
Access from the University of Nottingham repository:
http://eprints.nottingham.ac.uk/2802/1/GoodacrePCR.pdf

Copyright and reuse:

The Nottingham ePrints service makes this work by researchers of the University of Nottingham available open access under the following conditions.

This article is made available under the Creative Commons Attribution licence and may be reused according to the conditions of the licence. For more details see:
http://creativecommons.org/licenses/by/2.5/

A note on versions:

The version presented here may differ from the published version or from the version of record. If you wish to cite this item you are advised to consult the publisher's version. Please see the repository url above for details on accessing the published version and note that access may require a subscription.

For more information, please contact eprints@nottingham.ac.uk
Can Long-Range PCR Be Used to Amplify Genetically Divergent Mitochondrial Genomes for Comparative Phylogenetics? A Case Study within Spiders (Arthropoda: Araneae)

Andrew G. Briscoe1, Sara Goodacre2, Susan E. Masta3, Martin I. Taylor1, Miquel A. Arnedo4, David Penney5, John Kenny6, Simon Creer1

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.


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

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

Copyright: © 2013 Briscoe et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

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 systematists 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 RNA 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*, *Dienops* sp. and *Citharurus 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’- TACTCTCAC- TATCATAAAGACAGTTGG) and CHELR2 (5’- GGATGGCGC- GAAAAATCTAAAAATAATG) (COI) [33] and primers LR-N-13398 (5’- CGCCTGTATTAACAAAAAGCAT) and LR-J-12887 (5’- CCGGTCTGAACCTGACATCGAT) (16 s) [25]. PCR reactions comprised 1× PCR buffer, 1.5 mM MgCl₂, 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) Pelleti 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 

\[ Tm \] (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). In 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.

<table>
<thead>
<tr>
<th>Family</th>
<th>Genus</th>
<th>Specific epithet</th>
<th>Sample ID</th>
<th>Locality</th>
<th>COI</th>
<th>16 s</th>
<th>Sample storage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agelenidae</td>
<td>Malthonica</td>
<td>silvestris</td>
<td>464_SC_AB</td>
<td>Kent, UK</td>
<td>JQ412460</td>
<td></td>
<td>Frozen</td>
</tr>
<tr>
<td>Amaurobiidae</td>
<td>Amaurobius</td>
<td>similis</td>
<td>483_SC_AB</td>
<td>Lancashire, UK</td>
<td>JQ406635</td>
<td></td>
<td>Frozen</td>
</tr>
<tr>
<td>Anyphaenidae</td>
<td>Anyphaena</td>
<td>accentuata</td>
<td>522_SC_AB</td>
<td>Kent, UK</td>
<td>JQ412439</td>
<td>JQ406633</td>
<td>Frozen</td>
</tr>
<tr>
<td>Araneidae</td>
<td>Agalenatea</td>
<td>redivi</td>
<td>507_SC_AB</td>
<td>Kent, UK</td>
<td>JQ406637</td>
<td></td>
<td>Frozen</td>
</tr>
<tr>
<td>Araneidae</td>
<td>Araneus</td>
<td>diadematus</td>
<td>571_SC_AB</td>
<td>Dorset, UK</td>
<td>JQ412440</td>
<td>JQ406621</td>
<td>Frozen</td>
</tr>
<tr>
<td>Cithaeronidae</td>
<td>Cithaeron</td>
<td>praedonius</td>
<td>455_SC_AB</td>
<td>Kotu, Gambia</td>
<td>JQ412441</td>
<td></td>
<td>Ethanol</td>
</tr>
<tr>
<td>Clubionidae</td>
<td>Clubiona</td>
<td>terrestris</td>
<td>518_SC_AB</td>
<td>Kent, UK</td>
<td>JQ412442</td>
<td></td>
<td>Frozen</td>
</tr>
<tr>
<td>Corinnidae</td>
<td>Messopus</td>
<td>martini</td>
<td>453_SC_AB</td>
<td>Kerr Serign, Gambia</td>
<td>JQ406616</td>
<td></td>
<td>Ethanol</td>
</tr>
<tr>
<td>Corinnidae</td>
<td>Phrarolithus</td>
<td>festivus</td>
<td>503_SC_AB</td>
<td>Kent, UK</td>
<td>JQ406632</td>
<td></td>
<td>Ethanol</td>
</tr>
<tr>
<td>Ctenidae</td>
<td>Anahita</td>
<td>sp.</td>
<td>460_SC_AB</td>
<td>Kotu, Gambia</td>
<td>JQ406614</td>
<td></td>
<td>Ethanol</td>
</tr>
<tr>
<td>Cybaeidae</td>
<td>Argyroneta</td>
<td>aquatica</td>
<td>581_SC_AB</td>
<td>Dorset, UK</td>
<td>JQ412443</td>
<td></td>
<td>Frozen</td>
</tr>
<tr>
<td>Deinopidae</td>
<td>Deinopis</td>
<td>sp.</td>
<td>451_SC_AB</td>
<td>Gunjur, Gambia</td>
<td>JQ41244</td>
<td>JQ406617</td>
<td>Frozen</td>
</tr>
<tr>
<td>Dictynidae</td>
<td>Dictyna</td>
<td>latens</td>
<td>499_SC_AB</td>
<td>Kent, UK</td>
<td>JQ412444</td>
<td>JQ406629</td>
<td>Ethanol</td>
</tr>
<tr>
<td>Dysderidae</td>
<td>Dysdera</td>
<td>erythrina</td>
<td>479_SC_AB</td>
<td>Kent, UK</td>
<td>JQ412445</td>
<td>JQ406627</td>
<td>Ethanol</td>
</tr>
<tr>
<td>Eresidae</td>
<td>Stegodyphus</td>
<td>sp.</td>
<td>454_SC_AB</td>
<td>Bijilo, Gambia</td>
<td>JQ412459</td>
<td></td>
<td>Ethanol</td>
</tr>
<tr>
<td>Gnaphosidae</td>
<td>Zeolites</td>
<td>apricorum</td>
<td>462_SC_AB</td>
<td>Kent, UK</td>
<td>JQ412463</td>
<td></td>
<td>Frozen</td>
</tr>
<tr>
<td>Idiopidae</td>
<td>Gorgyrella</td>
<td>sp.</td>
<td>448_SC_AB</td>
<td>UK Pet Trade</td>
<td>JQ412447</td>
<td></td>
<td>Frozen</td>
</tr>
<tr>
<td>Linyphiidae</td>
<td>Unknown</td>
<td>Unknown</td>
<td>559_SC_AB</td>
<td>Gwynedd, UK</td>
<td>JQ142448</td>
<td>JQ406636</td>
<td>Frozen</td>
</tr>
<tr>
<td>Lycosidae</td>
<td>Pardosa</td>
<td>nigriceps</td>
<td>477_SC_AB</td>
<td>Kent, UK</td>
<td>JQ142452</td>
<td>JQ406631</td>
<td>Ethanol</td>
</tr>
<tr>
<td>Philodromidae</td>
<td>Philodromus</td>
<td>dispar</td>
<td>517_SC_AB</td>
<td>Kent, UK</td>
<td>JQ142453</td>
<td>JQ406634</td>
<td>Frozen</td>
</tr>
<tr>
<td>Pisauridae</td>
<td>Pisaursidea</td>
<td>mirabilis</td>
<td>502_SC_AB</td>
<td>Kent, UK</td>
<td>JQ142454</td>
<td>JQ406630</td>
<td>Ethanol</td>
</tr>
<tr>
<td>Pholidae</td>
<td>Pholcus</td>
<td>phialangioides</td>
<td>484_SC_AB</td>
<td>Lancashire, UK</td>
<td>JQ406625</td>
<td></td>
<td>Frozen</td>
</tr>
<tr>
<td>Salticidae</td>
<td>Salticus</td>
<td>scenicus</td>
<td>423_SC_AB</td>
<td>Gwynedd, UK</td>
<td>JQ142456</td>
<td>JQ406628</td>
<td>Ethanol</td>
</tr>
<tr>
<td>Segestridae</td>
<td>Segestria</td>
<td>senoculata</td>
<td>583_SC_AB</td>
<td>Gwynedd, UK</td>
<td>JQ142457</td>
<td>JQ406615</td>
<td>Ethanol</td>
</tr>
<tr>
<td>Selenopidae</td>
<td>Selenops</td>
<td>annulatus</td>
<td>449_SC_AB</td>
<td>Kotu, Gambia</td>
<td>JQ412458</td>
<td></td>
<td>Ethanol</td>
</tr>
<tr>
<td>Sparassidae</td>
<td>Micrommata</td>
<td>virescens</td>
<td>461_SC_AB</td>
<td>Kent, UK</td>
<td>JQ142451</td>
<td>JQ406618</td>
<td>Frozen</td>
</tr>
<tr>
<td>Tetragenathidae</td>
<td>Meta</td>
<td>menardi</td>
<td>481_SC_AB</td>
<td>Gwynedd, UK</td>
<td>JQ142449</td>
<td>JQ406620</td>
<td>Frozen</td>
</tr>
<tr>
<td>Theraphosidae</td>
<td>Eupalaestrus</td>
<td>campestratus</td>
<td>446_SC_AB</td>
<td>UK Pet Trade</td>
<td>JQ142446</td>
<td>JQ406626</td>
<td>Frozen</td>
</tr>
<tr>
<td>Theraphosidae</td>
<td>Grammostola</td>
<td>rosea</td>
<td>445_SC_AB</td>
<td>UK Pet Trade</td>
<td>JQ142446</td>
<td>JQ406624</td>
<td>Frozen</td>
</tr>
<tr>
<td>Theraphosidae</td>
<td>Psalmopoeus</td>
<td>cambridgei</td>
<td>447_SC_AB</td>
<td>UK Pet Trade</td>
<td>JQ142455</td>
<td>JW406623</td>
<td>Frozen</td>
</tr>
<tr>
<td>Theridiosomatida</td>
<td>Theridiosoma</td>
<td>gemmosum</td>
<td>489_SC_AB</td>
<td>Glamarog, UK</td>
<td>JQ142461</td>
<td></td>
<td>Frozen</td>
</tr>
<tr>
<td>Thomisidae</td>
<td>Xysticus</td>
<td>audax</td>
<td>521_SC_AB</td>
<td>Kent, UK</td>
<td>JQ142462</td>
<td>JQ406622</td>
<td>Frozen</td>
</tr>
<tr>
<td>Uloboridae</td>
<td>Miagrammopone</td>
<td>sp.</td>
<td>452_SC_AB</td>
<td>Bijilo, Gambia</td>
<td>JQ412450</td>
<td>JQ406619</td>
<td>Ethanol</td>
</tr>
</tbody>
</table>

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

Araneae Mitogenome Amplification Challenges

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, Pisaursidea 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 1/16th of a 454 Roche GSFLX Titanium platform (Centre for Genomic 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 (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: COI - JQ412439 - JQ412463 and 16 s - JQ406614 - JQ406637).

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

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
repeatedly, intermittently, or those that consistently failed to
differences between any combination of primers that amplified
454 Roche Sequencing
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
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 reproduc-
ibility (Meta menardi, Nysticus audax, Psalmopoeus cambridgei, a Gorgyrella
sp., Eupalaestra campestris, a Linyphiidae sp., Zyloites 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 phyloge-
etics. 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
ubab.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 compar-
ison 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 bioinfor-
matic 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 Tags,
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 amplifi-
cation. 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.

<table>
<thead>
<tr>
<th>Sample reference</th>
<th>Clontech</th>
<th>NEB LongAmp</th>
<th>Expand Long Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>448_SC_AB</td>
<td>Stochastic</td>
<td>No amplification</td>
<td>Good</td>
</tr>
<tr>
<td>462_SC_AB</td>
<td>Good</td>
<td>No amplification</td>
<td>Weak</td>
</tr>
<tr>
<td>464_SC_AB</td>
<td>Good</td>
<td>No amplification</td>
<td>No amplification</td>
</tr>
<tr>
<td>446_SC_AB</td>
<td>No amplification*</td>
<td>No amplification*</td>
<td>No amplification*</td>
</tr>
<tr>
<td>451_SC_AB</td>
<td>Stochastic</td>
<td>No amplification</td>
<td>Stochastic</td>
</tr>
</tbody>
</table>

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
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, Tm). 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 [46]. 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 organelar 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 organelar 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.

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

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
Data Accessibility
DNA sequences: Genbank accessions JQ412439 - JQ412463, JQ406614 - JQ406637 and SRA051190.1.

Primer sequences, Contigs & Sliding window analysis: DRYAD entry doi:10.5061/dryad.8d3n.

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 Graill (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 JF. Conceived and designed the experiments: AGB SC. Performed the experiments: AGB. Analyzed the data: AGB SC. Contributed reagents/materials/analysis tools: AGB SC JF. Wrote the paper: AGB SC.

References


