2, Clontech Advantage

Table 1. Taxonomic information, storage conditions and GenBank accession numbers for specimens.

Family	Genus	Specific epithet	Sample ID	Locality	COI	16 s	Sample storage
Agelenidae	*Malthonica*	*silvestris*	464_SC_AB	Kent, UK	JQ412460		Frozen
Amaurobiidae	*Amaurobius*	*similis*	483_SC_AB	Lancashire, UK		JQ406635	Frozen
Anyphaenidae	*Anyphaena*	*accentuata*	522_SC_AB	Kent, UK	JQ412439	JQ406633	Frozen
Araneidae	*Agalenatea*	*redii*	507_SC_AB	Kent, UK		JQ406637	Frozen
Araneidae	*Araneus*	*diadematus*	571_SC_AB	Dorset, UK	JQ412440	JQ406621	Frozen
Cithaeronidae	*Cithaeron*	*praedonius*	455_SC_AB	Kotu, Gambia	JQ412441		Ethanol
Clubionidae	*Clubiona*	*terrestris*	518_SC_AB	Kent, UK	JQ412442		Frozen
Corinnidae	*Messapus*	*martini*	453_SC_AB	Kerr Serign, Gambia		JQ406616	Ethanol
Corinnidae	*Phrurolithus*	*festivus*	503_SC_AB	Kent, UK		JQ406632	Ethanol
Ctenidae	*Anahita*	sp.	460_SC_AB	Kotu, Gambia		JQ406614	Ethanol
Cybaeidae	*Argyroneta*	*aquatica*	581_SC_AB	Dorset, UK		JQ406617	Frozen
Deinopidae	*Deinopis*	sp.	451_SC_AB	Gunjur, Gambia	JQ412443		Ethanol
Dictynidae	*Dictyna*	*latens*	499_SC_AB	Kent, UK	JQ412444	JQ406629	Ethanol
Dysderidae	*Dysdera*	*erythrina*	479_SC_AB	Kent, UK	JQ412445	JQ406627	Ethanol
Eresidae	*Stegodyphus*	sp.	454_SC_AB	Bijilo, Gambia	JQ412459		Ethanol
Gnaphosidae	*Zelotes*	*apricorum*	462_SC_AB	Kent, UK	JQ412463		Frozen
Idiopidae	*Gorgyrella*	sp.	448_SC_AB	UK Pet Trade	JQ412447		Frozen
Linyphiidae	Unknown	Unknown	559_SC_AB	Gwynedd, UK	JQ412448	JQ406636	Frozen
Lycosidae	*Pardosa*	*nigriceps*	477_SC_AB	Kent, UK	JQ412452	JQ406631	Ethanol
Philodromidae	*Philodromus*	*dispar*	517_SC_AB	Kent, UK	JQ412453	JQ406634	Frozen
Pisauridae	*Pisaura*	*mirabilis*	502_SC_AB	Kent, UK	JQ412454	JQ406630	Ethanol
Pholcidae	*Pholcus*	*phalangioides*	484_SC_AB	Lancashire, UK		JQ406625	Frozen
Salticidae	*Salticus*	*scenicus*	423_SC_AB	Gwynedd, UK	JQ412456	JQ406628	Ethanol
Segestriidae	*Segestria*	*senoculata*	583_SC_AB	Gwynedd, UK	JQ412457	JQ406615	Ethanol
Selenopidae	*Selenops*	*annulatus*	449_SC_AB	Kotu, Gambia	JQ412458		Ethanol
Sparassidae	*Micrommata*	*virescens*	461_SC_AB	Kent, UK	JQ412451	JQ406618	Frozen
Tetragnathidae	*Meta*	*menardi*	481_SC_AB	Gwynedd, UK	JQ412449	JQ406620	Frozen
Theraphosidae	*Eupalaestrus*	*campestratus*	446_SC_AB	UK Pet Trade	JQ412446	JQ406626	Frozen
Theraphosidae	*Grammostola*	*rosea*	445_SC_AB	UK Pet Trade		JQ406624	Frozen
Theraphosidae	*Psalmopoeus*	*cambridgei*	447_SC_AB	UK Pet Trade	JQ412455	JW406623	Frozen
Theridiosomatidae	*Theridiosoma*	*gemmosum*	489_SC_AB	Glamorgan, UK	JQ412461		Frozen
Thomisidae	*Xysticus*	*audax*	521_SC_AB	Kent, UK	JQ412462	JQ406622	Frozen
Uloboridae	*Miagrammopes*	sp.	452_SC_AB	Bijilo, Gambia	JQ412450	JQ406619	Ethanol

Sample storage indicates the methods in which the specimens were preserved on collection; either freshly frozen at −80°C (frozen) or stored in 70–100% ethanol at 4°C.

doi:10.1371/journal.pone.0062404.t001

Titanium, Clontech Advantage HD (Takara Biotech), NEB LongAmp (NEB), GeneAmp XL (Applied Biosystems) & Expand Long Range PCR (Roche), using the protocols and cycling conditions recommended by the manufacturers. All DNA polymerases were initially tested on five phylogenetically disparate samples prior to using the most successful ones on the remaining samples. Successfully amplified PCR products were electrophoresed and subsequently purified using Qiagen QiaQuick Gel extraction kits (Qiagen) or Origene Rapid PCR Purification System (Origene), dependent on fragment size. Following long-range PCR amplification, primers were assigned to groups based on amplification success (good, stochastic or no amplification) and subsequently tested for significant differences in primer length, GC content and T_m using a Mann-Whitney U test in the statistics package SPSS [39].

Amplicon Shearing, Library Construction and 454 Roche Sequencing

Amplicons from eleven purified samples (*Meta menardi*, *Xysticus audax*, *Psalmopoeus cambridgei*, a Gorgyrella sp., Eupalaestrus campestratus, a Linyphiidae sp., *Zelotes apricorum*, *Malthonica silvestris*, *Araneus diadematus*, *Pisaura mirabilis* & *Dysdera erythrina*) were fragmented using a Covaris DNA shearer (Covaris) at 10% duty cycle, intensity: 4 with 200 cycles per burst for 65 seconds. Following quantification using a Qubit 2.0 fluorometer (Invitrogen), MID barcoded adapted sequencing libraries were constructed using the NEBNext DNA sample prep master mix set in accordance with manufacturer's instructions (New England Biolabs), pooled in equimolar concentrations and sequenced at low putative coverage on $1/16^{th}$ of a 454 Roche GSFLX Titanium platform (Centre for Genomic Research, University of Liverpool).

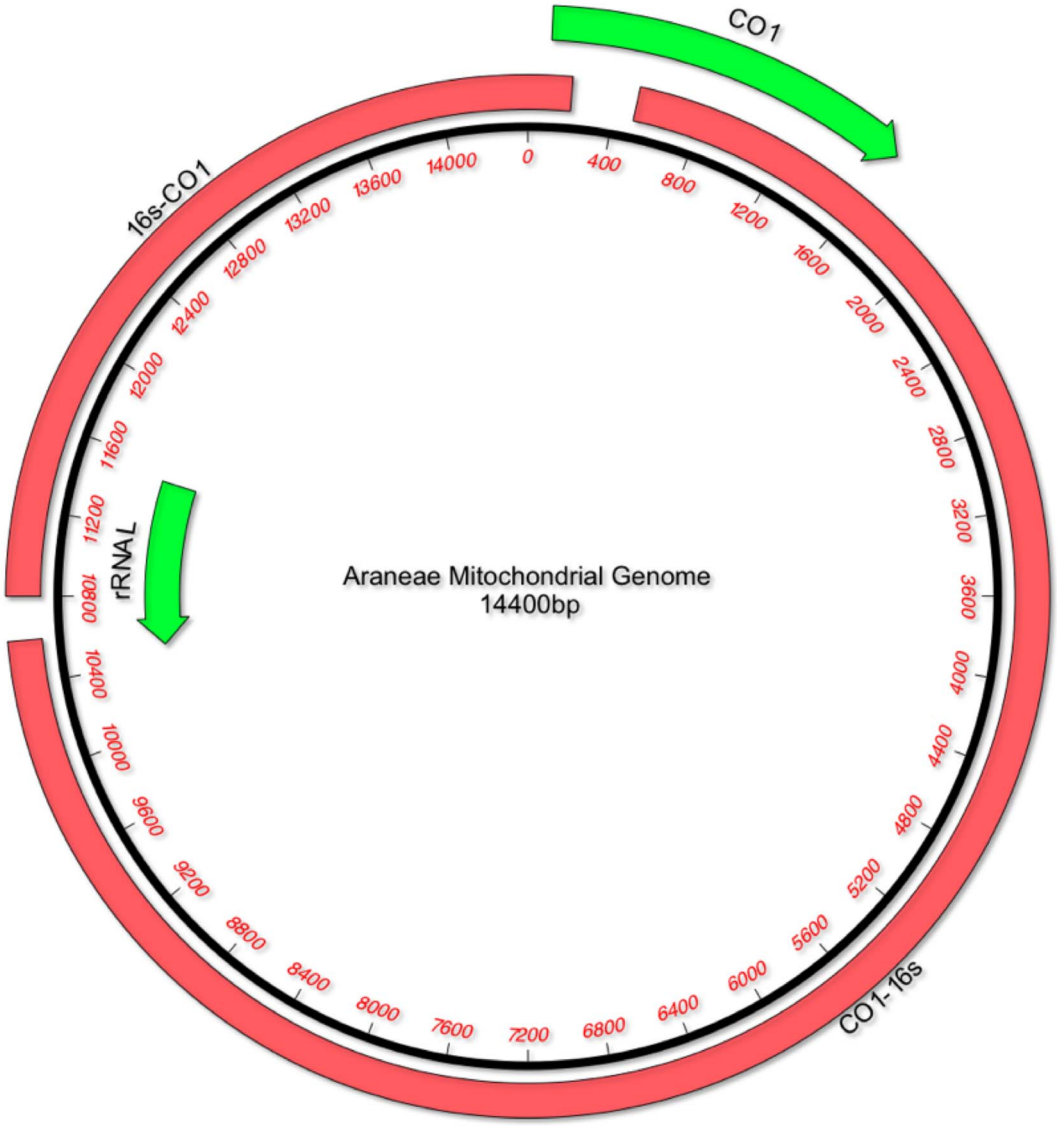


Figure 1. Araneae mitochondrial genome displaying anchoring regions. Spider mitochondrial genome highlighting the genes from which the two fragment long-range PCR strategy was designed (green arrows – direction indicates gene sequence 5′–3′). Solid red bars show the two long-range PCR products and approximate length.
doi:10.1371/journal.pone.0062404.g001

Please note that the sequencing step here was intended as a quality control measure and not intended to sequence to a depth required for full mitochondrial genome assembly (sequence quality data available on DRYAD entry doi:10.5061/dryad.8dd3n). The 454 Roche sequences were assembled using GS *De Novo* Assembler software (Roche). Following trials with Roche's GS *De Novo* software, MIRA (B. Chevreux) and CLC Genomics Workbench (CLC bio, Aarhus, Denmark), no substantial differences in number of contigs or length were present using any of the approaches. Therefore, in accordance with the proven functionality of the 454 Roche software, all sequences were assembled using the GS *De Novo* program. Following assembly, the resulting contigs where compared to existing Araneae mitochondrial DNA sequences within GenBank via BLAST [40] in order to investigate sequence homology (Genbank accession numbers: AY309258, AY309259, AY452691, NC_005942, NC_010777, NC010780).

Results

Anchoring Regions

Of the 33 species used, 27 and 24 taxa were successfully amplified for the COI and the 16 s regions respectively, yielding anchoring points for a total of 26 families (Table 1) (GenBank accession numbers: COI - JQ412439 - JQ412463 and 16 s - JQ406614–JQ406637).

Mitochondrial Genome PCR Amplifications

Initially, five specimens, covering a broad taxon range, were used to test the performance of the *Taq* polymerases for single fragment mitochondrial genome amplification (Table 2). Following repetition of the *Taq* testing when opting for a dual fragment approach, it was found that NEB LongAmp was equally as effective as Clontech Advantage 2 *Taq* polymerase. Subsequently, complete mitochondrial genomes were successfully amplified in c. 7–9 kb or c. 15 kb fragments using Clontech Advantage 2 (single

fragment) or NEB LongAmp (dual fragment) *Taq* polymerases. The probability of obtaining successful amplifications of complete mitochondrial genomes utilising either the 1 or 2 fragment amplification protocol was not demonstrably different (based upon presence/absence of bands), suggesting that the limiting factor in the amplification of the genomes was not necessarily the size of the target fragment. In some two-fragment amplifications (Samples *Pardosa nigriceps*, *Dysdera erythrina*, *Pisaura mirabilis*, *Phrurolithus festivus* & *Anyphaena accentuata*) the COI to 16 s region amplified consistently more often than the reverse, highlighting potential inhibition to amplification within the target sequence. However, of the initial 33 samples trialled, from which a subset of 22 (based on sequence availability and prior PCR amplification success) were used for long-range PCR, only 11 mitochondrial genomes could be amplified robustly and with a guaranteed level of reproducibility (*Meta menardi*, *Xysticus audax*, *Psalmopoeus cambridgei*, a *Gorgyrella* sp., *Eupalaestrus campestratus*, a Linyphiidae sp., *Zelotes apricorum*, *Malthonica silvestris*, *Araneus diadematus*, *Pisaura mirabilis* & *Dysdera erythrina*). The NEB LongAmp polymerase (NEB) amplified genomes in two fragments with greater reproducibility whilst Clontech Advantage 2 (Takara Biotech) amplified single fragment genomes with greater consistency. Statistical analysis of the primer properties (length, GC content and T_m) yielded no significant differences between any combination of primers that amplified repeatedly, intermittently, or those that consistently failed to amplify mitochondrial genome fragments.

454 Roche Sequencing

The sequencing of the 11 amplicon libraries provided a total of 27,892 tagged reads with an average length of approximately 278 bases (Table 3). Reads from sample 464_SC_AB (*Malthonica silvestris*) were unable to be recovered by the MID identifying software. *De novo* read assembly was unable to construct complete mitochondrial genomes but provided an average of an estimated 65% sequence assembly (assuming a ca. 15 kb mitogenome target; contigs and sequences available on DRYAD entry doi:10.5061/dryad.8dd3n).

In light of the likelihood of tRNA gene rearrangements and because only 6 mitochondrial genomes were available on Genbank at the time of analysis, a highly dissimilar BLAST search was performed on the nucleotide collection database. Of the 10 samples, all contigs created through short-read assembly had maximum identification values ranging from 78–100% Max ID for spider mitochondrial DNA, either through an unrestricted blast search or with searches focused on Araneae accessions. Nucleotide sequences for all reads were deposited in the GenBank short-read archive (SRA051390.1).

Discussion

We were able to amplify most of the target genes (COI and 16 s) across our taxon range, but were unable to identify or develop degenerate primers for any other potential anchoring region within the spider mitochondrial genome following sliding window analyses of all regions most commonly used in spider phylogenetics. Sliding window analysis was performed on all 13 protein coding genes and both ribosomal RNAs (rRNAs) of the six previously published spider mitochondrial genomes using the *Drosophila* Polymorphism database, SNP Graphics (http://dpdb.uab.es/dpdb/diversity.asp) (sliding window plots available on DRYAD entry doi:10.5061/dryad.8dd3n). This most likely highlights the high level of mitochondrial genetic diversity that can be found within the spiders along with the absence of universal primers available for no more than a handful of protein coding genes, resulting in few reference sequences available for comparison for large portions of the Araneae mitogenome [41].

Predictably, the 454 Roche low coverage sequencing step resulted in a wide variation of read coverage per amplicon pool, most likely due to differences in MID primer tag design and read recovery. Nevertheless, even highly covered mitochondrial genomes did not result in a full assembly, suggesting further optimisation may be required in either the shearing or bioinformatic steps of mitogenome assembly [42]. The BLAST search of the assembled contigs showed that spider mitochondrial genomes had indeed been amplified and that we had not inadvertently amplified nuclear DNA or DNA from a contaminating source such as the *Wolbachia* bacterium [43]. The need to use searches focused on Araneae accessions, in order to get positive matches, most likely highlights the genetic divergence between the trialled taxa and the complete mitogenomic data currently available for 6 out of the 112 described families of spiders.

Nevertheless, even following an extensive campaign of PCR strategies and optimization, using a variety of long range *Taqs*, we could not consistently amplify approximately two thirds of the species, a substantial proportion of our sample taxa. We have no reason to attribute PCR failure to degraded DNA resulting from sample storage conditions, since all specimens were preserved directly from living organisms using tried and tested preservation media. Moreover, of the three samples stored in 70% EtOH, although sample *Selenops annulatus* did not amplify, samples *Deinopis* sp. and *Cithaeron praedonius* could be amplified, albeit inconsistently, suggesting that 70% EtOH preservation over ca. three years may be sufficient to preserve mitogenome integrity for PCR amplification. However, we could not rule out the possibility that sub-optimal storage conditions can adversely affect the availability of

Table 2. Enzymatic information for successful long-range amplifications.

Sample reference	Clontech		NEB LongAmp	Expand Long Range
	Advantage2	AdvantageLA		
448_SC_AB	Stochastic	No amplification	No amplification	Good
462_SC_AB	Good	No amplification	Weak	No amplification
464_SC_AB	Good	No amplification	No amplification	No amplification
446_SC_AB	No amplification*	No amplification	No amplification*	No amplification
451_SC_AB	Stochastic	Stochastic	No amplification	Stochastic

Clontech Advantage Titanium and Roche GeneAmp XL are not shown but failed to amplify any samples.
*Indicates good amplification from a dual fragment PCR approach. The term Good indicates robust and consistent amplification. The term Stochastic indicates a non-consistently reproducible amplification.
doi:10.1371/journal.pone.0062404.t002

Table 3. Sequencing information following 454 GS FLX run and subsequent GS *De Novo* Assembler contig assembly.

Species Name	Sample Reference	Number of Reads	Number of Contigs	Average Contig Length	Roche MID Identifier	Longest Contig Length
Araneus diadematus	571_SC_AB	1,313	20	340	1	1,058
Psalmopoeus cambridgei	447_SC_AB	1,689	17	944	2	4,483
Eupalaestrus campestratus	446_SC_AB	380	5	849	3	1,916
Gorgyrella sp.	448_SC_SB	6,203	49	665	4	2,849
Xysticus audax	521_SC_AB	4,887	35	423	5	2,257
Pisaura mirabilis	502_SC_AB	1,192	14	505	6	872
Dysdera erythrina	479_SC_AB	1,554	16	733	7	1,523
Linyphiidae sp.	559_SC_AB	503	12	540	8	1,310
Zelotes apricorum	462_SC_AB	9,081	26	619	9	6,434
Meta menardi	481_SC_AB	1,090	18	532	11	1,452

Number of reads per sample represents only those from which the indicator MID sequence was recovered. Number of contigs represents only those formed with a length greater than 100 bases. Average contig length refers only to contigs over 100 bases in length. Longest contig length shows the largest single contig assembled.
doi:10.1371/journal.pone.0062404.t003

suitable templates for long-range PCR. It is also unlikely that the primers affected long-range PCR success rate, since comparative analysis between the primer sets revealed no discernable statistical differences in physical or chemical properties (i.e. GC content, length, T_m). The arachnid mitochondrial genome, as for most animal mitochondrial genomes, is a small extra-chromosomal genome comprising 37 genes including 22 coding for transfer RNAs (tRNAs), 13 coding for proteins and 2 coding for ribosomal RNAs (rRNAs) [1]. As well as the 37 coding genes, mitochondrial genomes also contain a small, approximately 1–2 kb, non-coding control region, named due to its perceived role in controlling the transcription and replication of the mtDNA molecule [44]. However, the protein coding gene arrangement of spiders is highly conserved and shared amongst many other Chelicerates and so is not likely to cause differences in amplification success. Previously published Araneae mitochondrial genomes are considered to be very A/T rich (64–76%), reflecting the low G/C content of the mitochondrial genomes of other arthropod orders [45]. In spite of the low G/C content of the DNA, no definitive reason is forthcoming for the failure to obtain more consistent long PCR amplifications from our target mitochondrial genomes. Whilst the larger region (COI-16 s) of the dual fragment approach amplified successfully more often than the shorter 16 s-COI counterpart, the G/C content of both fragments is comparable (approximate difference of 2.6% averaged across all currently published Araneae mitochondrial genomes) with no significant differences between the highest and lowest values of the comprising genes within each fragment. However the lesser-amplified small fragment (16 s-COI) does contain the control region, also known as the A+T rich region due to its low G/C content, which has been known to inhibit DNA polymerases in insects and platyhelminths [46]. However, enzyme inhibition is more likely to be caused by tandem repeats [47] present in and around the control region as comparative analysis of previously published arachnid mitochondrial genomes show other genes to also have lower G/C content.

Our results provide evidence for the successful amplification of several whole mitochondrial genomes, in one or two long fragments overlapping the COI and/or 16 s regions, using a variety of commercially available DNA polymerases optimised for large genomic amplifications. However, we have also highlighted the difficulty in obtaining amplifications with a guaranteed level of reproducibility, using multiple currently available polymerases and thus highlighted potential limitations to the feasibility of mitogenomic studies featuring large numbers of independently amplified taxa using single, or dual fragment approaches. Recent studies utilising mitogenomics to resolve phylogenetic incongruences have avoided this issue by focusing on the production of datasets comprising either incomplete genome sequences [10] or that are nominally intra-specific [9]. Whilst these data still provide a large number of informative phylogenetic characters, the lack of the whole mitochondrial genome precludes the acquisition of maximum phylogenetic resolution, including RGCs, from the mtDNA genome [48–50]. However, the importance of recovering complete mitogenomic sequences in order to more accurately reconstruct phylogenetic relationships has long been understood and so remains an important avenue of exploration in contemporary phylogenetics [51]. Although we present data on a single order (Araneae – Spiders) we believe that the results highlight the limitations to the feasibility of generating diverse, interspecific complete mitochondrial genome data sets from long-range PCR amplifications, in terms of both cost and efficiency. Whilst many studies have successfully used a multitude of PCR amplifications in order to generate mitochondrial genomes [48,52,53], this is both labour and time intensive. This highlights the potential need to utilise alternative, non-PCR based methods that are able to amplify complete mitochondrial genomes both quickly and cost effectively. Direct recovery of organellar genomes and mitochondrial gene partitions from whole shotgun genome sequencing and EST libraries is possible [54,55], but expedient mechanisms to isolate only the mtDNA locus are needed to optimise organellar coverage for large numbers of taxa. While methods such as Rolling Circle Amplification (RCA) [56] have been trialled successfully on a limited number of species, further investigations across a range of taxa will be desirable to investigate their full potential to create divergent, multi-taxon datasets for comparative mitogenomics. Such methods hold advantages over PCR-based strategies due to the non-specific nature of the amplification process. By using random hexamers, as opposed to synthesising bespoke taxon-specific oligonucleotides, it is possible to avoid the amplification failure shown by this study whilst allowing for the creation of large, diverse mitochondrial genome datasets from low amounts of starting material.

Data Accessibility

DNA sequences: Genbank accessions JQ412439 - JQ412463, JQ406614–JQ406637 and SRA051390.1.

Primer sequences, Contigs & Sliding window analysis: DRYAD entry doi:10.5061/dryad.8dd3n.

Acknowledgments

The authors would like to thank the members of the British Arachnological Society, especially Tony Russell-Smith, Jennifer Newton, John Harper, Rob Cummings & Richard Gallon for the provision of samples and specimen identification. We also thank Wendy Grail (Molecular Ecology & Fisheries Genetics Laboratory), Christiane Hertz-Fowler, Christian Bourne and the staff at the Centre for Genomic Research, University of Liverpool, for sequencing support.

Author Contributions

Critical review of the manuscript: SG SM MIT MAA DP JK. Conceived and designed the experiments: AGB SC. Performed the experiments: AGB. Analyzed the data: AGB SC. Contributed reagents/materials/analysis tools: AGB SC JK. Wrote the paper: AGB SC.

References

1. Boore JL (1999) Animal mitochondrial genomes. Nucleic Acids Research 27: 1767–1780.
2. Moritz C, Dowling TE, Brown WM (1987) Evolution of animal mitochondrial-DNA - relevance for population biology and systematics. Annual Review of Ecology and Systematics 18: 269–292.
3. Delsuc F, Brinkmann H, Philippe H (2005) Phylogenomics and the reconstruction of the tree of life. Nature Reviews Genetics 6: 361–375.
4. Okazaki Y, Furuno M, Kasukawa T, Adachi J, Bono H, et al. (2002) Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. Nature 420: 563–573.
5. Murchison EP, Tovar C, Hsu A, Bender HS, Kheradpour P, et al. (2010) The Tasmanian Devil Transcriptome Reveals Schwann Cell Origins of a Clonally Transmissible Cancer. Science 327: 84–87.
6. Dellaporta SL, Xu A, Sagasser S, Jakob W, Moreno MA, et al. (2006) Mitochondrial genome of *Trichoplax adhaerens* supports Placozoa as the basal lower metazoan phylum. Proceedings of the National Academy of Sciences of the United States of America 103: 8751–8756.
7. Dordel J, Fisse F, Purschke G, Struck TH (2010) Phylogenetic position of Sipuncula derived from multi-gene and phylogenomic data and its implication for the evolution of segmentation. Journal of Zoological Systematics and Evolutionary Research 48: 197–207.
8. Struck TH, Paul C, Hill N, Hartmann S, Hoesel C, et al. (2011) Phylogenomic analyses unravel annelid evolution. Nature 471: 95–U113.
9. Morin PA, Archer FI, Foote AD, Vilstrup J, Allen EE, et al. (2010) Complete mitochondrial genome phylogeographic analysis of killer whales (*Orcinus orca*) indicates multiple species. Genome Research 20: 908–916.
10. Timmermans MJTN, Dodsworth S, Culverwell CL, Bocak L, Ahrens D, Littlewood DT, Pons J, Vogler AP (2010) Why barcode? High-throughput multiplex sequencing of mitochondrial genomes for molecular systematics. Nucleic Acids Research 38: e197.
11. Zhang H, Chen L (2011) The complete mitochondrial genome of dhole *Cuon alpinus*: phylogenetic analysis and dating evolutionary divergence within canidae. Molecular Biology Reports 38: 1651–1660.
12. Rokas A, Holland PWH (2000) Rare genomic changes as a tool for phylogenetics. Trends in Ecology & Evolution 15: 454–459.
13. Boore JL, Brown WM (1998) Big trees from little genomes: mitochondrial gene order as a phylogenetic tool. Current Opinion in Genetics & Development 8: 668–674.
14. Venkatesh B, Ning Y, Brenner S (1999) Late changes in spliceosomal introns define clades in vertebrate evolution. Proceedings of the National Academy of Sciences of the United States of America 96: 10267–10271.
15. Masta SE, McCall A, Longhorn SJ (2010) Rare genomic changes and mitochondrial sequences provide independent support for congruent relationships among the sea spiders (Arthropoda, Pycnogonida). Molecular Phylogenetics and Evolution 57: 59–70.
16. Dermauw W, Vanholme B, Tirry L, Van Leeuwen T (2010) Mitochondrial genome analysis of the predatory mite *Phytoseiulus persimilis* and a revisit of the *Metaseiulus occidentalis* mitochondrial genome. Genome 53: 285–301.
17. Willerslev E, Gilbert MTP, Binladen J, Ho SYW, Campos PF, et al. (2009) Analysis of complete mitochondrial genomes from extinct and extant rhinoceroses reveals lack of phylogenetic resolution. BMC Evolutionary Biology 9: 95 doi:10.1186/1471-2148-9-95.
18. Glenn TC (2011) Field guide to next-generation DNA sequencers. Molecular Ecology Resources 11: 759–769.
19. Prosdocimi F, de Carvalho DC, de Almeida RN, Beheregaray LB (2012) The complete mitochondrial genome of two recently derived species of the fish genus *Nannoperca* (Perciformes, Percichthyidae). Molecular Biology Reports 39: 2767–2772.
20. Penney D, Selden PA (2011) Fossil spiders: the evolutionary history of a mega-diverse order. Manchester: Siri Scientific Press. 128 p.
21. Penney D, Dunlop JA, Marusik YM (2012) Summary statistics for fossil spider species taxonomy. Zookeys 192: 1–13.
22. Platnick NI (2012) The world spider catalog, version 13.0. American Museum of Natural History. Available: http://research.amnh.org/iz/spiders/catalog. DOI: 10.5531/db.iz.0001: American Museum of Natural History. Last accessed 2012 Dec 30.
23. Sunderland KD, Fraser AM, Dixon AFG (1986) Field and laboratory studies on money spiders (Linyphiidae) as predators of cereal aphids. Journal of Applied Ecology 23: 433–447.
24. Benjamin SP, Dimitrov D, Gillespie RG, Hormiga G (2008) Family ties: molecular phylogeny of crab spiders (Araneae : Thomisidae). Cladistics 24: 708–722.
25. Blackledge TA, Scharff N, Coddington JA, Szuts T, Wenzel JW, et al. (2009) Reconstructing web evolution and spider diversification in the molecular era. Proceedings of the National Academy of Sciences of the United States of America 106: 5229–5234.
26. Masta SE, Boore JL (2004) The complete mitochondrial genome sequence of the spider *Habronattus oregonensis* reveals rearranged and extremely truncated tRNAs. Molecular Biology and Evolution 21: 893–902.
27. Masta SE, Boore JL (2008) Parallel evolution of truncated transfer RNA genes in arachnid mitochondrial genomes. Molecular Biology and Evolution 25: 949–959.
28. Ayoub NA, Garb JE, Hedin M, Hayashi CY (2007) Utility of the nuclear protein-coding gene, elongation factor-1 gamma (EF-1 gamma), for spider systematics, emphasizing family level relationships of tarantulas and their kin (Araneae : Mygalomorphae). Molecular Phylogenetics and Evolution 42: 394–409.
29. Dimitrov D, Astrin JJ, Huber BA (2012) Pholcid spider molecular systematics revisited, with new insights into the biogeography and the evolution of the group. Cladistics 29: 132–146.
30. Dimitrov D, Lopardo L, Giribet G, Arnedo MA, Alvarez-Padilla F, et al. (2012) Tangled in a sparse spider web: single origin of orb weavers and their spinning work unravelled by denser taxonomic sampling. Proceedings of the Royal Society B-Biological Sciences 279: 1341–1350.
31. Bodner MR, Maddison WP (2012) The biogeography and age of salticid spider radiations (Araneae: Salticidae). Molecular Phylogenetics and Evolution 65: 213–240.
32. Barnes WM (1994) PCR amplification of up to 35-KB DNA with high-fidelity andhigh-yield from lambda-bacteriophage templates. Proceedings of the National Academy of Sciences of the United States of America 91: 2216–2220.
33. Barrett RDH, Hebert PDN (2005) Identifying spiders through DNA barcodes. Canadian Journal of Zoology-Revue Canadienne De Zoologie 83: 481–491.
34. Werle E, Schneider C, Renner M, Volker M, Fiehn W (1994) Convenient single-step, one tube purification of PCR products for direct sequencing. Nucleic Acids Research 22: 4354–4355.
35. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, et al. (2007) Clustal W and clustal X version 2.0. Bioinformatics 23: 2947–2948.
36. Hwang UW, Park CJ, Yong TS, Kim W (2001) One-step PCR amplification of complete arthropod mitochondrial genomes. Molecular Phylogenetics and Evolution 19: 345–352.
37. Gai Y, Song D, Sun H, Yang Q, Zhou K (2008) The complete mitochondrial genome of *Symphylella sp* (Myriapoda: Symphyla): Extensive gene order rearrangement and evidence in favor of Progoneata. Molecular Phylogenetics and Evolution 49: 574–585.
38. Rozen S, Skaletsky H (2000) Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S (eds) Bioinformatics Methods and Protocols: Methods in Molecular Biology. Totowa: Humana Press. 365–386.
39. Corp. I (Released 2011) IBM SPSS Statistics for Windows, Version 20.0. Armonk, NY: IBM Corp.
40. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic Local Alignment Search Tool. Journal of Molecular Biology 215: 403–410.
41. Hosseini R, Keller MA, Schmidt O, Framenau VW (2007) Molecular identification of wolf spiders (Araneae : Lycosidae) by multiplex polymerase chain reaction. Biological Control 40: 128–135.
42. Alkan C, Sajjadian S, Eichler EE (2011) Limitations of next-generation genome sequence assembly. Nature Methods 8: 61–65.
43. Goodacre SL, Martin OY, Thomas CFG, Hewitt GM (2006) *Wolbachia* and other endosymbiont infections in spiders. Molecular Ecology 15: 517–527.
44. Mardulyn P, Termonia A, Milinkovitch MC (2003) Structure and evolution of the mitochondrial control region of leaf beetles (Coleoptera : Chrysomelidae): A hierarchical analysis of nucleotide sequence variation. Journal of Molecular Evolution 56: 38–45.

45. Beard CB, Mills Hamm D, Collins FH (1993) The mitochondrial genome of the mosquito *Anopheles gambiae*: DNA sequence, genome organization, and comparisons with mitochondrial sequences of other insects. Insect Molecular Biology 2: 103–124.
46. Hu M, Jex AR, Campbell BE, Gasser RB (2007) Long PCR amplification of the entire mitochondrial genome from individual helminths for direct sequencing. Nature Protocols 2: 2339–2344.
47. Zhang DX, Hewitt GM (1997) Insect mitochondrial control region: A review of its structure, evolution and usefulness in evolutionary studies. Biochemical Systematics and Ecology 25: 99–120.
48. Ernsting BR, Edwards DD, Aldred KJ, Fites JS, Neff CR (2009) Mitochondrial genome sequence of *Unionicola foili* (Acari: Unionicolidae): a unique gene order with implications for phylogenetic inference. Experimental and Applied Acarology 49: 305–316.
49. Quinn TW, Wilson AC (1993) Sequence evolution in and around the mitochondrial control region in birds. Journal of Molecular Evolution 37: 417–425.
50. Wenink PW, Baker AJ, Tilanus MGJ (1994) Mitochondrial control-region sequences in 2 shorebird species, the Turnstone and the Dunlin, and their utility in population genetic-studies. Molecular Biology and Evolution 11: 22–31.
51. Cummings MP, Otto SP, Wakeley J (1995) Sampling properties of DNA-sequence data in phylogenetic analysis. Molecular Biology and Evolution 12: 814–822.
52. Sorenson MD, Ast JC, Dimcheff DE, Yuri T, Mindell DP (1999) Primers for a PCR-based approach to mitochondrial genome sequencing in birds and other vertebrates. Molecular Phylogenetics and Evolution 12: 105–114.
53. Jiang S-T, Hong G-Y, Yu M, Li N, Yang Y, et al. (2009) Characterization of the complete mitochondrial genome of the giant silkworm moth, *Eriogyna pyretorum* (Lepidoptera: Saturniidae). International Journal of Biological Sciences 5: 351–365.
54. Kane N, Sveinsson S, Dempewolf H, Yang JY, Zhang D, et al. (2012) Ultra-barcoding in cacao (*Theobroma spp.*; Malvaceae) using whole chloroplast genomes and nuclear ribosomal DNA. American Journal of Botany 99: 320–329.
55. Groenenberg DSJ, Pirovano W, Gittenberger E, Schilthuizen M (2012) The complete mitogenome of *Cylindrus obtusus* (Helicidae, Ariantinae) using Illumina next generation sequencing. BMC Genomics 13: 14 doi:10.1186/1471-2164-13-114.
56. Tang S, Hyman BC (2005) Rolling circle amplification of complete nematode mitochondrial genomes. Journal of Nematology 37: 236–241.