Mitochondrial genomes of arachnids - a new approach to address open phylogenetic questions

Dissertation
zur Erlangung des Grades
des Doktors der Naturwissenschaften
am Fachbereich Biologie, Chemie, Pharmazie
der Freien Universität Berlin

vorgelegt von
Kathrin Fahrein
aus Berlin

Disputationsdatum: 30.04.2010

Gutachter

Erster Gutachter: Prof. Dr. Klaus Hausmann
Institut für Biologie
Fachbereich Biologie, Chemie, Pharmazie
Freie Universität Berlin

Zweiter Gutachter: Prof. Dr. Thomas Bartolomaeus
Institut für Evolutionsbiologie und Ökologie
Mathematisch-Naturwissenschaftliche Fakultät
Rheinische Friedrich-Wilhelms-Universität Bonn

Erklärung

Hiermit erkläre ich, dass ich die vorliegende Arbeit ohne die unzulässige Hilfe Dritter und ohne die Verwendung anderer als der angegebenen Hilfsmittel angefertigt habe. Die aus anderen Quellen direkt oder indirekt übernommenen Daten und Konzepte sind unter Angabe der Quelle gekennzeichnet.

Die Arbeit wurde bisher weder im In- noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt.

Berlin, April 2010

Kathrin Fahrein

Danksagung

Prof. Dr. Klaus Hausmann danke ich für die freundliche und unkomplizierte Übernahme der Betreuung meiner Doktorarbeit und für die dadurch überaus wertvolle Möglichkeit, diese in Berlin zu ende schreiben zu können.

Prof. Dr. Thomas Bartolomaeus sei zum einen für sämtliche Unterstützung vor Beginn meiner Doktorarbeit zum anderen aber auch für die anfängliche Betreuung und das weitere Interesse am Fortschritt dieser gedankt.

PD Dr. Lars Podsiadlowski möchte ich zutiefst für die fachlich und menschlich hervorragende Betreuung über die letzten Jahre danken. Nicht nur das Projekt sondern auch die Zusammenarbeit mit ihm war stets eine Freude. Weiterhin möchte ich mich für sein Vertrauen bedanken, eine Morphologin und Arachnophobikerin mit diesem Projekt zu betrauen!

PD Dr. Thomas Stach möchte ich dafür danken, dass Sein Bleiben den Wegzug der Arbeitsgruppe so angenehm wie nur möglich gestaltet hat. Und ohne die dadurch gegebene Kontinuität wäre diese Arbeit jetzt wahrscheinlich noch nicht fertig.

Bei meiner jetzigen Arbeitsgruppe, vor allem bei Esther und Sabrina, bedanke ich mich für Ihre heiteren Gemüter und das entspannte Miteinander im täglichen Büroverkehr.

Allen Mitgliedern der ehemaligen Ag15 Systematik und Evolution der Tiere danke ich für die schöne Zeit, die wir hier in Berlin hatten.

Waltraud Brackwehr sei dafür gedankt, dass sie das mit Ihrer Übersicht, Kompetenz und Hilfsbereitschaft schafft, was viele überfordern würde: Sie hält `den Laden` am Laufen.

Meiner Familie und meinen Freunden sei für das Interesse und die Teilnahme an allem, was ich tue, gedankt. Besonders meinen Eltern und meinem Bruder bin ich für Ihre stets bedingungslose und unaufgeforderte Unterstützung in allen Lebenslagen zu tiefstem Dank verpflichtet. Sie sind unersetzlich für mein Leben und mehr an dem Gelingen dieser Arbeit beteiligt, als sie wahrscheinlich denken!

Meinem Freund danke ich nicht nur für jede Minute (auch für die Schlechten und Anstrengenden, denn die waren besonders lehrreich!), die wir zusammen verbracht haben, sondern auch für die Ruhe, den Humor und die Zuversicht, die er immer genau dann in mein Leben bringt, wenn ich es selbst mal nicht schaffe. Gerade die letzte Zeit war trotz der `Umstände` dann doch eher wie ein Spaziergang!

Zusammenfassung

Die Phylogenie der Arachnida ist bis heute umstritten, einerseits wegen widersprüchlicher Interpretationen von Verwandtschaftsverhältnissen der übergeordneten Taxa, andererseits durch die fragliche Eingliederung der marinen Xiphosura in diese vorwiegend terrestrische Tiergruppe. Die konträren Hypothesen beruhen hauptsächlich auf morphologischen Merkmalen und 18S-Sequenzen, deren Verwendung in diesem Zusammenhang wenig geeignet erscheint, da mehrere morphologische Merkmale als homoplastisch angenommen werden und sich das 18S-Gen während der Evolution der Chelicerata offenbar nur wenig verändert hat. Mitochondriale Genome stellen einen weiteren Datensatz phylogenetisch wertvoller Merkmale dar. Dabei können neben den Aminosäure- bzw Nucleotid-Sequenzen auch strukturelle Merkmale dieser eher kleinen Genome analysiert werden. Diese „Genom-morphologischen“ Merkmale umfassen vergleichende Untersuchungen der Genreihenfolge, der RNA-Sekundärstrukturen und der Kontroll-Region, aber auch z. B. die Nukleotidzusammensetzung von protein-kodierenden und ribosomalen Genen.

In dieser Arbeit werden neue mitochondriale Genomsequenzen mehrerer Arachniden präsentiert und vor dem Hintergrund der Verwendbarkeit dieser molekularen Methode für die Phylogenie der Arachniden diskutiert. In diese Überlegungen werden sowohl Sequenzanalysen als auch Vergleiche von „Genom morphologischen“ Merkmalsbefunden einbezogen. Insgesamt konnten zehn Genome der Arachniden-Taxa Ricinulei, Opiliones, Araneae, Scorpiones und Amblypygi komplett sequenziert werden. Rückschlüsse über die frühe Diversifikation übergeordneter Arachniden-Taxa konnten unter Verwendung von mitochondrialen Sequenzen bzw. Genreihenfolgen nicht gezogen werden. Allerdings konnte ein klarer Nutzen für Aussagen auf untergeordneten taxonomischen Ebenen gezeigt werden. Die gewonnenen Daten unterstützen u.a. die aus morphologischer Sicht etablierte Phylogenie übergeordneter Spinnen-Taxa. Vor allem waren aber Belege vielversprechend, welche eine Abgrenzung einzelner Gruppen innerhalb der opisthothelen Spinnen erlauben, da hier die Schwestergruppen-Beziehungen der Subtaxa längst nicht geklärt ist. Eine ähnliche Bedeutung mitochondrialer Daten zeigt sich auch bei Skorpionen. Auch wenn mitochondriale Genome derzeit keine Aufklärung der Verwandschaftsbeziehungen hochrangiger Arachniden-Taxa vermögen, so können sie dennoch auf etwas niedrigerem taxonomischem Niveau wertvolle Hilfe zur Aufklärung der Phylogenie darstellen.

Summary

The phylogeny of the predominantly terrestrial Arachnida is still controversial, due to conflicting interpretations of the relatedness of the major lineages as well as the questionable inclusion of the marine Xiphosura. The conflicting hypotheses were produced by using basically morphological and nuclear sequence-based data, in that case both with doubtful convenience as many characters must have emerged convergently and the prevalent used 18S rRNA gene appears to have changed too sparsely during chelicerate evolution. Another large dataset of phylogenetic information is provided by mitochondrial genomes, which have served as models for comparative genomics for some time. Besides the sequence information retained in mitochondrial genomes in form of amino acid or nucleotide sequences of protein-coding and rRNA genes, other features of those rather small genomes bear phylogenetic information. These structural characters are often considered as `genome morphology` and comprise gene order, secondary structures of transfer and ribosomal RNAs, control region features, codon usage patterns, and the nucleotide compositional strand bias of protein-coding and rRNA genes.

In this thesis several new mitochondrial genomes from arachnids are presented against the background of an evaluation of the usability and efficiency of these molecular character complexes concerning arachnid systematics. Included in these considerations are sequence analyses as well as comparisons of findings in terms of different characters belonging to `genome morphology`. Altogether, sequencing was successful in case of ten complete genomes from the arachnid orders Ricinulei, Opiliones, Araneae, Scorpiones, and Amblypygi. It appears difficult to deduce the progression of early diversification among arachnid orders from mitochondrial genome sequence data or gene rearrangements. However, a real benefit is gained on lower taxonomic levels. In case of Araneae, e.g., mitochondrial sequences and genome organisations largely support morphological based phylogenetic relationships of the three major clades. Furthermore, it even could be proven that gene rearrangements are useful for in-depth phylogenetic analysis of the Opisthothelae, in which sister group relationships of some taxa still deserve study. Similarily, inverted repeat sequences in the control region of Scorpiones demarcate clades on different levels. So even if mitochondrial genomes can hardly shed light on the relationships of major arachnid lineages, the data is promising for a resolution of disputed intraordinal relationships.

List of Publications

The dissertation is based on following publications:

Chapter 2: Fahrein, K., Talarico, G., Braband, A., Podsiadlowski, L. (2007): The complete mitochondrial genome of *Pseudocellus pearsei* (Chelicerata: Ricinulei) and a comparison of mitochondrial gene rearrangements in Arachnida. BMC Genomics **8**: 386. doi:10.1186/1471-2164-8-386.

Chapter 3: Fahrein, K., Masta, S.E., Podsiadlowski, L. (2009): The first complete mitochondrial genome sequences of Amblypygi (Chelicerata: Arachnida) reveal conservation of the ancestral arthropod gene order. Genome **52**: 456–466. doi:10.1139/G09-023. Copyright: NRC Research Press.

Chapter 4: Fahrein, K., Talarico, G., Podsiadlowski, L. (in preparation): Differences in mitochondrial genomes of Ricinulei: new mitogenomic data from the two major clades of New World Ricinulei.

Chapter 5: Fahrein, K., Podsiadlowski, L. (in preparation): New mitogenomic data from a member of Opiliones (*Opilio parietinus*) reveal a rearrangement hot-spot in Opiliones.

Chapter 6: Fahrein, K., Podsiadlowski, L. (in preparation): A new mitochondrial genome sequence of a member of Scorpiones (*Hadogenes bicolor*) highlights peculiar features in the reversed nucleotide composition and the non-coding region of this taxon.

Chapter 7: Fahrein, K., Arabi, J., Braband, A., Podsiadlowski, L. (in preparation): Rearrangements and the mitochondrial control region in spiders - mitogenomic data from three further Araneae species and the phylogenetic implication.

Contributions of the listed authors

Chapter 2: GT and LP performed initial PCRs, KF and LP did all further PCRs, sequencing, sequence analysis and annotation. KF, AB and LP performed phylogenetic analyses. KF and LP wrote the manuscript.

Chapter 3: KF and SEM did the laboratory work and manuscript writing. LP was supervisor and helped in writing the manuscript.

Chapter 4: KF did the laboratory work and manuscript writing. LP was supervisor and helped in writing the manuscript.

Chapter 5: KF did the laboratory work and manuscript writing. LP was supervisor and helped in writing the manuscript.

Chapter 6: KF did the laboratory work and manuscript writing. LP was supervisor and helped in writing the manuscript.

Chapter 7: KF and JA performed PCRs, sequencing, sequence analysis and annotation. KF did the manuscript writing. JA improved the manuscript. LP was supervisor and helped in writing the manuscript.

Table of Contents

1. General introduction

1.1 Mitochondrial genomics

The mitochondrion and its genome

The mitochondrion is a double membrane-enclosed cell organelle typically found in most eukaryote cell types. Seperated from the cytoplasm, mitochondria generate cellular energy in form of adenosine triphosphat (ATP). The bacterial origin of mitochondria was revealed about a century ago. After scrutinising chloroplasts and other cell organelles, an endosymbiotic origin of plastids due to a symbiotic union of two organisms was assumed by different scientists (reviewed in Altmann 1890; Schimper 1883). Years later, this idea was extended to mitochondria (Wallin 1923) and the endosymbiotic theory itself was then corroborated and popularized by Sagan (1967) and Margulis (1970). It postulates that mitochondria were derived from free living prokaryotes, almost certainly from an alpha-proteobacterium (Gray 1999), engulfed by the ancestor of eukaryotic cells. The time, since eukaryotes containing mitochondria are assumed to exist, is coupled with global changes in the Precambrian oxygen evolution about 2000 Ma ago: models indicate that following the evolution of oxygenic cyanobacterial photosynthesis, the partial oxygen pressure (pO_2) in the atmosphere increased in the Paleoproterozoic to concentrations allowing aerobic respiration (Knoll and Holland 1995; Knoll 1992). After some time the symbiotic relationship became obligate in both directions, neither the host cell is able to persist without the endosymbiont, nor can mitochondria survive outside the cell (McFadden 2001). At the internal membrane of mitochondria oxidative phosphorylation, also known as cell respiration, takes place. Protein complexes transfer electrons from NADH to exogenous oxygen (acting as electron acceptor) and establish a proton gradient across the inner membrane whose potential energy is used to drive the phosphorylation of ADP to ATP by the ATP synthase enzyme. Besides energy metabolism mitochondria play a pivotal role in important biochemical processes like apoptosis (Kroemer et al. 1998), disease (Graeber and Muller 1998), aging (Wei 1998), and other metabolic pathways.

In case of sexual reproduction mitochondria are inherited maternally only, reproduced by binary division similar to bacterial cell fission. Although there are few genes retained in mitochondrial genomes, some similarities of mitochondrial genomes to genomes of alpha-proteobacteria remind to the bacterial heritage (Andersson et al. 1998). There is an assumption of either a functional gene transfer to the nucleus or a functional take over by pre-existing nuclear genes as a replacement of primarily existing mitochondrial genes (Adams and Palmer 2003). As gene transfer and loss occurred with different rates in eukaryotic lineages, the gene content varies among extant eukaryotes (Blanchard and Lynch 2000; Martin and Herrmann

1998). But in all cases the great majority of extant eukaryote mitochondrial proteins are nuclear encoded and this is why these metabolically active organelles depend heavily on their host to function properly.

In multicellular animals the mitochondrial genome normally is a circular double stranded DNA molecule possessing, if any, few repetitive segments, transposable elements or non-coding intergenic sequences (Figure 1.1). It is sized between 13-19 kb and typically contains 37 genes plus at least one AT-rich non-coding region of variable length (Saccone et al. 1999). All genes can be located on the same strand, as reported e.g for *Sipunculus nudus* (Mwinyi et al. 2009), or they are distributed on both strands. In general, the genes encode 13 protein subunits necessary for oxidative phosphorylation (*atp6+8*, *cob*, *cox1-3*, *nad1-6* and *nad4L*), two ribosomal RNAs (*rrnS* and *rrnL*) and 22 transfer RNAs (Boore 1999; Wolstenholme 1992).

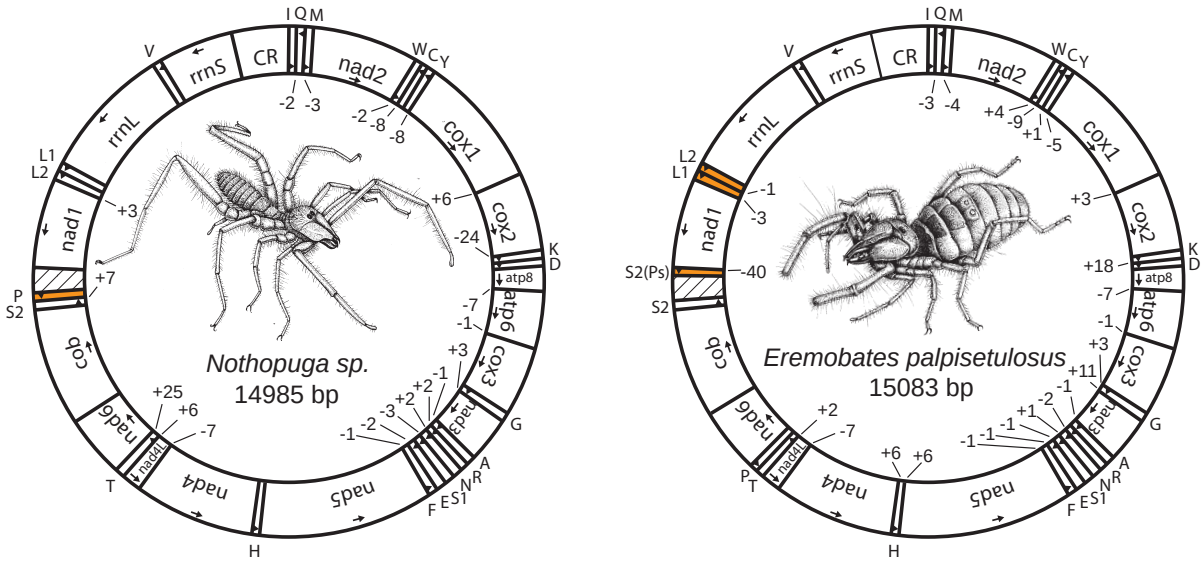


Figure 1.1: Mitochondrial genomes of the two Solifugae *Nothopuga* sp. and *Eremobates palpisetulosus* (according to Masta et al. 2008). tRNA genes are depicted by their one-letter code abbreviations; S2(Ps) indicates a pseudogene of *trnS2*$^{(UCN)}$. Numbers reflect non-coding (positive) or overlapping (negative) nucleotides between two adjacent genes; small arrows indicate the orientation of the genes. Orange shaded genes have a position relatively changed to the ancestral ground pattern. Horizontal stripes indicate regions of repeated sequences between *cob* and *nad1* appearing in both solpugid genomes.

The function of the non-coding AT-rich region was studied only in mammalian and insect mitochondrial genomes (Saito et al. 2005; Shadel and Clayton 1997; Taanman 1999; Zhang and Hewitt 1997; Zhang et al. 1995). Here regulatory elements for the initiation and control of transcription and replication were identified, and therefore it is referred to as the mitochondrial control region, respectively D-loop in vertebrates. It contains two major functionally independent transcription initiation sites, each representing the start of the transcription of one

strand. The two emerging large polycistrons are post-transcriptionally processed by specific enzymes into gene messages (Boore 1999). Furthermore located within the mammalian control region is one of the two distinct replication initiation sites, the origin of H-strand synthesis. The other one, the origin of L-strand synthesis, is situated in the mitochondrial DNA of vertebrates about two-thirds downstream of the control region (Clayton 1982). Replication therefore is an asymmetrical process, as the L-strand replication starts delayed and not until its initiation site is unveiled by the unidirectional replication of the H-strand. So only one strand occurs to be single-stranded for a longer timespan during replication. Generally, there are differences in nucleotide compositions of the two strands - one is typically more rich in C + A and the other in G + T (Perna and Kocher 1995; Reyes et al. 1998). Such a strand bias occurs in most metazoan species and is related to the implied asymmetric mutational constrains on the single-stranded parental H-strand during the replication process (Tanaka and Ozawa 1994), frequently causing deaminations of A and C nucleotides (Sancar and Sancar 1988). In insects, both replication origins are located within the control region, the origin of L-strand synthesis is situated in about 97% downstream of the initiation origin of H-strand synthesis (Saito et al. 2005), causing a likewise asynchronous replication process as described for vertebrate mitochondrial DNA.

Since the mitochondrial genome has evolved in close relation with the nuclear genome for a long time, it, too, has the ability to reflect the evolutionary history of organisms (Boore 1999; Moritz et al. 1987; Saccone et al. 1999) and has thus become object of comparative and phylogenetic studies.

Phylogenetically useful markers

The amount of sequenced mitochondrial genomes to adress phylogenetic questiones is growing rapidly (Carapelli et al. 2007; Podsiadlowski and Braband 2006). The fact that mitochondrial genomes in general are much smaller and more simply structured than nuclear ones, makes sequence data obtainable more easily and often much faster. Further helpful characters of mitochondrial DNA are a lack of recombination, the absence of paralogous genes, and a nucleotide substitution rate which is higher than in nuclear DNA (Moritz et al. 1987; Wolstenholme 1992). These facts also promoted the frequent use of mitochondrial DNA in studies dealing with phylogeny, phylogeography, or population genetics (Avise 2000).
However, in some points mitochondrial genomes exhibit features which may complicate phylogenetic analyses. There is an incidental occurance of nuclear copies of mitochondrial genes (Bensasson et al. 2001) as well as strong differences in nucleotide frequencies caused by e.g. a reversed strand bias (Hassanin et al. 2005). The latter condition may promote

homoplasious mutations, so that the phylogenetic signal can be erroneous due to long branch attraction (LBA, Felsenstein 1978), an artificial clustering of possibly unrelated taxa.

Besides the sequence data in form of nucleotides and amino acids mitochondrial genomes also offer non-sequence characters with phylogenetically useful information. These structural genomic features are also considered as 'genome morphology', as these discrete characters can be added to a morphological data matrix to resolve metazoan phylogeny (Dowton et al. 2002). To name just a few, gene rearrangments, tRNA and rRNA secondary stuctures, an unusual strand bias of the mitochondrial genome, changes in mitochondrial genetic code, and genome topology are valuable sources of data.
Surely, the best examined phylogenetic characters from 'genome morphology' are changes in the mitochondrial gene order due to **gene rearrangements** (Dermauw et al. 2009; Dowton et al. 2002; Fahrein et al. 2007; Kilpert and Podsiadlowski 2010). The comparison of gene order (Figure 1.1) is used on different taxonomic levels. Usually, it remains unchanged over long periods of time and thus is taxon specific (Boore et al. 1995; Shao et al. 2004). Contrariwise, distinctive changes in the gene order, which casually are assumed to be neutral to selection, can characterise subordinated clades, in fact due to the high number of possible recombination events and restrictions for a non-deleterious gene rearrangement (Dowton et al. 2002). Both facts make convergent changes rather unlikely, so changes in gene order are valuable information for phylogenetic questions (Boore et al. 1998; Lavrov et al. 2004). A certainly inhibiting effect on rearrangements is given by the sparsely appearing non-coding intergenic sequences, which enhances the risk of corrupting functional genes with all its negative consequences. Hence, excessively rearranged genomes, which do not allow to trace back the history of events leading to the highly derived gene order, appear rather seldom. Being the only part of mitochondrial genomes with longer non-coding sequence is one explanation why the control region is considered a "hot spot" for rearrangements, and genes located around the control region often show an accelerated mobility compared to other mitochondrial genes (Duarte et al. 2008). Furthermore, tRNA genes in general are translocated more frequently than protein-coding or ribosomal genes (Saccone et al. 1999), especially when located in tRNA clusters (Dowton and Austin 1999). Gene rearrangements are generally classified in local rearrangements, inversions, and long-range translocations. Local rearrangements, where genes move only a small number of gene boundaries or across the control region, are often consistent with the tandem duplication/random loss (TDRL) model (Dowton et al. 2002; Macey et al. 1997; Moritz et al. 1987), occasionally evidenced by the presence of pseudogenes or intergenic spacers (Fujita et al. 2007; Macey et al. 1998). This model is explained by slipped strand mispairing during complementary strand synthesis, generating a duplication of a gene or a gene block. Microdeletions accumulating over time provoke the loss of one redundant gene copy, finally causing the rearrangement of genes. Replication errors like strand slippage mispairing during replication are also accounted to be very likely responsible for frequently detectable copy number variation like tandem

repetitions or duplications in animal mitochondrial DNA (Saito et al. 2005; Zhang and Hewitt 1997). It is very unlikely that long-range translocations, without any additional change in the intervening gene order, are consistent with the TDRL model. Inversions are suggested to be the result of intramolecular recombination due to a twist of the mitochondrial genome followed by a double-strand break subsequently repaired (Dowton and Campbell 2001). In case of arthropod mitochondrial genomes the comparison of gene order was already successful in determining an ancestral arthropod ground pattern of gene order (Lavrov et al. 2000; Staton et al. 1997), which is extremely helpful for the inference of rearrangement events in newly sequenced species, and thus for phylogenetic considerations.

The influence of the replication process according to the 'strand-displacement model' on the nucleotide composition of the two mitochondrial DNA strands has already been depicted. Also the control region itself is under suspicion to have an influence on the distibution of nucleotides, as for different unrelated taxa a clearly **reversed mitochondrial strand bias**, appearing TG-rich rather than CA-rich, is reported. It might have emerged multiple times independently among Bilateria, and Hassanin et al. (2005) proposed it as the result of a control region inversion involving a reversal of mutational constraints for the two strands of the mitochondrial genome. Since unusual nucleotide frequencies afford an opportunity to link or to delineate the members of clades on different levels, the calculation of a strand bias by now is included in most examinations of phylogenetic relationships (Masta et al. 2009). Unusual nucleotide frequencies seem to affect sequence-based phylogenetic analyses (Hassanin et al. 2005; Hassanin 2006), therefore their convenience has to be revalued as the case arises.

The use of inferred **transfer RNA secondary structures** as structural data to reconstruct phylogeny is approved since aberrations of the typical cloverleaf structure within nematodes depicted a delimitation of some nematode lineages (Lavrov and Brown 2001; Okimoto et al. 1992; Wolstenholme et al. 1987). Also the structural change in *trnS*$^{(AGN)}$ of nearly all metazoans (Haen et al. 2007; Wolstenholme 1992) emphasizes the long-term stability of tRNA structural features, making it possible to trace back diversification. A first comprehensive analysis of transfer RNA genes to elucidate phylogenetic relationships of arachnids, has recently been published (Masta and Boore 2008), presenting promising results for following investigations.

Comparison of inferred **ribosomal RNA secondary structures** among arthropods also reveals structural differences in this otherwise well-conserved molecule (Dermauw et al. 2009; Domes et al. 2008; Fahrein et al. 2009; Gillespie et al. 2006; Masta 2010; Shao et al. 2006). If these findings are connected with on and off observable derived secondary structures of transfer RNAs, they at least deserve an increase of the few rRNA secondary structures published so far. Anyhow, Masta (2000) suggested that the normal functioning of tRNAs exhibiting a TV-replacement loop instead of a T-arm in several cases, may require changes in rRNA secondary structures, yet further studies are indispensible.

1.2 Arachnid phylogeny

From a textbook perspective the predominantly terrestrial Arachnida are situated together with the marine Xiphosura and Pycnogonida to form the clearly defined Chelicerata (Raw 1957; Weygoldt and Paulus 1979), altogether a taxon of a great diversity of predatory and parasitic arthropods. Chelicerates are often considered to be the sister group of Mandibulata (e.g. Regier et al. 2010; Snodgrass 1938), although some molecular-based studies join Chelicerata with Myriapoda, together forming a clade called "Paradoxopoda" or "Myriochelata" (Hwang et al. 2001a; Mallatt et al. 2004). Chelicerata comprise about 84.000 extant species and underwent their initial radiation in the Cambrian (Lindquist 1984). Basically two peculiarities characterize chelicerates: 1) the eponymous pre-oral **chelicerae** which serve as a primarily tripartite grasping organ and emerged from modified appendages of the second segment, and 2) a body which is composed of **two tagmata** - the prosoma comprising the first 7 segments, an opisthosoma of twelve segments, and a post-anal telson (Ax 1996; Selden and Dunlop 1998).

The phylogeny of Chelicerata is still controversial, partly due to the questionable inclusion of the marine Pycnogonida and partly regarding conflicting interpretations of the interrelations of major arachnid lineages. Regarding pycnogonids, many authors consider these bizarre and scarcely studied arthropods as primarily aquatic chelicerates with a basal position outside of Arachnida (Dunlop and Arango 2005; Giribet et al. 2005; Mallatt et al. 2004; Regier et al. 2005; Regier and Shultz 2001) or even as aquatic arachnids closly related to Acari (Park et al. 2007). However, their affinity to euchelicerates is questioned by others (Giribet et al. 2001; Maxmen et al. 2005), but difficulties in cladistic analyses are associated with the derived external morphology and the high amount of inapplicable (absent) data (Arango 2003). Even mitochondrial data (Park et al. 2007; Podsiadlowski and Braband 2006) and large sets of nuclear data (Regier et al. 2010) did not unequivocally reveal the phylogenetic position of pycnogonids. But that question is not covered by this thesis, I rather concentrate on the likewise disputed phylogeny of arachnids. Although some studies with morphological datasets (Giribet et al. 2002; Van der Hammen 1989) challenge the monophyly of arachnids by assuming the marine Xiphosura as close relatives of arachnid subtaxa, a common origin of all arachnids is widely accepted and supported by several morphological characters associated with terrestrial life (Ax 1996; Shultz 1990; Shultz 2007; Weygoldt and Paulus 1979). Some of the adaptations pertaining to the terrestrial living are the locomotion with 4 pairs of walking legs (due to the transformation of the former first walking leg to the tactile pedipalp), lungs or trachee for the respiration of atmospheric air, malpighian vessels for new excretory demands, indirect sperm transfer with spermatophores, and new sensory organs (e.g. trichobothria and slit organs).

Figure 1.2 shows a compilation of some morphological (Figure 1.2 a-c, f) and combined (Figure 1.2 d+e) analyses of arachnid phylogeny. Consistent with all analyses is the appearance

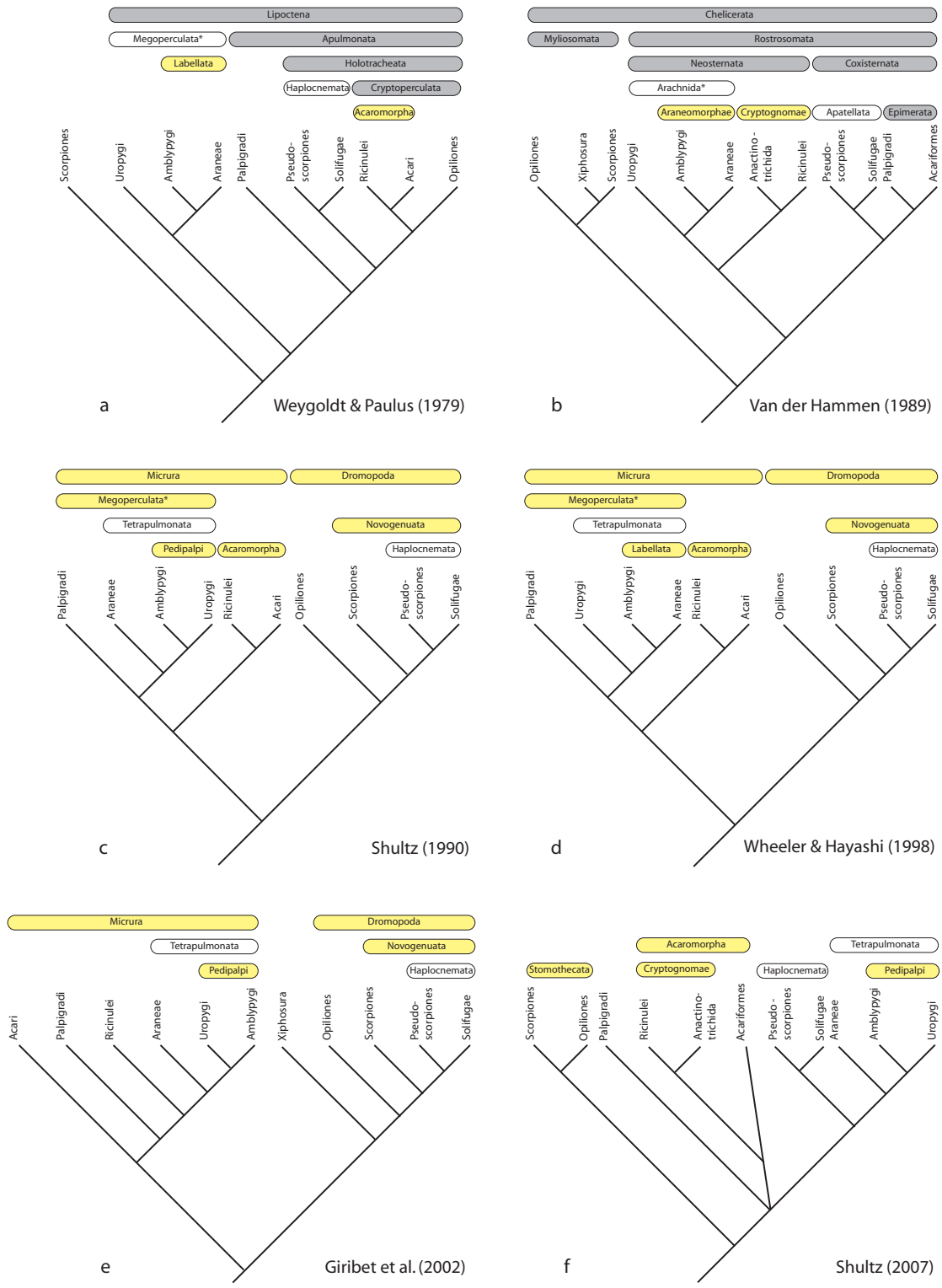


Figure 1.2: Analyses of several authors regarding the phylogeny of Arachnida (**a, c, d, f**) respectively Euchelicerata (**b+e**). **White** shaded taxa appear in all analyses, **yellow** ones appear in more than one analysis, and **grey** shaded taxa only appear in the corresponding analyses. * indicates that the taxa Megoperculata and Arachnida are used with different meaning by several authors: Shultz (1990) and Wheeler and Hayashi (1998) but not Weygoldt and Paulus (1979) enclose Palpigradi in Megoperculata;Arachnida *sensu* Van der Hammen (1989) comprises only Uropygi, Amblypygi and Araneae and not like in **a, c, d,** and **f** all terrestrial Chelicerata.

of two clades, one comprising Uropygi (vinegaroons), Amblypygi (whip spiders), and Araneae (true spiders) in a close relationship and the other containing Pseudoscorpiones and Solifugae (camel spiders) as sister groups. Apart from these relationships the proposed arachnid phylogenies are highly diverse. From a morphological perspective, in particular, two diverging hypotheses are established (Figure 1.2 a+c) which considerably vary regarding the position of Scorpiones. Weygoldt and Paulus (1979) suggest a basal position of the Scorpiones as sister taxon of the remaining arachnids, the so-called Lipoctena, which is also supported by others (e.g. Firstman 1973; Selden and Dunlop 1998; Weygoldt 1998). Contrary to this, Shultz (1990) favours a close relationship of scorpions to Opiliones (harvestman), Pseudoscorpiones, and Solifugae, together forming the clade Dromopoda, which also appears in the combined analyses of Wheeler and Hayashi (1998) and Giribet et al. (2002). A close relationship between scorpions and Opiliones is also affirmed by morphological studies of Giribet et al. (2002) Van der Hammen (1989), while a more recent analysis of Hassanin (2006) based on only molecular data reveals scorpions as one of the most primitive chelicerates (thus questioning the monophyly of arachnids).

Furthermore problematic is an evaluation of the intraordinal relationships of the Tetrapulmonata, respectively Megoperculata *sensu* Weygoldt and Paulus (1979). As mentioned above, the monophyly of Uropygi, Amblypygi, and Araneae is widely accepted, but phylogenetic relationships among these three taxa are controversial with two major competing hypotheses (see Figure 1.2 for “Labellata”, respectively ”Araneomorphae” hypothesis versus “Pedipalpi” hypothesis).

Another debatable issue of arachnid phylogeny concerns the Acari (mites and ticks). According to the morphological studies of Shultz (1990) and Weygoldt and Paulus (1979) and the combined morphological and molecular analysis of Wheeler and Hayashi (1998), Acari are considered to be the sister group of ricinuleids, together forming the clade “Acaromorpha”. Unlike this, in the combined data set of Giribet et al. (2002) the Acari have a basal position in a clade together with Palpigradi, Ricinulei, and Tetrapulmonata, although the authors mentioned the instability of this relationship as in their analysis of only the molecular data set Ricinulei appear to be the sister group to all remaining arachnids. However, the morphological studies of Van der Hammen (1989) and Shultz (2007) even question the monophyly of Acari by placing Ricinulei within the Acari as the sister group to Anactinotrichida (Parasitiformes s. str. + Opilioacariformes).

Especially striking is the relative position of Palpigradi, which happens to be highly unstable as well in morphological as in combined analyses.

So both, morphological analyses and combined analyses (which consist of nuclear sequence-based and morphological characters) produced conflicting hypotheses of arachnid phylogenetic relationships which thus remain unresolved. The difficulties in evaluating phylogenetic relationships within the Arachnida based on morphological data are probably caused by

homoplasy or reduction of anatomical characters, which frequently have to be correlated with particular requirements of the newly-captured ecological zone. Controversial results from combined analyses (Giribet et al. 2002; Wheeler and Hayashi 1998) may be caused due to the prevalent use of 18S rRNA sequences as molecular markers, because this gene appears to have changed sparsely during chelicerate evolution and thus, in this case, does not display a suitable data set for the inference of phylogeny. Hence, there obviously is a need for additional data sets attempting to reconstruct phylogeny.

1.3 Aims of this thesis

As debated before, using morphological and nuclear sequence-based characters, a largely agreed arachnid phylogeny could not be established so far, and since mitochondrial genomes have been proven useful for phylogenetic reconstructions, I chose these compact circular genomes to shed light on relationships among major lineages of arachnids. When work on this thesis was started, complete mt genomes were only available for a small amount of taxa. For the arachnid orders Ricinulei, Palpigradi, Pseudoscorpiones, Amblypygi, Uropygi, Opiliones, and Solifugae there was no or no complete mitochondrial genome sequence information available. So in the first instance I simply tried to fill some of these “gaps”. As a next step and for phylogenetic questioned analysis, the newly achieved mitochondrial genome data was used together with arachnid mitochondrial genome data deposited at GenBank, for sequence-based as well as for non-sequence-based analyses. Both are items of the publications on which this thesis is based. Above all, the following questions are included in considerations about arachnid phylogeny:

- *Is mitochondrial genomics applicable to infer arachnid phylogenetic relationships?*

- *Do rearrangements of mitochondrial genes provide phylogenetic information to reveal or evaluate intraordinal relationships of Arachnida?*

- *Is there any difference in the value of rearrangements for the phylogenetic inference of “higher” and “lower” ranking taxa?*

- *Are genome rearrangements distributed equally over the genome or do areas with an increased rate of translocations exist (= rearrangement-hotspots)?*

- *Are any other characters of ‘genome morphology’ useful to shed light on relationships of the major lineages of Arachnida?*

2. The complete mitochondrial genome of *Pseudocellus pearsei* (Chelicerata: Ricinulei) and a comparison of mitochondrial gene rearrangements in Arachnida

2.1 Abstract

Background: Mitochondrial genomes are widely utilized for phylogenetic and population genetic analyses among animals. In addition to sequence data the mitochondrial gene order and RNA secondary structure data are used in phylogenetic analyses. Arachnid phylogeny is still highly debated and there is a lack of sufficient sequence data for many taxa. Ricinulei (hooded tickspiders) are a morphologically distinct clade of arachnids with uncertain phylogenetic affinities.

Results: The first complete mitochondrial DNA genome of a member of the Ricinulei, *Pseudocellus pearsei* (Arachnida: Ricinulei) was sequenced using a PCR-based approach. The mitochondrial genome is a typical circular duplex DNA molecule with a size of 15,099 bp, showing the complete set of genes usually present in bilaterian mitochondrial genomes. Five tRNA genes (*trnW*, *trnY*, *trnN*, *trnL*(CUN), *trnV*) show different relative positions compared to other Chelicerata (e.g. *Limulus polyphemus*, *Ixodes* spp.). We propose that two events led to this derived gene order: (1) a tandem duplication followed by random deletion and (2) an independent translocation of *trnN*. Most of the inferred tRNA secondary structures show the common cloverleaf pattern except tRNA-Glu where the TψC-arm is missing. In phylogenetic analyses (maximum likelihood, maximum parsimony, Bayesian inference) using concatenated amino acid and nucleotide sequences of protein-coding genes the basal relationships of arachnid orders remain unresolved.

Conclusion: Phylogenetic analyses (ML, MP, BI) of arachnid mitochondrial genomes fail to resolve interordinal relationships of Arachnida and remain in a preliminary stage because there is still a lack of mitogenomic data from important taxa such as Opiliones and Pseudoscorpiones. Gene order varies considerably within Arachnida - only eight out of 23 species have retained the putative arthropod ground pattern. Some gene order changes are valuable characters in phylogenetic analysis of intraordinal relationships, e.g. in Acari.

2.2 Background

Due to their bacterial origin (Altmann 1890; Margulis 1970) mitochondria have retained a circular DNA double-helix, which in animals is sized between 12-30 kb. This is only a small part of the original bacterial chromosome, the majority was eliminated or transferred to the nucleus (Adams and Palmer 2003). The mitochondrial DNA of Bilateria typically contains 37 genes and one AT-rich non-coding part, which putatively bears regulatory elements for transcription and translation and is therefore referred to as the mitochondrial control region (Boore 1999). In general the genes encode 13 protein subunits necessary for oxidative phosphorylation (*atp6+8*, *cob*, *cox1-3*, *nad1-6* and *nad4L*), 22 transfer RNAs and two rRNAs (*rrnS* and *rrnL*) (Wolstenholme 1992). Except for the control region, mtDNA possesses only few non-coding sections between genes, even gene overlaps are common. E.g., in many species the last seven nucleotides of *atp8* are also the first seven nucleotides of *atp6*. A similar overlap is often seen on the boundary between *nad4* and *nad4L*. As a consequence, rearrangements in mitochondrial genomes most often disrupt genes and thus are deleterious - a possible reason for the stability of mitochondrial gene order (Boore 1999).
Mitochondrial genomes have been proven useful for phylogenetic analyses (Boore 1999). Nucleotide or amino acid sequences as well as rearrangements in mitochondrial gene order are used as phylogenetic markers (Boore et al. 1998; Boore et al. 1995). Gene rearrangements are considered to be valuable characters, because it is very unlikely that closely related taxa exhibit homoplastic translocations (Dowton et al. 2002). In addition the secondary structure of encoded tRNAs (Macey et al. 2000) and changes in the mitochondrial genetic code (Castresana et al. 1998) have also been used as characters in phylogenetic analysis.
Ricinulei (hooded tickspiders) are a small order of arachnids, comprising 3 genera with 55 described species (Bonaldo and Pinto-da-Rocha 2003; Cokendolpher and Enriquez 2004; Harvey 2003; Pinto-da-Rocha and Bonaldo 2007). They are predatory animals that live in humid caves or leaf litter of tropical regions (Adis et al. 1989; Cooke 1967; Mitchell 1970). Species of *Ricinoides* occur in West Africa whereas species of *Cryptocellus* and *Pseudocellus* live in Central and South America (Harvey 2003). Ricinuleids have body lengths of 3 to 10 mm (Moritz 1993) and their cuticle is strongly sclerotized and extraordinarily thick (Kennaugh 1968). Several peculiarities characterize ricinuleids - a moveable hood (cucullus) in front of the prosoma covering the mouthparts, two jointed chelicerae, chelate pedipalps, elongated second legs, a tarsal copulatory organ on the third pair of legs of adult males, a locking mechanism between pro- and opisthosoma, which can be unlocked during mating and egg-laying, a 6-legged larvae, the lack of distinct eyes, and tracheal respiration (Ax 1996; Hansen and Sørensen 1904; Legg 1976; Legg 1977; Pittard and Mitchell 1972; Selden and Dunlop 1998). According to morphological studies and some combined morphological and molecular analyses, ricinuleids are often considered to be the sister group of Acari (Lindquist 1984; Shultz 1990; Shultz 2007; Weygoldt and Paulus 1979), together forming

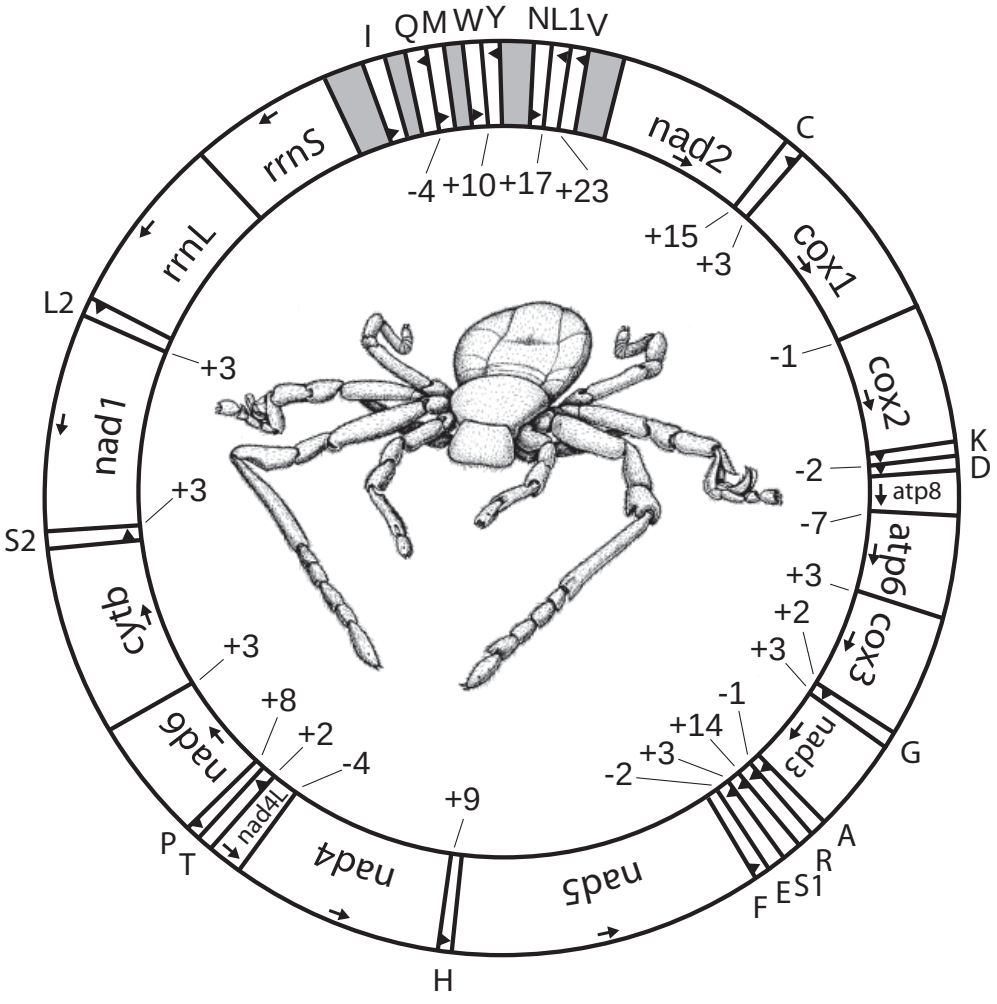
the Acaromorpha. This clade is characterized by a unique post-embryonic development: a hexapodal larva followed by three octapod nymphal instars (Lindquist 1984). A gnathosoma with medially fused palpal coxae is another unique character of Acaromporpha. However the definition of a "gnathosoma" varies and its presence in Ricinulei is questioned by some authors (e.g. Van der Hammen 1989). The analysis of Shultz (Shultz 2007) obtained seven additional "homoplasious synapomorphies" supporting Acaromorpha (= these characters are not exclusively found in Acari and Ricinulei). Van der Hammen (1972; 1982) however has questioned the monophyly of Acari and placed Ricinulei within the Acari as the sister group to Anactinotrichida (Parasitiformes s.str.+ Opilioacariformes) (Van der Hammen 1979). Dunlop (Dunlop 1996) suggested Ricinulei as sister group to the extinct Trigonotarbida which are collectively the sister of the extant Tetrapulmonata (Araneae, Uropygi, Amblypygi).
Because available sequence data from nuclear or mitochondrial genes is very limited for Ricinulei, recent molecular studies of arthropod phylogeny seldom included ricinuleids (e.g. Mallatt and Giribet 2006). With a combined data set (93 morphological characters and 18S + 28S rRNA sequences) Wheeler & Hayashi (1998) placed Ricinulei as sister group to Acari. In another combined analysis by Giribet et al. (2002; 253 morphological characters, 18S and 28S rRNA) Ricinulei appear as the sister group of Tetrapulmonata (Araneae + Amblypygi + Uropygi). When fossil taxa were included, a close relationship between the fossil Trigonotarbida and Ricinulei was recovered and both together were the sister group to Tetrapulmonata. This phylogenetic hypothesis corresponds with that of Dunlop (Dunlop 1996), but Giribet et al. mentioned the instability of this relationship: In their analysis of only the molecular data set Ricinulei is early branching within the arachnid tree as the sister group to all remaining arachnids (Giribet et al. 2002).
In this study, we present the first complete mitochondrial genome of a member of Ricinulei, the hooded tickspider *Pseudocellus pearsei* (Chamberlin & Ivie, 1938). The sequence data is used to unveil phylogenetic relationships between Arachnida. Furthermore, the gene order of mitochondrial genomes from all available arachnid species is compared in order to reconstruct the events leading to derived genome arrangements and to evaluate the phylogenetic significance of gene translocations within the Arachnida.

2.3 Results and Discussion

Genome organization, gene order and non-coding parts

The generation of overlapping PCR fragments and subsequent sequencing demonstrated that the mitochondrial genome of *P. pearsei* is a typical circular DNA molecule with a length of 15099 bp [GenBank:EU024483]. All 37 genes usually present in bilaterian mitochondrial genomes have been identified (Figure 2.1, Table 2.1). With the exception of five translocated

tRNAs, gene order is similar to that of the horseshoe crab, *Limulus polyphemus*, which is considered to represent the putative ground pattern of the Arthropoda (Lavrov et al. 2000; Staton et al. 1997). The tRNAs (*trnW*, *trnY*, *trnN*, *trnL*(CUN), *trnV*) changed their position to a new location between *trnM* and *nad2*. In this part of the genome there are five non-coding regions, which are ranging in size from 87 to 250 bp. The three largest non-coding regions are located between *rrnS* and *trnI* (250 bp), *trnY* and *trnN* (183 bp) and *trnV* and *nad2* (169 bp). The shorter ones are situated between *trnI* and *trnQ* (100 bp) and *trnM* and *trnW* (87 bp). The longest non-coding region (250 bp) is flanked by *rrnSI* and *trn*, and based on its similarity to other arthropods (Qiu et al. 2005; Wolstenholme 1992; Zhang and Hewitt 1997) can be identified as the putative mitochondrial control region. Part of this region is capable



Pseudocellus pearsei
15099 bp

Figure 2.1: Mitochondrial genome map of *P. pearsei*. Transfer RNA genes are depicted by their one letter amino acid code (L_1: *trnL*(CUN), L_2: *trnL*(UUR), S_1: *trnS*(AGN), S_2: *trnS*(UCN)). Numbers indicate non-coding nucleotide spacers between genes (positive values) or gene overlap (negative values). Arrows indicate orientation on (+)strand (clockwise) or (-) strand (counterclockwise). Grey shaded parts represent larger non-coding regions (>50bp). Line drawing of *P. pearsei* by Peter Adam. Body lengths of the animal is 4.6 mm.

Table 2.1: Genome organisation of *P. pearsei*. Complete circular mtDNA has a length of 15099 bp.

Gene	*Strand*	*Position*	*Length (nuc.)*	*CG-skew*	*Start-codon*	*Stop-codon*	*Intergenic nucleotides*
cox1	+	1 - 1548	1548	0.277	ATG	TAA	-1
cox2	+	1548- 2218	671	0.414	ATG	TA	0
trnK	+	2219- 2289	71				-2
trnD	+	2288- 2350	63				0
atp8	+	2351- 2503	153	0.745	ATT	TAA	-7
atp6	+	2497- 3171	675	0.482	ATG	TAA	+3
cox3	+	3175- 3954	780	0.341	ATG	TAA	+2
trnG	+	3957- 4019	63				+3
nad3	+	4023- 4356	334	0.543	ATA	T	0
trnA	+	4357- 4421	65				-1
trnR	+	4421- 4482	62				+14
trnS(AGN)	+	4497- 4552	56				+3
trnE	+	4556- 4618	63				-2
trnF	-	4617- 4677	61				0
Nad5	-	4678- 6364	1687	-0.650	ATA	T	0
trnH	-	6365- 6425	61				+9
Nad4	-	6435- 7752	1318	-0.588	ATA	T	-4
Nad4L	-	7749- 8021	273	-0.872	ATG	TAG	+2
trnT	+	8024- 8085	62				0
trnP	-	8086- 8149	64				+8
Nad6	+	8158- 8589	432	0.587	ATA	TAA	+3
Cob	+	8593- 9697	1115	0.503	ATG	T	0
trnS(UCN)	+	9698- 9762	65				+3
nad1	-	9766- 10651	886	-0.588	ATA	T	+3
trnL(UUR)	-	10655- 10720	66				0
rrnL	-	10721- 11970	1250	-0.496			0
rrnS	-	11971- 12713	743	-0.471			+250
trnI	+	12964- 13020	57				+100
trnQ	-	13121- 13184	64				-4
trnM	+	13181- 13247	67				+87
trnW	+	13335- 13400	66				+10
trnY	-	13411- 13472	62				+183
trnN	+	13656- 13716	61				+17
trnL(CUN)	-	13734- 13800	67				+23
trnV	-	13824- 13888	65				+169
nad2	+	14058- 15020	963	0.556	ATT	TAA	+15
trnC	+	15036- 15096	61				+3

of folding into a hairpin-like formation (Figure 2.2) with a loop consisting of 10 nucleotides and a stem composed of 21 paired nucleotides (five mismatches). Furthermore conserved motifs occur in the flanking sequences around the stem-loop structure: a TATA motif appears in the 5`-flanking sequence whereas the motif GA(A)T is found in the 3`-flanking sequence (Figure 2.2). Both motifs are also present in flanking sequences of other arthropods, e.g. in metastriate ticks (Black and Roehrdanz 1998), crustaceans (Kilpert and Podsiadlowski 2006),

and insects and are presumed to have functional significance in transcription and/or replication (Zhang et al. 1995). The other larger non-coding regions do not bear similar hairpin-like structures. Besides the above mentioned regions only smaller non-coding regions of 2-23 bp are found in the mitochondrial genome. Gene overlaps occur between *cox1* and *cox2* (1 bp), *nad4* and *nad4L* (4 bp), *atp6* and *atp8* (7 bp), and in four cases tRNA-genes are overlapping each other (Figure 2.1, Table 2.1).

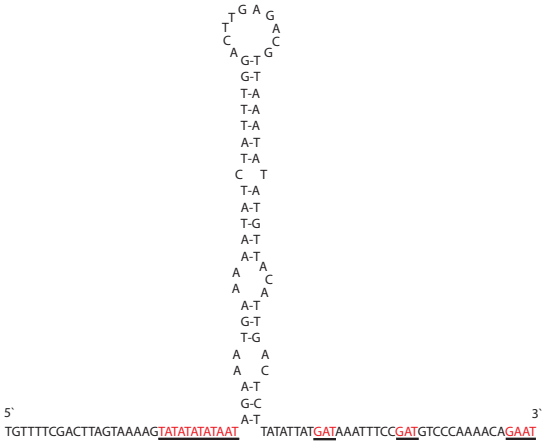


Figure 2.2: Stem-loop structure in the mitochondrial control region of *P. pearsei*. Underlined areas point out conserved motifs in 3`-and 5`-flanking sequences.

Protein-coding genes and nucleotide composition

All of the 13 identified protein-coding genes begin with one of the common start codons for mtDNA ATG, ATA or ATT (Table 2.1). Out of these 13 protein-coding genes, six show incomplete stop codons (*cox2*, *cob* and *nad1*, *3*, *4*, *5*). In *cox2* the stop codon is truncated and the gene terminates with TA, in the other five protein-coding genes only a single thymine serves as partial stop codon. Similar structural events have also been described for the mtDNA of other species where post-transcriptional polyadenylation completes a T or TA partial stop codon into a functional one (Ojala et al. 1981).

The CG-skew (= (%C-%G)/(%C+%G)) of mitochondrial genes is a good indicator of the strand specific nucleotide frequency bias (Hassanin et al. 2005; Perna and Kocher 1995). In *P. pearsei* the CG-skew is positive in all (+)strand encoded genes and negative in all (-)strand encoded genes (Table 2.1). We defined the (+)strand as the strand bearing the majority of coding sequence. The strand specific bias found in *P. pearsei* occurs in most other arthropods, while a reversal of that bias has been reported for only a few species (Hassanin et al. 2005; Hassanin 2006; Kilpert and Podsiadlowski 2006; Navajas et al. 2002; Podsiadlowski and Bartolomaeus 2006). Table 2.2 shows the CG-skews of third codon positions of *cox1* for all chelicerates and outgroup taxa included in our phylogenetic analyses. We chose *cox1* because this gene is found on (+)strand in all species examined. A reversal of CG-skew is seen in the two scorpions, in three Araneae (but not in *Heptathela*) and in the parasitiform mite *Varroa*.

Table 2.2: Nucleotide composition of third codon positions of the (+)strand encoded gene *cox1*, demonstrating strand bias in nucleotide frequencies within chelicerates.

Species	Taxon	A	C	G	T	CG skew	Accession number
Lithobius forficatus	Myriapoda	0.411	**0.200**	0.061	0.329	0.534	[GenBank:NC_002629]
Daphnia pulex	Crustacea	0.273	**0.211**	0.166	0.350	0.119	[GenBank:NC_000844]
Penaeus monodon	Crustacea	0.388	**0.133**	0.053	0.427	0.432	[GenBank:NC_002184]
Limulus polyphemus	Xiphosura	0.434	**0.229**	0.023	0.315	0.814	[GenBank:NC_003057]
Oltacola gomezi	Solifugae	0.407	**0.199**	0.055	0.339	0.569	[GenBank:EU024482]
Centruroides limpidus	Scorpiones	0.094	0.045	**0.264**	0.597	-0.709	[GenBank:NC_006896]
Mesobuthus martensii	Scorpiones	0.161	0.020	**0.309**	0.511	-0.881	[GenBank:NC_009738]
Heptathela hangzhouensis	Araneae	0.374	**0.096**	0.061	0.470	0.225	[GenBank:NC_005924]
Ornithoctonus huwena	Araneae	0.326	0.033	**0.166**	0.475	-0.667	[GenBank:NC_005925]
Habronattus oregonensis	Araneae	0.361	0.012	**0.101**	0.526	-0.792	[GenBank:NC_005942]
Nephila clavata	Araneae	0.379	0.016	**0.109**	0.496	-0.750	[GenBank:NC_008063]
Pseudocellus pearsei	Ricinulei	0.405	**0.281**	0.031	0.283	0.801	[GenBank:EU024483]
Leptotrombidium akamushi	Acariformes	0.356	**0.176**	0.084	0.384	0.354	[GenBank:NC_007601]
Leptotrombidium deliense	Acariformes	0.389	**0.123**	0.074	0.413	0.247	[GenBank:NC_007600]
Leptotrombidium pallidum	Acariformes	0.388	**0.114**	0.055	0.444	0.349	[GenBank:NC_007177]
Amblyomma triguttatum	Parasitiformes	0.392	**0.106**	0.043	0.459	0.421	[GenBank:NC_005963]
Haemaphysalis flava	Parasitiformes	0.433	**0.105**	0.043	0.419	0.421	[GenBank:NC_005292]
Rhipicephalus sanguineus	Parasitiformes	0.431	**0.086**	0.025	0.458	0.545	[GenBank:NC_002074]
Ixodes hexagonus	Parasitiformes	0.380	**0.197**	0.043	0.380	0.642	[GenBank:NC_002010]
Ixodes holocyclus	Parasitiformes	0.420	**0.093**	0.035	0.451	0.455	[GenBank:NC_005293]
Ixodes persulcatus	Parasitiformes	0.388	**0.103**	0.031	0.478	0.536	[GenBank:NC_004370]
Ixodes uriae	Parasitiformes	0.410	**0.173**	0.039	0.379	0.633	[GenBank:NC_006078]
Carios capensis	Parasitiformes	0.446	**0.185**	0.020	0.349	0.809	[GenBank:NC_005291]
Ornithodoros moubata	Parasitiformes	0.413	**0.164**	0.041	0.382	0.600	[GenBank:NC_004357]
Ornithodoros porcinus	Parasitiformes	0.378	**0.164**	0.064	0.394	0.436	[GenBank:NC_005820]
Metaseiulus occidentalis	Parasitiformes	0.400	**0.214**	0.037	0.349	0.708	[GenBank:NC_009093]
Varroa destructor	Parasitiformes	0.363	0.018	**0.062**	0.557	-0.560	[GenBank:NC_004454]

Bold numbers indicate higher values in comparison of G and C proportions. Only those taxa (24 chelicerate and 3 outgroup taxa) included in our phylogenetic analysis (Figure 2.4) are listed.

Secondary structure of transfer RNAs

The mitochondrial genome of *P. pearsei* bears all of the 22 tRNAs commonly found in metazoan mtDNA (Figure 2.1, Table 2.1). Except for tRNA-Glu, all tRNAs possess the typical cloverleaf secondary structure, though the TψC stem is shortened in several tRNAs (Figure 2.3). The TψC-arm of tRNA-Glu is entirely absent. It is shortened to a single pair of nucleotides in tRNA-Met and tRNA-Phe and it is composed of just two paired bases in tRNA-Gly, tRNA-His, tRNA-Ile, tRNA-Leu (CUN), tRNA-Ser (AGN) and tRNA-Thr.

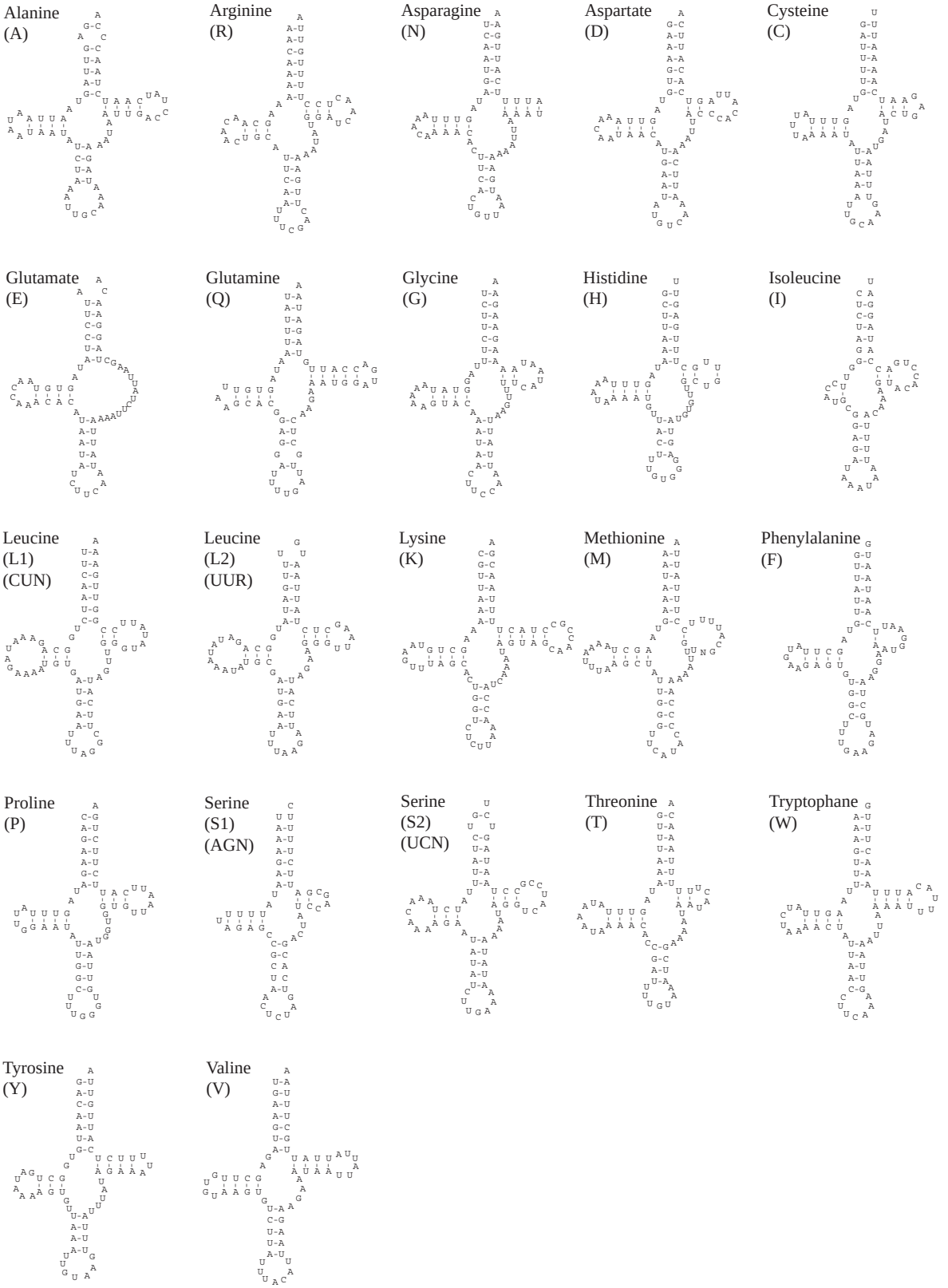


Figure 2.3: Putative secondary structures of mitochondrial tRNAs found in *P. pearsei.* All tRNAs can be folded into the usual cloverleaf secondary structure.

Phylogenetic analysis

We performed phylogenetic analyses with two different data sets: concatenated amino acid and concatenated nucleotide sequences (without third codon positions) from all protein-coding genes. Topologies with best likelihood scores from maximum likelihood (ML) analysis are shown (Figure 2.4). Topologies of the 50% majority rule consensus trees from ML bootstrapping and Bayesian Inference (BI) differ only slightly from the best topologies of ML analysis: BI with the amino acid dataset resulted in a basal split between *Oltacola* and the remainder of chelicerates including *Limulus*, but that node is not supported by bootstrapping or BI; in ML bootstrapping with the nucleotide dataset there was no resolution between Ricinulei, Scorpiones, Araneae and Acari (ML bootstrap <50%). In all performed analyses, good support was found for monophyly of Scorpiones, Opisthothelae (= all Araneae except Mesothelae, here represented by *Heptathela*), *Ixodes*, *Leptotrombidium*, Ornithodorinae (*Carios* + *Ornithodoros*), Ixodidae (*Ixodes* + *Amblyomma* + *Haemaphysalis* + *Rhipicephalus*), Metastriata (*Amblyomma* + *Haemaphysalis* + *Rhipicephalus*), and Dermanyssina (*Metaseiulus* + *Varroa*). Monophyly of Acari-Parasitiformes was also well supported by most analyses, except for the maximum parsimony (MP) bootstrapping (56%). Monophyly of Acari was recovered in the best topology of the nucleotide ML analysis and moderately supported by ML bootstrapping (80%), but not by BI or MP bootstrapping. Acari and Ricinulei form sister groups in the best ML topology in the analysis of the nucleotide alignment, but this clade found support only by BI (0.98), not by ML or MP bootstrapping. In contrast analyses of the amino acid dataset resulted in paraphyletic Acari, as Acariformes (*Leptotrombidium*) form a clade with Ricinulei, and both together form the sister group to Araneae. Both clades are well supported only by BI (Acariformes + Ricinulei: 1.0; Acariformes + Ricinulei + Araneae: 0.99).

Arachnid monophyly is only supported by ML analysis with the amino acid dataset (bootstrap: 74%). In all other analyses the three basal chelicerate branches - *Limulus* (Xiphosura), *Oltacola* (Solifugae), and the remainder of arachnids - are not well resolved. Monophyly of Arachnida excl. Solifugae (in this case Scorpiones + Araneae + Acari + Ricinulei) is well supported in ML and BI analyses of the amino acid alignment and weakly supported by ML analysis of the nucleotide alignment.

The classical, morphology based view of the phylogenetic position of Ricinulei is as the sister group to Acari (Lindquist 1984; Shultz 1990; Shultz 2007; Weygoldt and Paulus 1979). Because our analysis does not resolve the basal relationships of arachnid orders well, it is not in conflict with actual morphology based analyses of arachnid interrelationships (Shultz 2007; Wheeler and Hayashi 1998). A sister group relation between Acari and Ricinulei is neither well supported, nor refuted by our results.

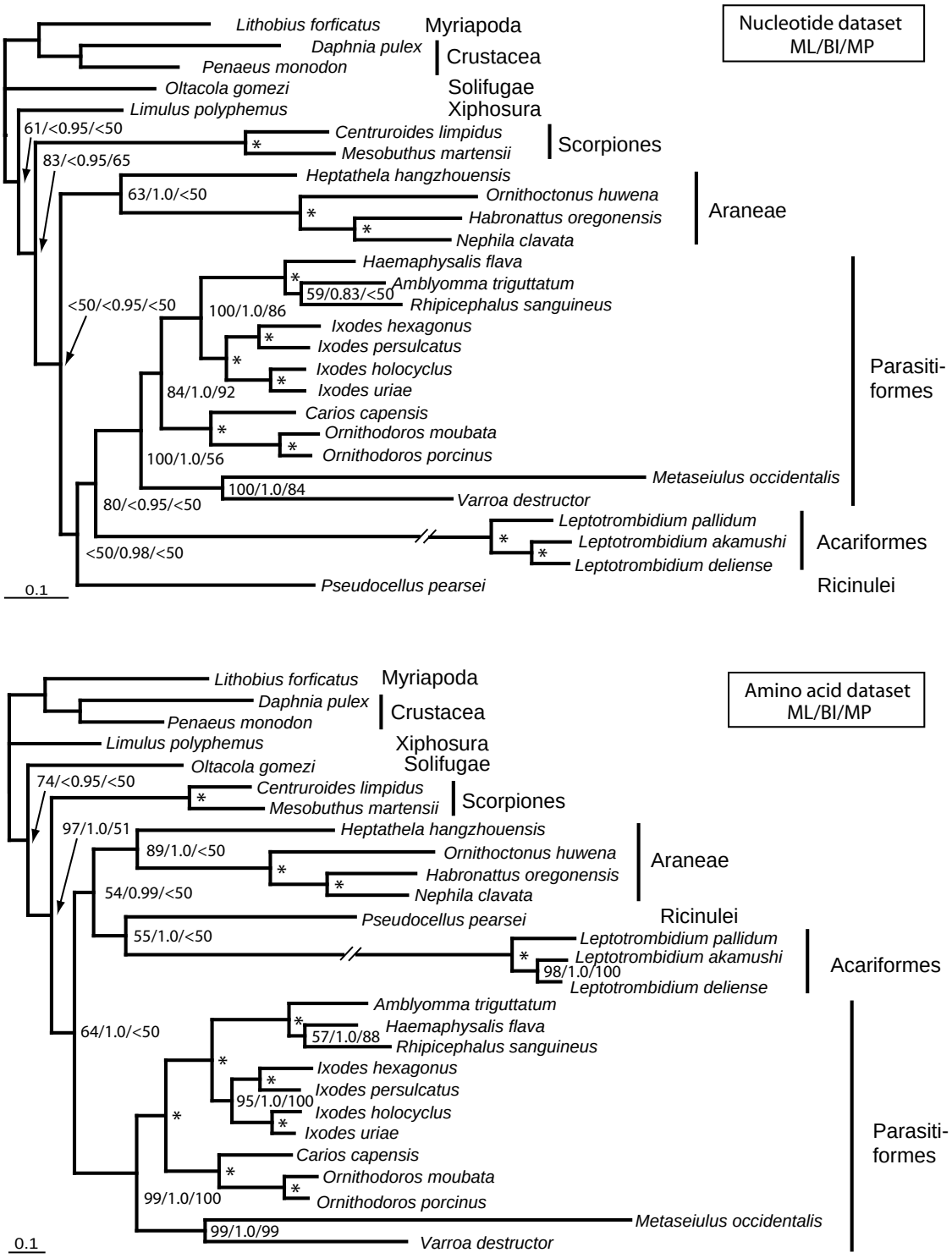


Figure 2.4: Phylogenetic trees of chelicerate relationships, inferred from nucleotide (upper) and amino acid (lower) datasets. All protein coding gene sequences were aligned and concatenated; ambiguously aligned regions were omitted by Gblocks. Trees were rooted with outgroup taxa (*Lithobius*, *Daphnia*, *Penaeus*). Topologies and branch lengths were taken from the best scoring trees of the maximum likelihood (ML) analyses. Numbers behind the branching points are percentages from ML bootstrapping (left), Bayesian posterior probabilities (BPP, middle) and maximum parsimony bootstrap percentages (MP, right). Stars indicate that values are 100 (ML), 1.0 (BI) and 100 (MP). See Table 2.2 for accession numbers.

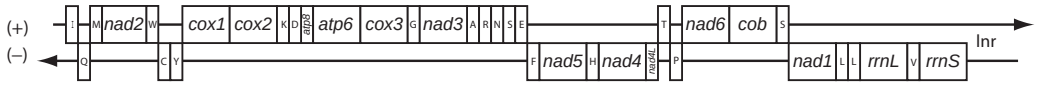
The phylogenetic analysis of mitochondrial genome data in chelicerate phylogenetics suffers from several problems. First, the currently incomplete taxon sampling (complete lack of data for Opiliones, Pseudoscorpiones, Palpigradi, Uropygi, and Amblypygi) highlights the preliminary nature of this analysis. While the problem of taxon sampling will be overcome in the near future, other problems lie in the nature of the sequence data itself. Several taxa show a reversed strand bias of nucleotide composition (Table 2.2), which complicates phylogenetic analyses (Hassanin et al. 2005; Hassanin 2006). This may strongly affect the phylogenetic position of Scorpiones in analyses with mitochondrial datsets (Jones et al. 2007). Another problem is the heterogeneous substitution rate among arachnids. Our analysis demonstrates great variability in branch lengths, with very long branches for some Acari species (*Leptotrombidium, Varroa, Metaseiulus*), and very short branches, e.g. from *Limulus* and *Oltacola*.

Mitochondrial gene order variation in Arachnida

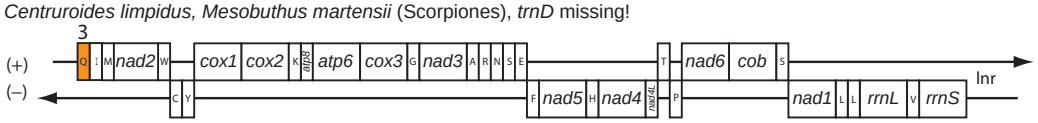
Apart from five translocated tRNA-genes, the mitochondrial gene order of *P. pearsei* does not differ from the putative arthropod ground pattern (Figure 2.5 and 2.6). A minimum of two events is required to lead to the derived gene order of *P. pearsei*: a tandem duplication / random deletion event and a single tRNA gene transposition. Tandem duplication and random loss of genes is widely accepted as a mode of genome shuffling in mitochondria (Moritz et al. 1987). In the present case a segment ranging from *trnL* to *trnY* (ground pattern) can be inferred to be involved in such a duplication event. Subsequently the first copy may have lost *trnL*, *trnV*, *nad1*, and *trnC*, the second copy may have lost *rrnL*, *rrnS*, *trnI*, *trnQ*, *trnM*, *trnW*, and *trnY,* as well as the control region. Besides the control region (between *rrnS* and *trnI*) there are four larger non-coding regions (87-183 bp) present between *trnI* and *nad2* (Figure 2.1, Table 2.1). This is a further hint towards a tandem duplication random deletion event, although there are no apparent sequence homologies to any of the lost genes. In addition, *trnN* is located in a novel relative position which is best explained by the transposition of this single gene, the second inferred event.

The comparison of mitochondrial gene order from Arachnida reveals a great variation (Figure 2.5). Only eight out of 23 species have retained the arthropod ground pattern as represented by *Limulus* (Staton et al. 1997). We present a parsimonious scenario of gene order changes and mapped these events on a phylogenetic tree (Figure 2.6). Three different modes leading to a change in gene order were assumed: (1) tandem duplication random deletion events (Moritz et al. 1987), (2) inversions, and (3) transposition of single tRNAs. Although the latter mode is currently not well explained by a molecular model, we have assumed that single tRNA transpositions have occurred in all cases in which tandem duplication and random deletion seems to be an implausible explanation, e.g. when only one tRNA gene

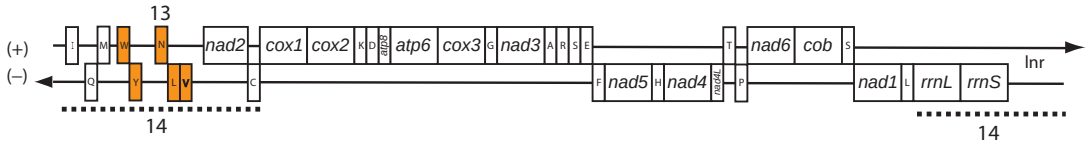
Arthropod ground pattern, retained in: *Limulus polyphemus* (Xiphosura), *Heptathela hangzhouensis* (Araneae), *Ixodes* (4 sp., Acari), *Ornithodoros* (2 sp., Acari), *Carios capensis* (Acari)



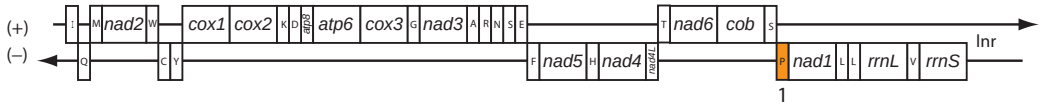
Centruroides limpidus, Mesobuthus martensii (Scorpiones), *trnD* missing!



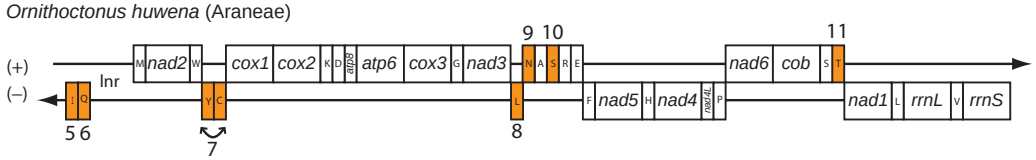
Pseudocellus pearsei (Ricinulei)



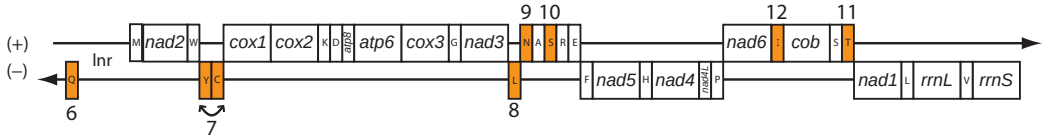
Oltacola gomezi (Solifugae)



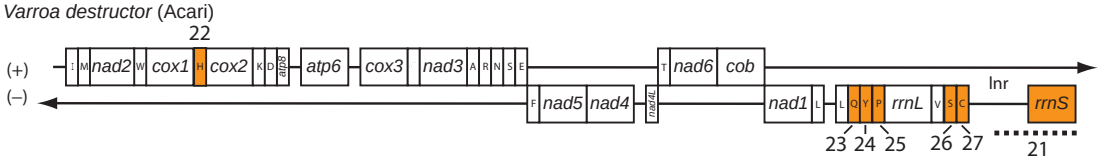
Ornithoctonus huwena (Araneae)



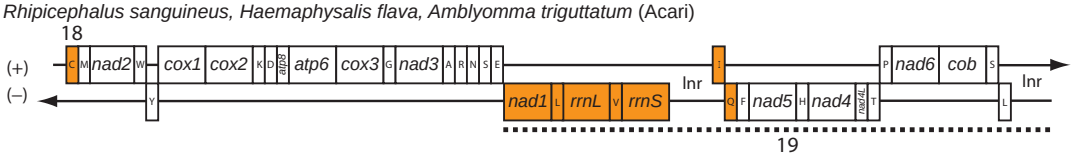
Habronattus oregonensis, Nephila clavata (Araneae)



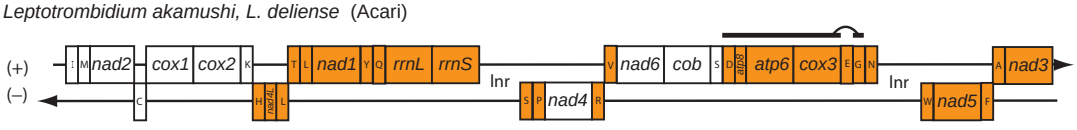
Varroa destructor (Acari)



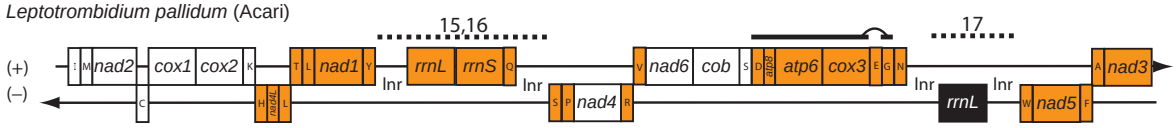
Rhipicephalus sanguineus, Haemaphysalis flava, Amblyomma triguttatum (Acari)



Leptotrombidium akamushi, L. deliense (Acari)



Leptotrombidium pallidum (Acari)



Metaseiulus occidentalis (Acari), *nad3*, *nad6* missing!

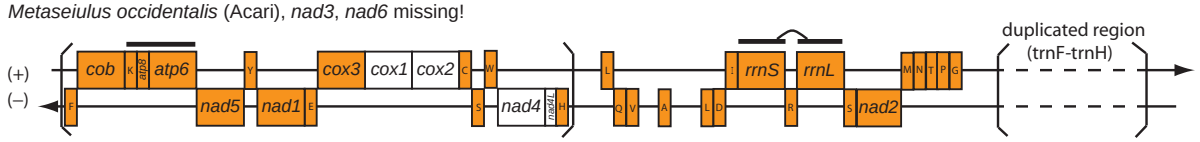


Figure 2.5: Changes in gene order in mitochondrial genomes of Arachnida compared to the putative ancestral arthropod gene order. Transfer RNA genes are labelled according to the one letter amino acid code. Genes marked white show the same relative position as in the arthropod ground pattern; genes marked orange have relative positions differing from the arthropod ground pattern; the gene marked black indicates a duplicated *rrnL* gene in *Leptotrombidium pallidum*. Horizontal lines above genes illustrate adjacent genes which were probably translocated together; dotted lines indicate regions where tandem duplication and random deletion events may have occurred; connected arrows show adjacent genes which have switched their position, making it difficult to assess which gene was translocated. Braces accentuate the duplicated regions in the mitochondrial genome of *Metaseiulus occidentalis*. lnr: large non-coding region, putative mitochondrial control region; other non-coding regions (> 50 bp) are illustrated by gaps between genes. Numbers refer to rearrangement events, compare Fig 2.6. For GenBank accession numbers see Table 2.2.

found a new position distant from its original position, or on the other strand. The tandem duplication random deletion model was proposed when more than one gene was involved and/or non-coding sequence was found between the genes involved. The tree we have used for mapping the events is the best scoring topology from the ML analysis of the amino acid alignment (Figure 2.4, lower tree), because this analysis provided bootstrap support for most of the branches. However, regarding gene order changes there is also no character conflict (= homoplastic characters / events) with the other topology obtained from the nucleotide alignment.

None of the gene order characters is capable to resolve interordinal relationships of the taxa studied (Figure 2.6). The arthropod ground pattern (Staton et al. 1997) is retained in some parasitiform mites (e.g. *Ixodes* (Black and Roehrdanz 1998), *Carios* and *Ornithodoros* (Shao et al. 2004)) (Figure 2.5, Figure 2.6), and in the spider *Heptathela* (Araneae: Mesothelae (Qiu et al. 2005)). The mitochondrial genomes of Araneae-Opisthothelae (*Ornithoctonus* (Qiu et al. 2005), *Nephila*, [GenBank:NC_008063], *Habronattus* (Masta and Boore 2004)) share seven translocated tRNA genes. A subsequent translocation of *trnI* is found in *Nephila* and *Habronattus* (Masta and Boore 2004). The camel spider *Oltacola gomezi* (Solifugae) has a mitochondrial genome arrangement almost unaltered from the arthropod ground pattern [GenBank:EU024482], only *trnP* was translocated to a new relative position. The scorpions *Centruroides* (Davila et al. 2005) and *Mesobuthus* (Choi et al. 2007) share a translocated and inverted *trnQ* and both lack *trnD*.

Several independently derived mitochondrial gene orders are found in Acari. Metastriata (Parasitiformes: Ixodidae) show a derived gene order probably caused by a tandem duplication random deletion event and an additional tRNA translocation (*Rhipicephalus* (Black and Roehrdanz 1998), *Haemaphysalis* (Shao et al. 2004), *Amblyomma*, [GenBank:NC_005963]). The honeybee mite *Varroa* shows translocation of six tRNA genes (Navajas et al. 2002). In addition the reversed arrangement of *rrnS* and the control region in this species is hypothesized to be the result of a tandem duplication and random deletion event.

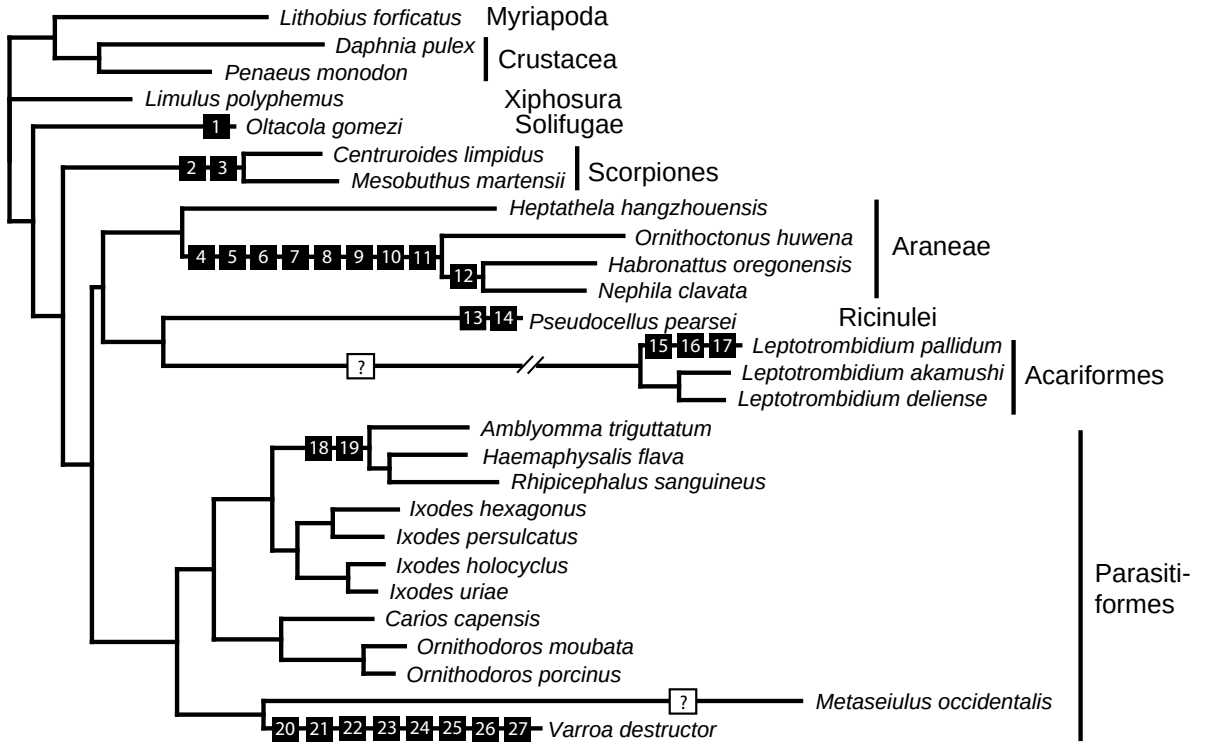


Figure 2.6: Hypothesized events leading to derived gene orders (compare Figure 2.5) and strand bias reversal (compare Table 2.2) both mapped as black squares on a topology based on the results of our phylogenetic analysis (compare Figure 2.4, amino acid data set). **1**: Translocation of *trnP*; **2**: Reversal in nucleotide bias; **3**: Translocation of *trnQ*; **4**: Reversal in nucleotide bias; **5**: Translocation of *trnI*; **6**: Translocation of *trnQ*; **7**: Translocation of *trnY* or *trnC*; **8**: Translocation of *trnL*; **9**: Translocation of *trnN*; **10**: Translocation of *trnS*; **11**: Translocation of *trnT*; **12**: Translocation of *trnI*; **13**: Translocation of *trnN*; **14**: Tandem duplication of a gene block ranging from *trnL* to *trnY* followed by random deletion; **15, 16**: Two subsequent tandem duplications of a gene block ranging from *trnQ* to the large non-coding region followed by random deletion; **17**: Duplication and inversion of *rrnL* and a large non-coding region, probably by nonhomologous intergenome recombination (15-17 according to (Shao et al. 2006)); **18**: Translocation of *trnC*; **19**: Tandem duplication of a gene block ranging from *trnF* to *trnQ* followed by random deletion; **20**: Reversal in nucleotide bias; **21**: Tandem duplication involving the putative control region and *rrnS* followed by random deletion; **22**: Translocation of *trnH*; **23**: Translocation of *trnQ*; **24**: Translocation of *trnY*; **25**: Translocation of *trnP*; **26**: Translocation of *trnS*; **27**: Translocation of *trnC*; "**?**": Probably due to multiple rearrangements, the history of events leading to the highly derived gene order of *Leptotrombidium* and *Metaseiulus* cannot reliably be reconstructed.

Metaseiulus (Parasitiformes: Mesostimata) possesses the most derived mitochondrial gene order among arachnids, probably due to multiple rearrangements (Jeyaprakash and Hoy 2007). This genome is the largest (25 kb) within the Chelicerata due to the presence of duplicate and triplicate regions. The duplicate region contains 18 genes plus a copy of the control region and partial *trnL*(UUR) sequence while the triplicate region comprises only the control region sequence and partial *trnL*(UUR) sequence (Figure 2.5). The mitochondrial

genome of *Metaseiulus* is also remarkable for the absence of *nad3* and *nad6,* because no other chelicerate is known to have lost any of the protein coding genes. Due to the magnitude of these changes, it is difficult to reliably reconstruct the events leading to the gene order found in *Metaseiulus* which is unique amongst arachnids.
The same is true for the genus *Leptotrombidium* (Acariformes) (Shao et al. 2005a; Shao et al. 2006). The three species share a common, derived arrangement with secondary rearrangements in *L. pallidum*. The secondarily derived gene order of *L. pallidum* (translocation of *trnQ*, duplication of *rrnL* and the presence of four large noncoding regions) is considered to have evolved by a combination of tandem duplication with random deletion plus intergenomic recombination between several genes with subsequent gene conversion (Shao et al. 2006). It is apparent that the two taxa showing the most complicated genome rearrangement history (*Leptotrombidium* and *Metaseiulus*) exhibit also the longest branches (=highest substitution rates) in phylogenetic analysis (Figure 2.4 and 2.6).

Reversals of strand bias in nucleotide frequency

A reversal of nucleotide strand bias was detected in six species (Table 2.2), most probably due to three independent events: (1) in the *Varroa* mite (Navajas et al. 2002), (2) in Scorpiones and (3) in Araneae-Opisthothelae (*Ornithoctonus* (Qiu et al. 2005), *Nephila*, [GenBank:NC_008063], *Habronattus* (Masta and Boore 2004)). These changes are also mapped on the tree in Figure 2.6 (characters 2, 4, and 20). The comparatively low substitution rate in the mitochondrial genome of *Heptathela* (Araneae: Mesothelae; Figure 2.6) argues against a re-reversal of nucleotide bias within Araneae and suggests that the reversal is in fact a synapomorphy of the Opisthothelae (Figure 2.6).

2.4 Conclusions

The first complete mitochondrial sequence of a hooded tickspider (*P. pearsei*, Ricinulei, Arachnida) reveals a typical circular duplex DNA molecule with a compact gene organisation as found in other bilaterians. In comparison to the putative arthropod ground pattern we observed a derived gene order with five tRNA genes found in different relative positions compared to the gene order of *Limulus*. Probably two events led to the derived gene order of *Pseudocellus*: (1) a duplication and random deletion event may be responsible for the translocation of four of these tRNA genes and (2) an additional translocation of the *trnN*. The putative mitochondrial control region is situated in the same position as in other arthropod

mitochondrial genomes. Part of the putative control region can be folded into a characteristic stem-loop structure with conserved flanking sequences as found in other arthropods. All tRNAs, except tRNA-Glu, can be folded into a typical cloverleaf secondary structure.
Aligments of nucleotide and amino acid sequences from mitochondrial protein-coding genes were used in a phylogenetic analysis of arachnid relationships. In the best scoring topologies from ML analyses, Ricinulei appear as sister taxon to Acari (nucleotide alignment) or Acariformes (amino acid alignment), but in both cases without sufficient support from bootstrapping or Bayesian inference. Within the Acari, monophyly of Parasitiformes is well supported, while Acariformes are represented by only a single genus (*Leptotrombidium*). Monophyly of Acari was only recovered in the analysis of the nucleotide alignment and found strong support by BI and only weak support by ML/MP bootstrapping.
Because of highly divergent substitution rates amongst arachnid species, phylogenetic analyses may be generally biased due to long-branch attraction. Another complicating factor is the reversal of nucleotide bias, which has occurred independently in the arachnid clades Scorpiones, Acari (*Varroa*) and Araneae (Opisthothelae). Due to the lack of support for interordinal relations within Arachnida, the phylogenetic analysis of mitochondrial genome data reveals no strong conflict with recent morphological analyses.
Investigation of mitochondrial gene rearrangements across the range of taxa for which complete mitochondrial genomes are available on GenBank, reveals high variability in gene order within arachnids. Among arachnids, 15 of the 23 species investigated have derived features in gene order. When mapped onto the tree, the hypotheses of events leading to gene order changes are not in conflict with our phylogenetic analyses of sequence data. None of the hypothesized events is useful as a phylogenetic character to resolve interordinal relationships within arachnids, but some of these characters are promising in analyses of intraordinal relationships. Especially in Acari the comparison of mitochondrial genomes from a larger taxon sampling would be promising for the detection of gene order changes, which will be valuable in a phylogenetic analysis. Altogether, further mitogenomic data of a broader taxon sampling are necessary, especially from Opiliones, Pseudoscorpiones, Palpigradi, Uropygi and Amblypygi.

2.5 Methods

Samples and DNA extraction

P. pearsei specimens were collected by Gerd Alberti in the Gruta Sabac-Ha, near Merida, province of Yucatán, México (20°10’18”N, 89°16’03”W). DNA extraction was done with two legs obtained from a freshly killed animal. We used “DNeasy blood and tissue kit” (Qiagen, Hilden, Germany) according to the manufacturers protocol, except for reduction of the elution volume to 50 µl.

PCR and sequencing

Initial PCRs were done with a published primer set designed for crustacean mitochondrial genomes (Yamauchi et al. 2004). All PCRs were performed on an Eppendorf Mastercycler and Mastercycler gradient. The Eppendorf 5-prime-Taq kit (Eppendorf, Germany) was used in 50 µl volumes (5 µl buffer; 1 µl dNTP mix, 10 µM; 0.25 µl Taq polymerase; 1 µl template DNA, 40,75 µl water, 1 µl primer mix, 10 µM each). PCR conditions were: initial denaturation (94°C, 1 min), 40 cycles of denaturation (94°C, 30 sec), annealing (50°C, 30 sec), and elongation (68°C, 1 min), followed by a final elongation step (68°C, 1 min). Successful PCR amplification was obtained with six primer pairs: S8(*cox1*, about 650 bp), S10(*cox1*, about 1050 bp), S15(*cox2-atp6*, about 830 bp), S20(*cox3-trnF*, about 950 bp), S37(*nd4l-cob*, about 1600 bp), and S46(*rrnL-rrnS*, about 1000 bp), see (Yamauchi et al. 2004) for primer sequences. PCR products were visualized on 1% agarose gels. Before sequencing PCR products were purified using the Bluematrix DNA purification kit (EURx, Gdansk, Poland). Except for S8 and S10 there was no overlap between these initial sequences. Species specific primer pairs were designed to bridge the five gaps between these initial sequences (see Table 2.3 for primer information). Long PCRs were performed with Takara LA Taq kit (Takara) in 50 µl volumes (5 µl Buffer, Mg+; 8 µl dNTP mix, 2.5 µM; 1 µl DNA; 1 µl primer mix, 10 µM each; 0.5 µl Takara LA Taq; 32 µl water). PCR conditions were: initial denaturation (94°C, 1 min), 40 cycles of denaturation (94°C, 30 sec), annealing (primer specific temperature see Table 2.3, 1 min), and elongation (68°C, 3 min), followed by a final elongation step (68°C, 2 min). Long PCR products were sequenced by primer walking. All sequencing was performed on a CEQ 8000 capillary sequencer (Beckmann-Coulter) using CEQ DCTS kits (Beckmann-Coulter) with 10 µl reaction volumes (4 µl DCTS master mix, 1 µl primer, 10 µM, 1-5 µl DNA, 0-4 µl water). Sequencing reaction was done with 30 cycles of denaturation (94°C, 30 sec), annealing (primer specific temperature, 30 sec) and elongation (60°C, 2 min).

Table 2.3: Species specific PCR primer pairs used for amplification of fragments from the mitochondrial genome of *Pseudocellus pearsei*.

Primer name	***Nucleotide sequence (5'-3')***	***Annealing temp.***
Rici-12s-co1	CCACATTACAACATAGTAACTCATTTC	56°C
Rici-co1-12s	AAGTTCACCCTGTTCCTGCTC	56°C
Rici-co1-co2	ATTACGTTGTAGCACACTTCCAC	56°C
Rici-co2-co1	AGATGGTAATGTTAGTAATATTTGGTG	56°C
Rici-co2-co3	GTAACATCAACTCTAGCACTAACAC	50°C
Rici-co3-co2	GTGTCGTGGAAATTGGGA	50°C
Rici-co3-cob	CTATCAATCTAACCCAACAAAAAAG	54°C
Rici-cob-co3	GTCAGAAGATAGTTTATTGGAATATTGGC	54°C
Rici-cob-12s	CCACCATTAACACCCAAAGCC	56°C
Rici-12s-cob	TTGGATTTAATAGTAAGGAAGTATTAGATAGG	56°C

Sequence assemblage and annotation

Primary sequence analysis was performed with the CEQ software (quality check). Sequence assemblage was done with Bioedit 7.0.1 (Hall 1999). Protein-coding and ribosomal RNA genes and gene boundaries were identified by BLAST search and in comparison with alignments from other chelicerate species. Genomic position and secondary structure from 20 out of 22 transfer RNAs were identified by tRNA-scan SE (Lowe and Eddy 1997), the remaining two by eye inspection of the regions under suspect. Sequence data was deposited at NCBI database [GenBank:EU024483].

Phylogenetic analysis

Alignments from all protein-coding genes were used in phylogenetic analysis. Alignments of amino acid sequences and nucleotide sequences were performed with ClustalW (Thompson et al. 1994), as implemented in Bioedit 7.0.1 (Hall 1999). Nucleotide sequences were translated to amino acid sequence prior to the multiple alignment and translated back afterwards (making use of the "toggle" function of Bioedit). Sequence data was obtained from 24 chelicerate species and three outgroup taxa (for a list of taxa with GenBank accession numbers see Table 2.2). Ambiguously aligned regions were omitted using Gblocks ver. 0.91b (Castresana 2000), using default settings, except for changing "allowed gap positions" to "with half". The final amino acid alignment consisted of 2786 amino acids. For the nucleotide dataset the "codons" option was used, so that only complete codons (not single nucleotides) were omitted from the alignment by Gblocks. A saturation analysis (Xia et al. 2003) was performed with DAMBE 4.2.13 (Xia and Xie 2001) using subsets of the nucleotide dataset representing first, second and third codon positions. Due to significant saturation, third codon positions were omitted from the final nucleotide alignment (6490 nucleotides).

Maximum parsimony (MP) analysis was performed with PAUP* ver. 4.0b10 (Swofford 1993), 1,000 bootstrap replicates were performed, each with 10 replicates with random addition of taxa. Maximum likelihood (ML) analysis was done with Treefinder (Jobb 2007). Model selection was done with ProtTest ver. 1.3 (Abascal et al. 2005). According to the Akaike information criterion, the mtArt+G+I model was optimal for ML analysis. The tree with the best likelihood value was chosen out of 10 independent analyses with random taxon order. Bootstrapping was performed with 100 pseudoreplicates. Bayesian inference (BI) with MrBayes ver. 3.1.2 (Huelsenbeck and Ronquist 2001) was done with the model mtRev+G+I, as the mtArt model is not implemented in the current version of MrBayes. Eight chains ran for 1,000,000 generations, while tree sampling was done every 1,000 generations; we examined likelihood values of the sampled trees and decided to omit the first 100 trees as burn-in, while the remaining 900 were used to calculate Bayesian posterior probabilities (BPP).

2.6 Abbreviations

A, adenine; *atp6* and *8*, genes encoding ATPase subunit 6 and 8; BI, Bayesian inference; bp, base pairs; BPP, Bayesian posterior probability; *cox1-3*, genes encoding cytochrome oxidase subunits I-III; *cob*, gene encoding cytochrome b; C, cytosine; G, guanine; ML, maximum likelihood; mtDNA, mitochondrial DNA; *nad1-6* and *nad4L*, genes encoding NADH dehydrogenase subunits 1-6 and 4L; PCR, polymerase chain reaction; rDNA, ribosomal DNA; rRNA, ribosomal RNA; *rrnL*, large (16S) rRNA subunit (gene); *rrnS*, small (12S) rRNA subunit (gene); T, thymine; TDRD, tandem duplication and random deletion; tRNA-Xxx (where Xxx is replaced by three letter amino acid code of the corresponding amino acid), transfer RNA; *trnX* (where *X* is replaced by one letter amino acid code of the corresponding amino acid), tRNA gene; TψC-arm, T-loop and T-stem of a tRNA secondary structure.

2.7 Authors` contributions

GT sampled and determined animals and isolated DNA. KF, GT and LP performed PCRs, KF and LP did all sequencing, sequence analysis and annotation. KF, AB and LP performed phylogenetic analyses. KF and LP wrote the manuscript.

2.8 Acknowledgements

The authors thank Thomas Bartolomaeus (Freie Universität Berlin, Germany) for his kind support through all stages of this work. We are also grateful to Gerd Alberti (Universität Greifswald), Gerhard Scholtz (Humboldt-Universität Berlin), and J.G. Palacios-Vargas (México) for further support. GT wishes to express his thanks to Tanja Müther and Thomas Schierl for their kind hospitality. We also thank three anonymous reviewers for improvement of the manuscript. This study was supported by German Research Foundation (DFG) grants Ba 1520/10-1 (to LP) and DFG Scho 442/8-1 (to AB), both from Priority Program 1174 (Deep Metazoan Phylogeny), and Al 138/13-1 (to GT).

3. The first complete mitochondrial genome sequences of Amblypygi (Chelicerata: Arachnida) reveal conservation of the ancestral arthropod gene order

3.1 Abstract

Amblypygi (whip spiders) are terrestrial chelicerates inhabiting the subtropics and tropics. In morphological and rRNA based phylogenetic analyses, Amblypygi cluster with Uropygi (whip scorpions) and Araneae (spiders) to form the taxon Tetrapulmonata, but there is controversy regarding the interrelationship of these three taxa. Mitochondrial genomes provide an additional large dataset of phylogenetic information (sequences, gene order, RNA secondary structure), but in arachnids mitochondrial genome data is missing for some of the major orders. In the course of an ongoing project concerning arachnid mitochondrial genomics we present the first two complete mitochondrial genomes from Amblypygi. Both genomes were found to be typical circular duplex DNA molecules with all 37 genes usually present in bilaterian mitochondrial genomes. In both species, gene order is identical to that of *Limulus polyphemus* (Xiphosura), which is assumed to reflect the putative arthropod ground pattern. All tRNA gene sequences have the potential to fold into structures that are typical of metazoan mt tRNAs, except for tRNA-Ala, which lacks the D-arm in both amblypygids, suggesting the loss of this feature early in amblypygid evolution. Phylogenetic analysis resulted in weak support for Uropygi being the sistergroup of Amblypygi.

3.2 Introduction

Whip spiders (Amblypygi) are a small order of terrestrial chelicerates, common in humid regions of the tropics and subtropics all over the world, with some species also occurring in more temperate to arid regions (Weygoldt 2000). Taxonomically they are divided into two suborders: the Paleoamblypygi are represented by a single extant species, the small (7 mm) and blind *Paracharon caecus* (Hansen 1921), and the larger (6-36 mm) Euamblypygi (Weygoldt 1996), comprising of four families, 16 genera and at least 157 described species (Harvey 2002, 2003, 2007). Amblypygi are bizarre animals due to their strong and spinous raptorial pedipalps and their thin and multi-segmented first walking legs, which serve as sensory and communicatory organs. Thus amblypygids show functional hexapody. Phylogenetically amblypygids are well characterized as a monophylum by various apomorphies from morphology: a pretarsal depressor muscle without a patella head, a vestigial labrum, large anterior coxal apodemes on all walking legs, divided tibiae, and nearly immovable patellotibial joints due to the fusion of these two segments (Shultz 1990). Based on morphological characters, most arachnologists agree there is likely a close phylogenetic relationship between Amblypygi, Araneae and Uropygi (e.g. Weygoldt and Paulus 1979; Van der Hammen 1989; Shultz 1989, 1990). This group (Tetrapulmonata, resp. Megoperculata sensu Weygoldt and Paulus 1979) is supported by the existence of two-segmented chelicerae hinged ventrolaterally and an unusual microtubule arrangement in their sperm axonemes. The phylogenetic relationships among these three taxa is more controversial, with two major competing hypotheses. Many authors favour Uropygi as sister group to Amblypygi ("Pedipalpi" hypothesis), due to the presence of raptorial pedipalps and antenniform first walking legs in both taxa (Shear et al. 1987; Shultz 1989, 1990, 1999, 2007). In contrast, a sister group relationship between Amblypygi and Araneae ("Labellata" hypothesis) is recognized by other authors (Petrunkevitch 1955; Weygoldt and Paulus 1979; Van der Hammen 1989), with support provided by a postcerebral pharynx and a pedicel in both taxa (Ax 1996).

The difficulties in evaluating phylogenetic relationships within the Tetrapulmonata based on morphological data are probably caused by homoplasy or reduction of anatomical characters. Controversial results from nuclear sequence data and from combined analyses (Wheeler and Hayashi 1998; Giribet 2002) hint to the need for additional data sets for phylogenetic reconstructions, such as mitochondrial genomes (mt genomes). In animals these circular double-stranded DNA molecules are about 16 kb long and contain 37 genes plus one AT-rich non-coding region (Wolstenholme 1992; Boore 1999). In this publication, we provide the first two complete mitochondrial genome sequences covering two families of the Amblypygi, *Damon diadema* (Phrynichidae) and *Phrynus* sp. (Phrynidae). We discuss general features of the genomes, compare inferred secondary structures of tRNAs and rRNAs, nucleotide frequency bias and codon usage and provide a phylogenetic analysis of arachnid interrelationships.

3.3 Materials and Methods

Animals

A specimen of *D. diadema* (Simon 1876) was obtained from a commercial source. Species determination was done with the morphological key and in comparison to *cox1* sequences according to Prendini et al. (2005). A gift from M. Hedin of a specimen of *Phyrnus sp.* that was collected in Baja California, Mexico, was identified using the key provided in Weygoldt (2000). Total DNA was extracted from one leg by using Qiagen extraction kits (Qiagen, Hilden, Germany) following the manufacturers` protocol.

PCR

The whole mt genome of *D. diadema* was amplified in two overlapping fragments by using the primer pairs Art-HPK16SA/B (Kambhampati and Smith 1995; Simon et al. 1994) and Art-HPK16Saa/bb (Hwang et al. 2001b). Long-range PCR with primers Art-HPK16Saa/bb was performed with Takara LA Taq kit (Takara) in 50 µl volumes (5µl buffer, 8 µl dNTP solution, 0.5 µl Takara LA Taq, 1µl DNA, 1µl primer mix (10 µM), 34.5 µl water). This yielded a PCR fragment of about 15 kb size. Conserved primers for crustaceans (Yamauchi et al. 2004) were used to amplify smaller mitochondrial fragments from the long PCR product. Successful amplification was performed with primer pairs S1, S2, S5, S7-S11, S13, S15, S24, S25, S29, S30, S35, S36, S42, S46 and S48. Finally, additional primer pairs were designed to amplify larger fragments in order to bridge the gaps between S13/S15, S15/S24, S25/S29, S30/S35, S35/S42 and S48/S2 (for primer sequences and annealing temperatures see Supplementary file 3.1). Secondary PCRs were performed on an Eppendorf Mastercycler and Mastercycler gradient using the Eppendorf 5-prime-Taq kit (Eppendorf, Germany) in 50 µl volumes (41,75 µl molecular grade water; 5 µl buffer; 1 µl dNTP mix, 10 µM; 1 µl template DNA (= 1:100 dilution of the long PCR fragment); 1 µl primer mix, 10 µM each; 0,25 µl Taq polymerase). PCR products were visualized on 1% agarose gels and purified using the Bluematrix DNA purification kit (EURx, Gdansk, Poland). If extra bands were present, a Gel extraction was performed, following the manufacturers` protocol (Qiagen). The mt genome sequence of *Phrynus* was amplified with taxon-specific primers (see Supplementary file 3.1), which were designed based upon a region of the *cob* gene that was amplified with the primers CobF and CobR (Boore and Brown 2000). For details on long-PCR amplification, sequencing, and sequence assembly see Masta and Boore (2008).

Sequencing and genome assemblage

Sequencing of the *D. diadema* was performed on a CEQ 8000 capillary sequencer using the CEQ DCTS kit (both Beckmann-Coulter). Sequencing reactions were performed in Eppendorf Mastercycler and Mastercycler gradient. The quality of the sequences was checked with CEQ software. Sequence assembly was performed with BioEdit 7.0.1. (Hall 1999). Protein-coding and ribosomal genes were identified by BLAST searches on NCBI databases. To determine boundaries the sequences were also compared with alignments from other chelicerate species. We assumed the start and ending of the ribosomal RNA genes and the control regions to be extended to the boundaries of flanking genes. The boundary of the 12S rRNA gene to the control region was inferred by comparison with 12S rRNA genes of other arachnids. Hairpin structures in the control regions were identified by eye inspection. Genomic position and secondary structure of tRNAs were identified using tRNAscan-SE (Lowe and Eddy 1997) and ARWEN (Laslett and Canbäck 2008). Secondary structures of the rRNAs were made in comparison to published structure models of the honeybee *Apis mellifera* (Gillespie et al. 2006). Complete mt genome sequences were deposited at NCBI database [Genbank: FJ204233 (*Damon diadema*), EU520641 (*Phrynus* sp.)]. Nucleotide frequency and relative synonymous codon usage were determined using DAMBE ver. 4.2.13 (Xia and Xie 2001).

Phylogenetic analysis

Phylogenetic analysis was performed with concatenated amino acid alignments of 11 protein-coding genes (omitting the shortest and least conserved genes *atp8* and *nad4L*). Sequences were retrieved from the Mitome database (www.mitome.inf; Lee et al. 2008). Alignments were done with ClustalW (Chenna et al. 2003) under default conditions and in some cases corrected after inspection by eye. Phylogenetic analysis of this dataset was performed in two ways: (1) Maximum likelihood analysis with RaxML 7.0 (Stamatakis et al. 2008) was performed using the mtREV+G+I model, with data partitions according to the eleven genes. In addition to a search for the best tree, 100 bootstrap replicates were performed, making use of the CIPRES Portal web server (www.phylo.org/sub_sections/portal/). (2) Maximum likelihood analysis with Treefinder - version Oct. 2008 (Jobb et al. 2004) was performed using the mtART+G+I model with individual optimization of data partitions (according to the eleven genes). Edge support (=an approximation of bootstrapping) was computed with 1000 replicates.

3.4 Results and discussion

Genome organization and gene order

Both mt genomes have a typical circular organization with a length of 14764 bp in *Phrynus* sp. and 14786 bp in *D. diadema*. For both genomes, we identified all 37 genes usually present in bilaterians: 13 protein-coding genes, two genes for rRNA subunits and 22 tRNA genes (Figure 3.1, Supplementary files 3.2 and 3.3). The gene order in both species is identical to that found in the mt genome of the horseshoe crab, *Limulus polyphemus* (Xiphosura), which is thought to represent the putative arthropod ground pattern (Staton et al. 1997; Lavrov et al. 2000). Between *rrnS* and *trnI* we detected one major non-coding region in the mt genome of both species (see next section). This region has a higher A+T content than the rest of the genome. In vertebrates, a similar region bears signal sequences that initiate H-strand synthesis in replication, as well as H- and L-strand transcription (Clayton 1991) and thus is referred to as the mitochondrial control region. Beside the putative control region only much smaller non-coding sequences occur, extending up to 21 bp (between *rrnL* and *trnV*) in *Phrynus* sp. and 23 bp (between *trnS*-UGA and *nad1*) in *D. diadema* (see Supplementary files 3.2 and 3.3). Both mt genomes show overlap at 11 gene boundaries with up to 20 shared nucleotides (between *nad4* and *nad4L*) in *Phrynus* sp. and up to 15 nucleotides (between *trnH* and *nad4*) in *D. diadema* (Figure 3.1, Supplementary files 3.2 and 3.3).

Control region

In both amblypygid mt genomes the largest non-coding sequence (364 bp in *D. diadema* and 359 bp in *Phrynus* sp.) is located between *rrnS* and *trnI*. The location of the putative control region in invertebrates shows great variability, but in many arthropods showing a conserved mitochondrial gene order it has the same relative location as in Amblypygi (Wolstenholme 1992; Zhang and Hewitt 1997; Saito et al. 2005). Part of the putative control region can be folded into stable stem-loop structures due to inverted repeat sequences. In *D. diadema* the stem is composed of 23 paired nucleotides without any mismatches and the loop consists of 15 nucleotides (Figure 3.1). The stem-loop formation in the control region of *Phrynus* sp. shows a larger loop consisting of 35 nucleotides, the stem consists of 22 paired nucleotides (one mismatch and a side loop). While their exact function is unclear, inverted repeat sequences often occur in arthropod mitochondrial control regions (Kilpert and Podsiadlowski 2006; Fahrein et al. 2007; Masta et al. 2008) and are probably located at or near the replication origin of the L-strand (Zhang and Hewitt 1997). Conserved motifs in the flanking sequences around the stem-loop structure are also found in the two amblypygids: both species exhibit a TATA motif in the 5′-flanking sequence whereas a $G(A)_nT$ motif only appears in the 3′-flanking

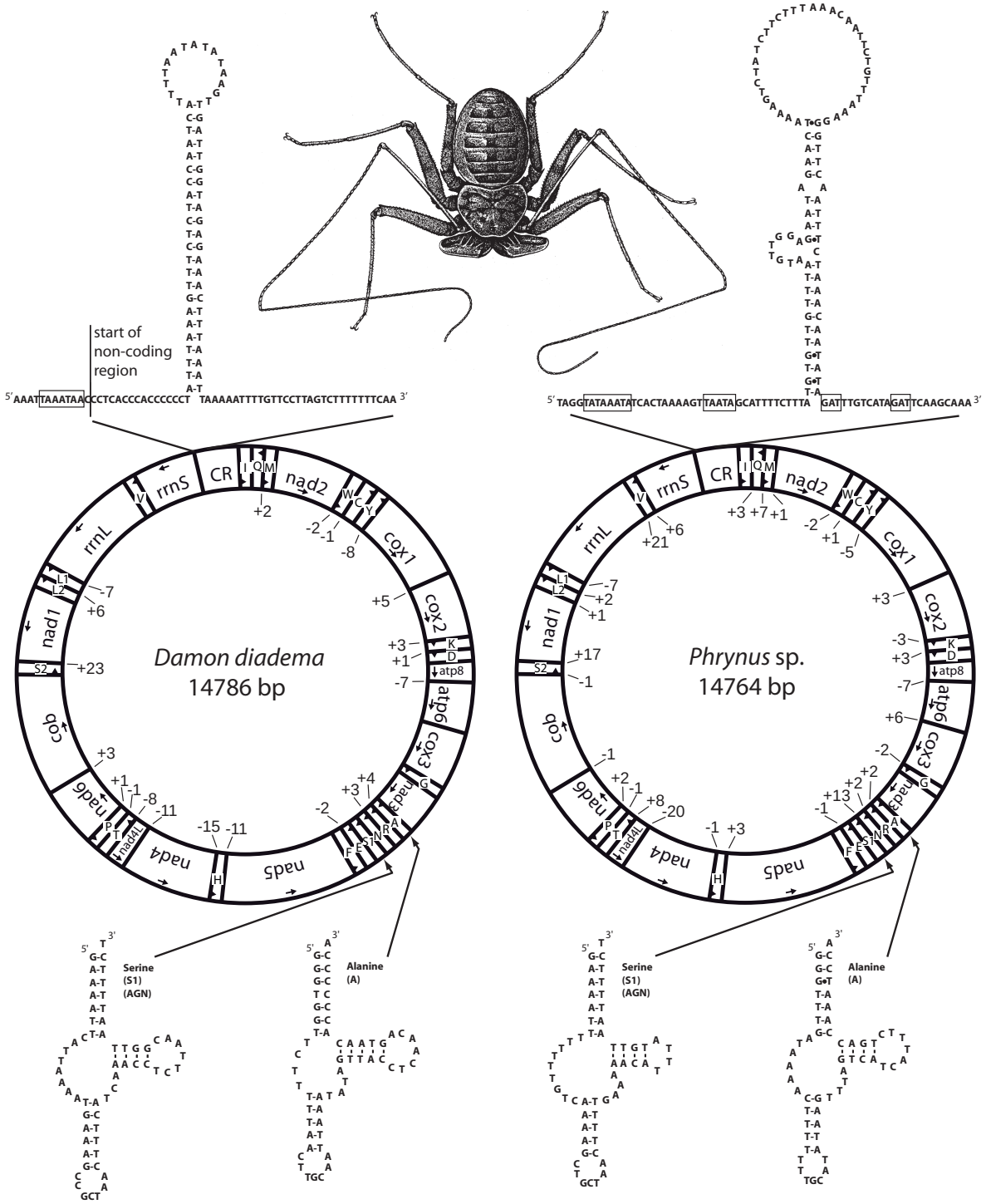


Figure 3.1: Mitochondrial genomes and control region loop secondary structure of Amblypygi. tRNA genes are depicted by their one-letter code abbreviations. Numbers reflect non-coding (positive) or overlapping (negative) nucleotides between two adjacent genes; small arrows indicate the orientation of the genes. Above the circular maps stem-loop structures found in the largest non-coding part (= the putative control region, CR) are shown. Small boxes in sequence highlight putative signal motifs. Below the circular maps plots of inferred secondary structures of tRNA-Ala and tRNA-Ser(AGN) are shown. These are the only tRNAs in Amblypygi lacking the DHU stem. For all *Phrynus* sp. tRNAs see Masta and Boore (2008), for all *Damon diadema* tRNAs see Supplementary file 3.4. Dots illustrate the base pairing of the pyrimidine base uracil with the purine base guanine. The depicted animal is *Phrynus* sp.

sequence of *Phrynus* sp. (Figure 3.1). The latter motif also occurs in the origin of L-strand replication in vertebrates and the plant *Petunia hybrida* (Zhang et al. 1995). Both motifs are presumed to play an important role in the initiation of transcription and/or replication of the mitochondrial genome (Zhang et al. 1995; Zhang and Hewitt 1997).

Secondary structure of tRNAs

In the mt genomes of both amblypygids all 22 tRNAs typical for bilaterian mt genomes were found (see Supplementary file 3.4). As in many other animals the inferred secondary structure of tRNA-Ser(AGN) lacks a DHU-stem (Wolstenholme 1992). While most of the other tRNA genes have inferred secondary structures with a canonical cloverleaf shape, in both, *Phrynus* (Masta and Boore 2008) and *Damon* (Figure 3.1), tRNA-Ala lacks the DHU-stem. This loss is not shared by other chelicerates, but instead is probably a synapomorphy shared by members of Amblypygi.

Ribosomal RNA genes

Both ribosomal RNA genes of Amblypygi have lengths similar to those found in the xiphosuran *Limulus polyphemus*. The inferred lengths of the *rrnL* genes in these two amblypigid taxa are 1211 nt for *Phrynus* and 1294 nt for *Damon*, while *Limulus* possesses a *rrnL* gene of 1296 nt (Lavrov et al. 2000). In contrast, the jumping spider *Habronattus oregonensis* has a considerably shorter *rrnL* gene, with a length of only 1018 nt (Masta 2000). Likewise, the inferred sizes for the *rrnS* genes are similar in the two amblypygids (786 nt in *Phrynus* and 750 nt in *Damon*) and in *Limulus* (799 nt) (Lavrov et al. 2000). The *rrnS* gene in *Habronattus* is much shorter, with a length of only 693 nt (Masta and Boore 2004). Together, these data suggest that the RNA components of the ribosome became reduced in size in spiders, after their divergence from their common ancestor with amblypygids and *Limulus*. Secondary structures of 12S and 16S rRNA were inferred for *Damon diadema* (Figures 3.2 and 3.3). There is much more similarity to the secondary structures proposed for the insect *Apis mellifera* (Gillespie et al. 2006) than to the other published rRNA secondary structure analyses of mites, the chigger mite *Leptotrombidium pallidum* (Shao et al. 2006), and the oribatid mite *Steganacarus magnus* (Domes et al. 2008). The differences in length between *Damon* and the mites are primarily due differences in the number of nucleotides that comprise certain helices. In particular, the 5`-end of *rrnL* appears to be truncated in mites such that they have lost helices H183, H235, H461, whereas these regions are present in *Apis* and amblypygids. Other differences in size can be attributed to the fact that the helices between H2043 and H2455, which are flanked by highly conserved sequence motifs in all four species, are

substantially shorter in mites (115 nt in *Steganacarus* and 114 nt in *Leptotrombidium*) than they are in *Damon* (196 nt) and *Apis* (187 nt). Likewise, helix H837 of *rrnL* is shorter in the two mites (36 nt in *Steganacarus* and 31 nt in *Leptotrombidium*) than in *Damon* (55 nt), or in *Apis (*52 nt). The *rrnS* gene also exhibits noticeable length differences between Damon and the two mites. This is most apparent in the 5`-end and in the helices between H769 and H885. Together, this suggests that amblypygids have retained a ribosome structure that is typical for arthropods, but that secondary reductions in helix sizes have occurred in mites and spiders. Further comparative work is necessary to more fully understand the evolution of structural differences among arachnids.

The termination signal for lrRNA transcription in animal mitochondria is believed to be the motif TGGCAGA (Valverde et al. 1994). In insects and crustaceans, this heptamer is located downstream of the lrRNA gene in the tRNA-Leu(CUN) gene, while in vertebrates it is found within the downstream tRNA-Leu(UUR) gene (Valverde et al. 1994). We find this same motif downstream from the lrRNA genes in both of these amblypygid mt genomes. However, in both amblypygids an identical motif is located in the tRNA-Leu(UUR) gene, while in Damon, the motif is also present in the tRNA-Leu(CUN) gene. Phrynus possesses a modification of this motif, TGACAGA, in its tRNA-Leu(CUN) gene. We find that amblypygids, *Limulus*, and *Habronattus* (Masta 2000) share this same gene location whereby the lrRNA termination signal is located within the tRNA-Leu(CUN) gene. It is known that the gene in which the rRNA transcription termination motif is located has changed over evolutionary time (Valverde et al. 1994), but we conclude that in amblypygids, the location of the rRNA transcription termination motif represents the ancestral location for chelicerates.

There are alternative models for *rrnL* termination signals (Cameron and Whiting 2008) but in our case we favour the above mentioned model (Valverde et al. 1994) as being more parsimonious due to the highly conserved motif, its location closer to the boundary *rrnL/ trnL*(UUR) and a comparable situation in other chelicerates.

Protein-coding genes

Except for one gene, in both amblypygids all of the 13 identified protein-coding genes start with one of the usual start codons for arthropod mt DNA (ATA, ATC, ATG, ATT). Only *cox1* from *Phrynus* sp. starts with the exceptional codon TTA (see Supplementary file 3.3). Most genes possess the stop codon TAA, except for the single occurrence of the stop codon TAG, which terminates *nad4L* (see Supplementary file 3.2 and 3.3). Truncated stop codons consisting only of a T or a TA are observed in three genes from *Phrynus* sp. and five genes from *D. diadema*. Such incomplete stop codons are also reported from the mt genomes of many animal species and it is assumed that these partial stop codons are completed by post-transcriptional polyadenylation (Ojala 1981).

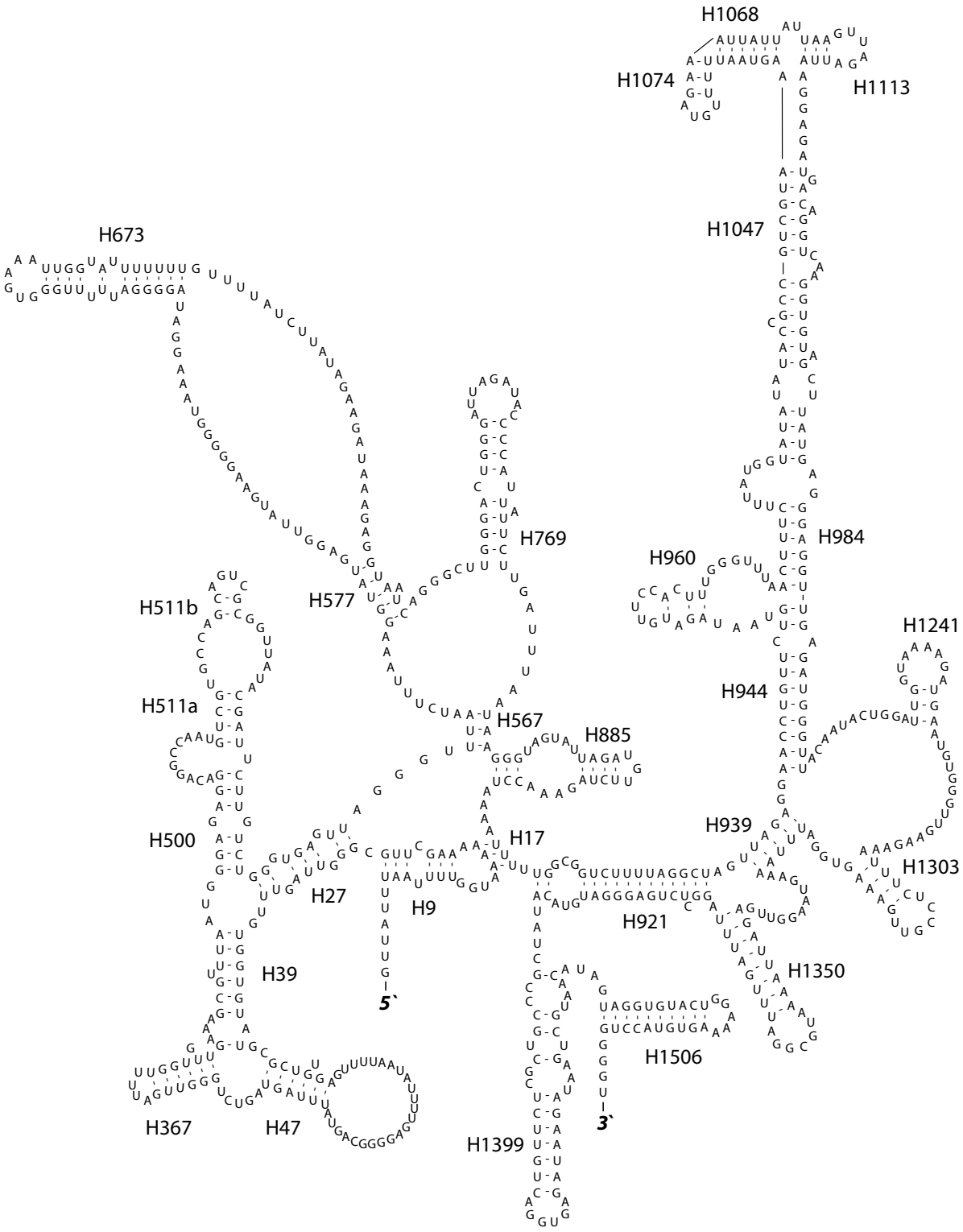


Figure 3.2: Plot of inferred secondary structure of SSU rRNA (12S) from the mitochondrial genome of *Damon diadema*. Helix numbering according to Gillespie et al. (2006).

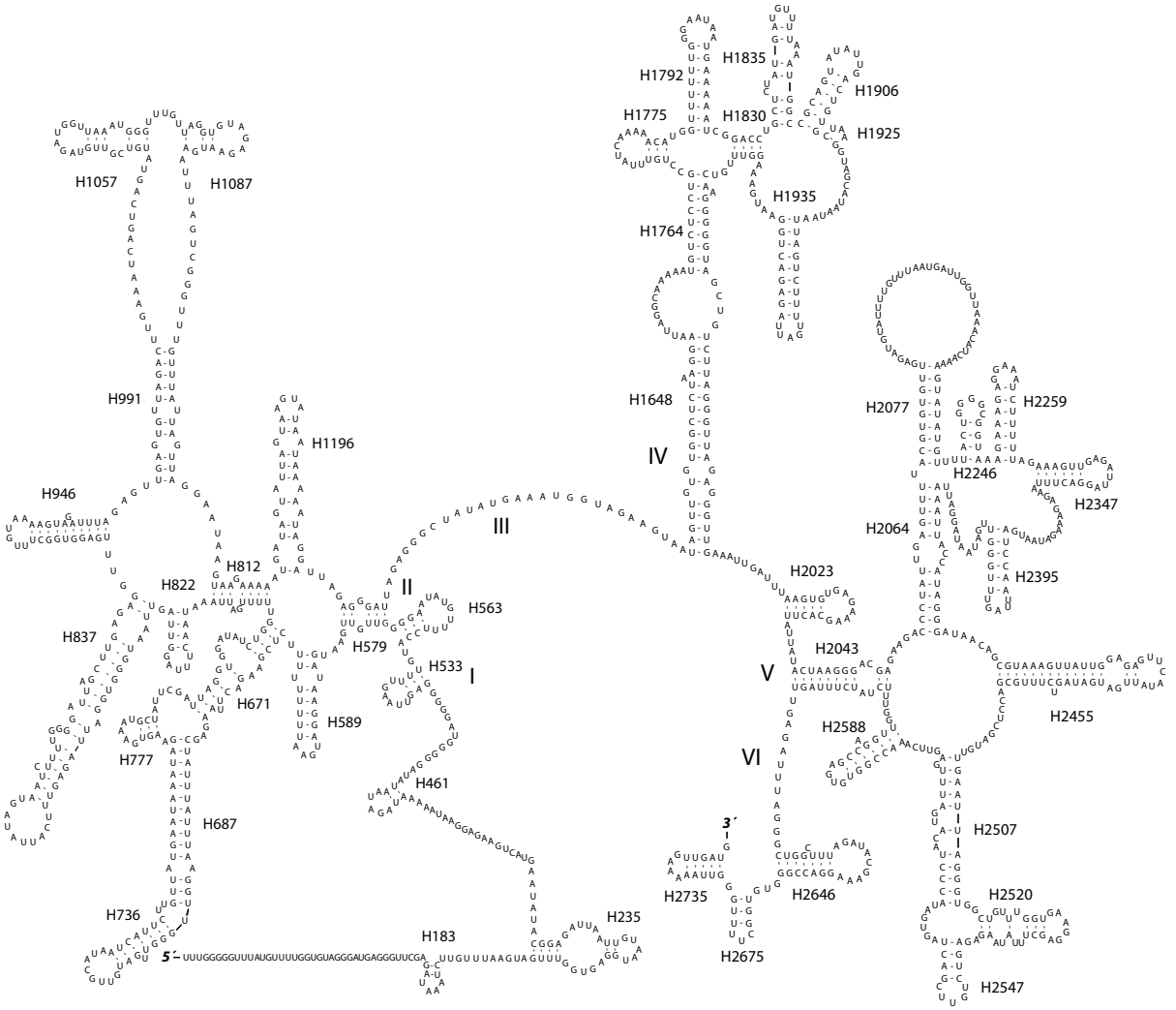


Figure 3.3: Plot of inferred secondary structure of LSU rRNA (16S) from the mitochondrial genome of *Damon diadema*. Helix numbering according to Gillespie et al. (2006).

Nucleotide frequency and codon usage

In both species, genes coded on the plus strand have a negative GC skew, while genes coded on the minus strand have a positive GC skew (see Supplementary files 3.2 and 3.3). We compared nucleotide frequencies and strand bias among species from Amblypygi, Uropygi, and Araneae (Table 3.1). Except for *Heptathela*, all spiders have a reversed GC skew compared to Amblypygi and Uropygi. AT skew is not so - *Damon* is the only species with a slightly positive AT skew (0.102), while all other species considered have a slightly negative one, with the maximal value in *Habronattus* (-0.113). AT content of the complete genomes is lower in Amblypygi (63.2% in *Damon* and 67.5% in *Phrynus*) than in Uropygi and Araneae (69.8-76%). The reversal in GC skew is also reflected in codon usage. As an example table 3.1 shows the relative synonymous codon usage (RSCU) values for Leucine and Serine in different arachnids (Amblypygi, Uropygi, Araneae). In the case of Leucine

the most frequently used codon in all cases is UUA, but besides this, in plus-strand encoded genes there is UUG second-most often in use in taxa with a positive GC skew, while in taxa with a negative GC skew it is CUA or CUU which is more often found. Genes encoded on minus-strand show a reversed codon usage for Leucine. Similar differences in usage of C and G containing codons is seen for Serine. Genes which are positively GC skewed (plus strand genes in those taxa with positive GC bias and minus-strand genes in those with negative GC bias) use predominantly AGA (but also UCU), while those genes with a negative GC skew predominantly use UCU, UCC and UCA codons.

Table 3.1: Relative synonymous codon usage of assorted amino acids in different arachnid species.

			Nephila Araneae	*Habronattus* Araneae	*Ornithoctonus* Araneae	*Heptathela* Araneae	*Mastigoproctus* Uropygi	*Phrynus* Amblypygi	*Damon* Amblypygi
GC-skew			0.242	0.301	0.344	-0.235	-0.297	-0.453	-0.473
AT-skew			-0.053	-0.113	-0.083	-0.023	-0.014	-0.029	0.102
AT %			76	74.4	69.8	72.2	70.6	67.5	63.2
strand	codon	aa							
(+)strand	UUG	L	0.983	**1.413**	**2.588**	0.247	0.301	0.053	0.145
	UUA	L	**4.198**	**3.993**	**2.679**	**3.152**	**2.589**	**1.567**	**1.745**
	CUA	L	0.246	0.287	0.229	**1.215**	**1.043**	0.979	**1.455**
	CUC	L	0.02	0	0.023	0.285	0.542	**1.175**	**1.364**
	CUG	L	0.041	0.061	0.115	0.171	0.161	0.089	0.2
	CUU	L	0.512	0.246	0.366	0.93	**1.365**	**2.136**	**1.091**
(-)strand	UUG	L	0.19	0.03	0.145	0.882	**1.053**	**2.038**	**2.941**
	UUA	L	**3.714**	**3.149**	**2.754**	**3.796**	**3.658**	**3.538**	**2.525**
	CUA	L	0.698	**1.99**	**1.507**	0.416	0.553	0.113	0.089
	CUC	L	0.159	0.238	0.348	0.024	0	0	0.03
	CUG	L	0.032	0.03	0.058	0.171	0.132	0.057	0.149
	CUU	L	**1.206**	0.564	**1.188**	0.71	0.605	0.255	0.267
(+)strand	AGC	S	0	0.122	0.077	0.214	0.134	0.335	0.652
	AGU	S	**1.089**	0.975	0.423	0.286	0.504	0.502	0.326
	AGG	S	0.749	0.244	0.462	0	0	0.033	0
	AGA	S	**2.077**	**2.315**	**2.692**	**1.25**	**1.076**	0.837	0.688
	UCG	S	0.102	0.244	0.423	0.071	0.067	0.033	0.109
	UCU	S	**2.689**	**3.168**	**3.115**	**2.5**	**1.916**	**2.343**	**1.484**
	UCC	S	0.034	0.325	0.115	**1.286**	**1.445**	**1.941**	**2.172**
	UCA	S	**1.26**	0.609	0.692	**2.393**	**2.857**	**1.975**	**2.57**
(-)strand	AGC	S	0.133	0.127	0.123	0.105	0.26	0.125	0.2
	AGU	S	0.133	0.254	0.205	0.737	0.715	1	**1.6**
	AGG	S	0.133	0.042	0	0.053	0.195	1	**1.467**
	AGA	S	**1.156**	**1.439**	**1.19**	**2.526**	**3.187**	**2.5**	**1.4**
	UCG	S	0	0.085	0.164	0.263	0.325	0.313	0.333
	UCU	S	**2.711**	**1.693**	**1.559**	**2.684**	**2.146**	**1.5**	**2**
	UCC	S	0.667	0.974	**1.518**	0.684	0	0.188	0.067
	UCA	S	**3.067**	**3.386**	**3.241**	0.947	**1.171**	**1.375**	0.933

Bold numbers indicate values above 1 (=predominantly used codons). GC- and AT-skew is for complete genomes, plus strand. For accession numbers see Figure 3.4.

Phylogenetic interrelationships

Gene order is conserved in Amblypygi, so therefore does not help in resolving interrelationships of the Tetrapulmonata. Our phylogenetic analysis of sequence data from mitochondrial protein-coding genes (Figure 3.4) weakly supports a sister group relationship of Uropygi and Amblypygi, as well as monophyly of Tetrapulmonata. Our result for the interrelationships of the Tetrapulmonata corresponds well to a previous analysis of a similar dataset (Masta et al. 2009; without data from *Damon diadema*), where some maximum likelihood and Bayesian analyses of amino acid datasets also found support for a clade of Uropygi and Amblypygi, although support varied with the model of evolution and type of analysis performed.

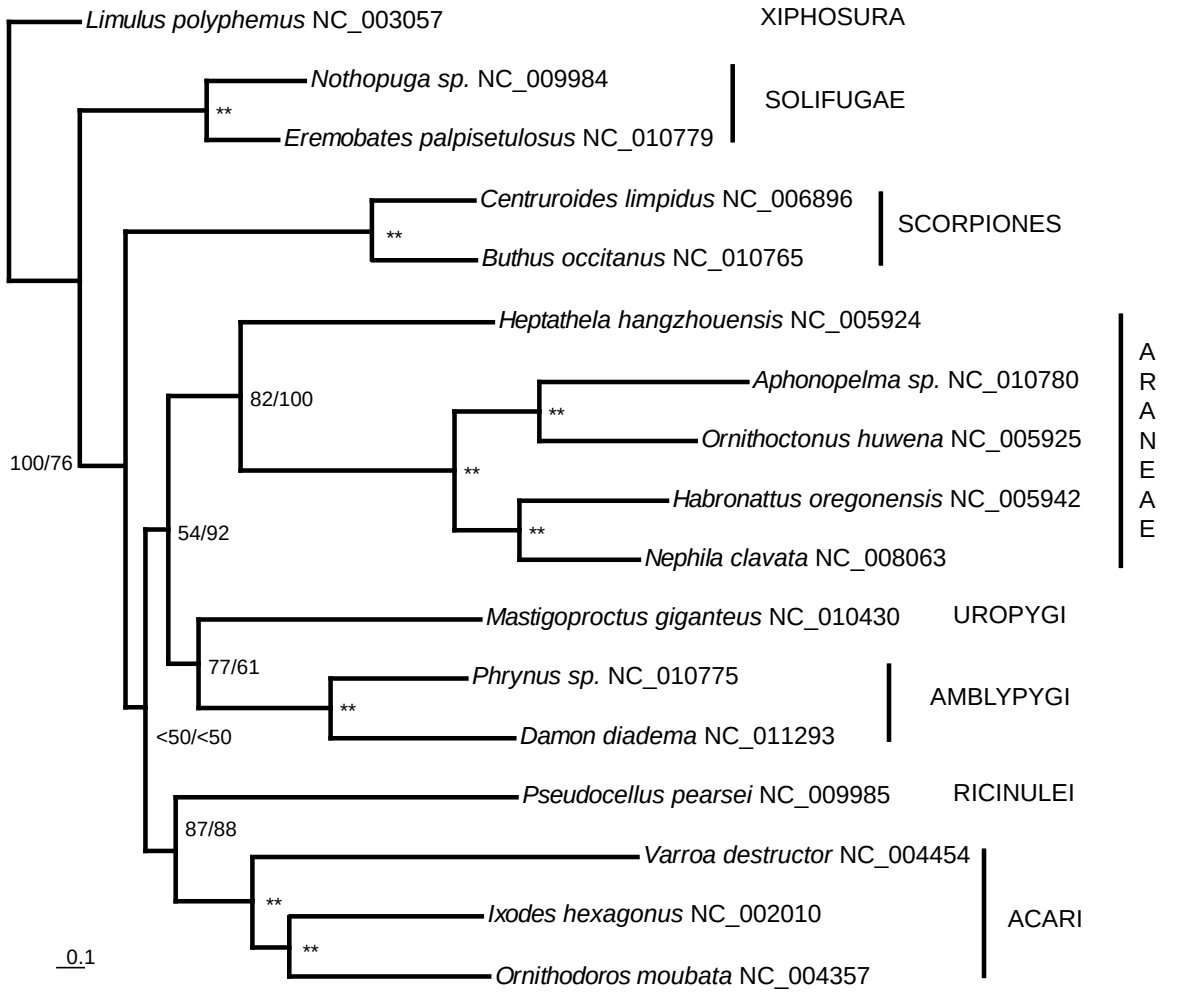


Figure 3.4: Phylogenetic analysis of arachnid relationships based on mitochondrial sequence data (concatenated aminoacid alignment of protein-coding genes). The best tree from RAxML analysis (mtREV+G+I) is shown. Numbers next to nodes reflect bootstrap percentages from RAxML analysis (mtREV+G+I, partitioned optimization; left number) and edge support percentages from Treefinder analysis (mtART+G+I, partitioned optimization; right number). Two asterisks depict maximal support from all three methods. Accession numbers of GenBank entries are given after the species name. Scalebar reflects aminoacid substitutions per site.

3.5 Acknowledgements

We thank Peter Adam (FU Berlin) for the line drawing of *Phrynus* sp.; K.F. and L.P. would like to thank Professor Thomas Bartolomaeus (FU Berlin) for his kind support of our work and the German Science Foundation for financial support through grant DFG Ba 1520/10-1,2 (Priority programme "Deep Metazoan Phylogeny"). K.F. was supported by NaFöG (Berlin). S.E.M. was funded by the National Science Foundation award no. DEB-0416628.

3.6 Additional material

Additional files 3.1 to 3.4 are provided with the online version and can also be found in the appendix of this thesis (Chapter 10).

4. Differences in mitochondrial genomes of Ricinulei: new mitogenomic data from the two major clades of New World Ricinulei

4.1 Abstract

We present entire mitochondrial genome sequences of two ricinuleid species, *Cryptocellus narino* and *Pseudocellus gertschi*, representing the two major clades of New World Ricinulei. In respect of genome size, strand bias, inferred secondary structures of tRNAs, and length of protein-coding as well as ribosomal RNA genes we found results similar to many other arthropod mt genomes. In *P. gertschi* the 37 genes usually present in the mitochondrial genome of Bilateria show a similar gene order as in the genome of *L. polyphemus*, and therefore resemble the putative arthropod ground pattern. On the contrary, in *C. narino* we detected a translocation of a tRNA (*trnY*). A phylogenetic analysis of chelicerate relationships based on an amino acid dataset appears interesting in terms of relationships among the three ricinuleid genera. The new insights imply that the traditionally strict division of the New World Ricinulei into the genera *Pseudocellus* and *Cryptocellus* may be reconsidered as C. *narino* appears as closest relative to one of the two *Pseudocellus* species.

4.2 Background

Comprising 64 living species, the Ricinulei comprise an order of rare species (Botero-Trujillo 2008). These predatory and rather small animals (3 to 11 mm) inhabit leaf litter or humid caves of tropical regions (Adis et al. 1989). Besides a characteristically strongly sclerotized cuticle (Kennaugh 1968) the eponymous apomorphy is represented by a moveable hood (cucullus) in front of the prosoma covering the mouthparts (Ax 1996). Extant representatives are divided into the three genera *Ricinoides* (Ewing 1929), *Cryptocellus* (Westwood 1874), and *Pseudocellus* (Platnick 1980). The Old World genus *Ricinoides* occurs exclusively in Western Africa while New World Ricinulei species are found in continental America and are traditionally divided into the genus *Cryptocellus* ("southern group") characteristically found in South America and *Pseudocellus* ("northern group") in Central America (Platnick 1980). Recent morphological data dealing with the sperm structure (Talarico et al. 2008) leads to the assumption of *Cryptocellus* and *Pseudocellus* being more closely related than previously thought.
Within arachnids, a sister group relationship of Ricinulei and Acari, or at least of Ricinulei and a clade of the Acari, is widely accepted (Fahrein et al. 2007; Masta 2010; Shultz 1990; Van der Hammen 1972; Van der Hammen 1982; Weygoldt and Paulus 1979; Wheeler and Hayashi 1998). Morphological characters supporting this so-called taxon Acaromorpha are a unique post-embryonic development including a hexapodal larva followed by three octapod nymphal instars and a gnathosoma with medially fused palpal coxae. However, arguing against the Acaromorpha is e.g. morphological data concerning the ultrastructure of spermatozoa (Alberti and Palacios-Vargas 1984; Talarico et al. 2008).

In studies of animal phylogeny mitochondrial genomes can be used as an advanced dataset. Besides sequence information in form of nucleotides or amino acids as well as changes in the mitochondrial genetic code (Castresana et al. 1998), also rearrangements in the mitochondrial gene order have been proven useful (Boore et al. 1998; Boore et al. 1995), because homoplastic translocations in closely related taxa are presumed to be very unlikely (Dowton et al. 2002). In addition, comparative work concerning the control region (Oliveira et al. 2007) and the inferred secondary structure of encoded tRNAs and rRNAs are more and more included in phylogenetic considerations (Fahrein et al. 2009; Masta 2010; Masta and Boore 2008).
Using mitochondrial data of different arachnids, various phylogenetic and gene rearrangement analyses lead to the assumption that rearrangements occured basically during the early diversification of each order, which restricts the use of these events to phylogenetic questions of intraordinal rather than interordinal relationships of arachnids (Masta 2010).

Here we present two new complete mitochondrial genome sequences from the Ricinulei species *Cryptocellus narino* and *Pseudocellus gertschi*. We focus on intraordinal relationships of Ricinulei, using sequence data as well as structural features of the mitochondrial genomes.

4.3 Results & Discussion

Genome organization, non-coding parts and gene order

The mitochondrial genomes of the three Ricinulei sequenced for this study appear to be typical circular DNA molecules each containing all 37 genes usually present in bilaterian mitochondrial genomes (Figure 4.1). The length varies between *P. gertschi* (14477 bp) and *C. narino* (14554 bp) and thus is rather shortened relative to *Pseudocellus pearsei* (15099 bp; Fahrein et al. 2007) and to other chelicerates. The overall AT-content in *P. gertschi* (68,03%) and *C. narino* (67.59%) is likewise low in *Limulus polyphemus* (67.6%) (Navajas et al. 2002).
In each mitochondrial genome we found a large non-coding region ranging in size from 388 to 404 bp (Figure 4.1, Supplementary files 4.1 a+b). We identified each of these large non-coding sequences as the putative mitochondrial control region, due to its conserved location between *rrnS* and *trnI* (Saito et al. 2005; Wolstenholme 1992; Zhang and Hewitt 1997). In the investigated species it is more AT-rich (*C.n.* 72,03%, *P.g.* 70,36%) compared to the A+T composition of the complete genome (see above). Moreover, part of each region is capable to be folded into a hairpin-like formation, with a stem containing a small number of mismatches (Figure 4.1). Such putative stem-loop secondary structures are known from many mitochondrial control regions and one was also detected in the largest non-coding region of *P. pearsei*. Conspiciously, in all Ricinulei these stem-loops are located within the first 150 nucleotides of the control region, overspanning the same area. Furthermore conserved motifs, also characterized in other arthropod mitochondrial control regions, occur in the flanking sequences around each stem-loop structure: a TATA motif appears in the 5'-flanking sequence whereas the motif GAT is found in the 3'-flanking sequence. Besides, all ricinuleid control regions begin with a poly-T stretch upstream of the hairpin structure. Hairpins, poly-T stretches, TA(A) and G(A)nT motifs are presumed to play an important role in the initiation of transcription and (or) replication of the mitochondrial genome (Zhang et al. 1995).

In *P. gertschi* the detected gene order is similar to that found in the horseshoe crab *L. polyphemus*, which is considered to reflect the ancestral gene order also retained in many arthropods (Lavrov et al. 2000; Staton et al. 1997). Rearrangements are detected in *C. narino*

in form of a translocation of *trnY* from its usual downstream position within the *WCY*-cluster of tRNAs to a upstream-position of this now *YWC* distribution of tRNAs (Figures 4.1 and 4.2). As reported in an earlier study, in *P. pearsei* five tRNAs (*trnW*, *trnY*, *trnN*, *trnL*$^{(CUN)}$, *trnV*) changed their position to a new location between *trnM* and *nad2* and we assumed a minimum of two events causing the derived gene order (Fahrein et al. 2007).

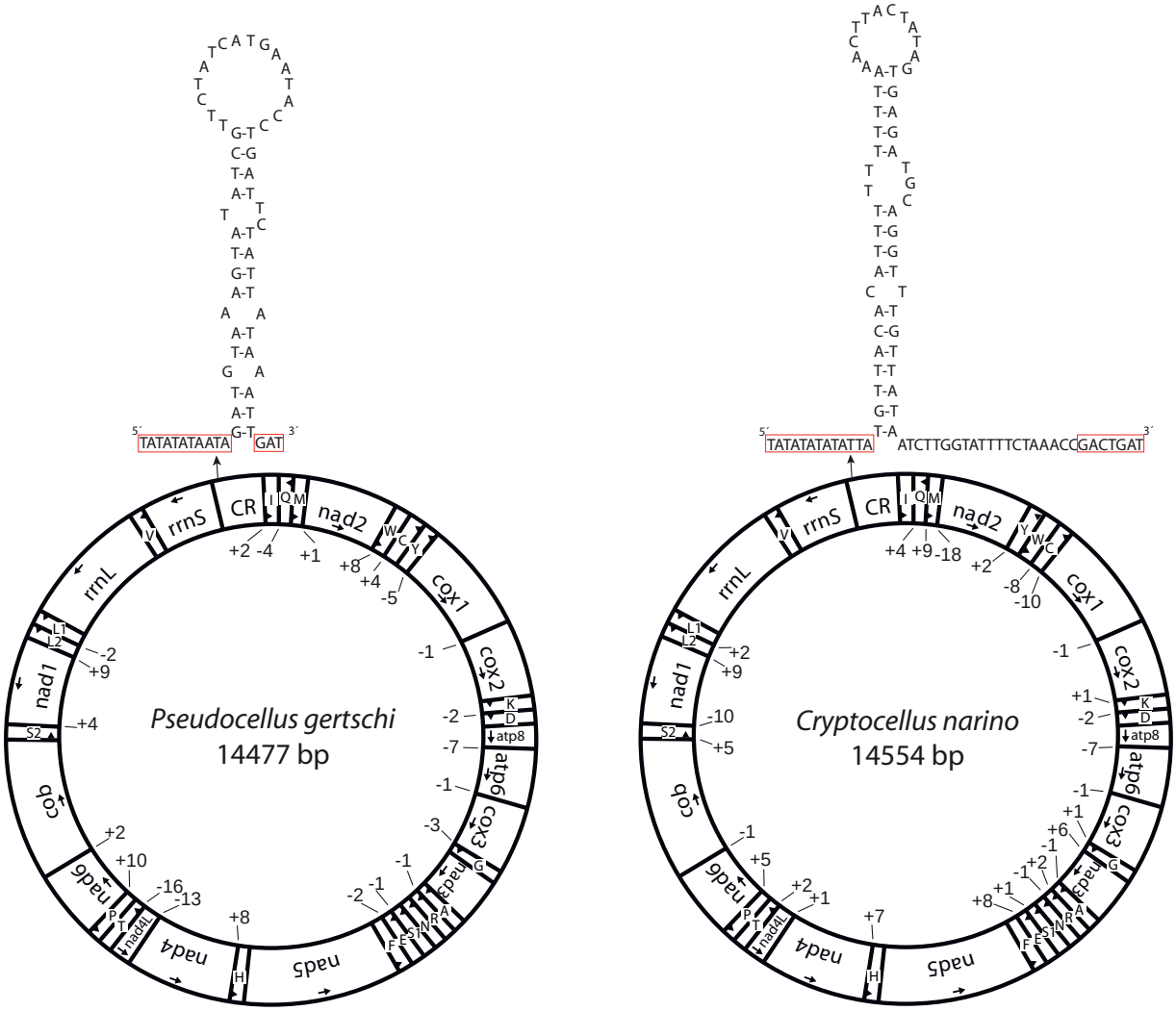


Figure 4.1: Mitochondrial genome maps and control region loop secondary structures of *Cryptocellus narino* and *Pseudocellus gertschi.* Above the circular maps, stem-loop structures found in the largest non-coding part (= the putative control region (CR)) are shown. Small boxes in sequence highlight putative signal motifs. tRNA genes are depicted by their one-letter code abbreviations. Numbers reflect non-coding (positive) or overlapping (negative) nucleotides between two adjacent genes; small arrows indicate the orientation of the genes on (+)strand (clockwise) or (-)strand (counterclockwise).

Figure 4.2 shows a compilation of gene orders found in New World Ricinulei. Apparently, the discovered rearrangements in the mitochondrial genomes of *P. pearsei* and *C. narino* cannot be associated with each other. We rather assume them as results of events which took place independently in each ricinuleid clade: 1) one single translocation of *trnY* as characteristic for *C. narino* or for all members of the genus *Cryptocellus*, and 2) a tandem duplication/ random deletion event plus a single tRNA gene transposition leading to the derived gene order of *P. pearsei* which is not a characteristic of all members of the genus *Pseudocellus*.

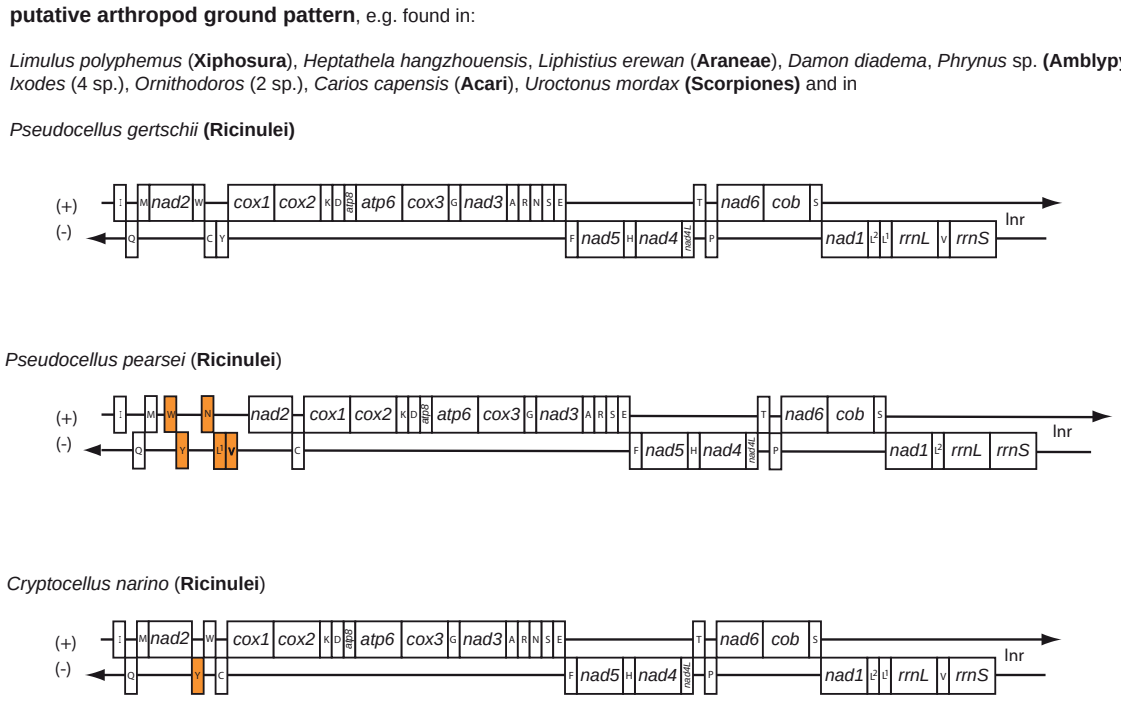


Figure 4.2: Compilation of gene order in mitochondrial genomes of Ricinulei compared to the putative ancestral arthropod gene order detected in *P. gertschi*. Transfer RNA genes are labelled according to the one letter amino acid code. Genes marked white show the same relative position as in the arthropod ground pattern; genes marked orange have relative positions differing from the arthropod ground pattern.

Protein-coding genes and nucleotide frequencies

Most protein-coding genes (Supplementary files 4.1 a+b) have a length comparable between the two ricinuleids and similar to their counterparts in *P. pearsei* and *L. polyphemus*. The major difference is found in *cox1*, which contains in *C. narino* as in *P. pearsei* 4 amino acids more than in *P. gertschi* respectively five more than in *L. polyphemus*. In the majority of gene boundaries smaller overlaps (up to 18 bp in *C. narino*) or non-coding sequences (up to 10 bp in *P. gertschi*) between adjacent genes are found.

Most of the identified protein-coding genes begin with one of the common start codons for mitochondrial DNA (ATN, GTG, TTG, GTT; Wolstenholme 1992). Exceptional start codons were identified for *cox1* in *P. gertschi* (TTA) and in *C. narino* (CTA), as well as for the initiation of *atp8* in *C. narino* (GTA). For the majority of protein-coding genes complete termination codons are inferred. Solely differing are *nad3*, *nad5*, and *nad4* which in both species terminate with a single thymine, just like *cob* and *cox3* in *P. gertschi*. The truncated stop codon TA is assumed to be the gene terminus of *cox2* in *P. gertschi*. In several cases, truncated stop codons are also reported for similar genes in the mitochondrial genome of *P. pearsei* (Fahrein et al. 2007). Partial stop codons have also been described for protein-coding genes in mitochondrial genomes of other species, where a T or TA is assumed to be complemented by post-transcriptional polyadenylation into a functional termination codon (Ojala et al. 1981).

To describe the strand specific nucleotide frequency bias of mitochondrial genes, the CG-skew [(%C-%G)/(%C+%G)] is often used, so in this study (Supplementary files 4.1 a+b). As the (+)strand usually is defined as the strand bearing the majority of coding sequence, the detected CG-skew is positive in all (+)strand encoded genes and negative in all (-)strand encoded genes (including the two ribosomal RNAs). This strand specific bias found in both Ricinulei occurs in most other arthropods including *P. pearsei*, while a reversal of that bias has been reported for only a few species (Hassanin et al. 2005; Hassanin 2006; Navajas et al. 2002).

rRNA genes

Compared to the ribosomal subunits of *L. polyphemus* (*rrnL*:1296, *rrnS*:799), the length of both ribosomal RNA genes appear to be significantly shortened in the investigated ricinuleids: *rrnL* is abbreviated more than 100 nucleotides (*C.n.* 1187 bp, *P.g.* 1162 bp) and *rrns* more than 50 nucleotides (*C.n.* 740 bp, *P.g.* 743 bp). Regarding *P. Pearsei*, a similar short length is reported for *rrnS* (743 bp) while *rrnL* (1250 bp) shows an in-between size. Contingently the longer *rrnL* of *P. pearsei* compared to the other Ricinulei results from annotation problems, depending on possibly existing and undetected remains of *trnV* between *rrnL* and *rrnS*.

As often found in animal mitochondrial DNA, the two rRNA genes are separated by *trnV*, thus resembling the ancestral condition which is found to be derived in *P. pearsei*. The termination signal for *rrnL* transcription in animal mitochondria is believed to be the conserved heptamer motif TGGCAGA (Valverde et al. 1994). Anymore, slight modifications of this motif in form of single nucleotide variations were found in different arthropods (Fahrein et al. 2009). In all ricinuleids the conserved termination signal is found downstream of

rrnL in both *trnL*, though it appears to be modificated to a TGGCAAA in *trnL*$^{(UUR)}$ of *C. narino* as well as in *trnL*$^{(CUN)}$ of *P. gertschi*. Since the same location of the *rrnL* termination signal downstream of the large rRNA gene within the *trnL*$^{(CUN)}$ is reported for insects and crustaceans (Valverde et al. 1994) as well as for chelicerates like amblypygids, *Limulus*, and *Habronattus* (Fahrein et al. 2009; Masta 2000), while it is found within the downstream *trnL*$^{(UUR)}$ only in vertebrates, we assume the location in *trnL*$^{(CUN)}$ in Ricinulei representing the ancestral location for chelicerates.

Secondary structure of tRNAs

Most tRNAs possess the typical cloverleaf secondary structure, though for all three species in several tRNAs a shortened TψC-arm is inferred (Supplementary files 4.2 a+b) which is also the case in *P. pearsei* (Fahrein et al. 2007). An entirely absence of the TψC-arm is detected only in *trnF* of *P. gertschi* and in *trnC* of *C. narino*. If so, it is replaced by a TV-replacement loop. The loss of D-arm or the T-arm sequences is often found in metazoan mitochondrial tRNAs (Boore and Brown 2000; Helfenbein et al. 2001), with a remarkable intensity reported for several chelicerate clades (e.g. Masta and Boore 2008) and one nematode clade (Okimoto et al. 1992; Wolstenholme et al. 1987). As typical for many metazoan *trnS*$^{(AGN)}$, the D-arm sequence is found to be missing in *trnS*$^{(AGN)}$ of the two ricinuleids investigated for this study, which is consistent with an assumed loss early in the evolution of Metazoa (Wolstenholme 1992).

Phylogenetic Analysis

In our analysis of chelicerate phylogenetic relationships based on amino acids Ricinulei come out as sister group to a clade combining paraphyletic Acari (Figure 4.3) with monophyletic Pycnogonida. The monophyly of the Acari was questioned by different authors before (Fahrein et al. 2007; Masta 2010; Regier et al. 2010; Van der Hammen 1972; Van der Hammen 1982). Familiar is the split between the clades Acariformes and Parasitiformes, but new is a close relationship of Acariformes and pycnogonids. However, in terms of the mentioned relationship, the revealed topology has to be regarded carefully due to missing support by ML bootstrapping. A better support is shown regarding the intraordinal relationships of Ricinulei, especially for the deepest branching in the American clade. Taking a closer look at the Central American genus *Pseudocellus*, it appears not as a monophyletic group with the *Cryptocellus* species being sister to *P. gertschi*. Thus, the traditional and strict division (e.g. made by Platnick 1980) of the New World Ricinulei in a southern and a northern group, more precisely in the genera *Cryptocellus* and *Pseudocellus*, possibly is unsustainable.

Certainly, for a concluding evaluation of these findings more genomic data is indispensable. From aspects of morphology considerations concerning this matter have already been made by Talarico et al. (2008), but also here more morphological data investigating especially New World Ricinulei is essential.

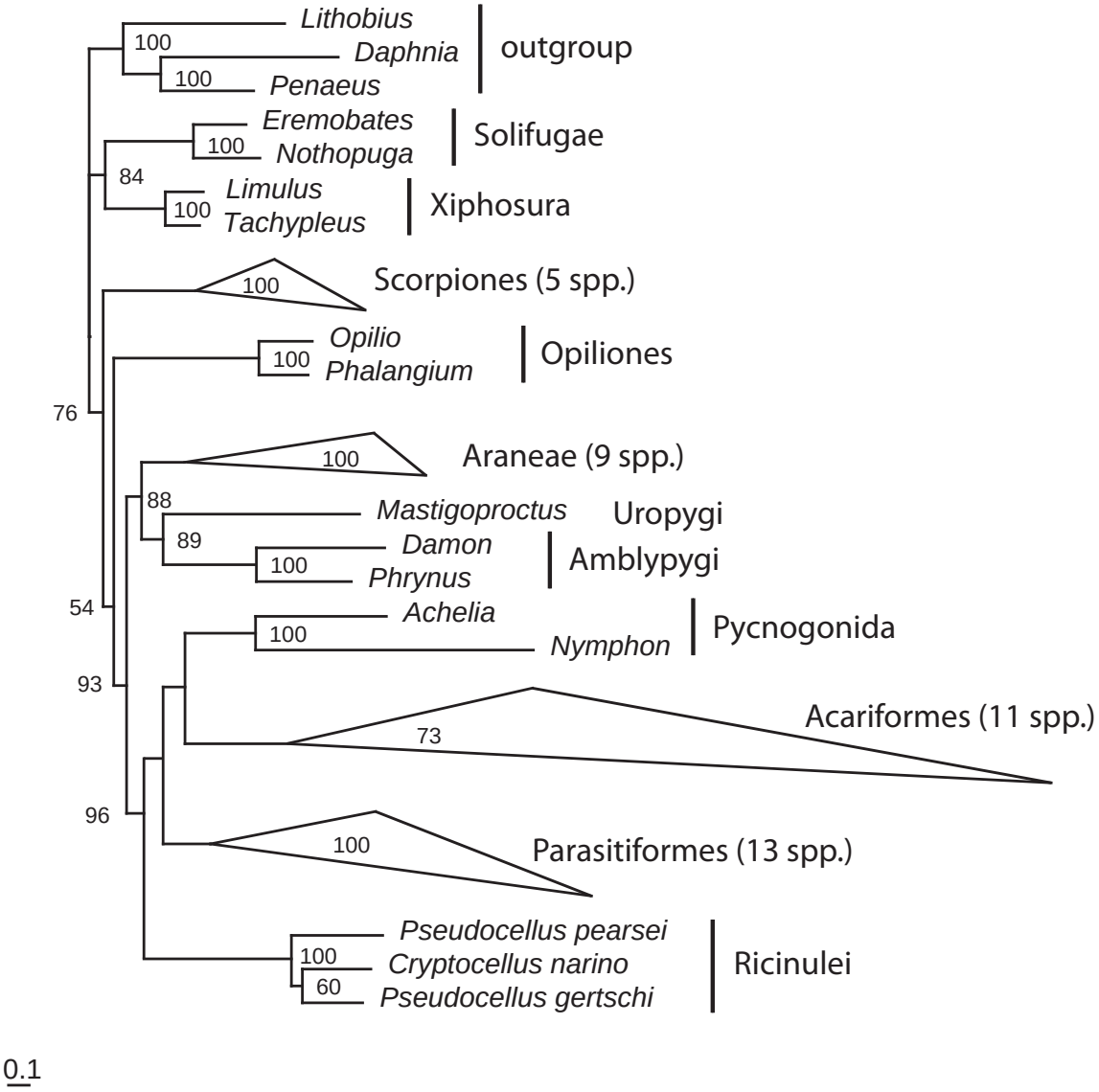


Figure 4.3: Phylogenetic tree of chelicerate relationships, inferred from amino acid dataset. Protein-coding gene sequences from all complete chelicerate mitochondrial genomes available in NCBI Refseq plus our own new data were aligned and concatenated. The tree was rooted with three outgroup taxa (*Lithobius*, *Daphnia*, *Penaeus*), summing up to a total of 55 taxa included in the analysis. Topology and branch lengths were taken from the best scoring tree of the maximum likelihood (ML) analysis. Numbers behind the branching points are percentages from ML bootstrapping.

4.4 Conclusions

We found the analysed mitochondrial genomes of members of both New World ricinuleid genera being similar to many arthropod mitochondrial genomes regarding genome size and content, strand bias, and secondary structure of encoded tRNAs. Also the detected gene order in *P. gertschi* resembles the arrangement of genes known from many arthropod mitochondrial genomes, while *C. narino* has a derived gene order due to one translocated tRNA (*trnY*). By reason of the conserved gene order in parts of the Ricinulei as also known to be true in parts of other arachnid orders, it is not possible to infer phylogenetic relationships based on genome arrangements. In a phylogenetic analysis of chelicerate relationships Ricinulei are found in a close relationship to an Acari-Pycnogonida clade, where Acari appears to be paraphyletic with respect to Pycnogonida. Owing to bad support values in this case, these findings should be interpreted with caution. Furthermore, our phylogenetic analysis prefigures the genus *Pseudocellus* not being monophyletic but more genomic as well as morphological data is necessary for a well-funded new hypothesis of ricinuleid intraordinal relationships.

4.5 Material and Methods

Specimens and DNA extraction

The specimens of *Cryptocellus narino* and *Pseudocellus gertschi* were collected in Puerto Boyacá, Kolumbien (*C.n.*) by an anonymous collector and in Veracruz, Mexico (*P.g.*) by J. Kral. In each case total DNA was extracted from one leg conserved in 100% ethanol using the DNeasy Blood & Tissue kit (Qiagen, Hilden, Germany) and following the manufacturers`protocol.

PCR conditions, Sequencing and Gene annotation

Sequencing strategy, PCR conditions, sequencing, sequence assembly and gene annotation were nearly identical to those described in the publication about the mitochondrial genome of the Ricinulei *Pseudocellus pearsei*, we here refer to the material and method section of the according open access publication (Fahrein et al. 2007) where details can be looked up.

Phylogenetic Analysis

For the phylogenetic analyses we used concatenated amino acid alignments of all 13 protein-coding genes. All complete mitochondrial genome sequences from chelicerates found in NCBI Refseq were used in addition to our own new data. As a root for the tree three species from outgroups were included, a centipede (*Lithobius*), a shrimp (*Penaeus*) and a water flea (*Daphnia*), summing up to a total of 55 taxa included in the analysis. Alignment of protein-coding genes was done with MAFFT using the FFT-NS-i option (Katoh et al. 2009). The concatenated alignment consists of 3943 amino acid positions. Maximum likelihood analysis (ML) was conducted using Treefinder (Jobb 2007). After evaluating different model for amino acid substitutions in mt genomes with the model evaluation tool implemented in Treefinder, we chose the mtArt+G model (Abascal et al. 2007). In addition to the best tree, 100 bootstrap replicates were generated and analysed.

4.6 Abbreviations

A, adenine; *atp6* and *8*, genes encoding ATPase subunit 6 and 8; bp, base pairs; *cox1-3*, genes encoding cytochrome oxidase subunits I-III; *cob*, gene encoding cytochrome b; C, cytosine; G, guanine; ML, maximum likelihood; *nad1-6* and *nad4L*, genes encoding NADH dehydrogenase subunits 1-6 and 4L; PCR, polymerase chain reaction; rRNA, ribosomal RNA; *rrnL*, large (16S) rRNA subunit (gene); *rrnS*, small (12S) rRNA subunit (gene); T, thymine; tRNA-Abc (where Abc is replaced by three letter amino acid code of the corresponding amino acid), transfer RNA; *trnX* (where *X* is replaced by one letter amino acid code of the corresponding amino acid), tRNA gene; TψC-arm, T-loop and T-stem of a tRNA secondary structure.

4.7 Acknowledgements

We thank the anonymous collector of *Cryptocellus narino* and J. Kral for the kind and straightforward provision of the two Ricinulei specimen.

4.8 Additional material

Supplementary files 4.1 and 4.2 can be found in the appendix of this thesis (Chapter 10).

5. New mitogenomic data from a member of Opiliones (*Opilio parietinus*) reveal a rearrangement hot-spot in Opiliones

5.1 Abstract

Here we present the complete mitochondrial genome of *Opilio parietinus* (Eupnoi, Phalangioidea, Opilioniinae), the second complete sequence from a member of the Opiliones. Concerning strand bias, inferred secondary structures of tRNAs, overall genome size and length of protein-coding as well as ribosomal RNA genes, the mitochondrial genome of *O. parietinus* is similar to conditions known for many arthropod mitochondrial genomes. Differences were detected regarding the gene order, which in fact reflects the gene order widely spread among arthropods in the vast majority of gene boundaries, but rearrangements in a delimited genome area are obvious. Beside a duplication of the control region a new location of *trnI* and *trnQ* is also evidenced. In comparison to the previously published mitochondrial genome of the opilionid *Phalangium opilio* conformity could be proven for the genome section bearing the rearrangements (between *trnL*$^{(UUR)}$ and *trnM*), but not for the rearrangements themselves. The presented data on gene rearrangements and on other features of the mitochondrial genome of *O. parietinus* provides no reliable basis to finally resolve the still debated sister taxon of Opiliones. Nevertheless, in the shape of detected rearrangements, we found features useful for considerations of intraordinal relationships of Opiliones.

5.2 Background

Within the Arachnida, Opilionida are a very diverse order - with more than 6000 described species they are excelled only by spiders and mites (Pinto-da-Rocha et al. 2010). Beside various convergent features, harvestmen (or “daddy-longlegs”) can be clearly distinguished from other arachnids due to e.g. paired tracheal stigmata as well as characteristics concerning reproduction, males possess a penis and females an ovipositor. In many morphological or combined (morphological and nuclear-sequence based) analysis of arachnid phylogenetic relationships Opiliones group with camel spiders, scorpions, and pseudoscorpions, together forming the clade Dromopoda (Giribet et al. 2002; Shultz 1990; Wheeler and Hayashi 1998). Though, it has to be mentioned that, regarding strict molecular (nuclear-sequence based) or strict morphological analysis, some authors favour Ricinulei and Acari, or part of the latter ones, as closest relatives (Regier et al. 2010; Weygoldt and Paulus 1979). Thus, further data might elucidate the relatedness of these lineages. With regard to molecular analysis, not only comparative studies of nuclear genes are more and more conducted and helpful to infer phylogenetic relationships but also mitochondrial genomes. These, compared to the nuclear genome, well-arranged molecules are often easier to gain and have, beside the pure sequence information in form of nucleotides and amino acids, a bulk of additional features considered to bear phylogenetic useful information. Well proven are, in addition to the gene order, the also up-coming comparative work concerning the control region and inferred secondary structures of transfer and ribosomal RNAs.
In this study we present the mitochondrial genome of *Opilio parietinus* (Eupnoi, Phalangioidea, Opilioniinae), the second complete genome of a member of the Opiliones. In the context of arachnid phylogeny, analyzed features in terms of gene order, control region, protein-coding and ribosomal RNA genes, and inferred tRNA secondary structures are mainly compared with the condition found in *Phalangium opilio* (Eupnoi, Phalangioidea, Phalangiinae), the other completely sequenced Opiliones.

5.3 Results & Discussion

Genome organization

We found the mitochondrial genome of *O. parietinus* to be a typical circular doublestrand DNA molecule of 15400 bp length, bearing all 37 genes usually present in bilaterian mitochondria (Figure 5.1, lower part). With 72,29% the overall AT content of the complete genome is close to that in *P. opilio* and lies well within the range described for Chelicerata (Navajas et al. 2002). Regarding gene boundaries overlaps appear less frequent than non-coding sequences (Table 5.1). Besides the three biggest non-coding sequences of more than

50 nt (514 bp between *trnL*$^{(UUR)}$/*trnL*$^{(CUN)}$, 56 bp between *trnQ*/*rrnL*, 336 bp (between *rrnS*/*trnM*) we detected 15 smaller ones of two to 20 bp. Only at ten gene boundaries we found overlaps of one to nine basepairs, the remaining genes are immediately contiguous.

Table 5.1: Genome organisation of *Opilio parietinus* (15400 bp).

Gene	***Strand***	***Position***		***Length (nuc.)***	*CG-skew*	***Start-codon***	***Stop-codon***	***Intergenic nucleotides***
cox1	+	1-	1539	1539	0,121	ATT	TAA	-1
cox2	+	1539-	2222	684	0,215	ATG	TAA	+11
trnK	+	2234-	2299	66				+3
trnD	+	2303-	2369	67				0
atp8	+	2370-	2540	171	0,561	ATC	TAA	-7
atp6	+	2534-	3208	675	0,385	ATG	TAA	-1
cox3	+	3208-	3990	783	0,173	ATG	TAT	-2
trnG	+	3989-	4054	66				0
nad3	+	4055-	4393	339	0,263	ATT	TAA	+20
trnA	+	4414-	4479	66				-1
trnR	+	4479-	4546	68				+3
trnN	+	4550-	4602	53				+11
trnS1-AGN	+	4614-	4669	56				0
trnE	+	4670-	4737	68				+3
trnF	-	4741-	4805	65				+5
nad5	-	4811-	6511	1701	-0,332	GTG	TAA	0
trnH	-	6512-	6578	67				+3
nad4	-	6582-	7910	1329	-0,167	ATG	TAG	-7
nad4L	-	7904-	8194	291	-0,6	ATG	TAG	+10
trnT	+	8205-	8265	61				+2
trnP	-	8268-	8330	63				+5
nad6	+	8336-	8785	450	0,316	ATT	TAA	0
cob	+	8786-	9922	1137	0,196	ATG	TAA	-2
trnS2-UCN	+	9921-	9984	64				+14
nad1	-	9999-	10896	898	-0,225	ATT	TAT	0
trnL2-UUR	-	10897-	10962	66				*
non-coding		10963-	11476	514				*
trnL1-CUN	-	11477-	11545	69				+12
trnI	-	11558-	11622	65				+5
trnQ	-	11628-	11695	68				*
non-coding		11696-	11751	56				*
rrnL	-	11752-	12993	1242	-0,314			*
trnV	-	12994-	13063	70				*
rrnS	-	13064-	13849	786	-0,262			*
non-coding		13850-	14185	336				*
trnM	+	14186-	14256	71				+1
nad2	+	14258-	15217	960	0,235	ATG	TAA	-2
trnW	+	15216-	15278	63				-9
trnC	-	15270-	15332	63				+13
trnY	-	15346-	8	63				-8

* Borders determined according to adjacent gene boundaries

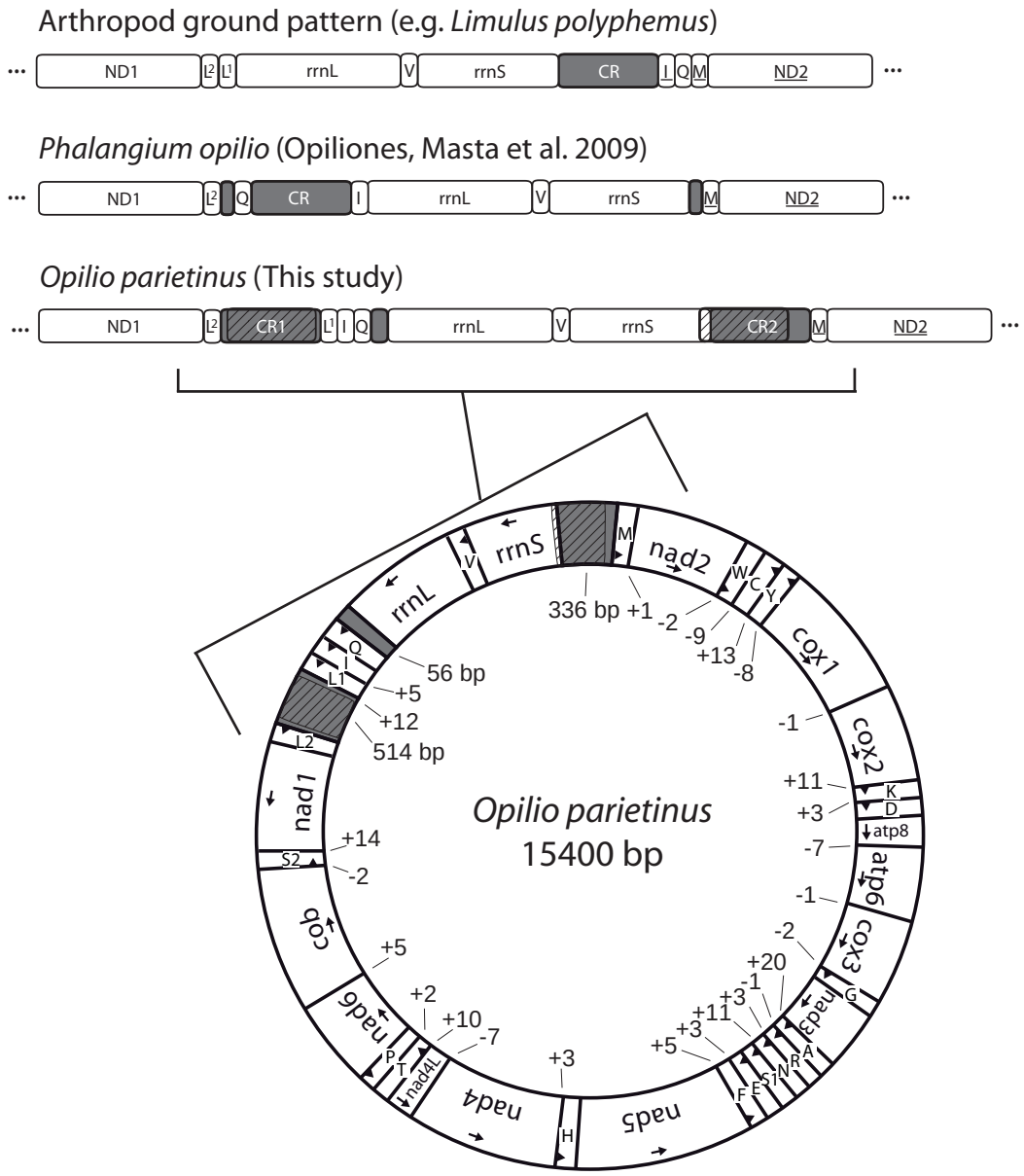


Figure 5.1: Gene order around the mitochondrial control region of Opiliones (above) and circular map of the mitochondrial genome of *O. parietinus* (below). Illustrated in the upper part is just the linearised gene section between ND1 and ND2 of the mitochondrial genome of the two completely sequenced Opiliones species *O. parietinus* and *P. opilio*. Non-coding areas are depicted in grey, tRNA, protein-coding, and ribosomal RNA genes in white. Horizontal underlines indicate the orientation of the genes on (+)strand. CR means control region, in *O. parietinus* CR1 and CR2 are the two control region candidates. The horizontal stripes in the mitochondrial genome section of *O. parietinus* highlight nearly identical genome sequences (each 484 bp in length with two mismatches in comparison). In the lower circular genome of *O. parietinus* tRNA genes are depicted by their one-letter code abbreviations. Numbers reflect non-coding (positive) or overlapping (negative) nucleotides between two adjacent genes; small arrows indicate the orientation of the genes on (+)strand (clockwise) or (-)strand (counterclockwise); grey-shaded areas illustrate non-coding sequences of more than 50 nucleotides; horizontal stripes indicate nearly identical sequences; flanking dots illustrate adjacent genes of the mitochondrial genome.

Control region

We detected two large non-coding sequences, one (CR1) is 514 bp in length and located between *trnL*$^{(UUR)}$/*trnL*$^{(CUN)}$, the other (CR2) comprises 336 bp between *rrnS*/*trnM* (Figure 5.1). Both have a higher AT content of 78,21% (CR1) and 81,55% (CR2) in comparison to the complete genome (72,29%). Even though the location of the putative control region in invertebrates shows great variability (Wolstenholme 1992), it was demonstrated having the same relative location between *rrnS* and *trnI* in most arthropods with a conserved mitochondrial gene order (Saito et al. 2005; Zhang and Hewitt 1997). Thus, we consider CR2 as putative control region in the mitochondrial genome of *O. parietinus* due to its more conserved location downstream of *rrnS*. Moreover, we assume CR1 as a copy of *rrnS* and CR2 because 484 nt (position 10973 - 11456) of the 514 nt long CR1 are congruent to a sequence spanning the 5`end of *rrnS* and more than three-quarters of CR2 (position 13647 - 14130). We assume that this duplication event might have occurred recently, so that the duplicated sequence had no time for deletion, because an additional control region is only rarely found in mitochondrial genomes and with only two mismatches (99,5% identity).

Gene rearrangements

The gene order in the mitochondrial genome of *O. parietinus* largely resembles the putative arthropod ground pattern, e.g. found in *Limulus polyphemus* and many other arthropods (Lavrov et al. 2000; Staton et al. 1997). As in *P. opilio*, these differences are restricted to a region between *trnL*$^{(UUR)}$ and *trnM*, which in both species is flanked by non-coding sequences (Figure 5.1, upper part). Comparing the two Opiliones, these non-coding sequences are of different length: 514 bp (between *trnL*$^{(UUR)}$/*trnL*$^{(CUN)}$) and 336 bp (between *rrnS*/*trnM*) in *O. parietinus* respectively 23 bp (*trnL*$^{(UUR)}$/*trnQ*) and 31 bp (*rrnS*/*trnM*) in *P. opilio*. In *O. parietinus* only *trnI* and *trnQ* have a new location between *trnL*$^{(CUN)}$ and *rrnL* upstream of their common location, including an inversion of *trnI*. Further peculiarities are the already mentioned non-coding sequences between *trnL*$^{(UUR)}$/*trnL*$^{(CUN)}$ (514 bp) and between *trnQ*/*rrnL* (56 bp). In *P. opilio* the detected rearrangements concern *trnQ*, the putative control region, and *trnI*, which are located upstream of the for chelicerates typical position at their new position downstrem of *rrnL*. The absence of *trnL*$^{(CUN)}$ is another deviation in the mitochondrial genome of *P. opilio*. A possible explanation for the findings concerning the non-coding sequences could be a duplication of a gene-block including minimum the control region and parts of the *rrnS* and consistent with the tandem duplication/random loss modell (Fujita et al. 2007; Macey et al. 1997) which occurred at least in the lineage leading to the two Opiliones. A following translocation of *trnQ* and translocation and inversion of *trnI*,

which both took place independently in the lineage leading to each of the two species *P. opilio* (Eupnoi, Phalangiinae) and *O. parietinus* (Eupnoi, Opilioninae), would explain the differences in the detected gene orders in Opiliones. A duplication of the mitochondrial control region is well known among animals, although it is mainly described for vertebrates (Amer and Kumazawa 2005; Eberhard et al. 2001). Furthermore, some authors pointed out that tRNA genes next to non-coding regions are generally more mobile than other tRNAs (Duarte et al. 2008), while others found genes close to the ancestral position of the control region (the *trnI-trnQ-trnM*-cluster) the most mobile ones within mitochondrial genomes (Dowton et al. 2009).

Protein-coding genes and nucleotide composition

For all protein subunits a legitimate start codon (ATN, GTG, TTG, GTT; Wolstenholme 1992) could be identified as initiation codon starting the open reading frame (ORF). Furthermore, one of the commonly found complete stop codons TAA, TAG, or TAT are shown for termination in all protein-coding genes (Table 5.1). Concerning the size of protein-coding genes and compared to their counterparts in *P. opilio* we find a comparable situation in all but *atp8* (*O.p.* 171 nt) and *nad5* (*O.p.* 1701 nt), which appear to be slightly shorter in *P. opilio* (*atp8* 159 nt, *nad5* 1684 nt). The same is true for *atp8* in *L. poyphemus* (156 nt), while *nad5* is longer (1714 nt).

The asymmetric mode of DNA replication causes a different time span of leading and lagging strand remaining in the single-stranded state, what is often mentioned to explain the skewed base compositions of the two DNA strands (Clayton 1982; Faith and Pollock 2003; Reyes et al. 1998). To describe the strand specific bias of nucleotide composition in mitochondrial genomes an often used approach is the CG-skew [(C-G/C+G], which reflects the relative number of guanine and cytosine. In *O. parietinus* the CG-skew is positive for all (+)strand encoded genes while it is negative in all genes encoded on (-)strand (Table 5.1). That implies a clear bias toward cytosine in the (+)-strand genome sequence and vice versa. Such a strand skew is well known from many other arthropods (Hassanin 2006) while a reversion of it is infrequent and within Chelicerates only known from Scorpiones (Choi et al. 2007; Davila et al. 2005; Jones et al. 2007), opisthothele spiders (Masta and Boore 2004; Masta et al. 2009; Qiu et al. 2005), and one member of the Acari, *Varroa destructor* (Navajas et al. 2002).

rRNA genes

Both ribosomal RNA genes have a length (*rrnS*: 786 nt, *rrnL*: 1242 nt) nearly identical to their counterparts in the mitochondrial genome of another member of the Opiliones, *P. opilio* (*rrnS*: 790 nt, *rrnL*: 1243 nt; Masta 2010). Regarding the ribosomal subunits of *L.*

polyphemus (*rrnS*: 799 nt, *rrnL*: 1296 nt; Lavrov et al. 2000) we find a similar size of the small subunits while the large subunits appears to be slightly shortened of about 50 nt in Opiliones. Masta (2010) inferred secondary structures of the large and small ribosomal RNAs of *P. opilio* and showed that the *rrnL* is more complex with more helices in the E domain than in insects (Gillespie et al. 2006). Although we did not infer an overall *rrnL* secondary structure, helices E1-E8 in the corresponding domain also seem to be present in the *rrnL* of *O. parietinus*.

The conserved heptamer motif TGGCAGA is believed to be the termination signal for *rrnL* transcription in animal mitochondrial genomes (Valverde et al. 1994), and also single nucleotide variations were found in different arthropods (Fahrein et al. 2009; Valverde et al. 1994). We found the conserved motif located in *trnL*$^{(UUR)}$ 797 nt away and a slight modification of this motif (TGTCAGA) in *trnL*$^{(CUN)}$ 214 nt away from the 3`end of *rrnL* (Figure 5.2). Although it is assumed that the gene in which the termination signal for rRNA transcription is located can change over evolutionary time, its position within *trnL*$^{(CUN)}$ is conserved among arthropods (Valverde et al. 1994) and is also found in several chelicerates (Fahrein et al. 2009; Masta 2000). Since the given distance between the 3`end of *rrnL* and the location of the termination motif in *trnL*$^{(CUN)}$, it may not serve as a practical termination signal for rRNA transcription. Anyway, another modification (GGACAGA) of the conserved motif was discovered 22 nt away from the 3`end of *rrnL* in a 56 long non-coding sequence between *rrnL* and *trnQ*, thus this location may be more suitable for a precise termination signal.

Secondary structure of tRNAs

All of the 22 transfer RNAs typically present in mitochondrial genomes of Bilateria were detected (Figure 5.2). With the exception of *trnS*$^{(AGN)}$ and *trnN*, which do not possess a T-respectively D-arm, all tRNA gene sequences can be folded into the usual cloverleaf-shaped secondary structure consisting of a D-arm, a anticodon arm bearing the anticodon, a variable loop, a T arm and an acceptor stem. As in *O. parietinus* it is also known from many metazoan mitochondrial *trnS*$^{(AGN)}$, that this gene lacks the potential to form a D-arm which instead is replaced by a D-replacement loop (Wolstenholme 1992). A truncation of the T-arm sequence is in general well-known from several metazoans (Helfenbein et al. 2001; Macey and Verma 1997), but in case of *trnN* is not prevalent among animals. As *trnN* in the mitochondrial genome of *P. opilio* exhibits the sequence coding for a T-arm (Masta 2010), its lack might not be a peculiarity of Opiliones or Eupnoi but of Opilioninae. Reciprocally the same is true concerning the complete lack of *trnL*$^{(CUN)}$ in the genome of *P. opilio*.

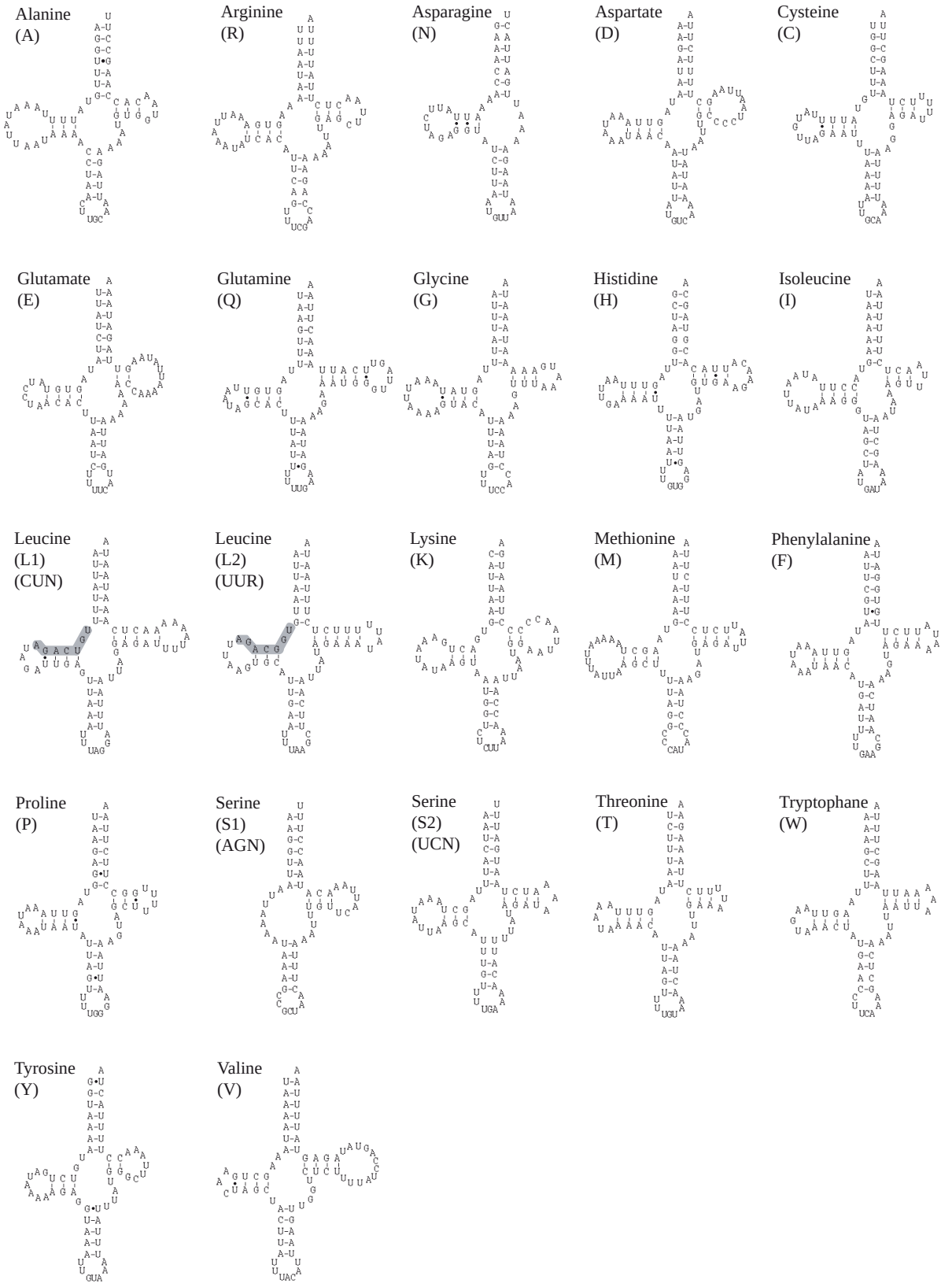


Figure 5.2: Inferred secondary structures of tRNAs from the mitochondrial genome of *O. parietinus*. The grey shaded parts in the secondary structures of the leucine genes mark the putative termination signals for *rrnL* transcription or modifications of this motif. Dots illustrate the base pairing of the pyrimidine base uracil with the purine base guanine.

5.4 Conclusions

The mitochondrial genome of *O. parietinus* mainly resembles conditions known for many arthropod mitochondrial genomes regarding strand bias, overall genome size and length of protein-coding as well as ribosomal RNA genes. Derived secondary structures of tRNAs due to truncated sequences as typical for different arachnid clades like spiders, scorpions, vinagaroons, and some mites (Choi et al. 2007; Domes et al. 2008; Masta 2000; Masta and Boore 2008; Shao et al. 2005a) are not found in *O. parietinus*. Conspicuous differences concern the gene order which varies from the putative arthropod ground pattern, although modifications are restricted to one genome section, while the remaining gene order is similar to other arthropods. Variations concern two tRNA rearrangements and a duplicated control region. Findings in *O. parietinus* are mostly similar to those described for *P. opilio*, with the exception of the exact rearrangements. Indeed, detected rearrangements occur within the same section in both species but with differing results concerning the new tRNA location. Additionaly, *P. opilio* has no duplicated control region nor does *O. parietinus* show a lack of $trnL^{(CUN)}$. If these changes initially are based on one event and furthermore, if this event took place in the stem line of Opiliones or rather at any splitting inside this clade, has to be clarified. Therefore, additional mitogenomic data from members of the opilionid clades Cyphophthalmi, Laniatores, or Dyspnoi would be very helpful. In any case, this region bearing rearrangements in Opiliones is interesting for further gene order comparisons among arachnids.

5.5 Material and Methods

Specimen and DNA extraction

The specimen of *O. parietinus* was collected in Montseny, Spain (41°45’45N 2°21’52.8E) by G. Giribet, E. Mateos, and N. Lopez. Total DNA was extracted from one leg conserved in 100% ethanol using the DNeasy Blood & Tissue kit (Qiagen, Hilden, Germany) and following the manufacturers`protocol.

PCR, sequencing, gene annotation

As PCR conditions, sequencing, sequence assembly and gene annotation were nearly identical to those described for the Araneae, we here refer to the material and method section in Chapter 7 where details can be looked up.

Differences have to be mentioned regarding the initial PCRs, which were successful in case of S7, S15, S28, S35, S43 (Yamauchi et al. 2004). On basis of the successfully obtained sequences, taxon-specific primer pairs were designed to amplify missing fragments in order to bridge the gaps between the sequences (visible in upcoming publication).

Another exception concern the annotation of the ribosomal RNA genes. Usually boundaries of flanking genes are used to assume the start and ending of the ribosomal RNA genes and the extension of control regions. In *O. parietinus* the ribosomal RNA genes are flanked by non-coding regions so the 5`end of *rrnS* and the 3´end of *rrnL* was inferred by comparison with *rrnS* respectively *rrnL* genes of other arachnids. Helpful in this case were the corresponding ribosomal RNA gene sequences of *P. opilio* [Genbank: NC_010766.1].

5.6 Abbreviations

A, adenine; *atp6* and *8*, genes encoding ATPase subunit 6 and 8; bp, base pairs; *cox1-3*, genes encoding cytochrome oxidase subunits I-III; *cob*, gene encoding cytochrome b; C, cytosine; D-arm, D-loop and D-stem of a tRNA secondary structure; G, guanine; *nad1-6* and *nad4L*, genes encoding NADH dehydrogenase subunits 1-6 and 4L; nt, nucleotides; PCR, polymerase chain reaction; rRNA, ribosomal RNA; *rrnL*, large (16S) rRNA subunit (gene); *rrnS*, small (12S) rRNA subunit (gene); T, thymine; tRNA-Abc (where Abc is replaced by three letter amino acid code of the corresponding amino acid), transfer RNA; *trnX* (where *X* is replaced by one letter amino acid code of the corresponding amino acid), tRNA gene; T-arm, T-loop and T-stem of a tRNA secondary structure.

5.7 Acknowledgements

We thank G. Giribet for his kind provision of the *O. parietinus* specimen.

6. A new mitochondrial genome sequence of a member of Scorpiones (*Hadogenes bicolor*) highlights peculiar features in the reversed nucleotide composition and the non-coding region of this taxon

6.1 Abstract

Scorpiones still have a highly disputed phylogenetic position within the Chelicerata. Some authors regard them as true arachnids either early branching as sister group to the remaining Arachnida or together with Opliliones, Pseudoscorpiones, and Solifugae in a clade named Dromopoda. However, also a basal position within the Chelicerata or a sister group relationship to the marine Xiphosura was discussed, thereby questioning monophyly of Arachnida.
Here we present the mitochondrial genome of the scorpion *Hadogenes bicolor*, which mainly resembles a situation known for most Bilateria in terms of genome organisation, content, and structure. One difference concerns the occurrence of an additional copy of *trnW* and identical sequences flanking both those tRNA copies. We assume this slightly derived gene order as a result of a single local rearrangement consistent with the tandem duplication/ random deletion model. But, as in general, the gene order is conserved among scorpions, no phylogenetic affinity of this taxon to any other arachnid clade can be derived from gene order. Another peculiarity is found regarding the unusual nucleotide composition of the genome. Just like in all other scorpions, also in *H. bicolor* an exceptional dispersal of positively and negatively skewed protein-coding and rRNA genes is found: all genes encoded on the (+) strand as well as (-)strand encoded *nad1* and the two also (-)strand encoded rRNA genes show a negative CG-skew while the remaining majority of protein-coding genes encoded on the (-)strand have a positive CG-skew. A sliding window analysis of CG-skew along the plus-strand genome sequence demonstrates that upstream the ARNSEF-cluster of tRNA genes the CG-skew is gradually decreasing to a clear negative value (control region: -0.4) while it is positive upstream the control region up to the ARNSEF-cluster. Since the CG-skew is normally stable (positive or negative) over the complete genome, the uncommon change of CG-skew in scorpion mitochondrial genomes implicates a special importance of the ARNSEF-cluster during mitochondrial genome replication.

6.2 Background

Extant scorpions (comprising 1279 described species in 16 families; Fet et al. 2000) are terrestrial predatores inhabiting tropical and temperate parts all over the world. Scorpions are the oldest arachnid group with a fossil record going back to the early Silurian (*Dolichophonus loudonensis*, Laurie 1889). Regarding the Arachnida, a diverse taxon of predominatly terrestrial chelicerates, scorpions have a controversial phylogenetic position. Besides the highly disputed opinion that scorpions are more closely related to the fossil eurypterids than to the rest of the arachnids (e.g. Grasshoff 1978) there also is, considering scorpions as true arachnids, a discussion of the phylogenetic affinities of scorpions within arachnids. From some authors, based on morphological data, a basal position of the Scorpiones as sister taxon of the remaining arachnids is supported (e.g. Firstman 1973; Selden and Dunlop 1998; Weygoldt 1998; Weygoldt and Paulus 1979). Regarding molecular and/or morphological data, others favour a close relationship of scorpions to Opiliones (harvestman), Pseudoscorpiones and Solifugae (camel spiders), together forming the clade Dromopoda (e.g. Giribet et al. 2002; Shultz 1990; Wheeler and Hayashi 1998). Van der Hammen (e.g. 1989) even hypothesized a sister-group relationship of scorpions and xiphosurans and by this denied a monophyletic origin of arachnids. In a more recent analysis of Hassanin (2006) based on molecular data, scorpions come out to be one of the most primitive chelicerates which also questions the monophyly of arachnids. So both morphological as well as sequence-based and combined analyses produced conflicting hypotheses of arachnid phylogenetic relationships which thus remain unresolved.

Due to the fast-growing amount and easily accessible data, molecular sequences provide a valuable source of data for analysing animal phylogeny (Dunn et al. 2008; Halanych 2004; Mallatt and Winchell 2002). Concerning the use of molecular data to infer phylogenetic relationships, mitochondrial (mt) genomes have become a popular tool of molecular sequence analysis during the last years. This is to explain not at least by the small size and the more simple organisation of mt genomes, which usually simplifies their retrieval compared to the nuclear genome. In metazoan animals mt genomes are single circular double strand DNA molecules, that generally range in size from 13 to 19 kb (Saccone et al. 1999) and contain 13 protein-coding genes, two ribosomal RNA (rRNA) genes, 22 transfer RNA (tRNA) genes (Boore 1999), and at least one major AT-rich non-coding region, referred to be the control region (CR) for transcription and/or DNA replication of the mt genome (Taanman 1999; Wolstenholme 1992). Mt genomes offer a large variety of phylogenetically useful information which could be used as discrete or additional characters to resolve metazoan phylogeny. The comparison of gene order is used on different taxonomic levels. It often remains unchanged over long periods of time and by this is taxon specific (Boore et al. 1995; Shao et al. 2004). On the other hand, distinctive changes in the gene order can characterise intra-taxonomic

clades due to the high number of possible recombinations and restrictions for a succesful gene rearrangement (Dowton et al. 2002) which both make convergent changes rather unlikely - so changes in gene order are valuable information for phylogenetic questions (Boore et al. 1998; Lavrov et al. 2004). Apart from the gene order, mt genomes also provide sequence information in the form of amino acid or nucleotide sequences of protein-coding and rRNA genes to be analysed in aspects of animal phylogeny (Carapelli et al. 2007; Fahrein et al. 2007; Hassanin 2006; Podsiadlowski and Braband 2006). Also comparative work concerning the control region, secondary structures of rRNA and tRNA genes, codon usage patterns, and the nucleotide compositional strand-bias of protein-coding and rRNA genes is more and more upcoming for phylogenetic studies (e.g. Choi et al. 2007; Fahrein et al. 2009; Haen et al. 2007; Masta et al. 2009; Masta and Boore 2008).

Here, we describe the mitochondrial genome of the giant banded flat rock scorpion *Hadogenes bicolor* (Liochelidae) and compare it to the five published scorpionid genomes (four Buthidae and one Chactidae). Furthermore, inferred secondary structures of scorpionid tRNAs are discussed with regard to truncated and completely missing tRNA genes. Also, characteristics in scorpionid non-coding regions and the nucleotide compositional bias of protein-coding genes are highlighted.

6.3 Results & Discussion

Genome organisation

The mitochondrial genome of *Hadogenes bicolor* is found to be a common circular duplex molecule with a length of 15083 bp and an AT content of 70,39%. In bilaterian mt genomes usually 37 genes can be detected (Boore 1999): 13 genes coding for protein and two genes coding for ribosomal subunits as well as 22 tRNA genes. In the mt genome of *H. bicolor* all of these genes were identified plus one additional sequence copy encoding tRNA-W (Figure 6.1, Supplementary file 6.1). Both of these sequences are identical and situated between *nad2* and *trnC* (first copy - $trnW^1$) respectively *trnC* and *trnY* (second copy - $trnW^2$). Moreover, also the flanking nucleotides of the two *trnW*s display exactly the same sequences: the last 114 nucleotides of *nad2* and the 112 nt long non-coding sequence between *trnC* and $trnW^2$ are identical, just like the first 31 nt of *trnC* and *trnY*. The two nt shorter pseudogene sequence (between *trnC* and $trnW^2$) can be explained by the two nt gene overlap between *nad2* and $trnW^1$. In this context also noticeable is a 42 nt long inverted repeat sequence in the second half of both, the pseudogene and the end of *nad2*, which can be folded into a stable stem-loop structure. The stem is composed of 15 paired nucleotides with one unpaired nucleotide and the loop consists of 11 nucleotides (Figure 6.1 c, Figure 6.2 h).

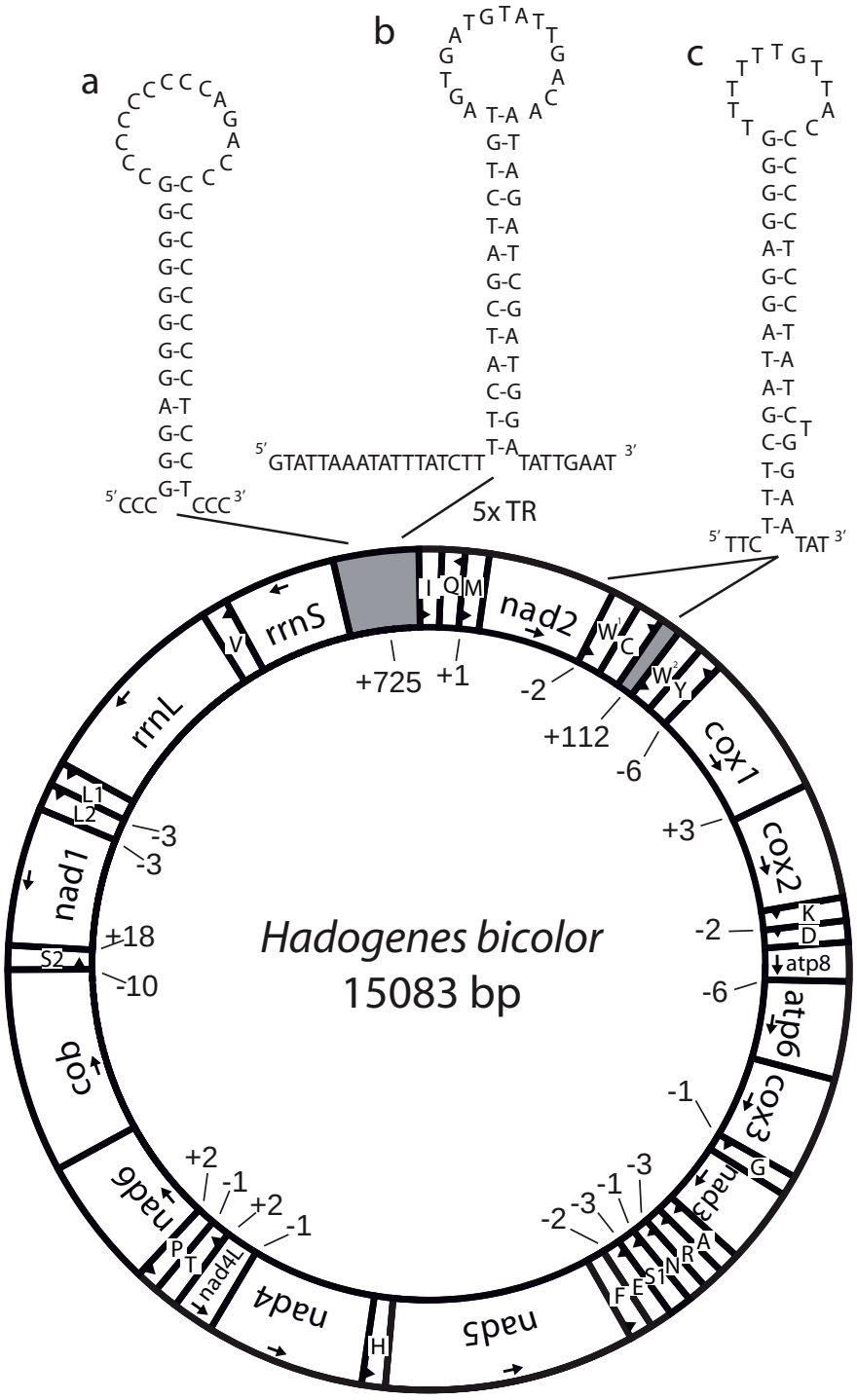


Figure 6.1: Mitochondrial genome and stem-loop secondary structure of *Hadogenes bicolor*. tRNA genes are depicted by their one-letter code abbreviations. Numbers reflect non-coding (positive) or overlapping (negative) nucleotides between two adjacent genes; small arrows indicate the orientation of the genes; grey shaded parts illustrate non-coding sequences. Above the circular map stem-loop structures found in the largest non-coding part (a & b) and near the end of *nad2* and a smaller non-coding region (c) are shown. 5x TR means, that (b) is repeated 5 times.

In addition to the mentioned non-coding sequence between *trnC* and *trnW²* we found one major non-coding sequence flanked by *rrnS* and *trnI*. This putative control region is 725 bp in length and comprises several inverted repeats discussed below. Besides these two only few and much smaller non-coding sequences occur, extending up to 18 bp between *trnS²* and *nad1*. The vast majority of intergenic sequences are overlaps at 14 gene boundaries ranging from one to ten shared nucleotides between *cob* and *trnS²* (Figure 6.1, Supplementary file 6.1).

Except from the occurrance of *trnW²*, the observed gene order is identical to that found in *Limulus polyphemus* (Xiphosura), which is supposed to reflect the ancestral arthropod ground pattern (Lavrov et al. 2000; Staton et al. 1997). We see the gene order in *H. bicolor* as result of a single local rearrangement consistent with the duplication/random loss model (Macey et al. 1997; Moritz et al. 1987) as the most parsimonious explanation with strong evidence from the pseudogene (Fujita et al. 2007; Macey et al. 1998) detected between *trnC* and *trnW²*.

Control region

The major non-coding region (lnr or control region) has a length of 725 bp and is located between *rrnS* and *trnI* (Figure 6.1, Supplementary file 6.1). Although the relative location of the lnr varies greatly among invertebrates, its location found in *Hadogenes*, is conserved between arthropods (Boore 1999). The lnr has an AT content of 69,24% which is lower than that of the rest of the mt genome (70,45%). A comparable situation is found in all other scorpions (and some Acari), which also exhibit a lower AT content in the lnr than in the whole genome (Choi et al. 2007).

Within the lnr we could identify several inverted repeat sequences, each capable of forming a stable stem-loop structure. The first one is a single inverted repeat with a length of 38 bp at position 84-121 (Figure 6.1 a, Figure 6.2 f and 6.2 g) followed by a fivefold inverted repeat, each of 67 bases length and the first repeat overstretching position 293-359 (Figure 6.1 b). None of the stem-loop structures are flanked by a poly-A or poly-T strech which are associated with the origin of replication (OR) in insect mitochondrial DNA (Saito et al. 2005; Zhang and Hewitt 1997). Other structural elements known from arthropod mitochondrial control regions are GA(A)T- or TA(A)-motifs (Black and Roehrdanz 1998; Zhang and Hewitt 1997; Zhang et al. 1995), which presumably play an important role in the replication and/or transcription of the mitochondrial genome. Both are present flanking the stem-loop structures of the fivefold inverted repeat - the TA(A)-motif downstream and the GA(A)T-motif upstream of each sem-loop (Figure 6.1 b). Multiple inverted repeats in the mitochondrial control region are also known from other scorpions. Choi et al. (2007) pictured stem-loop structures of inverted repeats found in two *Mesobuthus* species and pointed out the similarity between them (Figure 62 i and 6.2 j). As in *Mesobuthus martensii* the first

inverted repeat (*Mm*1) slightly differs from the adjacent ones (*Mm*2), both sequences are considered when comparing sequence similarities (*Mm*1+*Mg*: 75,86%; *Mm*2+*Mg*: 68,85%). In *Buthus occitanus* a fourfold inverted repeat, very similar to the *Mesobuthus* repeats, is present (sequence similarity: *Bo*+*Mm*1: 83,05%; Bo+*Mm*2: 75%; *Bo*+*Mg*: 74,58%). The first repeat has a length of 59 nt and is capable of forming a stable stem-loop structure (Figure 6.2 a). Another buthid scorpion, *Centruroides limpidus*, also displays three different inverted repeat sequences in the control region (Figure 6.2 b+c). One of these single inverted repeats (14334-14392, Figure 6.2 c) has a slightly increased sequence similarity to the inverted repeats of the other buthids (*Bo*: 41,94%; *Mg*: 46,67%; *Mm*1: 42,62%; *Mm*2: 41,27%). It is possible that this single inverted repeat in *C. limpidus* and one of the manifold inverted repeats respectively in *B. occitanus* and the two *Mesobuthus* species are a result from shared sequence evolution. Figure 6.3 shows an alignment of the manifold inverted repeats (in all but *C. limpidus*) which are found in buthid control regions. Further inverted repeats are found in the control region of the chactide *Uroctonus mordax* (Figure 6.2 d+e), but neither the single nor the fourfold repeated one has a considerable sequence similarity to the inverted repeat of any of the buthid scorpions (>37%; Table 6.1). The same is true for the single inverted repeat found in *H. bicolor*, while the fivefold inverted repeat shows a slightly heightened sequence similarity to three buthids (*Bo*: 43,48%; *Mm*2: 44,93%; *Mg*: 44,77%). The latter sequence similarity might have emerged due to convergent or shared sequence evolution in buthids and *H. bicolor* while a manifold repetition of the inverted repeat more likely evolved independently, at least two times in each lineage leading to a clade exhibiting tandem repeats. But if we have a concerted sequence evolution of the tandem repeats of *Buthus* and the *Mesobuthus* species or if there are other structural features or mechanisms present which enhance sequence duplication is not clear. However, another elevated sequence similarity of 58,14% is found in the single inverted repeats of *U. mordax* and *H. bicolor* (*Hb*: 12938-12975, Figure 6.1 b and Figure 6.2 f; *Um*:14347-14394, Figure 6.2 d). This similarity might be caused due to distinctive guanine and cytosine stretches in both sequences rather than to a common origin in evolution.

Protein-coding genes

From the 13 identified protein-coding genes in the mt genome of *H. bicolor*, all but one start with for arthropod mt DNA usual start codons (ATG, ATT or ATA). Only *nad2* begins with the exceptional codon GTT (Supplementary file 6.1). Five genes (*cox2*, *nad3* - *nad6*) terminate with the truncated stop codon T, the remaining genes end with one of the common stop codons TAA or TAG. Incomplete stop codons are known from several other species and assumed to be completed to functional ones by post-transcriptional polyadenylation (Ojala et al. 1981).

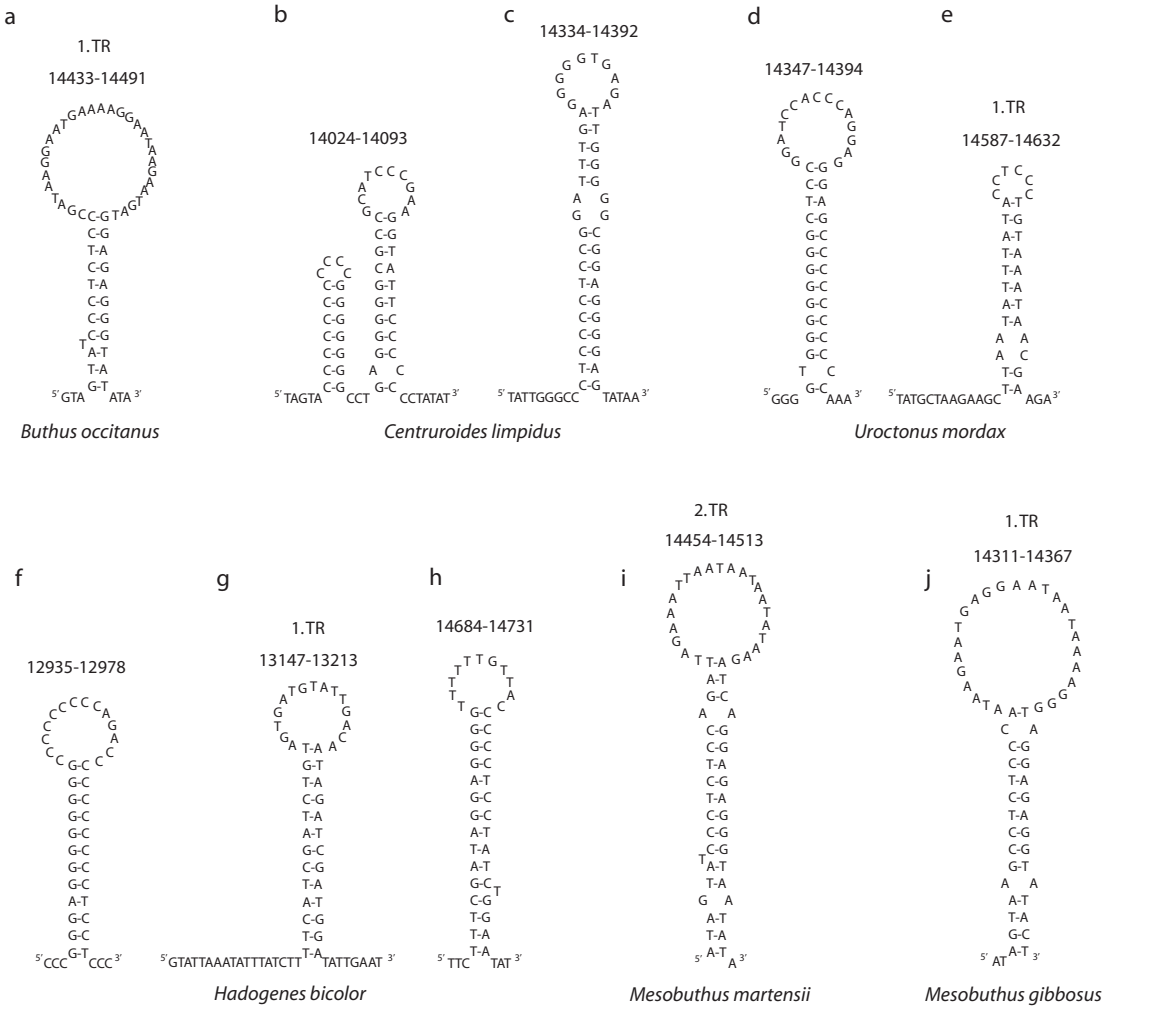


Figure 6.2: Stem-loop structures found in the largest non-coding part (= the putative control region, CR) of the mitochondrial genomes of the scorpions *Buthus occitanus* (a), *Centruroides limpidus* (b + c), *Uroctonus mordax* (d + e), *Mesobuthus martensii* (i), and *Mesobuthus gibbosus* (j), as well as stem-loops found in non-coding parts of *Hadogenes bicolor* (f - h). The numbers above stand for nucleotides and reflect the position within the respective genome. In case of *B. occitanus*, *U. mordax*, *H. bicolor*, and the two *Mesobuthus* species the first (1. TR) or the second (2.TR) inverted repeat of tandem repeats is illustrated.

```
                10        20        30        40        50        60
      ....|....|....|....|....|....|....|....|....|....|....|....|...
Mm1   ---TTAGTATGCCTCTCCCGATTAG-AATGAGAAAGAATAAGAAGGAAGGAGAGG-CAATATA
Mm2   ---A......C.......A......A...T.AT..T....TA.GATC........GT......
Mg    ---A.....A.........A..A..-.......G..T.....A.G..T........-T.-..CT
Bo    ---G......C...........A..G.....A..G.........T.T........GT.-....
Cl    TAT.GG.CCCT..C....G....T.AGGG.GT.G...T.GG.GGC.G...G...TAT..----
```

Figure 6.3: Comparison of aligned inverted repeats found in the control region of buthid scorpiones. The alingmnet shows the first repeat unit of the tandem repeat in *B. occitanus* and *M. gibbosus* (*Bo* and *Mg*), the first (*Mm*1) and the second (*Mm*2) repeat unit of the tandem repeat *in M. martensii*, and a single repeat unit in *C. limpidus*. Dots indicate nucleotides identical to those of the first line in the alignment.

Table 1: Sequence similarities of inverted repeats in scorpionid control regions (TR marks multiple copies of inverted repeats, in each case the first inverted repeat is considered for similarity analyses; species are abbreviated by the first letter of each genus and epithet; bold numbers indicate similarities >40%).

Species	*Bo* (TR)	*Mm*1 (TR)	*Mm*2 (TR)	*Mg* (TR)	*Cl* (single)	*Um* (TR)	*Um* (single)	*Hb* (TR)
Bo (TR)								
*Mm*1 (TR)	**83,05**							
*Mm*2 (TR)	**75**	**73,33**						
Mg (TR)	**74,58**	**75,86**	**68,85**					
Cl (single)	**41,94**	**42,62**	**41,27**	**46,67**				
Um (TR)	27,03	25,35	30,64	36,21	25,42			
Um (single)	23,73	22,41	20,97	19,67	25,81	20		
Hb (TR)	**43,48**	34,32	**44,93**	**44,77**	36,76	34,29	14,92	
Hb (single)	15,39	13,79	15,15	17,24	20,97	21,74	**58,14**	10,45

rRNA genes

The genes for the mitochondrial encoded ribosomal subunits in *H. bicolor* have a length of 791 bp (*rrnS*) and 1162 bp (*rrnL*), which lies within the size range of rRNA genes in chelicerates. Compared to the rRNA genes of the horseshoe crab *L. polyphemus* (*rrnS*: 799 bp and *rrnL*: 1294 bp; Lavrov et al. 2000), the *rrnL* gene in *Hadogenes* is considerably shorter while the *rrnS* gene has a similar length. In comparison with rRNA genes of other scorpions, the inferred length of *rrnS* and *rrnL* in *Hadogenes* is respectively similar. The location on the (-)strand encompassing *trnV*, and surronded by *trnL*^{1} downstream of *rrnL* and a large non-coding region upstream of *rrnS* is equal to other scorpionid mt genomes and reflects the putative arthropod ground pattern (Figure 6.4).

In animal mitochondria the motif TGGCAGA is believed to be the termination signal for *rrnL* transcription (Valverde et al. 1994). This heptamer is located downstream of *rrnL* either in the tRNA-L^{1}(CUN) gene (in insects and crustaceans) or in the tRNA-L^{2}(UUR) gene (vertebrates). In *H. bicolor* we find a modification of this motif, TAGCAAA, in both genes coding for leucine (Supplementary file 6.2). We recovered the same conditions in the mt genomes of the buthid scorpion *C. limpidus* (Davila et al. 2005) and the iruid scorpion *U. mordax* (Masta and Boore 2008), whereas in the mt genomes of the two published *Mesobuthus* species (*M. martensii*: Choi et al.2007; *M. gibbosus*: Jones et al. 2007) and *B. occitanus* (Masta and Boore 2008) the identic modificated motif is only found in the tRNA-L^{2}(UUR) gene. The tRNA-L^{1}(CUN) gene of these three buthid scorpions possesses the conserved motif TGGCAGA. In arthropods, a modification of the motif TGGCAGA in the tRNA-L^{1}(CUN) is also known from e.g. the amblypygid *Phrynus* sp. (TGACAGA - Fahrein et al.2009), the sun spider *Eremobates* cf. *palpisetulosus* (TAGCAGA - Masta et al. 2008), the opilionid *Opilio parietinus* (TGTCAGA - Chapter 5) and the honeybee *Apis melifera*

(TAGCATA - Valverde et al. 1994), thus it appears that our findings are not unusual. For Scopiones, it is most parsimonious to conclude that the modificated transcription termination motif found in *C. limpidus*, *U. mordax* and *H. bicolor* represents the primary character state whereas the seemingly conserved motif found in the tRNA-L[1](CUN) gene of the remaining buthid scorpions is derived. This is congruent with current hypotheses about phylogenetic relationships within the Buthida (e.g. Fet et al. 2003).

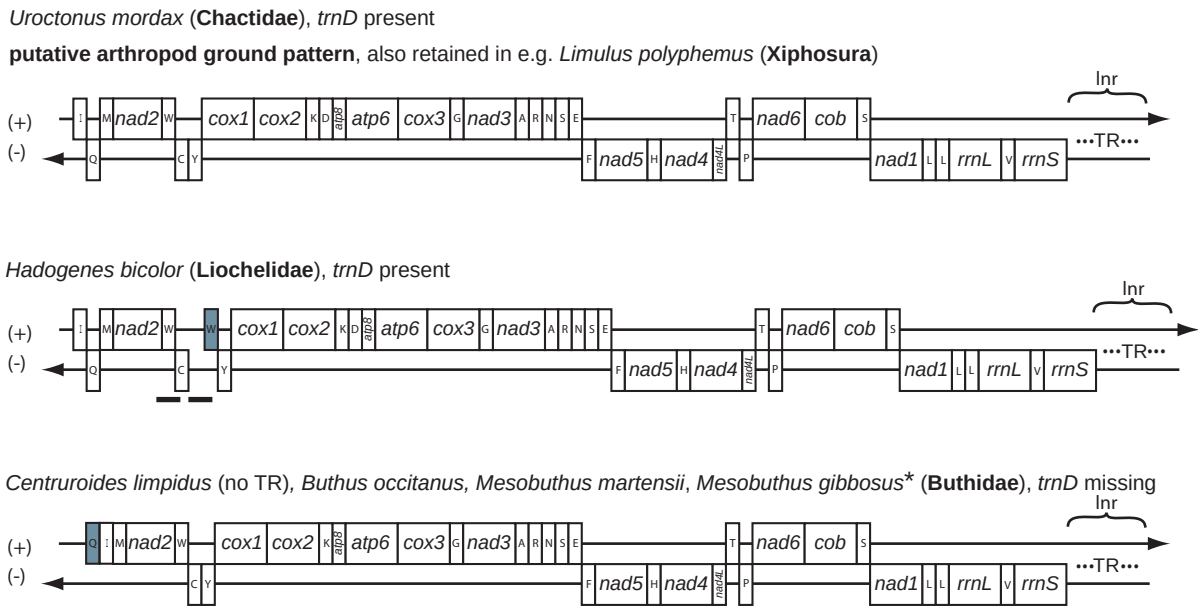


Figure 6.4: Gene order in linearised mitochondrial genomes of scorpions. Transfer RNA genes are labelled according to the one letter amino acid code. Genes marked white show the same relative position as in the arthropod ground pattern; genes marked orange have relative positions differing from the arthropod ground pattern. Black horizontal lines under the genome of *H. bicolor* illustrates duplicated regions. (*) The gene order of *M. gibbosus* used in this study differs from that in Jones et al. (2007) and the entry at NCBI [NC_006515].

Nucleotide composition

In *H. bicolor* all genes encoded on the (+)strand show a negative CG-skew while the majority of protein-coding genes encoded on the (-)strand have a positive CG-skew. Only (-)strand encoded *nad1* is negative CG-skewed, just like the two also (-)strand encoded rRNA genes are (Table 6.2). The same exceptional dispersal of positively and negativly skewed protein-coding and rRNA genes is found in all other scorpiones (Table 6.2). A sliding window analysis of CG skew along the plus-strand genome sequence (Figure 6.5 a) demonstrates that there is positive CG skew (scattering around 0.4) in the part of the genome ranging from

nad2 to the ARNSEF-cluster of tRNA genes, while starting from this cluster downstream to the control region the CG skew is gradually decreasing to a clear negative value (control region: -0.4). In contrast the majority of arachnids (Figure 6.5 b gives four examples) have a stable negative CG skew in the complete genome, with only slight differences between the genome regions. Opisthothelae spiders show a complete reversal of CG skew in all parts of the genome, probably due to a reversal of the control region. The uncommon change of CG skew in scorpion mt genomes implicates a special importance of the ARNSEF-cluster during mt genome replication. While in *Drosophila* plus- and minus-strand replication origin are both located in the non-coding control region, this is different e.g. in vertebrates, where minus-strand origin is located several kb apart from the plus-strand origin, in a tRNA cluster. This may be similar in scorpions, but needs further detailed studies for mt genome replication in Scorpiones.

Table 6.2: Nucleotide frequencies (CG-skew: C-G/C+G) of all protein-coding genes plus RNL and RNS (behind the genes the respective coding strand is listed; the bold numerary indicates unusual CG-skews).

species / gene	*U. mordax*	*H. bicolor*	*B. occitanus*	*C. limpidus*	*M. gibbosus*	*M. martensii*	*L. polyphem.*	*V. destructor*
ATP6 (+)	-0,243	-0,414	-0,427	-0,409	-0,398	-0,447	0,526	-0,173
ATP8 (+)	-0,484	-0,689	-0,640	-0,544	-0,708	-0,692	0,721	**0,077**
Cox1 (+)	-0,167	-0,326	-0,316	-0,288	-0,299	-0,354	0,210	-0,199
Cox2 (+)	-0,284	-0,444	-0,392	-0,391	-0,368	-0,334	0,331	-0,238
Cox3 (+)	-0,284	-0,462	-0,414	-0,408	-0,413	-0,447	0,334	-0,255
CYTB (+)	-0,099	-0,125	-0,078	-0,092	-0,022	-0,020	0,427	-0,162
ND1 (-)	**-0,132**	**-0,066**	**-0,095**	**-0,057**	**-0,083**	**-0,043**	-0,511	0,044
ND2 (+)	-0,174	-0,308	-0,516	-0,444	-0,534	-0,536	0,570	-0,452
ND3 (+)	-0,383	-0,534	-0,545	-0,609	-0,619	-0,615	0,524	-0,347
ND4 (-)	0,132	0,313	0,443	0,470	0,397	0,477	-0,485	0,153
ND4L (-)	0,016	0,014	0,131	0,297	0,105	0,273	-0,637	**-0,077**
ND5 (-)	0,255	0,422	0,459	0,488	0,440	0,486	-0,396	0,264
ND6 (+)	-0,236	-0,391	-0,160	-0,268	-0,334	-0,353	0,629	-0,143
RNL (-)	**-0,140**	**-0,232**	**-0,221**	**-0,194**	**-0,205**	**-0,238**	-0,447	0,093
RNS (-)	**-0,209**	**-0,271**	**-0,330**	**-0,337**	**-0,366**	**-0,364**	-0,330	0,338

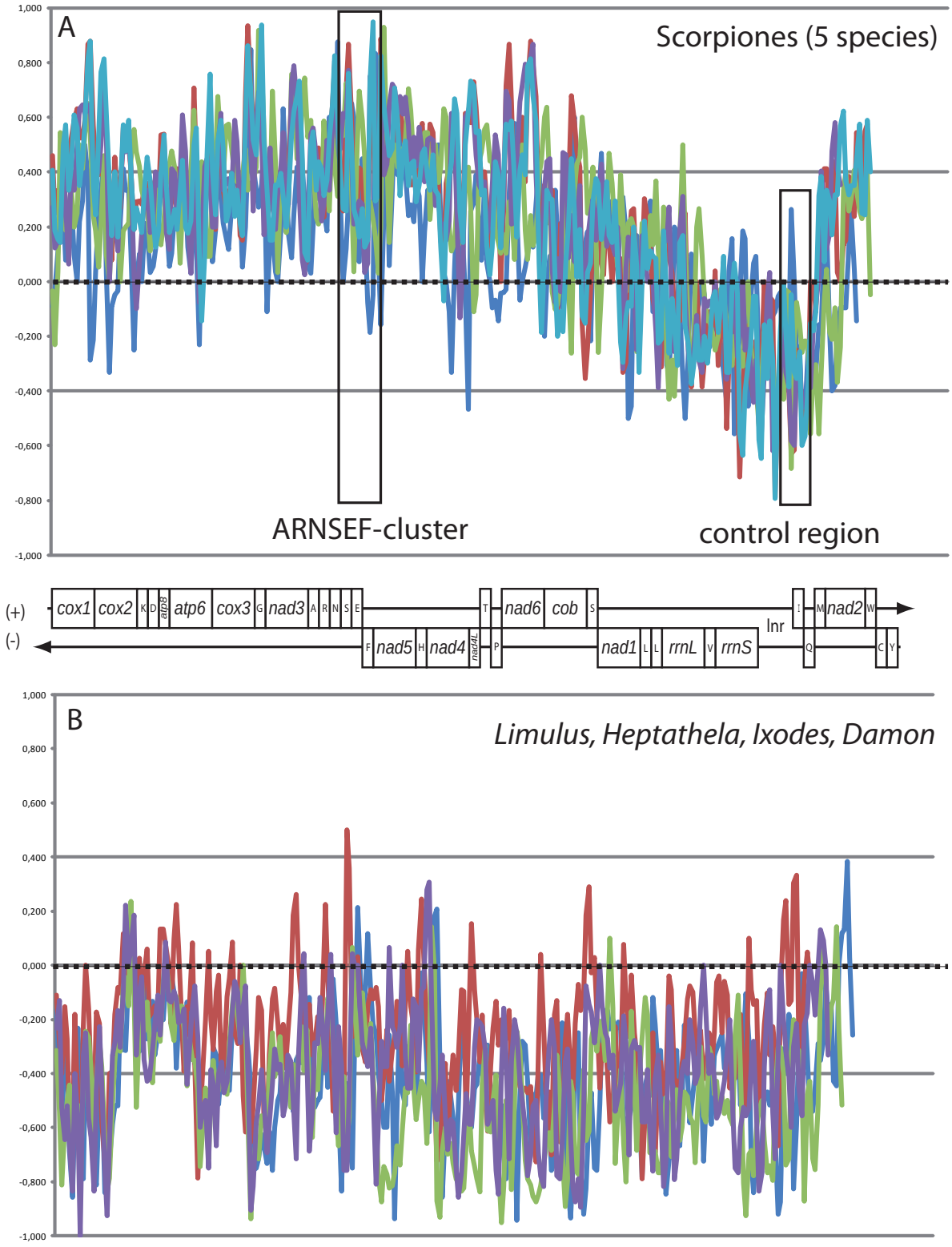


Figure 6.5: Sliding window analysis of CG skew (CG skew of 100 nucleotides was calculated, with a step width of 50 nucleotides). A) CG skews along plus-strand mt genome of five scorpion species. Below a almost proportional linear map of the mt genome is given. B) CG skew along plus-strand mt genomes of four chelicerate species with a “normal” negative CG skew.

Secondary structure of tRNAs

In *H. bicolor* all 22 tRNAs typically present in bilaterian mt genomes are present plus one additional copy of tRNA-W (= *trnW*^{2}). The vast majority of the inferred tRNA secondary structures (Supplementary file 6.2) resembles a canonical cloverleaf shape, which is known to be highly conserved among organisms. Usually it consists of a T-arm, a D-arm, a anticodon arm bearing the anticodon, and an aminoacyl acceptor stem but for some metazoans it is already known that their mt tRNAs lack the sequences for the D-arm and/or the T-arm (e.g. Boore and Brown 2000; Helfenbein et al. 2001). In *Hadogenes* only the sequences coding for tRNA-H and tRNA-S1(AGN) are not capable of being folded into the typical cloverleaf structure. In tRNA-H the sequence for the T-arm is missing while tRNA-S1(AGN) lacks, as in many other metazoans (Wolstenholme 1992), the D-arm sequence. Surprisingly, for the buthid scorpions *C. limpidus* and *M. gibbosus* a D-arm is suggested to exist in the tRNA-S1(AGN), but many mismatches weaken that hypothesis. Concerning tRNA-H, for all scorpions published so far but *M. martensii*, it is asserted that the sequence for the T-arm is missing. Due to a single nucleotide between the anticodon arm and the T-arm which moreover has only two paired bases in its stem, we propose that the tRNA-H in *M. martensii* also lacks the T-arm and the that the variable loop and the T-arm are replaced by a TV-raplacement loop of 9 nt length (Figure 6.6). Nonetheless, this loss is disputatious as a probable synapomorphy shared by all scorpions because it is also shared by some other chelicerates (see Masta and Boore 2008 for a compilation).

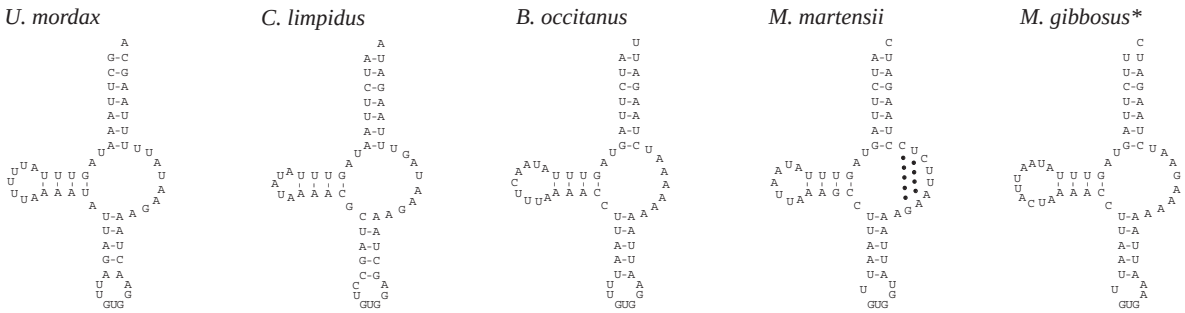


Figure 6.6: Compilation of putative secondary structures of the tRNA histidine in scorpions. The arrow indicates a single nucleotide between the anticodon- and the T-arm of *Mesobuthus martensii* (after Choi et al. 2007) - we prefer a secondary structure with a TV-replacement loop. Also we reannotated the putative secondary structure of the tRNA-H in *M. gibbosus* (*) which differs from that in Jones et al. (2007).

In general, a truncation of tRNA sequences in scorpions, particularly in buthid ones, is obvious (Choi et al. 2007; Davila et al. 2005; Jones et al. 2007; Masta and Boore 2008). Although it seems to us that the situation in *M. gibbosus* (we assume T-arms also in tRNA-G, L1, L2, S1, P, V, M, and Q) and *C. limpidus* (we assume T-arms also in tRNA-M, F, P, S2, Q, and V) is overestimated. By contrast we see an underestimation in *M. martensii* where we favour a missing T-arm at least in tRNA-H, F, and T. Within chelicerates it is also known from several species belonging to the Acari (Shao et al. 2005b; Shao et al. 2006), the Araneae (Masta and Boore 2004; Masta and Boore 2008; Qiu et al. 2005), the Amblypygi (Fahrein et al. 2009; Masta and Boore 2008), and the Thelyphonida (Masta and Boore 2008) that they possess many tRNA genes that lack either D-arm or T-arm sequences. It is suggested that among arachnids these structural changes evolved multiple times independently, but it is not clarified what causes them (Masta and Boore 2008).

Mitochondrial gene order in Scorpiones

Mitochondrial genomes of scorpions reveal a conserved gene order. Figure 6.4 illustrates a compilation of the gene-rearrangements found in scorpions. In the mitochondrial genome of the western forest scorpion *Uroctonus mordax* (Iruida, Chactidae) the gene order is identical to that present in the xiphosuran *Limulus polyphemus* and several other arthropods. Due to the appearance in distinct arthropod clades this gene order is regarded to reflect the putative arthropod ground pattern (Lavrov et al. 2000; Staton et al. 1997). The gene order found in *Hadogenes bicolor* (Iruida, Liochelidae) only differs slightly from the ancestral gene order and can easily be established with one tandem duplication/random loss event, which caused the duplicated *trnW*^{2}.

Before allegorising the situation found in buthid scorpions, we have to point out that after carefully checking the mitochondrial genome sequence of *M. gibbosus* at NCBI [NC_006515] we came to the conclusion that *trnD* is missing. Desisive for this assumption is the position of *trnD* within *rrnS*, which is extremely unusual for a gene, in combination with the inferred secondary structure (Jones et al. 2007), which has many mismatches in both the amino-acyl and the anticodon stem. Our findings are consistent with those of Choi et al. 2007 who also suggested that *trnD* is missing in the mt genome of all four buthid scorpions sequenced so far. The most parsimonious explanation for the missing *trnD* and the inverted and translocated *trnQ* displayed in buthid mt genomes is the assumption of two translocation events in the lineage leading to buthid scorpions. Thus, these two translocations can be seen as synapomorphies of at least a monophyletic clade containing the three buthid groups *Buthus*, *Mesobuthus* and *Centruroides*. Missing genes in mitochondrial genomes of chelicerates are only known from two other arachnids: in *Metaseiulus occidentalis nd3* and *nd6* are missing (Jeyaprakash and Hoy 2007) while in *Phalangium opilio* (Masta and Boore

2008) no tRNA-L1(CUN) can be identified. In the mt genome of several metazoan animals a convergent loss of *atp8* is proposed (*Mytilus edulis*: Hoffmann et al. 1992; Plathyelminthes: Le et al. 2000; Nematoda: Okimoto et al. 1992) and it is assumed that the function is taken over by another ATPase subunit. An alternative explanaition for missing genes in mt genomes is their transfer to the nuclear genome which than requires an import mechanism for gene products into the mitochondrion.

6.4 Concluding remarks

As the gene order is conserved in Scorpiones, no phylogenetic affinity of this taxon to any other arachnid clade can be concluded by it. However, the few but present translocations of tRNA genes underline their helpful character to infer or strengthen intra-taxonomic relationships of Scorpiones. In the control region there is a unique repeat structure in several scorpion species. Further comparative studies may reveal the origin and evolution of these repreats, probably in conjunction with the plus-strand replication origin.There is also a unique change in CG skew over the mt genome of scorpions, probably indicating a replication scheme different from other arthropods, but this hypothesis needs some experimental support.

6.5 Material and Methods

Specimen and DNA extraction

Total DNA of the specimen of *H. bicolor* (Purcell 1899) was extracted from one leg by using Qiagen extraction kits (Qiagen, Hilden, Germany) following the manufacturers` protocol.

PCR conditions, primers and sequencing

Via initial PCRs, the whole mitochondrial genome of *H. bicolor* was amplified in two overlapping fragments by using the primer pairs Art-HPK16SA with Art-HPK16SB (Kambhampati et al. 1995; Simon et al. 1994) and Art-HPK16Saa with Art-HPK16Sbb (Hwang et al. 2001b). The long amplification product, produced by the latter primer pair, was used as a template for 454 pyrosequencing. This method was developed by Ronaghi (1996) and is based on the principle “sequencing by synthesis”. Unlike the detection of chain termination with didesoxynucleotides as in the case of the DNA sequencing method developed by Sanger (1977), in pyrosequencing the pyrophosphate release on nucleotide incorporation is detected. The initial PCR of the long fragment was performed on an

Eppendorf Mastercycler with Takara LA Taq kit (Takara) in 50 µl volumes (5µl buffer, 8 µl dNTP solution, 0.5 µl Takara LA Taq, 1µl DNA, 1µl primer mix (10 µM), 34.5 µl water). Reaction conditions were: Initial denaturation (1 min, 94°C); 40 cycles of denaturation (1 min, 94°C), primer annealing (1 min, 65°C), elongation (15 min, 65°C); final elongation (10 min, 65°C).

Afer 454 pyrosequencing taxon-specific primer pairs were designed to amplify missing fragments in order to bridge the gaps between 16S and 12S rRNA as well as the gaps within Cox1 and within the control region (CR) (Supplementary file 6.3). These secondary PCRs were performed on an Eppendorf Mastercycler and Mastercycler gradient using the Eppendorf 5-prime-Taq kit (Eppendorf, Germany) in 50 µl volumes (41,75 µl molecular grade water; 5 µl buffer; 1 µl dNTP mix, 10 µM; 1 µl template DNA (= 1:100 dilution of the long PCR fragment); 1 µl primer mix, 10 µM each; 0,25 µl Taq polymerase). Conditions for the secondary PCRs were: initial denaturation (94°C for 1 min), 30 cycles of denaturation (94° C for 30 sec), annealing (30 sec, primer specific temperature according to Supplementary file 6.3) and elongation (68°C for 1 min, respectively 3 min in case of the larger fragments), followed by a final elongation step (68°C for 1 min, respectively 3 min in case of the larger fragments).

PCR products were visualized on 1% agarose gels and purified using the Bluematrix DNA purification kit (EURx, Gdansk, Poland). If extra bands were present, a Gel extraction was performed, following the manufacturers` protocol (Qiagen). Purified templates were sent to LGC AGOWA genomics, where DNA sequencing according to the dideoxy chain-termination method of Sanger was performed. Each sample contained 10 µl template DNA and 4 µl Primer (1:10 dilution).

Gene annotation

Sequence assembly from sequences obtained through both DNA sequencing methods was performed with BioEdit 7.0.1. (Hall 1999). To determine the protein-coding and ribosomal gene boundaries BLAST searches on NCBI Blast databases were performed. The sequences were also compared with alignments from other chelicerate species. We assumed the start and ending of the ribosomal RNA genes and the control regions extended to the boundaries of flanking genes. The boundary of the 12S rRNA gene to the control region was inferred by comparison with 12S rRNA genes of other arachnids. Hairpin structures in the control regions were identified by eye inspection. Genomic position and secondary structure of tRNAs were identified by tRNAscan-SE (Lowe and Eddy 1997) and ARWEN (Laslett and Canbäck 2008). Sequence data was deposited at NCBI database [Genbank: XXX]. Nucleotide

frequencies and sequence similarities were determined using BioEdit 7.0.1. (Hall 1999). Sliding window analysis of CG skew was performed with the GenSkew tool (http://www.helmholtz-muenchen.de/en/mips/services/analysis-tools/genskew/index.html). A sliding window of 100 nucleotides was calculated for CG skew along the complete genome sequence, in steps of 50 nucleotides.

6.6 Abbreviations

A, adenine; *atp6* and *8*, genes encoding ATPase subunit 6 and 8; bp, base pairs; *cox1-3*, genes encoding cytochrome oxidase subunits I-III; *cob*, gene encoding cytochrome b; C, cytosine; D-arm, D-loop and D-stem of a tRNA secondary structure; G, guanine; mt genome, mitochondrial genome; *nad1-6* and *nad4L*, genes encoding NADH dehydrogenase subunits 1-6 and 4L; nt, nucleotide; PCR, polymerase chain reaction; rRNA, ribosomal RNA; *rrnL*, large (16S) rRNA subunit (gene); *rrnS*, small (12S) rRNA subunit (gene); T, thymine; tRNA-X (where X is replaced by one letter amino acid code of the corresponding amino acid), transfer RNA; *trnX* (where *X* is replaced by one letter amino acid code of the corresponding amino acid), tRNA gene; T-arm, T-loop and T-stem of a tRNA secondary structure.

6.7 Additional material

Additional files 6.1 to 6.3 can be found in the appendix of this thesis (Chapter 10).

7. Rearrangements and the mitochondrial control region in spiders - mitogenomic data from three further Araneae species and the phylogenetic implication

7.1 Abstract

For this study the entire mitochondrial genomes of three araneid species (*Liphistius erewan* (Mesothelae), *Phyxioschema suthepium* (Opisthothelae, Mygalomorphae), and *Pholcus phalangioides* (Opisthothelae, Araneomorphae)), one out of each major clade, were sequenced. All of the three presented mitochondrial genome sequences are rather short compared to other chelicerates that can possibly be associated with two further findings: I) Especially in the two Opisthothelae most sequences coding for transfer RNAs are reduced in size so that the inferred secondary structures are often lacking the D-arm or the T-arm and fully paired acceptor stems are rare. II) Also exhibited in the investigated species are shortened ribosomal RNA sequences. In all three species we found the 37 genes usually present in the mitochondrial genome of Bilateria, only *P. suthepium* exhibits a second copy of *trnI* pseudogene. In the genome of *L. erewan*, a representative of the Mesothelae, the gene order resembles the putative arthropod ground pattern, also retained in the mitochondrial genome of *Limulus polyphemus*. In both, *P. suthepium* and *P. phalangioides* we detected several rearrangements of which most are also exhibited in the gene order of other Opisthothelae, a similar gene order is only known from the araneomorph *Hypochilus thorelli*. Finally we present a scenario which may explain these identical rearrangements found in one mygalomorph and two araneomorph spiders.

7.2 Background

Araneae (spiders) are a very diverse arachnid group and within the chelicerates probably the most familiar and also best investigated one. The generally most striking morphological characters of the Araneae are taxon specific and probably contributed immensely to the spiders evolutionary success: the ability of silk spinning due to opisthosomal silk glands and spin warts as well as a venom gland in the distal segment of the bipartite chelicerae. Another apomorphic feature is the male pedipalp transformed to a copulatory organ enabling a direct sperm transfer (Ax 1996). Currently there are 109 valid spider families with almost 3800 genera described, comprising more than 41.000 extant species (Platnick 2009). They inhabit a wide range of different habitats from high mountains to deserts and forests in temperate to tropical zones, and are even common in anthropogenic affected areas. Considered as primarily terrestrial arachnids some species spend parts of their lives in marine or limnic waters.
Various analyses produced different hypotheses of phylogenetic relationships among the 11 arachnid orders (Giribet et al. 2002; Shultz 1989; Shultz 1990; Shultz 2007; Shultz and Regier 2000; Van der Hammen 1989; Weygoldt 1998; Weygoldt and Paulus 1979; Wheeler and Hayashi 1998). In most analyses based on morphological data the Araneae group with Amblypygi (whip-spiders) and Uropygi (vinegaroons, often treated as the sister groups Thelyphonida and Schizomida), together forming the clade mostly known as Tetrapulmonata. This group is characterized as a monophylum due to an unusual microtubule arrangement in the sperm axonemes and two-segmented chelicerae hinged ventrolaterally. Controversially seen from many authors are the phylogenetic relationships among these three taxa. There are two major competing hypotheses: the “Pedipalpi” hypothesis (Shear et al. 1987; Shultz 1989; Shultz 1990; Shultz 1999; Shultz 2007) sees Uropygi as sister-group of Amblypygi supported by raptorial pedipalps and antenniform first walking legs appearing in both the taxa, and the “Labellata” hypothesis (Petrunkevitch 1955; Van der Hammen 1989; Weygoldt and Paulus 1979) which favours a closer relationship of Araneae and Amblypygi owing to a postcerebral pharynx and a pedicel in both taxa (Ax 1996). Less divergent are considerations about the phylogenetic relationships of the major clades of the Araneae, traditionally divided into the Mesothelae and the Opitothelae (containing araneomorph and mygalomorph spiders).
In case of morphological data sets a possible explanation for the difficult evaluation of sister group relationships within the Tetrapulmaonata can be the presence of homoplasy or the reduction of anatomical structures. But also nuclear sequence data and combined analyses produce conflicting results (Giribet et al. 2002; Wheeler and Hayashi 1998) which shows the relevance of other or additional data sets for analysing phylogenetic relationships.

During the last years mitochondrial genomes have become a popular tool for reconstructing the phylogeny of animal taxa. As mitochondrial genomes in general are much smaller and more simply structured than nuclear ones, sequence data can be obtained more easily and

often much faster. Thus, the amount of mitochondrial sequence data for analysis to adress phylogenetic questiones is growing rapidly. Beyond that, also stuctural features can be of use. These characters are often considered as genome morphology and besides the comparison of the gene order there is also done e.g. comparative work concerning tRNAs, rRNAs, and the control region (Dowton et al. 2002; Fahrein et al. 2009; Masta and Boore 2008; Oliveira et al. 2007).
Mitochondrial genomics of Araneae lagely support morphological based phylogenetic relationships on an intrataxonomic level. Mesothele spiders show the plesiomorphic gene arrangement in their mitochondrial genomes, which is found in many arthropods. Opisthothele spiders display variations of this putative arthropod ground pattern. Thus, some of these rearrangements in the mitochondrial gene order are considered to be characteristics for opisthothele mitochondrial genomes in general, while the remaining ones are suggested to border araneomorph from mygalomorph spiders. Due to the conserved gene order, at least in parts of the Araneae, it does not help in resolving interrelationships of the Tetrapulmonata.

Here we present complete mitochondrial genomes of three further species of the Araneae: *Liphistius erewan* (Mesothelae), *Phyxioschema suthepium* (Opisthothelae, Mygalomorphae), and *Pholcus phalangioides* (Opisthothelae, Araneomorphae).
As our results differ in one point from the common view of mitochondrial genomes of opisthothele spiders, the actual state of affairs concerning gene rearrangements in the mitochondrial genome of Araneae will be a subject of discussion. In doing so, we then focus on the mitochondrial control region which most likely is associated with some of the gene rearrangements and with the revearsed strand bias found in the mitochondrial genome of opisthothele spiders.

7.3 Results & Discussion

Genome organization

In the mitochondrial genomes of the three investigated Araneae species we found all 37 genes commonly known from mitochondrial genomes of Bilateria (Figure 7.1). Normally present are 13 protein subunits (*atp6+8*, *cob*, *cox1-3*, *nad1-6*, *nad4L*), 22 transfer RNAs and 2 ribosomal RNAs (*rrnL* and *rrnS*) encoded on a doublestranded and usually circular DNA molecule (Boore 1999; Wolstenholme 1992). In *P. suthepium* we detected an additional pseudogene sequence for *trnI* (position 12720 - 12771) upstream of *trnM*. The length of the complete mitochondrial genomes of *L. erewan* (14789 bp), *P. suthepium* (13931 bp), and *P. phalangioides* (14459 bp) is rather small compared to other chelicerates and even within the

for Bilateria typical size of about 13 to 19 kb (Saccone et al. 1999) the detected sequence lengths range in the lower span. In chelicerate mitochondrial genomes sizes smaller than 14000 bp are only known from the mygalomorph spider *Ornithoctonus huwena* (13874 bp) and from some members of the Acariformes (*Tetranychus urticae*: 13103 bp, *Panonychus ulmi*: 13115 bp, *Leptotrombidium akamushi* and *deliense*: 13698 and 13731 bp, *Steganacarus magnus*: 13818 bp). The AT-content in *P. phalangioides* (65.9%) and *P. suthepium* (67.4%) is lower than that in all other chelicerates (Navajas et al. 2002). Only *L. erewan* (67.7%) shows a slightly higher AT-content than *L. polyphemus* (67,6% ; Lavrov et al. 2000).

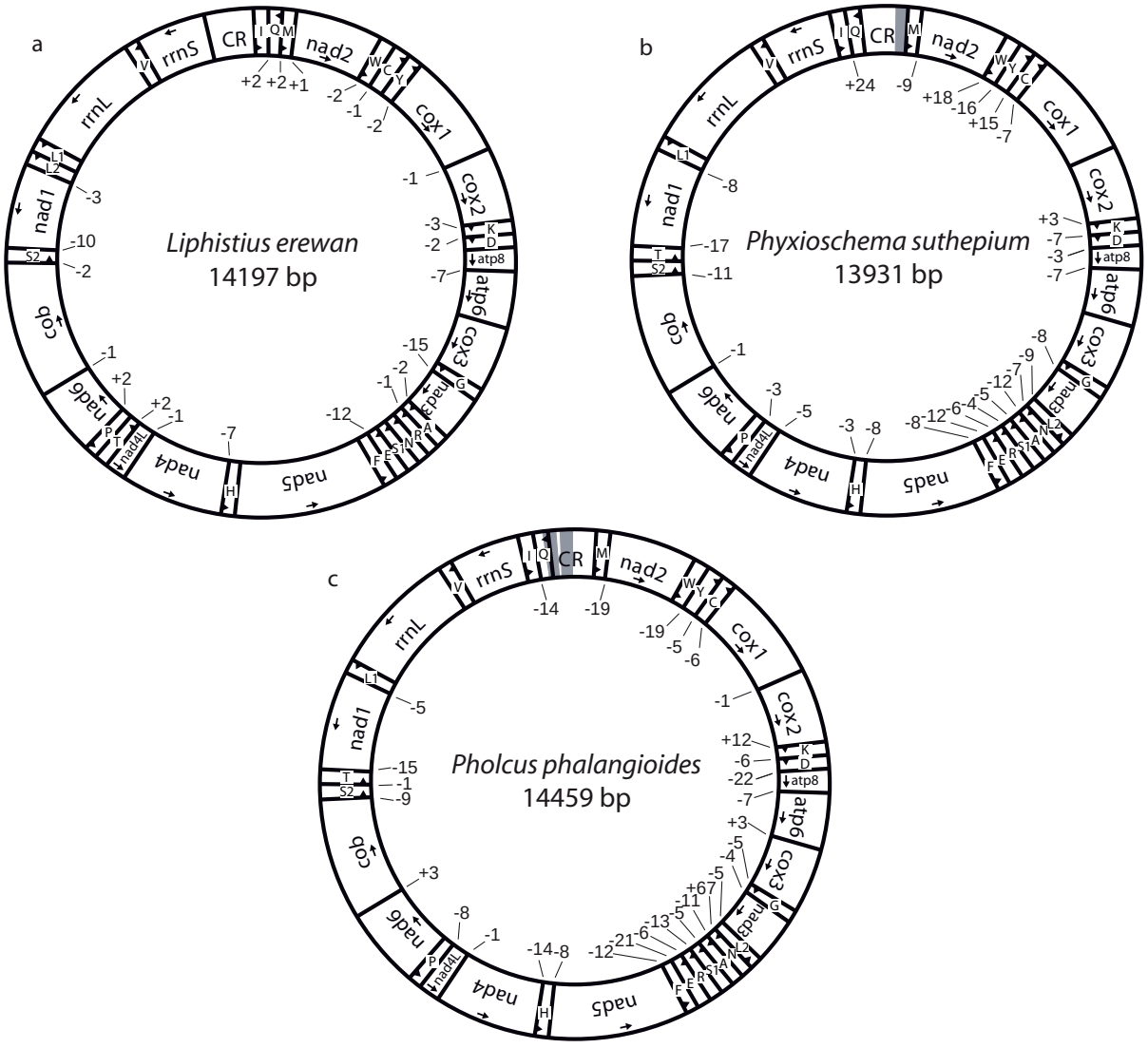


Figure 7.1: Circular maps of mitochondrial genomes of *Liphistius erewan* (a), *Phyxioschema suthepium* (b) and *Pholcus phalangioides* (c). tRNA genes are depicted by their one-letter code abbreviations. Numbers reflect non-coding (positive) or overlapping (negative) nucleotides between two adjacent genes; small arrows indicate the orientation of the genes on (+)strand (clockwise) or (-)strand (counterclockwise); grey shaded area in (b) illustrate pseudogene of *trnI* in the control region (CR) of *P. suthepium*; grey shaded parts in (c) illustrate tandem repeats in the control region (CR) of *P. phalangioides*.

In the case of *L. erewan* the gene order we found (Figure 7.1 a) resembles the distribution of genes found in the mitochondrial genomes of the xiphosuran *L. polyphemus* and many other arthropods, thus considered as the ancestral arthropod ground pattern (Lavrov et al. 2000; Staton et al. 1997). The gene order we detected in *P. suthepium* and in *P. phalangioides* (Figure 7.1 b+c) differs in several points from the putative arthropod ground pattern. Both show the same gene rearrangements without any gene inversions that led to the gene order already noted for the araneomorph spider *Hypochilus thorelli* (Masta and Boore 2008): *trnC* and *trnY* have changed their positions, *trnT* is translocated between *trnS*-(UCN) and *nad1*, *trnL*-(UUR) has a new position between *nad3* and *trnN* where further rearrangements in the so-called ARNSEF-cluster of tRNAs shuffled around some tRNAs, why we now find a LNASREF distribution of tRNAs in that area. Furthermore *trnI* and *trnQ* are located upstream of the control region next to the gene coding for the small ribosomal subunit (*rrnS*). Because of its relevance on a putative gene order characteristic for all Opisthothelae, we will discuss that rearrangement detailed later on in the context of gene rearrangements within Araneae.

The majority of genes in the three genomes presented here are either immediately contiguous or overlapping (Supplementary file 7.1 a-c). In *L. erewan* we found overlaps from one to 15 bp, in *P. phalangioides* from one to 22 bp and in *P. suthepium* from one to 17 bp. The putative control region in *L. erewan* is 338 bp in length and is flanked by *rrns* and *trnI*, while it is located between *trnQ* and *trnM* in both *P. phalangioides* and *P. suthepium* and is 1076 bp and 387 bp in length respectively. Besides the control region non-coding sequences are infrequent and dispersed over the whole genome. They are rather short (*L. erewan*: 1-15 bp, *P. phalangioides*: 1-12 bp, and *P. suthepium*: 3-24 bp), only in *P. phalangioides* one longer non-coding sequence of 67 bp between *trnL*-(UUR) and *trnN* occurs.

Protein-coding genes and nucleotide composition

Most protein-coding genes of *L. erewan*, *P. phalangioides*, and *P. suthepium* (Supplementary file 7.1 a-c) are either identical in length or slightly shorter than their equivalent in *L. polyphemus*. At the 5`end most proteins begin with one of the prevalent start codons ATN, GTG or TTG (Wolstenholme 1992). The only exception is *cox1* in *P. suthepium*, which begins with the infrequent start codon TTA and *cox1* in *L. erewan*, which uses the unusual start codon CTG. Both start codons also appear to be used in protein-coding genes of other chelicerates (Castresana et al. 1998; Lavrov et al. 2000; Qiu et al. 2005). The vast majority of the protein-coding genes in each species are terminated by one of the commonly found stop codons TAA, TAG or TAT. The abbreviated form T only appears at the end of *cox2, cox3* and *nad5* in *L. erewan*, of *cox2* and *nad5* in *P. phalangioides*, and of *cox3* and *nad4* in *P. suthepium*. Such truncated stop codons are widespread among animals and are thought to be polyadenylated after transcription to complete forms (Ojala et al. 1981).

In animals the two strands of the mitochondrial genome differ in their nucleotide compositions, one is typically more rich in C + A and the other in G + T (Perna and Kocher 1995; Reyes et al. 1998). Such a strand asymmetry or strand bias is shown in most metazoan species and it is related to asymmetric mutational constrains, frequently causing deaminations of A and C nucleotides during the replication process (Hassanin et al. 2005). The CG-skew (C-G/C+G) is often used to describe the strand-specific bias in reflecting the relative number of C + G. In *L. erewan* the protein-coding genes encoded on the (+)strand show a positive CG-skew, that means a bias towards C, while all (-)strand encoded genes are biased towards G, showing a negative CG-skew. The other two spiders, *P. suthepium* and *P. phalangioides* show a nucleotide bias with exactly an opposite pattern: all protein-coding genes encoded on the (+)strand show a negative CG-skew, while all (-)strand encoded protein-coding and both ribosomal RNA genes have a positive CG-skew (Supplementary file 7.1 a-c). Among arachnids, such a reverse strand bias is also known from all other Opisthothelae, scorpions and a mite (Davila et al. 2005; Jones et al. 2007; Masta and Boore 2004; Masta et al. 2009; Navajas et al. 2002; Qiu et al. 2005) and is also found in other metazoans (e.g. Hassanin et al. 2005; Helfenbein et al. 2001). Hassanin et al. (2005) suggests that a reversal of the strand bias emerged multiple times independently among Bilateria, probably through an inversion of the control region involving a reversal of mutational constraints for the two strands of the mitochondrial genome.

Ribosomal RNA genes

The rRNA genes in *L. erewan, P. phalangioides,* and in *P. suthepium* are seperated by *trnV* (Figure 7.1 a-c), representing an ancestral condition among arthropods.

Both rRNA genes of the three investigated Araneae (Supplementary file 7.1 a-c) have a length considerably shorter than those found in *L. polyphemus* (*rrnL*: 1296 bp, *rrnS*: 799 bp; Lavrov et al. 2000). The same is true for all other spiders (Masta and Boore 2004; Masta and Boore 2008; Qiu et al. 2005). The inferred rRNA size of *L. erewan* (*rrnL*: 1142 bp, *rrnS*: 690 bp) is the longest reported for Araneae, together with those of *H. hangzhouensis* (*rrnL*: 1119 bp, *rrnS*: 698 bp; Qiu et al. 2005), another member of the clade Mesothelae. With a length of 1005 bp the *rrnL* gene of *P. phalangioides* (*rrnS* 690 bp) is the shortest within the Araneae, even shorter than that in *Habronattus oregonensis* (*rrnL*: 1018 bp; Masta 2000; Masta and Boore 2004). A similar situation is found for the *rrnS* gene of *P. suthepium* (*rrnS*: 662 bp, *rrnL*: 1065 bp) which is shorter than its shortest counterpart in *Ornithoctonus huwena* (*rrnS*: 666 bp; Qiu et al. 2005).

These data, together with the truncated tRNA sequences found in this study and in others describing mitochondrial genomes of Araneae (Masta and Boore 2004; Masta and Boore 2008; Qiu et al. 2005), underlines the suggestion of Masta (2000), that normal functioning

of tRNAs exhibiting a TV-replacement loop instead of a T-arm in several cases, may require changes in rRNA secondary structures. As mentioned in Fahrein et al. (2009) the RNA components of the ribosome in spiders may have become reduced in size after their divergence from their common ancestor with amblypygids and *Limulus*.

The motif TGGCAGA is believed to be the termination signal for *rrnL* transcription in animal mitochondria (Valverde et al. 1994). In many insects and crustaceans, this heptamer is located downstream of the large rRNA gene in the *trnL*-(CUN) gene, while in vertebrates, it is found within the downstream *trnL*-(UUR) gene (Valverde et al. 1994). In all three investigated spiders we detected this motif or slight modifications of it in both *trnL* genes (Supplementary file 7.2 a-c). Modifications are found in the downstream *trnL*-(CUN) gene of *P. phalangioides* (TGGCGGA). A comparable situation is also found in other chelicerates e.g. amblypygids, other Araneae, and *L. polyphemus*.

Secondary structure of tRNAs

All 22 tRNAs usually present in mitochondrial genomes of Bilateria were found in the genomes of the three araneid species under study. In addition, in *P. suthepium* a second sequence for *trnI* was detected. Many of these tRNA genes, especially in *P. suthepium* and *P. phalangioides*, have a length distinctively shorter compared to those in *L. polyphemus* - correlated to the small size of rRNA genes and small sized mitochondrial genome sequences. Also unusual are the inferred secondary stuctures of many tRNAs (Supplementary file 7.2 a-c), which can be linked to the truncation of tRNA sequences. The typical and among animals highly conserved canonical cloverleaf shape of tRNA secondary structures, which comprises a T-arm, a D-arm, an anticodon arm bearing the anticodon, and an aminoacyl acceptor stem, was detected in only 12 tRNAs of *L. erewan*, and only in four tRNAs of both *P. suthepium* and *P. phalangioides*. In all of the remaining tRNAs the sequence of the D-arm or the T-arm is reduced or missing.

Furthermore, mismatches in the aminoacyl acceptor stems of the tRNAs in *P. suthepium* and *P. phalangioides* are prevalent. Masta and Boore (2004) pointed out that the absence of fully paired acceptor stems together with extensive overlaps with adjacent genes suggest that in this case the sequences that encode the 3`aminoacyl acceptor stems are lacking. They postulated that due to an RNA editing mechanism the 3`aminoacyl acceptor stems are restored and therfore functional tRNAs are recovered. Because it is difficult to ascertain the exact extent of the 3`end of tRNA genes, we annotated them conservative with the assumption of eight nucleotides constructing the 3`aminoacyl acceptor stems and by this allowed the mentioned overlaps. A lack of D-arm or the T-arm sequences is often found in metazoan mitochondrial tRNAs (Boore and Brown 2000; Helfenbein et al. 2001), e.g. the D-arm sequence is missing in *trnS*-(AGN) of most metazoans (Wolstenholme 1992). Regarding

chelicerates, the described truncation of tRNA sequences is obvious in Araneae (Masta and Boore 2004; Masta and Boore 2008; Qiu et al. 2005), Scorpiones (Choi et al. 2007; Davila et al. 2005; Jones et al. 2007; Masta and Boore 2008), Amblypygi (Fahrein et al. 2009; Masta and Boore 2008), Acari (Shao et al. 2005b; Shao et al. 2006), and Thelyphonida (Masta and Boore 2008). Masta & Boore (2008) suggested that these structural changes most likely have evolved multiple times independently among arachnids, but it is still unclear which (if any) evolutionary pressure or constraints cause or faciliate such a parallel evolution.
Despite the commonly derived secondary structure of *trnS*-(AGN) we found further tRNAs with a sequence loss in the mitochondrial genomes of all Araneae investigated to date: all araneids have in common that in the tRNAs W, F, G, and C the T-arm is missing and the variable loop and the T-arm are replaced by a TV-replacement loop. Hence, it is most easily explained that the derived secondary structures of these four tRNAs were already present in the last common ancestor of the Araneae.

Control region and its correlation to the strand bias

Although the position of the putative control region in the mitochondrial genome of invertebrates shows a great variability, it is conserved among those arthropods which have retained the ancestral arthropod gene order (Saito et al. 2005; Wolstenholme 1992; Zhang and Hewitt 1997). This ancestral position between *rrnS* and *trnI* is found in *L. erewan* (position 12558 - 12895), while in *P. phalangioides* (position 12278 - 13353) and *P. suthepium* (position 12408 - 12794) the putative control region has a different relative position between *trnQ* and *trnM*, perhaps due to tRNA translocations (see next section: “gene arrangements within Araneae”). In *P. suthepium*, and significantly in *L. erewan*, it is more AT-rich (71,32% and 78,11%) compared to the A+T composition of the remaining genome (67,29% and 67,47%). By contrast, *P. phalangioides* shows a slightly lowered A+T content of the control region (64,41%) in comparison to the rest of the genome (65,97%).
The control region is believed to be associated with the initiation of transcription and replication (Taanman 1999). Replication errors (like strand slippage mispairing during replication) are very likely responsible for copy number variation like tandem repetitions or duplications in animal mitochondrial DNA. A hairpin-like putative secondary structure can often be identified in control region sequence. Besides the hairpin structure some other conserved motifs were characterized in arthropod mitochondrial control regions - poly-T and/or TA(A) stretches upstream and GA-rich and/or G(A)nT sequences downstream of the hairpin structure (Zhang and Hewitt 1997). Hairspin and TA(A) and G(A)nT motifs are presumed to play an important role in the initiation of transcription and (or) replication of the mitochondrial genome (Zhang and Hewitt 1997; Zhang et al. 1995). In both *L. erewan*

and *P. suthepium* stable stem-loop structures are present (Figure 7.2). In *L. erewan* the stem is composed of 18 base-pairs without mismatches and the loop consists of 11 bases and in *P. suthepium* the stem is build of 19 base-pairs with one mismatch while the loop counts 13 bases. Those inverted repeat sequences often occur in arthropod mitochondrial control regions (Fahrein et al. 2007; Kilpert and Podsiadlowski 2006; Masta et al. 2008) and they are probably located at or near the replication origin of the L strand (Zhang and Hewitt 1997), but their exact function is still unclear.

Two further peculiarities are present in the control region of the investigated opisthothele spiders: A twofold tandem repeat of 252 nucleotides is exhibited in the first half of the control region of *P. phalangioides* (position 12278 - 12503 and 12548 - 12773) including the end of the upstream *trnQ*, and in the control region of *P. suthepium* a pseudogene of *trnI* (position 12720 - 12771) upstream of *trnM* can be detected. Both these findings mean influence on considerations about rearrangement events within the Araneae, which will be discussed in the following chapter.

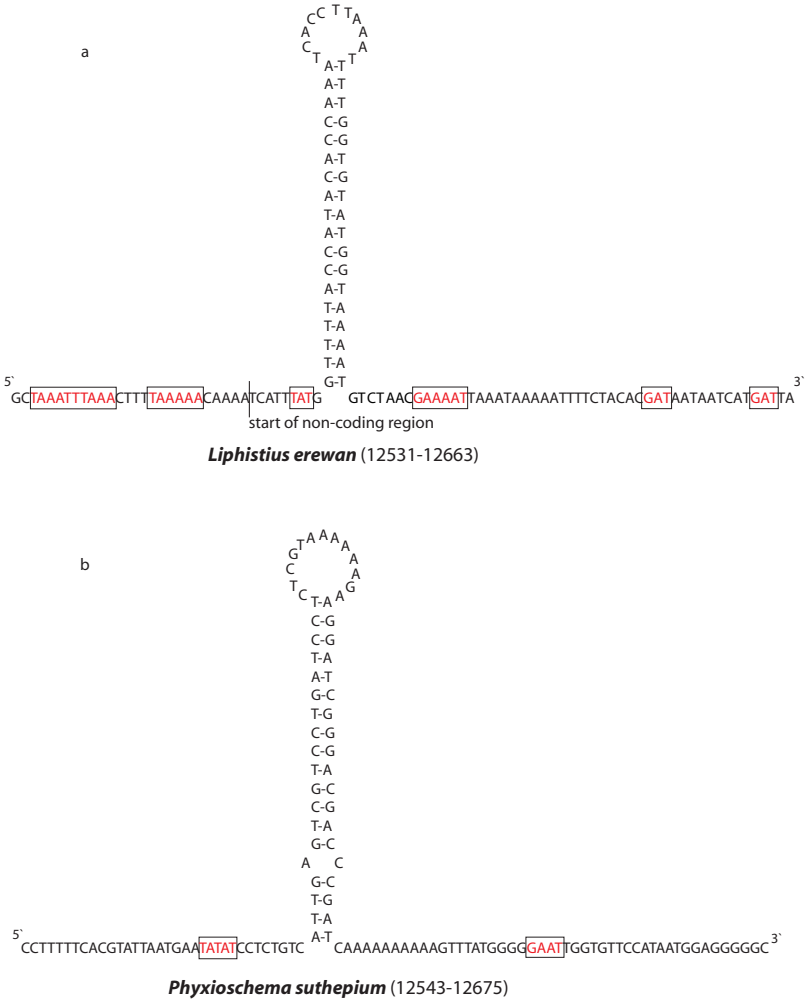


Figure 7.2: Stem-loop structures found in the largest non-coding part (= the putative control region, CR) of the mitochondrial genomes of the spiders *Liphistius erewan* (a) and *Phyxioschema suthepium* (b). The numbers behind the species stand for nucleotides and reflect the position of the illustrated sequence within the respective genome.

Gene rearrangements within Araneae

Within the mitochondrial genomes of Araneae gene rearrangements are only known from clades belonging to Opisthothelae. Neither *H. hangzhouensis* nor *L. erewan*, both representatives of the suborder Mesotheleae, possess any gene rearrangements in their mitochondrial genome compared to *L. polyphemus*, which is also thought to represent the common gene arrangement of Arthropoda. So it is most parsimonious to assume an arrangement of genes similar to the putative arthropod ground pattern being the primary state for the mitochondrial genomes of Araneae.

Most of the rearrangements detected in *P. suthepium* and in *P. phalangioides* are common for all opisthothele spiders: *trnC* and *trnY* have changed their positions, *trnT* is translocated between *trnS*-(UCN) and *nad1*, *trnL*-(UUR) is translocated away from *nad1* and *trnL*-(CUN) to its new position between *nad3* and *trnN* where further rearrangements of some tRNAs of the so-called ARNSEF-cluster of tRNAs can be detected. The new position of both *trnL*-(UUR) and *trnT* is suggested to be the result of duplicated large regions followed by gene loss (Qiu et al. 2005), even if some authors underline that single long-range gene translocations are highly unlikely under a duplication/random loss model when additional changes in the intervening gene order are missing (Dowton and Austin 1999; Dowton and Campbell 2001). The derived position of tRNAs in the ARNSEF-cluster as well as the interchanged position of *trnC* and *trnY* can indeed be explained by a local rearrangement in the particular area consistent with the tandem duplication/random deletion (TD/RD) model (Macey and Verma 1997; Moritz et al. 1987). Unique for the mygalomorph spider *P. suthepium* and for the two araneomorphs spiders *H. thorelli* and *P. phalangioides* is the upstream translocation of *trnI* and *trnQ* across the non-coding region without any further translocation or inversion of *trnI*. That rearrangement can also be declared as a local rearrangement (Dowton et al. 2009; Macey et al. 1998) with strong evidence coming from the sequence of the *trnI*-pseudogene upstream of *trnM* in the genome of *P. suthepium*. In other mygalomorph spiders (*O. huwena*, *Calisoga longitarsis*,) there is also an upstream translocation of *trnI* and *trnQ* but with an additional inversion of *trnI*. In the araneomorphs *H. oregonensis* and *Nephila clavata* only an upstream gene movement of *trnQ* is detectable, while *trnI* is translocated to another relative position, between *nad6* and *cob*.

All of those findings concerning gene rearrangements around the control region of opisthothele spiders can be explained by a tandem duplication event affecting a geneblock of at least the control region plus *trnI* and *trnQ* (Figure 7.3). That tandem duplication event was followed by a random loss of the upstream copy of the control region and the downstream copies of *trnI* and *trnQ*, and the whole TD/RD process took place in the lineage leading to opisthothele spiders. That would explain the same gene order found in parts of both mygalomorph (*P. suthepium*) and araneomorph (*P. phalangioides* and *H. thorelli*) spiders.

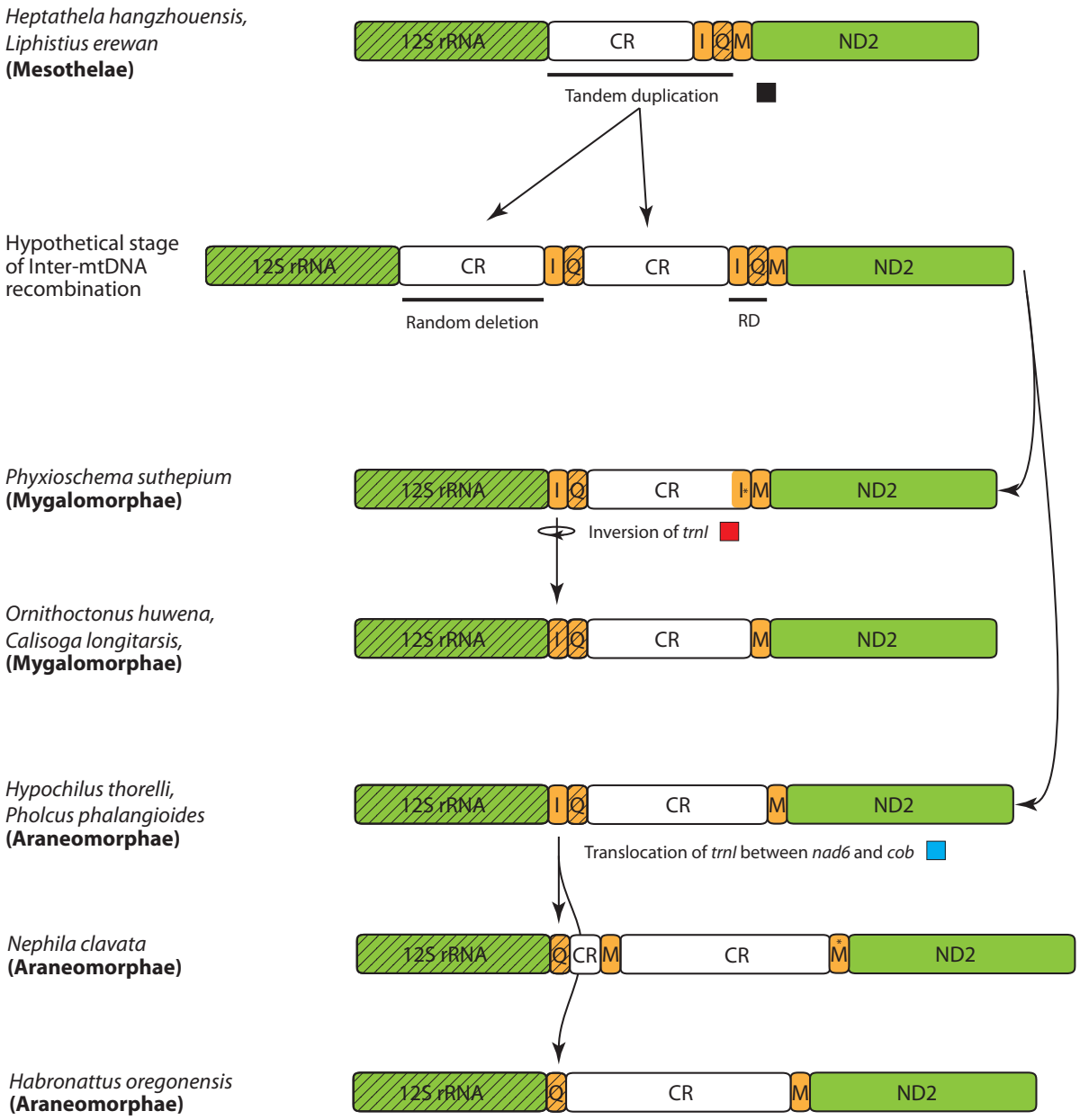


Figure 7.3: Gene order and rearrangements around the mitochondrial control region of Araneae. Illustrated is a section of the mitochonrial genome of all completely sequenced Araneae species, showing the control region (white) and all tRNAs (orange) which occur between the flanking genes coding for 12S rRNA and ND2 (light green). Horizontal stripes indicate the orientation of the genes on (-)strand. This hypothetical scenario issues a possible formation of gene orders around the control region found in different species of the Araneae considering a tandem duplication/random deletion event (TD/RD, black square) in the lineage leading to the Opistothelae followed by an inversion of *trnI* within mygalomorph spiders (red square) and accordingly a translocation of *trnI* away from the control region and between *nad6* and *cob* within the Araneomorphae (blue square).

The inversion of *trnI* as well as its translocation between *nad6* and *cob* are events that most likely took place within, and not in the lineage leading to, the Mygalomorphae and Araneomorphae respectively. Figure 7.4 shows those events plus all other translocations including the reversal of strand bias found in spiders, altogether mapped on a phylogenetic analysis. Potentially another tandem duplication of the control region plus *trnQ* and *trnM* took place in the lineage leading to the Araneomorphae or within the Araneomorphae in a lineage at least leading to *N. clavata* and *P. phalangioides*, namely the Araneoclada (after Jocque and Dippenaar-Schoeman 2007). This would explain both, the *trnM* (position 9510 - 9573) within the control region of *N. clavata* (Lee unpubl. [GenBank:NC_008063]) as we found a further sequence of *trnM* (Figure 7.5) downstream of the control region next to *nad2* (position 10194 - 10247), where it is also located in all other Araneae. This is also consistent with the twice-repeated 252 bp sequence comprising the end of *trnQ* and the first half of the control region of *P. phalangioides*. Compared to the length of the control region in Mesothelae (~340 bp) and Mygalomorpha (387 - 478 bp), an elongated control region as found in *P. phalangioides* (1076 bp) is also present in the araneomorphs *N. clavata* (157 and 672 bp after Lee unpubl. [GenBank:NC_008063]), *H. thorelli* (688 bp) and *H. oregonensis* (717 bp), which points to a duplication event in the lineage leading to the Araneomorphae, but with no evidence from pseudogene or tandemly repeated sequences. The elongation of the control region may be the reason why the tendency to shortened mitochondrial genome sequences in Araneae and in particular in opisthothele spiders seems to be broken at first sight in *N. clavata*, *H. oregonensis* and *P. phalangioides*.

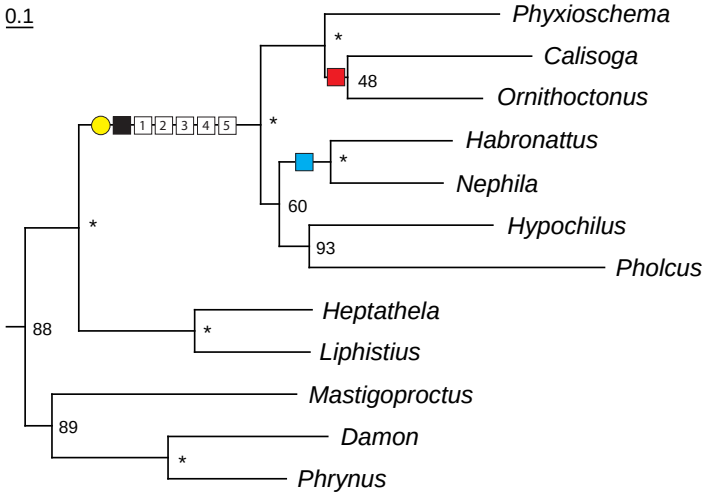


Figure 7.4: Hypothesized events (squares) and strand bias reversal (yellow circle) leading to derived gene orders known for different spiders mapped on a topology based on the results of a phylogenetic analysis (not published here). Coloured squares according to Figure 7.3 accentuate events around the control region: tandem duplication/random deletion event (black square) in the lineage leading to the Opistothelae; inversion of *trnI* within the Mygalomorphae (red square); translocation of *trnI* away from the control region and between *nad6* and *cob* within the Araneomorphae (blue square). Numbered squares represent further translocation events detected the mitochondrial genomes of all Opistothelae: 1 - translocation of *trnY* or *trnC*; 2 - translocation of *trnL*(UUR); 3 - translocation of *trnN*; 4 - translocation of *trnS*; 5 - translocation of *trnT* (numbering independent of possible succession chronology).

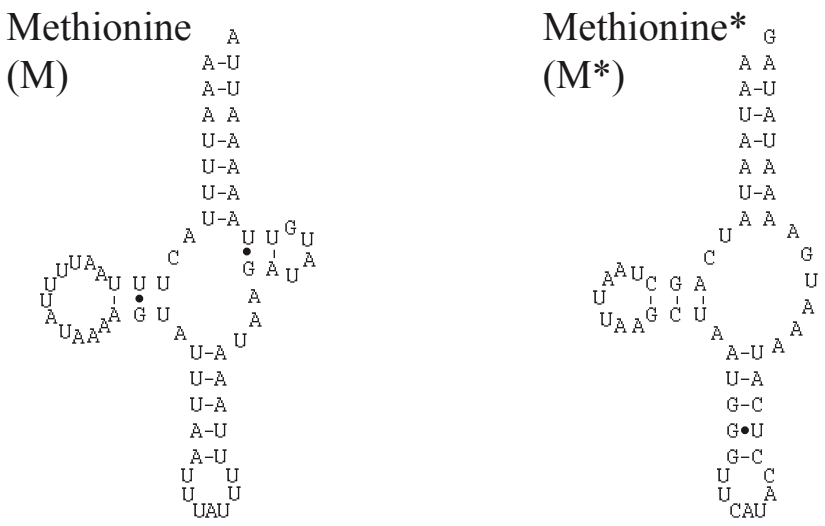


Figure7.5: Inferred tRNA secondary structure of *trnM* in the mitochondrial genome of *Nephila clavata* (at position 9510 - 9573 sensu Lee unpubl.) and a second *trnM* (*) we detected at position 10194 - 10247. Dots illustrate the base pairing of the pyrimidine base uracil with the purine base guanine.

7.4 Conclusions

Mitogenomic sequence data from other Araneae already revealed a conserved gene order in one clade, the mesothele spiders. Concerning the gene order our results from *L. erewan* are consistent with this, while our observations in the Opisthothelae spiders *P. phalangioides* and *P. suthepium* approve the reversed strand bias and several rearrangements found in other members of the Opisthothelae. In this context, basic findings of both rearrangement events around the control region of Opisthothelae spiders and differences in the sequence length of the control region give information about happenings within the Opisthothelae. Furthermore, a truncation of sequences coding for mitochondrial tRNAs as known from all other Araneae and especially characteristic for opisthothele spiders, were also found in the investigated Araneae and are foremost detectable in the Opisthotheleae. Therewith interrelated are the shortest *rrnL* and *rrnS* sequences reported for chelicerates, in our case found in *P. phalangioides* and *P. suthepium* respectively, and finally as a consequence the length of the mitochondrial genome sequences which was overall found to be rather short.

Due to the conserved conditions in mesothele spiders, the derived features found in *P. phalangioides* and *P. suthepium* and also in other members of the Opisthotheleae do not facilitate considerations about phylogenetic relationships of the three Megoperculata taxa to one another. But they can be used to make suggestions or to underline established hypotheses about araneid intra-relationships. And finally they provide further molecular evidence for the traditional division into Mesothelae and Opisthothelae.

7.5 Material and Methods

Specimen and DNA extraction

L. erewan and *P. suthepium* were obtained from Peter Schwendinger (Univerity of Innsbruck), *P. phalangioides* was collected by Anne Zakrzewski (University of Bonn). From each individuum total DNA was extracted from one leg by using Qiagen extraction kits (Qiagen, Hilden, Germany) following the manufacturers` protocol.

PCR conditions, primers and sequencing

Initial PCRs were done with a published primer set designed for crustacean mitochondrial genomes (Yamauchi et al. 2004). They were performed on an Eppendorf Mastercycler with Eppendorf 5-prime-Taq kit (Eppendorf, Germany) in 50 µl volumes (5 µl buffer; 1 µl dNTP mix, 10 µM; 0.25 µl Taq polymerase; 1 µl template DNA, 40,75 µl molecular grade water, 1 µl primer mix, 10 µM each). Reaction conditions were: Initial denaturation (1 min, 94°C); 40 cycles of denaturation (30 sec, 94°C), primer annealing (1 min, 50°C), elongation (1 min, 68°C); final elongation (1 min, 68°C). This was successful in case of S2, S4, S5, S7, S9, S10, S11, S15, and S38 (in *L. erewan*), S8, S10, S43, and S44 (in *P. suthepium*), and S9, S26, and S28 (in *P. phalangioides*). On the basis of the successfully obtained sequences, taxon-specific primer pairs were designed to amplify missing fragments in order to bridge the gaps between the sequences (Supplementary file 7.3). These secondary PCRs were also performed on an Eppendorf Mastercycler and Mastercycler gradient and in case of smaller fragments (up to about 2000 bp) using the Eppendorf 5-prime-Taq kit (Eppendorf, Germany) in 50 µl volumes as described above. Conditions for the secondary PCRs were: initial denaturation (1 min, 94°C), 30 cycles of denaturation (30 sec, 94° C), annealing (30 sec, primer specific temperature according to Supplementary file 7.3) and elongation (1 min, 68°C, respectively 3 min in case of the larger fragments), followed by a final elongation step (1 min, 68°C, respectively 3 min in case of the larger fragments). PCRs of longer fragments (over 2000 bp) were performed with Takara LATaq kit (Takara) in 50 µl volumes (5 µl Buffer, Mg+; 8 µl dNTP mix, 2.5 µM; 1 µl DNA; 1 µl primer mix, 10 µM each; 0.5 µl Takara LA Taq; 32 µl water). PCR conditions were: initial denaturation (1 min, 94°C), 40 cycles of denaturation (30 sec, 94°C), annealing (primer specific temperature see Supplementary file 7.3, 1 min), and elongation (5 min, 68°C), followed by a final elongation step (3 min, 68°C). Long PCR products were sequenced by primer walking.

All PCR products were visualized on 1% agarose gels and purified using the Bluematrix DNA purification kit (EURx, Gdansk, Poland). If extra bands were present, a Gel extraction was performed, following the manufacturers` protocol (Qiagen). Purified templates were

sent to LGC AGOWA genomics, where DNA sequencing according to the dideoxy chain-termination method of Sanger was performed. Each sample contained 10 µl template DNA and 4 µl Primer (1:10 dilution).

Gene annotation

Sequence assembly from obtained sequences was performed with BioEdit 7.0.1. (Hall 1999). BLAST searches on NCBI BLAST (Basic Local Alignment Search Tool) databases were performed to determine the protein-coding and ribosomal gene boundaries. Furthermore, the sequences were compared with alignments from other chelicerate species. Boundaries of flanking genes were used to assume the start and ending of the ribosomal RNA genes and the extension of control regions. In *L. erewan* the boundary of the *rrnS* gene to the control region was inferred by comparison with *rrnS* genes of other arachnids. Hairpin structures in the control regions were identified by eye inspection. Genomic position and secondary structure of tRNAs were identified by tRNAscan-SE (Lowe and Eddy 1997) and ARWEN (Laslett and Canbäck 2008), but due to tRNA sequence truncation especially in *P. suthepium* and *P. phalangioides* the rate of yield was low and the missing tRNAs were also identified by eye inspection. Sequence data was deposited at NCBI database [*L.e.*: Genbank: XXX], [*P.s.*: Genbank: XXX] and [*P.p.*: Genbank: XXX]. The determination of nucleotide frequencies was also performed with BioEdit 7.0.1. (Hall 1999).

7.6 Abbreviations

A, adenine; *atp6* and *8*, genes encoding ATPase subunit 6 and 8; bp, base pairs; *cox1-3*, genes encoding cytochrome oxidase subunits I-III; *cob*, gene encoding cytochrome b; C, cytosine; D-arm, D-loop and D-stem of a tRNA secondary structure; G, guanine; *nad1-6* and *nad4L*, genes encoding NADH dehydrogenase subunits 1-6 and 4L; PCR, polymerase chain reaction; rRNA, ribosomal RNA; *rrnL*, large (16S) rRNA subunit (gene); *rrnS*, small (12S) rRNA subunit (gene); T, thymine; tRNA-X (where Abc is replaced by one letter amino acid code of the corresponding amino acid), transfer RNA; *trnX* (where *X* is replaced by one letter amino acid code of the corresponding amino acid), tRNA gene; T-arm, T-loop and T-stem of a tRNA secondary structure.

7.7 Additional material

Additional files 7.1 to 7.3 can be found in the appendix of this thesis (Chapter 10).

8. Concluding discussion

The remarks made here refer to the questions raised in the introduction:

Is mitochondrial genomics applicable to infer arachnid phylogenetic relationships?

A first preliminary phylogenetic analysis (ML, MP, BI) using concatenated amino acid and nucleotide datasets of arachnid mitochondrial genomes failed to resolve interordinal relationships of Arachnida due to a lack of mitogenomic data from important taxa such as Opiliones, Pseudoscorpiones, Palpigradi, Uropygi, and Amblypygi (Chapter 2). A later phylogenetic analysis of sequence data from mitochondrial protein-coding genes, that time also containing sequence information of Uropygi and Amblypygi, weakly supports a sister group relationship of Ricinulei and Acari and of Uropygi and Amblypygi as well as the monophyly of Tetrapulmonata (Chapter 3). The result of the interrelationships of Tetrapulmonata corresponds well with another recent analysis of a nearly similar data set, although support varied with the model of evolution and type of analysis performed (Masta et al. 2009). So again, both analyses provide little resolution of the relationships among the major lineages of arachnids. A more recent analysis of arachnid relationships inferred from amino acid data (Masta 2010) recovered scorpions as sister to all other chelicerates with good support from maximum likelihood bootstrap values and Bayesian posterior probabilities (=99/1), while Micruran chelicerates (Araneae, Amblypygi, Uropygi, Ricinulei, and Acari) are supported as a clade only by Bayesian posterior probabilities (=1). Unfortunately no support for resolving relationships of any further chelicerate orders was yielded. Figure 7.1 depicts a most recent analysis using concatenated amino acid alignments of all 13 protein-coding genes from all complete chelicerate mitochondrial genome sequences found in NCBI Refseq and, additionally, our own not published data from three Ricinulei, one Opiliones, and three Araneae. With a support value of 96.3 from maximum likelihood bootstrapping, the best supported branch places Ricinulei as sister to a clade containing the paraphyletic Acari plus Pycnogonida. The affinity of Pycnogonida to Acari may be due to long-branch attraction (Podsiadlowski and Braband 2006), a frequent problem not restricted to phylogenetic analyses with mitochondrial genome data.

Hence, no interordinal sister group relation can be proposed, since for all other higher-ranking relationships even less phylogenetic resolution is given. A similarly unsatisfying result was recently obtained with a large dataset of 62 nuclear encoded genes from 75 arthropod species, covering all traditional arachnid orders (Regier et al. 2010).

Anyhow, in our analysis almost all of the traditional chelicerate orders as well as several clades within these orders are strongly supported. The only exception are the Acari which appear to be paraphyletic, with best support for both recognized subtaxa, Acariformes and Parasitiformes.

Thus it appears difficult to deduce the progression of early diversification among arachnid orders from mitochondrial genome sequence data. Also the reversals of strand bias as documented by Hassanin (2006) and the use of different evolutionary models can affect the usefulness of multigene phylogenetic analyses (Jones et al. 2007; Masta et al. 2009). Altogether, contentious issues of arachnid phylogenetic relationships can hardly be answered at present using mitochondrial sequence data. In addition, there still is a lack of sequence information from Pseudoscorpiones, Palpigradi, and Opilioacariformes.

Do rearrangements of mitochondrial genes provide phylogenetic information to reveal or evaluate intraordinal relationships of Arachnida? And is there any difference in the value of rearrangements for the phylogenetic inference of "higher" and "lower" ranking taxa?

Rearrangements of mitochondrial genes indeed are a valuable source of data - at least to test additional support or conflict with phylogenetic relationships derived from sequence data (Figure 7.1). In this regard, however, an evaluation is practicable on different taxonomic levels excluding a resolution of the relatedness of major arachnid lineages (Chapter 4 to 7). That is enforced partly by the conserved gene order retained in leastwise some members of the majority of arachnid lineages (Xiphosura, Scorpiones, Araneae, Amblypygi, Uropygi, Ricinulei, Acari). None of the investigated species of Solifugae, Opiliones, and Pycnogonida reveal the ancestral distribution of genes in their mitochondrial genomes, but the detected rearrangements are (with the current taxon sampling) restricted to single species and thus not suitable for the deduction of sister group relationships. Anyhow, a benefit for the phylogenetic inference of lower-ranking arachnid taxa can be attributed to rearrangements in the following cases (compare numbers in Figure 7.1): all Buthida can be separated from the remaining Scorpiones due to the same derived gene order (**5**); several rearrangement events support the Opisthothelae (**9**); inside the Mygalomorphae an inversion of *trnI* defines the Crassitarsae (**10**); inside the Araneomorphae a clade within the entelegyne spiders is characterised by a translocation fron *trnI* (**11**); inside the Acariformes a clade within the Prostigmata is seperated (**16**); inside the hardbacked ticks (Ixodidae) a clade comprising Amblyomminae, Haemaphysalinae, and Rhipicephalinae is marked-off the Ixodinae (**21**). Taking a closer look at the Acariformes there are cases of rearrangements even in congeneric species. In the Prostigmata three species of the genus *Leptotrombidium* and in the Astigmata two species of the genus *Dermatophagoides* share an almost identical derived gene order, but in both cases completely different from the ground pattern (marked with a question mark). In each of the two genera further nameable rearrangements (**17** in case of *Leptotrombidium* and **20** in case of *Dermatophagoides*) differentiate one species from the remaining. The last two examples also accentuate that extensive changes in gene order due to multiple and not traceable rearrangements must not necessarily imply an incapability of phylogenetic

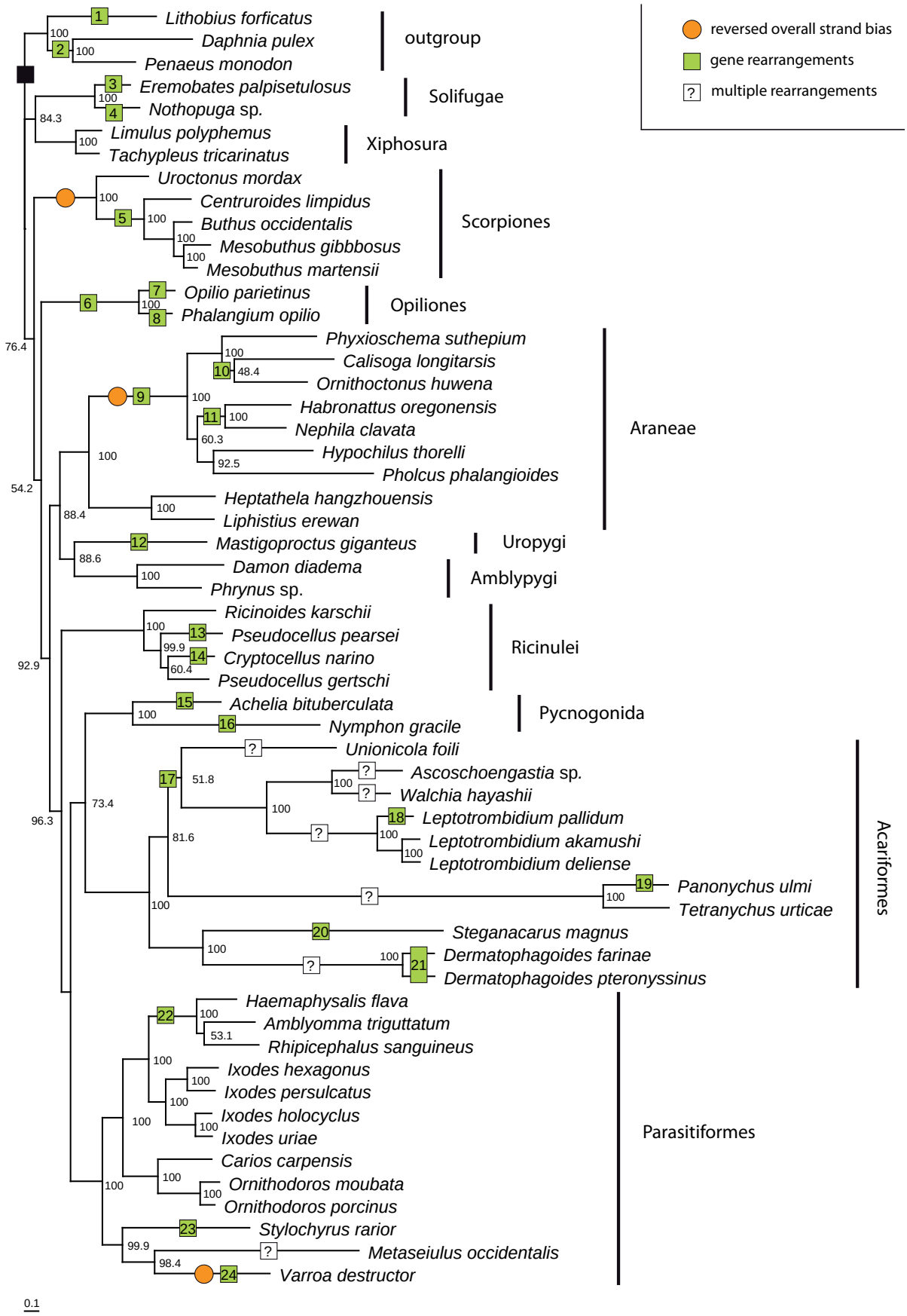

reversed overall strand bias
gene rearrangements
multiple rearrangements
Lithobius forficatus
Daphnia pulex
Penaeus monodon
outgroup
Eremobates palpisetulosus
Nothopuga sp.
Solifugae
Limulus polyphemus
Tachypleus tricarinatus
Xiphosura
Uroctonus mordax
Centruroides limpidus
Buthus occidentalis
Mesobuthus gibbbosus
Mesobuthus martensii
Scorpiones
Opilio parietinus
Phalangium opilio
Opiliones
Phyxioschema suthepium
Calisoga longitarsis
Ornithoctonus huwena
Habronattus oregonensis
Nephila clavata
Hypochilus thorelli
Pholcus phalangioides
Araneae
Heptathela hangzhouensis
Liphistius erewan
Mastigoproctus giganteus
Uropygi
Damon diadema
Phrynus sp.
Amblypygi
Ricinoides karschii
Pseudocellus pearsei
Cryptocellus narino
Pseudocellus gertschi
Ricinulei
Achelia bituberculata
Nymphon gracile
Pycnogonida
Unionicola foili
Ascoschoengastia sp.
Walchia hayashii
Leptotrombidium pallidum
Leptotrombidium akamushi
Leptotrombidium deliense
Panonychus ulmi
Tetranychus urticae
Steganacarus magnus
Dermatophagoides farinae
Dermatophagoides pteronyssinus
Acariformes
Haemaphysalis flava
Amblyomma triguttatum
Rhipicephalus sanguineus
Ixodes hexagonus
Ixodes persulcatus
Ixodes holocyclus
Ixodes uriae
Carios carpensis
Ornithodoros moubata
Ornithodoros porcinus
Stylochyrus rarior
Metaseiulus occidentalis
Varroa destructor
Parasitiformes
0.1

Figure 7.1: Hypothesised gene rearrangement events (gene transpositions and reversal of strand biases) mapped on a molecular tree depicting chelicerate relationships. The black shaded box illustrates the putative ancestral gene order. Question marks (?) indicate multiple rearrangements in the detected gene order, whereas single events cannot be traced back. Orange circles denote an inversion of the control region, hypothesised due to an overall reversed strand bias in the genomes of the respective species. Green boxes depict assumed rearrangement events like translocations (TL), tandem duplication followed by random loss (TDRL), and inversions (I) in the respective lineage; thereby, numbers imply the hypothesised events as follows: **1** - TL of *trnC*; **2** - TL of *trnL2*$^{(UUR)}$; **3** - TL of *trnL1*$^{(CUN)}$ or *trnL2*$^{(UUR)}$; **4** - TL of *trnP*; **5** - TL and I of *trnQ*, loss of *trnD*; **6** - TDRL of a segment comprising parts of *rrnS* and *lnr*; **7** - TL of *trnQ*, TL and I of *trnI*; **8** - TL of *trnQ*, TL and I of *trnI*, loss of *trnL1*$^{(CUN)}$; **9** - TDRL of a segment ranging from *lnr* to *trnQ*, TL of *trnY* or *trnC*, TL of *trnL2*$^{(UUR)}$, *A*, *R*, *T*; **10** - I of *trnI*; **11** - TL of *trnI*; **12** - TL of *trnA* and *trnR*; **13** - TDRL of a segment ranging from *trnL1*$^{(CUN)}$ to *trnY*, TL of *trnN*; **14** - TL of *trnY*; **15** - TL of *trnQ*, I of *trnT*, I of both *trnL1*$^{(CUN)}$ and *trnL2*$^{(UUR)}$; **16** - I of a gene block ranging from *trnI* to *trnD*, TL of *trnE* or *trnR*, TL of *trnY*, *Q*, *S1*$^{(AGN)}$, TL and I of *trnN*, *G*, *P*, *M*, *A*, *V*, TL of *lnr*; **17** - besides not traceable rearrangements a TL and I of *trnR* and *trnV* between *nad4* and *nad6* plus a TL of *nad4l*, *trnT*, and *trnP* away from this position can be determined, originating a conserved gene-cluster (*nad4*, *trnR*, *trnV*, *nad6*, *cob*, *trnS2*$^{(UCN)}$) detectable in all prostigmate mites of this clade; **18**- 2x TDRL of a segment ranging from *lnr* to *trnQ*, TD (without RL) and I from a gene block ranging from *rrnL* to *lnr*; **19** - TL and I of *trnF*, *K*; **20** - TL of *nad1*, *nad2*, *rrnL*, *trnQ*, TL and I of *trnW*, *trnL1*$^{(CUN)}$, *trnS2*$^{(UCN)}$, loss of *trnK*, *D*, *G*, *T*, *L2*$^{(UUR)}$, *V*, *I*, *M*, *C*, *Y*, all tRNAs of the *ARNSEF*-cluster; **21** - TL of *trnC*, *Y*, *A* in either *D. pteronyssinus* or *D. farinae*, TL of *trnV* in *D. pteronyssinus*; **22** - TL and I of *trnC*, TDRL from a segment ranging from *trnF* to *trnQ*; **23** - TL and I of *trnE*, TL of a gene block ranging from *nad1* to *trnL1*$^{(CUN)}$; **24** - TDRL involving *rrnS* and *lnr*, TL and I of *trnH*, *S2*$^{(UCN)}$, TL of *trnQ*, *Y*, *P*, *C*. See Fahrein et al. (2007, also chapter 2) and Masta et al. (2010) for description of the hypothesised events; events within Araneae and Opiliones are re-evaluated.

inferences as critisised by several authors (e.g. Dermauw et al. 2009). So even if largely rearranged, gene order data is useful for inferring phylogeny at lower taxonomic levels, what was already pointed out by Helfenbein (2001) and Masta (2010).

Although there is still a lack of mitochondrial genome data from Pseudoscorpiones, Palpigradi, as well as from Opilioacariformes, I do not expect that gene order information from these taxa would finally clarify the relatedness of the major arachnid lineages. Regarding Opilioacariformes, one of the three major Acari lineages beside Acariformes and Parasitiformes, their data is important prior to considerations about a monophyletic origin of Acari, which certainly has an influence on arachnid phylogeny. Thus, for comprehensive analysis, the missing mitochondrial gene order information just as mentioned for the missing sequence information is indispensable.

Since unusual strand biases are ascribed as result of an inversion of the control region (Hassanin et al. 2005), that event may also be considered a character in the reconstruction

of arachnid phylogeny. Reversed nucleotide frequencies are proved to be present in several arachnids (Figure 7.1), as for different members of Araneae, Acari, and Scorpiones. In two out of these three cases it most likely has emerged within the order, thus, it is most parsimoniously to suggest an inversion of the control region multiple times independently in at least three different arachnid lineages. Surely, for Scorpiones and Opisthothelae it can be interpreted as an apomorphy.

Are genome rearrangements distributed equally over the genome or do areas with an increased rate of translocations exist (= rearrangement-hotspots)?

In mitochondrial genomes of arachnids predominantly tRNA genes are involved in translocation events. Other rare events are inversions and events consistent with the TDRL model. These "larger" rearrangements often comprise gene blocks containing several genes (including protein-coding and rRNA genes). Only in the astigmate mite *Steganacarus magnus* and in some prostigmate mites single protein-coding genes are assumed to be rearranged due to a single translocation event with an unknown mechanism. The single translocation events are prevalent detectable in the range of the *ARNSEF*-cluster of tRNAs, as shown in all Opisthothelae (Chapter 7), the Uropygi *Mastigoproctus giganteus*, the Ricinulei *Pseudocellus pearsei* (Chapter 2), the Pycnogonida *Nymphon gracile*, and nearly all Acari with a derived gene order. In that connection both, gene transloctions away from this cluster to a new position and originating from another location to this cluster, are embraced. Concerning the accelerated mobility of tRNA genes compared to protein-coding and rRNA genes in general and the more frequently affected tRNA-cluster in particular, our findings are in agreement with former studies (Dowton and Austin 1999; Saccone et al. 1999). Another hotspot of gene order variation is found regarding the control region of Acari, Opisthothelae, Opiliones, Scorpiones, Pycnogonida, and Ricinulei. Genes being originally located upstream or downstream the control region show an increased mobility, either within this area or away from it. This correlation was already mentioned by Duarte et al. (2008), stressing an enhancing effect of the non-coding control region on a successful gene rearrangement as a basis for explanation. This may also explain why genes originating from other locations often find a new position close to the control region. Furthermore we found evidence for several independent TDRL events involving this region. Hence, in most Acari and all Opisthothelae, Scorpiones, Ricinulei and also most likely in Opiliones, rearrangements around the control region can be accounted for by partial genome duplications as a result of a slipped strand misspairing during replication (Chapter 2, 6, and 7).

Gene order variations between *cob* and *nad1* detected in Opisthothelae and Solifugae are not insignificant. Although pertaining to different events in both members of Solifugae these peculiarities are in close connection to extensive repeat regions. Partly sequenced

mitochondrial genomes of four further solpugid genera (not published) revealed non-coding sequences of about 30 bp length occuring between *trnS2*$^{(CUN)}$ and *nad1*, which might promote gene order changes or the origination of tandem repetitions in this section.
Anyhow, areas with a low rate of rearrangement do exist. Although being more or less affected by rearrangements in several Acari and in *Nymphon gracile*, the most conserved part of the genome ranges from *cox1* to *nad3*. Furthermore, found to be highly stable among Arachnida is the adjoining appearance of *nad6*/*cob* and *trnF/nad5/trnH/nad4/nad4L*, the latter gene block embracing slight deviations only in some Acari.

Are any other characters of 'genome morphology' useful to shed light on relationships of the major lineages of the Arachnida?

Unfortunately, despite the failure of gene rearrangements or genome sequence data to resolve the early diversification of arachnid orders, at that time no other character of 'genome morphology' is able to do so. However, it becomes apparent that some features of 'genome morphology'are indeed characters bearing phylogenetic signals. Especially comparative work of transfer RNA and ribosomal RNA secondary structures is promising. So far, comparisons of inferred tRNA structures are certainly more established, since it is much easier to deduce a secondary structure from a sequence of about 60 bp in length than from one embracing about 1300 bp (*rrnL*), respectively 750 bp (*rrnS*).
Within arachnids, best investigated are tRNA secondary structures from spiders, in which for all Opisthothelae a lack of sequences was proven to encode a 3`aminoacyl acceptor stem at least in some tRNA genes. Furthermore, all Araneae have in common that in the *trnW*, *trnF*, *trnG*, and *trnC* the T-arm is missing and both the variable loop and the T-arm are replaced by a TV-replacement loop (Chapter 7). Hence, it is possible that the sequence loss of these four tRNAs evolved in the lineage leading to the Araneae. Another example of potentially concerted sequence evolution is found regarding the aberrant secondary structure of *trnH* in scorpions (Chapter 6). Because a missing T-arm in the secondary structure of *trnH* is not known from other arachnids or even arthropods, it very likely is an apomorphy of Scorpiones. Since a truncation of many tRNAs is exhibited also in mitochondrial genomes of Araneae, Scorpiones, Amblypygi, and Thelyphonida, it is suggested that among arachnids these structural changes independently evolved multiple times, but, up to now, it is neither known what causes them, nor what are the consequences for protein biosynthesis in mitochondria. Anyway, RNA secondary structures have to be treated with respect due to the high degree of subjectivity when inferring them.
Comparison of inferred ribosomal RNA secondary structures among arachnids also reveals structural differences in this otherwise well-conserved molecule. For example, in *rrnS* of Opiliones some helices are considerably shorter than in Amblypygi (Masta 2010). Data

from other arthropods are more similar to the findings in Amblypygi, which leads to the assumption of a reduction of these helices within Opiliones after their divergence from a common ancestor with Amblypygi. For Araneae it could be shown that both rRNA genes have a considerably shorter length than their counterparts in *L. polyphemus* and *Damon diadema* (Chapter 3 and 7). It is assumed that rRNA sequences in spiders may have become reduced in size after their divergence from their common ancestor with amblypygids and *Limulus*, but for a detailed evaluation of rRNA structure there is too little information. rRNA structure is not routinely analysed in publications of mitochondrial genomes.

In rare cases tandemly repeated sequences occur in mitochondrial genomes, thus their presence/absence represents phylogenetically useful information. In this context, inverted repeats capable of forming stable stem-loop structures are exhibited in the control regions of all Scorpiones. High sequence similarities of buthid inverted repeat sequences imply that at least some of the detected inverted repeats in *C. limpidus*, *B. occitanus* and two *Mesobuthus* species could be a result of shared sequence evolution (Chapter 6). But again, these data are applicable only for the evaluation intraordinal phylogenetic relationships.

In conclusion, it appears difficult to deduce the progression of early diversification among arachnid orders from mitochondrial genome sequence data or gene rearrangements - at the moment. Certainly, this has to be seen partly in conjunction with missing data from Pseudoscorpiones, Palpigradi, and Opilioacariformes. However, a suitability on lower taxonomic levels as presented, e.g. for in-depth phylogenetic analysis of the Opisthothelae, has undoubtedly been proved. So even if mitochondrial genomes can hardly shed light on the relationships among major arachnid lineages, the data is promising for a resolution of disputed intraordinal relationships.

9. Collected References

Abascal,F., Zardoya,R., Posada,D., 2005. ProtTest: Selection of best-fit models of protein evolution. Bioinformatics 21, 2104-2105.

Abascal,F., Posada,D., Zardoya,R., 2007. MtArt: A new model of amino acid replacement for arthropoda. Mol. Biol. Evol. 24, 1-5.

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

Adis,J.U., Platnick,N.I., de Morais,J.W., Rodrigues,J.M.G., 1989. On the abundance and ecology of Ricinulei (Arachnida) from Central Amazonia, Brazil. J. N.Y. Entomol. Soc. 97, 133-140.

Alberti,G., Palacios-Vargas,J.G., 1984. Fine structure of spermatogenesis and mature spermatozoa in *Cryptocellus boneti* Bolivar y Pieltain, 1941 (Arachnida, Ricinulei). J. Ultrastr. Res. 87, 1-12.

Altmann,R., 1890. Die Elementarorganismen und ihre Beziehungen zu den Zellen. Veit & Comp, Leipzig.

Amer,S.A., Kumazawa,Y., 2005. Mitochondrial genome of *Pogona vitticepes* (Reptilia; Agamidae): control region duplication and the origin of Australasian agamids. Gene 346, 249-256.

Andersson,S.G.E., Zomorodipour,A., Andersson,J.O., Sicheritz-Ponten,T., Alsmark,U.C., Podowski,R.M., Naslund,A.K., Eriksson,A.S., Winkler,H.H., Kurland,C.G., 1998. The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. Nature 396, 133-140.

Arango,C.P., 2003. Molecular approach to the phylogenetics of sea spiders (Arthropoda: Pycnogonida) using partial sequences of nuclear ribosomal DNA. Mol. Phylogenet. Evol. 28, 588-600.

Avise,J.C., 2000. Phylogeography the history and formation of species. Harvard University Press, Cambridge, Mass.

Ax,P., 1996. Multicellular animals. A new approach to the phylogenetic order in nature. Springer, Berlin, Heidelberg, New York.

Bensasson,D., Zhang,D., Hartl,D.L., Hewitt,G.M., 2001. Mitochondrial pseudogenes: evolution‘s misplaced witnesses. Trends Ecol. Evol. 16, 314-321.

Black,W.C., Roehrdanz,R.L., 1998. Mitochondrial gene order is not conserved in arthropods: prostriate and metastriate tick mitochondrial genomes. Mol. Biol. Evol. 15, 1772-1785.

Blanchard,J.L., Lynch,M., 2000. Organellar genes: why do they end up in the nucleus? Trends Genet. 16, 315-320.

Bonaldo,A.B., Pinto-da-Rocha,R., 2003. On a new species of *Cryptocellus* from the Brasilian Amazonas (Arachnida, Ricinulei). Revista Ibérica de Aracnología 7, 103-108.

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

Boore,J.L., 1999. Animal mitochondrial genomes. Nucl. Acids Res. 27, 1767-1780.

Boore,J.L., Brown,W.M., 2000. Mitochondrial genomes of *Galathealinum*, *Helobdella*, and *Platynereis*: sequence and gene arrangement comparisons indicate that Pogonophora is not a phylum and Annelida and Arthropoda are not sister taxa. Mol. Biol. Evol. 17, 87-106.

Boore,J.L., Collins,T.M., Stanton,D., Daehler,L.L., Brown,W.M., 1995. Deducing the pattern of arthropod phylogeny from mitochondrial DNA rearrangements. Nature 376, 163-165.

Botero-Trujillo,R., 2008. A new species of *Cryptocellus* (Arachnida, Ricinulei) from northwestern Colombia. J. Arachnol. 36, 468-471.

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

Carapelli,A., Lio,P., Nardi,F., van der Wath,E., Frati,F., 2007. Phylogenetic analysis of mitochondrial protein coding genes confirms the reciprocal paraphyly of Hexapoda and Crustacea. BMC Evol. Biol. 7, S8.

Castresana,J., Feldmaier-Fuchs,G., Pääbo,S., 1998. Codon reassignment and amino acid composition in hemichordate mitochondria. Proc. Natl. Acad. Sci. U.S.A 95, 3703-3707.

Castresana,J., 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol. Biol. Evol. 17, 540-552.

Chenna,R., Sugawara,H., Koike,T., Lopez,R., Gibson,T.J., Higgins,D.G., Thompson,J.D., 2003. Multiple sequence alignment with the Clustal series of programs. Nucl. Acids Res. 31, 3497-3500.

Choi,E.H., Park,S.J., Jang,K.H., Hwang,W., 2007. Complete mitochondrial genome of a Chinese scorpion *Mesobuthus martensii* (Chelicerata, Scorpiones, Buthidae). DNA Seq. 18, 459-471.

Clayton,D.A., 1982. Replication of animal mitochondrial DNA. Cell 28, 693-705.

Clayton,D.A., 1991. Replication and transcription of vertebrate mitochondrial DNA. Annu. Rev. Cell Biol. 7, 453-478.

Cokendolpher,J., Enriquez,T., 2004. A new species and records of *Pseudocellus* (Arachnida: Ricinulei: Ricinoididae) from caves in Yucatán, Mexico and Belize. Texas Memorial Museum, Speleological Monographs 6, 95-99.

Cooke,J.A.L., 1967. Observations on the biology of Ricinulei (Arachnida) with description of two new species of *Cryptocellus*. J. Zool. 151, 31-42.

Davila,S., Pinero,D., Bustos,P., Cevallos,M.A., Davila,G., 2005. The mitochondrial genome sequence of the scorpion *Centruroides limpidus* (Karsch 1879) (Chelicerata; Arachnida). Gene 360, 92-102.

Dermauw,W., Van Leeuwen,T., Vanholme,B., Tirry,L., 2009. The complete mitochondrial genome of the house dust mite *Dermatophagoides pteronyssinus* (Trouessart): a novel gene arrangement among arthropods. BMC Genomics 10, 107.

Domes,K., Maraun,M., Scheu,S., Cameron,S.L., 2008. The complete mitochondrial genome of the sexual oribatid mite *Steganacarus magnus*: genome rearrangements and loss of tRNAs. BMC Genomics 9, 532.

Dowton,M., Austin,A.D., 1999. Evolutionary dynamics of a mitochondrial rearrangement “hot spot” in the Hymenoptera. Mol. Biol. Evol. 16, 298-309.

Dowton,M., Campbell,N.J.H., 2001. Intramitochondrial recombination - is it why some mitochondrial genes sleep around? Trends Ecol. Evol. 16, 269-271.

Dowton,M., Cameron,S.L., Dowavic,J.I., Austin,A.D., Whiting,M.F., 2009. Characterization of 67 Mitochondrial tRNA Gene Rearrangements in the Hymenoptera Suggests That

Mitochondrial tRNA Gene Position Is Selectively Neutral. Mol. Biol. Evol. 26, 1607-1617.

Dowton,M., Castro,L.R., Austin,A.D., 2002. Mitochondrial gene rearrangements as phylogenetic characters in the invertebrates: The examination of genome 'morphology'. Invertebr. Syst. 16, 345-356.

Duarte,G.T., De Azeredo-Espin,A.M., Junqueira,A.C., 2008. The mitochondrial control region of blowflies (Diptera: Calliphoridae): a hot spot for mitochondrial genome rearrangements. J. Med. Entomol. 45, 667-676.

Dunlop,J.A., 1996. Evidence for a sister group relationship between Ricinulei and Trigonotarbida. Bull. Br. Arachnol. Soc.10, 193-204.

Dunlop,J.A., Arango,C.P., 2005. Pycnogonid affinities: a review. J. Zool. Syst. Evol. Res. 43, 8-21.

Dunn,C.W., Hejnol,A., Matus,D.Q., Pang,K., Browne,W.E., Smith,S.A., Seaver,E., Rouse,G.W., Obst,M., Edgecombe,G.D., Sorensen,M.V., Haddock,S.H.D., Schmidt-Rhaesa,A., Okusu,A., Kristensen,R.M., Wheeler,W.C., Martindale,M.Q., Giribet,G., 2008. Broad phylogenomic sampling improves resolution of the animal tree of life. Nature 452, 745-749.

Eberhard,J.R., Wright,T.F., Bermingham,E., 2001. Duplication and concerted evolution of the mitochondrial control region in the parrot genus Amazona. Mol. Biol. Evol. 18, 1330-1342.

Fahrein,K., Talarico,G., Braband,A., Podsiadlowski,L., 2007. The complete mitochondrial genome of *Pseudocellus pearsei* (Chelicerata : Ricinulei) and a comparison of mitochondrial gene rearrangements in Arachnida. BMC Genomics 8.

Fahrein,K., Masta,S.E., Podsiadlowski,L., 2009. The first complete mitochondrial genome sequences of Amblypygi (Chelicerata: Arachnida) reveal conservation of the ancestral arthropod gene order. Genome 52, 456-466.

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

Felsenstein,J., 1978. Cases in which Parsimony or Compatibility Methods Will be Positively Misleading. Syst. Zool.27, 401-410.

Fet,V., Sissom,D.W., Lowe,G., Braunwalder,M.E., 2000. Catalog of the scorpions of the World (1758-1998). New York.

Fet,V., Gantenbein,B., Soleglad,M.E., Vignoli,V., Salomone,N., Fet,E., V, Schembri,P.J.,

2003. New molecular and morphological data on the "*Euscorpius carpathicus*" species complex (Scorpiones: Euscorpiidae) from Italy, Malta, and Greece justify the elevation of *E. c. sicanus* (C. L. Koch, 1837) to the species level. Revue Suisse de Zoologie 110, 355-379.

Firstman,B., 1973. The Relationship of the Chelicerate Arterial System to the Evolution of the Endosternite. J. Arachnol.1, 1-54.

Fujita,M.K., Boore,J.L., Moritz,C., 2007. Multiple Origins and Rapid Evolution of Duplicated Mitochondrial Genes in Parthenogenetic Geckos (*Heteronotia binoei*; Squamata, Gekkonidae). Mol. Biol. Evol. 24, 2775-2786.

Gillespie,J.J., Johnston,J.S., Cannone,J.J., Gutell,R.R., 2006. Characteristics of the nuclear (18S, 5.8S, 28S and 5S) and mitochondrial (12S and 16S) rRNA genes of *Apis mellifera* (Insecta: Hymenoptera): structure, organization, and retrotransposable elements. Insect Mol. Biol.15, 657-686.

Giribet,G., Edgecombe,G.D., Wheeler,W.C., Babbitt,C., 2002. Phylogeny and systematic position of Opiliones: a combined analysis of chelicerate relationships using morphological and molecular data. Cladistics 18, 5-70.

Giribet,G., Richter,S., Edgecombe,G.D., Wheeler,W.C., 2005. The position of crustaceans within Arthropoda - evidence from nine molecular loci and morphology. Crustacean Issues 16, 307-330.

Graeber,M.B., Muller,U., 1998. Recent developments in the molecular genetics of mitochondrial disorders. J. Neurol. Sci. 153, 251-263.

Grasshoff,M., 1978. A model of the evolution of the main chelicerate groups. Symp. Zool. Soc. London 42, 273-284.

Gray,M.W., 1999. Evolution of organellar genomes. Curr. Opin. Genet. Dev. 9, 678-687.

Haen,K.M., Lang,B.F., Pomponi,S.A., Lavrov,D.V., 2007. Glass Sponges and Bilaterian Animals Share Derived Mitochondrial Genomic Features: A Common Ancestry or Parallel Evolution? Mol. Biol. Evol. 24, 1518-1527.

Halanych,K.M., 2004. The new view of animal phylogeny. Ann. Rev. Ecol. Evol. Syst. 35, 229-256.

Hall,T.A., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl. Acids. Symp. Ser. 41, 95-98.

Hansen,H.J., Sørensen,W., 1904. The order Ricinulei. In: Hansen,H.J., Sorensen,W. (Eds.), On two Orders of Arachnida. Cambridge University Press, Cambridge, pp. 114-157.

Harvey,M.S., 2002. The neglected cousins: What do we know about the smaller Arachnid orders? J. Arachnol. 30, 357-372.

Harvey,M.S., 2003. Catalogue of the Smaller Arachnid Orders of the World: Amblypygi, Uropygi, Schizomida, Palpigradi, Ricinulei and Solifugae. CSIRO Publishing.

Harvey,M.S., 2007. The smaller arachnid orders: diversity, descriptions and distributions from Linnaeus to the present (1758 to 2007). Zootaxa 363-380.

Hassanin,A., Leger,N., Deutsch,J., 2005. Evidence for multiple reversals of asymmetric mutational constraints during the evolution of the mitochondrial genome of metazoa, and consequences for phylogenetic inferences. Syst. Biol. 54, 277-298.

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. Mol. Phylogenet. Evol. 38, 100-116.

Helfenbein,K.G., Brown,W.M., Boore,J.L., 2001. The complete mitochondrial genome of the articulate brachiopod *Terebratalia transversa*. Mol. Biol. Evol. 18, 1734-1744.

Hoffmann,R.J., Boore,J.L., Brown,W.M., 1992. A Novel Mitochondrial Genome Organization for the Blue Mussel, *Mytilus edulis*. Genetics 131, 397-412.

Huelsenbeck,J.P., Ronquist,F., 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17, 754-755.

Hwang,U.W., Park,C.J., Yong,T.S., Kim,W., 2001b. One-step PCR amplification of complete arthropod mitochondrial genomes. Mol. Phylogenet. Evol. 19, 345-352.

Hwang,U.W., Friedrich,M., Tautz,D., Park,C.J., Kim,W., 2001a. Mitochondrial protein phylogeny joins myriapods with chelicerates. Nature 413, 154-157.

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.

Jobb,G., von,H.A., Strimmer,K., 2004. TREEFINDER: a powerful graphical analysis environment for molecular phylogenetics. BMC Evol. Biol. 4, 18.

Jobb,G., 2007. Treefinder. Version of Feb. 2007. Munich, Germany. Distributed by the author at www.treefinder.de.

Jocque,R., Dippenaar-Schoeman,A.S., 2007. Spider Families of the World, Second Edition.

Royal Mus. Centr. Afr.

Jones,M., Gantenbein,B., Fet,V., Blaxter,M., 2007. The effect of model choice on phylogenetic inference using mitochondrial sequence data: Lessons from the scorpions. Mol. Phylogenet. Evol. 43, 583-595.

Kambhampati,S., Smith,P.T., 1995. PCR primers for the amplification of four insect mitochondrial gene fragments. Insect Mol. Biol. 4, 233-236.

Kennaugh,J.H., 1968. An examination of the cuticle of three species of Ricinulei (Arachnida). J. Zool. 156, 393-404.

Kilpert,F., Podsiadlowski,L., 2006. The complete mitochondrial genome of the common sea slater, *Ligia oceanica* (Crustacea, Isopoda) bears a novel gene order and unusual control region features. BMC Genomics 7, 241.

Kilpert,F., Podsiadlowski,L., 2010. The Australian fresh water isopod (Phreatoicidea: Isopoda) allows insights into the early mitogenomic evolution of isopods. Comparative Biochemistry and Physiology Part D: Genomics and Proteomics 5, 36-44.

Knoll,A.H., Holland,H.D., 1995. Oxygen and Proterozoic evolution: an update. In: S. Stanley (Ed.), Effects of Past Global Change on Life. Natl. Acad. Press, Washington, DC, pp. 21-33.

Knoll,A.H., 1992. The Early Evolution of Eukaryotes: A Geological Perspective. Science 256, 622-627.

Kroemer,G., Dallaporta,B., Resche-Rigon,M., 1998. The mitochondrial death/life regulator in apoptosis and necrosis. Ann. Rev. Physiol. 60, 619-642.

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

Laurie,M., 1889. On a Silurian scorpion and some additional eurypterid remains from the Pentland Hills. Trans. R. Soc. Edinb., Earth Sci. 39, 515-590.

Lavrov,D.V., Boore,J.L., Brown,W.M., 2000. The complete mitochondrial DNA sequence of the horseshoe crab *Limulus polyphemus*. Mol. Biol. Evol. 17, 813-824.

Lavrov,D.V., Brown,W.M., 2001. *Trichinella spiralis* mtDNA: a nematode mitochondrial genome that encodes a putative ATP8 and normally structured tRNAs and has a gene arrangement relatable to those of coelomate metazoans. Genetics 157, 621-637.

Lavrov,D.V., Brown,W.M., Boore,J.L., 2004. Phylogenetic position of the Pentastomida and (pan)crustacean relationships. Proc. R. Soc. Lond B Biol. Sci. 271, 537-544.

Le,T.H., Blair,D., Agatsuma,T., Humair,P.F., Campbell,N.J.H., Iwagami,M., Littlewood,D.T., Peacock,B., Johnston,D.A., Bartley,J., Rollinson,D., Herniou,E.A., Zarlenga,D.S., McManus,D.P., 2000. Phylogenies Inferred from Mitochondrial Gene Orders - A Cautionary Tale from the Parasitic Flatworms. Mol. Biol. Evol. 17, 1123-1125.

Lee,Y.S., Oh,J., Kim,Y.U., Kim,N., Yang,S., Hwang,U.W., 2008. Mitome: dynamic and interactive database for comparative mitochondrial genomics in metazoan animals. Nucl. Acids Res. 36, 938-942.

Legg,G., 1976. The external morphology of a new species of ricinuleid (Arachnida) from Sierra Leone. Zool. J. Linn. Soc. 59, 1-58.

Legg,G., 1977. Sperm transfer and mating in *Ricinoides hanseni* (Ricinulei: Arachnida). J. Zool. 182, 51-61.

Lindquist,E.E., 1984. Current theories on the evolution of major groups of Acari and on their relationships with other groups of Arachnida, with consequent implications for their classification, 1 edition. John Wiley, New York.

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

Macey,J.R., Verma,A., 1997. Re: Homology in phylogenetic analysis: alignment of transfer RNA genes and the phylogenetic position of snakes. Mol. Phylogenet. Evol. 7, 272-279.

Macey,J.R., Schulte,J.A., Larson,A., Papenfuss,T.J., 1998. Tandem duplication via light-strand synthesis may provide a precursor for mitochondrial genomic rearrangement. Mol. Biol. Evol. 15, 71-75.

Macey,J.R., Schulte,J.A., Larson,A., 2000. Evolution and phylogenetic information content of mitochondrial genomic structural features illustrated with acrodont lizards. Syst. Biol. 49, 257-277.

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

Mallatt,J., Giribet,G., 2006. Further use of nearly complete 28S and 18S rRNA genes to classify Ecdysozoa: 37 more arthropods and a kinorhynch. Mol. Phylogenet. Evol. 40, 772-794.

Mallatt,J.M., Garey,J.R., Shultz,J.W., 2004. Ecdysozoan phylogeny and Bayesian inference: first use of nearly complete 28S and 18S rRNA gene sequences to classify the arthropods

and their kin. Mol. Phylogenet. Evol. 31, 178-191.

Mallatt,J., Winchell,C.J., 2002. Testing the new animal phylogeny: First use of combined large-subunit and small-subunit rRNA gene sequences to classify the protostomes. Mol. Biol. Evol. 19, 289-301.

Margulis,L., 1970. Origin of eukaryotic cells - evidence and research implications for a theory of the origin and evolution of microbial, plant, and animal cells on the Precambrian earth. Yale University Press, New Haven.

Martin,W., Herrmann,R.G., 1998. Gene Transfer from Organelles to the Nucleus: How Much, What Happens, and Why? Plant Physiol. 118, 9-17.

Masta,S.E., 2000. Mitochondrial sequence evolution in spiders: intraspecific variation in tRNAs lacking the TPsiC Arm. Mol. Biol. Evol. 17, 1091-1100.

Masta,S.E., Boore,J.L., 2004. The Complete Mitochondrial Genome Sequence of the Spider *Habronattus oregonensis* Reveals Rearranged and Extremely Truncated tRNAs. Mol. Biol. Evol. 21, 893-902.

Masta,S.E., 2010. Mitochondrial rRNA secondary structures and genome arrangements distinguish chelicerates: comparisons with a harvestman (Arachnida: Opiliones: *Phalangium opilio*). Gene 449, 9-21.

Masta,S.E., Boore,J.L., 2008. Parallel Evolution of Truncated Transfer RNA Genes in Arachnid Mitochondrial Genomes. Mol. Biol. Evol. 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., Longhorn,S.J., Boore,J.L., 2009. Arachnid relationships based on mitochondrial genomes: Asymmetric nucleotide and amino acid bias affects phylogenetic analyses. Mol. Phylogenet. Evol. 50, 117-128.

Maxmen,A., Browne,W.E., Martindale,M.Q., Giribet,G., 2005. Neuroanatomy of sea spiders implies an appendicular origin of the protocerebral segment. Nature 437, 1144-1148.

McFadden,G.I., 2001. Primary and secondary endosymbiosis and the origin of plastids. J. Phycol. 37, 951-959.

Mitchell,R.W., 1970. Population size and dispersion and species associations of a Mexican cavernicole ricinuleid (Arachnida). Ciencia, Mexico 27, 63-74.

Moritz,C., Dowling,T.E., Brown,W.M., 1987. Evolution of Animal Mitochondrial-DNA - Relevance for Population Biology and Systematics. Annu. Rev. Ecol. Syst. 18, 269-292.

Moritz,M., 1993. Überklasse Chelicerata, Fühlerlose. In: Gruner,H.-E. (Ed.), Lehrbuch der speziellen Zoologie/ Band I: Wirbellose Tiere/ 4 Teil: Arthropoda (ohne Insecta). Gustav Fischer Verlag, Stuttgart, Jena, New York, pp. 73-442.

Mwinyi,A., Meyer,A., Bleidorn,C., Lieb,B., Bartolomaeus,T., Podsiadlowski,L., 2009. Mitochondrial genome sequence and gene order of *Sipunculus nudus* give additional support for an inclusion of Sipuncula into Annelida. BMC Genomics 10, 27.

Navajas,M., Le Conte,Y., Solignac,M., Cros-Arteil,S., Cornuet,J.M., 2002. The complete sequence of the mitochondrial genome of the honeybee ectoparasite mite Varroa destructor (Acari: Mesostigmata). Mol. Biol. Evol. 19, 2313-2317.

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

Okimoto,R., Macfarlane,J.L., Clary,D.O., Wolstenholme,D.R., 1992. The Mitochondrial Genomes of Two Nematodes, *Caenorhabditis elegans* and *Ascaris suum*. Genetics 130, 471-498.

Oliveira,M.T., Zeredo-Espin,A.M.L., Lessinger,A.C., 2007. The mitochondrial DNA control region of muscidae flies: Evolution and structural conservation in a dipteran context. J. Mol. Evol. 64, 519-527.

Park,S.J., Lee,Y.S., Hwang,U., 2007. The complete mitochondrial genome of the sea spider *Achelia bituberculata* (Pycnogonida, Ammotheidae): arthropod ground pattern of gene arrangement. BMC Genomics 8, 343.

Perna,N.T., Kocher,T.D., 1995. Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes. J. Mol. Evol. 41, 353-358.

Petrunkevitch,A., 1955. Arachnida. In: Moore,R.C. (Ed.), Treatise on Invertebrate Paleontology, Part P, Arthropoda 2. Geol. Soc. Americ. and Univ. Kansas Press, Lawrence, Kansas, pp. 42-162.

Pinto-da-Rocha,R., Bonaldo,A.B., 2007. A new species of *Cryptocellus* (Arachnida, Ricinulei) from Oriental Amazonia. Zootaxa 1386, 47-51.

Pinto-da-Rocha,R., Machado,G., & Giribet,G., 2010. Harvestmen - The Biology of Opiliones. Harvard University Press.

Pittard,K., Mitchell,R.W., 1972. Comparative morphology of the life stages of *Cryptocellus*

pelaezi (Arachnida, Ricinulei). Grad. Stud. Texas Tech. Univ. 1, 3-77.

Platnick,N.I., 1980. On the phylogeny of Ricinulei. Proc. 8th Intern. Congr. Arachnol, 349-353.

Platnick,N.I., 2009. The world spider catalog, version10.0. http://research.amnh.org/entomology/spiders/catalog/index.html. Am. Mus. Nat. Hist.

Podsiadlowski,L., Braband,A., 2006. The complete mitochondrial genome of the sea spider *Nymphon gracile* (Arthropoda: Pycnogonida). BMC Genomics 7, 284.

Podsiadlowski,L., Bartolomaeus,T., 2006. Major rearrangements characterize the mitochondrial genome of the isopod *Idotea baltica* (Crustacea: Peracarida). Mol. Phylogenet. Evol. 40, 893-899.

Prendini,L., Weygoldt,P., Wheeler,W.C., 2005. Systematics of the *Damon variegatus* group of African whip spiders (Chelicerata: Amblypygi): Evidence from behaviour, morphology and DNA. Org. Div. Evol. 5, 203-236.

Purcell,W.F., 1899. New South African scorpions in the collection of the South African Museum. Ann. South Afr. Mus. 1, 433-438.

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

Raw,F., 1957. Origin of Chelicerates. J. Paleont. 31, 139-192.

Regier,J.C., Shultz,J.W., 2001. Elongation factor-2: a useful gene for arthropod phylogenetics. Mol. Phylogenet. Evol. 20, 136-148.

Regier,J.C., Shultz,J.W., Kambic,R.E., 2005. Pancrustacean phylogeny: hexapods are terrestrial crustaceans and maxillopods are not monophyletic. Proc. R. Soc. Lond B Biol. Sci. 272, 395-401.

Regier,J.C., Shultz,J.W., Zwick,A., Hussey,A., Ball,B., Wetzer,R., Martin,J.W., Cunningham,C.W., 2010. Arthropod relationships revealed by phylogenomic analysis of nuclear protein-coding sequences. Nature 463(7284):1079-1083.

Reyes,A., Gissi,C., Pesole,G., Saccone,C., 1998. Asymmetrical directional mutation pressure in the mitochondrial genome of mammals. Mol. Biol. Evol. 15, 957-966.

Ronaghi,M., Karamohamed,S., Pettersson,B., Uhlqn,M., Nyrqn,P., 1996. Real-Time DNA Sequencing Using Detection of Pyrophosphate Release. Analyti. Biochem. 242, 84-89.

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.

Sagan,L., 1967. On the origin of mitosing cells. J. Theoret. Biol. 14, 225-IN6.

Saito,S., Tamura,K., Aotsuka,T., 2005. Replication origin of mitochondrial DNA in insects. Genetics 171, 1695-1705.

Sancar,A., Sancar,G.B., 1988. DNA Repair Enzymes. Ann. Rev. Biochem. 57, 29-67.

Sanger,F., Nicklen,S., Coulson,A.R., 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. 74, 5463-5467.

Schimper,A.F.W., 1883. Über die Entwicklung der Chlorophyllkörner und Farbkörper. Bot. Z. 41, 105-14, 121-31, 137-46, 153-62.

Selden,P.A., Dunlop,J.A., 1998. Fossil taxa and relationships of chelicerates. In: Edgecombe,G. (Ed.), Arthropod Fossils and Phylogeny. New York, pp. 303-331.

Shadel,G.S., Clayton,D.A., 1997. Mitochondrial DNA maintenance in vertebrates. Ann. Rev. Biochem. 66, 409-435.

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 Mol. Biol. 13, 219-224.

Shao,R., Mitani,H., Barker,S.C., Takahashi,M., Fukunaga,M., 2005a. Novel mitochondrial gene content and gene arrangement indicate illegitimate inter-mtDNA recombination in the chigger mite, *Leptotrombidium pallidum*. J. Mol. Evol. 60, 764-773.

Shao,R., Barker,S.C., Mitani,H., Aoki,Y., Fukunaga,M., 2005b. Evolution of duplicate control regions in the mitochondrial genomes of metazoa: a case study with Australasian *Ixodes* ticks. Mol. Biol. Evol. 22, 620-629.

Shao,R., Barker,S.C., Mitani,H., Takahashi,M., Fukunaga,M., 2006. Molecular mechanisms for the variation of mitochondrial gene content and gene arrangement among chigger mites of the genus *Leptotrombidium* (Acari: Acariformes). J. Mol. Evol. 63, 251-261.

Shear,W.A., Selden,P.A., Rolfe,W.D.I., Bonamo,P.M., Grierson,J.D., 1987. New terrestrial arachnids from the Devonian of Gilboa, New York (Arachnida, Trigonotarbida). Am. Mus. Novit. 2901, 1-74.

Shultz,J.W., 1989. Morphology of Locomotor Appendages in Arachnida - Evolutionary Trends and Phylogenetic Implications. Zool. J. Linn. Soc. 97, 1-56.

Shultz,J.W., 1990. Evolutionary morphology and phylogeny of Arachnida. Cladistics 6, 1-38.

Shultz,J.W., 1999. Muscular anatomy of a whipspider, *Phrynus longipes* (Pocock) (Arachnida : Amblypygi), and its evolutionary significance. Zool. J. Linn. Soc. 126, 81-116.

Shultz,J.W., 2007. A phylogenetic analysis of the arachnid orders based on morphological characters. Zool. J. Linn. Soc. 150, 221-265.

Shultz,J.W., Regier,J.C., 2000. Phylogenetic analysis of arthropods using two nuclear protein-encoding genes supports a crustacean + hexapod clade. Proc. R. Soc. Biol. Sci. Ser. B 267, 1011-1019.

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. Ann. Ent. Soc. Am. 87, 651-701.

Snodgrass,R.E., 1938. Evolution of Annelida, Onychophora and Arthropoda. Smithosian Misc Collection 97, 1-77.

Stamatakis,A., Hoover,P., Rougemont,J., 2008. A rapid bootstrap algorithm for the RAxML Web servers. Syst. Biol. 57, 758-771.

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. Mol. Biol. Evol. 14, 867-874.

Swofford,D.L., 1993. Paup - A Computer-Program for Phylogenetic Inference Using Maximum Parsimony. J. Gen. Physiol.102, A9.

Taanman,J.W., 1999. The mitochondrial genome: structure, transcription, translation and replication. Biochim. Biophys. Acta 1410, 103-123.

Talarico,G., Garcia Hernandez,L.F., Michalik,P., 2008. The male genital system of the New World Ricinulei (Arachnida): ultrastructure of spermatozoa and spermiogenesis with special emphasis on its phylogenetic implications. Arthropod Struct. Dev. 37, 396-409.

Tanaka,M., Ozawa,T., 1994. Strand asymmetry in human mitochondrial DNA mutations. Genomics 22, 327-335.

Thompson,J.D., Higgins,D.G., Gibson,T.J., 1994. Clustal-W - Improving the Sensitivity of Progressive Multiple Sequence Alignment Through Sequence Weighting, Position-Specific Gap Penalties and Weight Matrix Choice. Nuc. Acids Res. 22, 4673-4680.

Valverde,J.R., Marco,R., Garesse,R., 1994. A Conserved Heptamer Motif for Ribosomal-Rna Transcription Termination in Animal Mitochondria. Proc. Natl. Acad. Sci. U.S.A. 91, 5368-5371.

Van der Hammen,L., 1972. A revised classification of the mites (Arachnidea, Acarida) with diagnoses, a key, and notes on phylogeny. Zool. Meded. Leiden 47, 273-292.

Van der Hammen,L., 1979. Comparative studies in Chelicerata I. The Cryptognomae (Ricinulei, Architarbi and Anactinotrichida). Zool. Verh. Leiden 174, 1-62.

Van der Hammen,L., 1982. Comparative studies in Chelicerata II. Epimerata (Palpigradi and Actinotrichida). Zool. Verh. Leiden 196, 1-70.

Van der Hammen,L., 1989. An introduction to comparative arachnology. SPB Academic Publishing, The Hague.

Wallin,I.E., 1923. The Mitochondria Problem. The American Naturalist 57, 255-261.

Wei,Y.H., 1998. Oxidative stress and mitochondrial DNA mutations in human aging. Proc. Soc. Exp. Biol. Med. 53-63.

Weygoldt,P., Paulus,H.F., 1979. Studies on the Morphology, Taxonomy and Phylogeny of the Chelicerata .2. Cladograms and Development of Chelicerata. Z. Zool. Syst. Evol. Forsch. 17, 177-200.

Weygoldt,P., 1996. Evolutionary morphology of whip spiders: Towards a phylogenetic system (Chelicerata: Arachnida: Amblypygi). J. Zool. Syst. Evol. Res. 34, 185-202.

Weygoldt,P., 1998. Evolution and systematics of the Chelicerata. Exp. Appl. Acarol. 22, 63-79.

Weygoldt,P., 2000. Whip Spiders (Amblypygi): Biology, Morphology and Systematics. Apollo Books, Stenstrup, Denmark.

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

Wolstenholme,D.R., Macfarlane,J.L., Okimoto,R., Clary,D.O., Wahleithner,J.A., 1987. Bizarre tRNAs inferred from DNA sequences of mitochondrial genomes of nematode worms. Proc. Natl. Acad. Sci. U.S.A. 84, 1324-1328.

Wolstenholme,D.R., 1992. Animal mitochondrial DNA: structure and evolution. Int. Rev. Cytol. 141, 173-216.

Xia,X., Xie,Z., 2001. DAMBE: software package for data analysis in molecular biology and evolution. J. Hered. 92, 371-373.

Xia,X., Xie,Z., Salemi,M., Chen,L., Wang,Y., 2003. An index of substitution saturation and its application. Mol. Phylogenet. Evol. 26, 1-7.

Yamauchi,M.M., Miya,M.U., Machida,R.J., Nishida,M., 2004. PCR-based approach for sequencing mitochondrial genomes of decapod crustaceans, with a practical example from kuruma prawn (*Marsupenaeus japonicus*). Mar. Biotechnol. (NY) 6, 419-429.

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

Zhang,D.X., Szymura,J.M., Hewitt,G.M., 1995. Evolution and structural conservation of the control region of insect mitochondrial DNA. J. Mol. Evol. 40, 382-391.

10. Appendix

The appendix contains all additional figures and tables of Chapter 3, 4, 6 and 7. Supplementary data for the article of Chapter 3 is also available on the journal Web site (http://genome.nrc.ca).

Supplementary file 3.1: PCR primers used in this study.

Primer	Sequence (5`- 3`)	Annealing temperature (°C)
Art-HPK16Saa [1]	ATGCTACCTTTGCACRGTCAAGATACYGCGGC	65
Art-HPK16Sbb [1]	CTTATCGAYAAAAAAGWTTGCGACCTCGATGTTG	65
Art-HPK16SA [2]	CGCCTGTTTATCAAAAACAT	50
Art-HPK16SB [3]	CCGGTTGAACTCAGATCA	50
Dd_2f-48	GGGCAATGAAAGAGATAGTG	48
Dd_48-2f	TCACCCAAAAATCCCC	48
Dd_42-36	GGGTGGTGACTTGTGTTGC	50
Dd_36-42	CCTCCGCCTCCTCCAC	50
Dd_35-30	TTGTGTGTGGTTGGGAGG	50
Dd_30-35	CCCACTGAAAAATAAGCAATC	50
Dd_29-25	AATTTATTGTGTAGAGTATATAGAGGG	50
Dd_25-29	CCCAACAATAGACATAAAAACC	50
Dd_24-15	ATTCATGGATTAAGCCAACTAG	50
Dd_15-24	TTCCATTACTCCCTTATCACC	50
Dd_15-13	GTAGAGGGGTCGAAGATGG	50
Dd_13-15	CTAAAAATAGACGCTGTCCC	50
Dd_13-11	CCCAGGTGTTAGGGTTCAG	50
Dd_11-13	AACTCCTCGATTGACATTACC	50
Phr-cytb-UF	CTTCCCTTTGTCATCTTTGCCTTA	65
Phr-cytb-LR	TATCCTCCTCAAAGTCATTGGACG	65

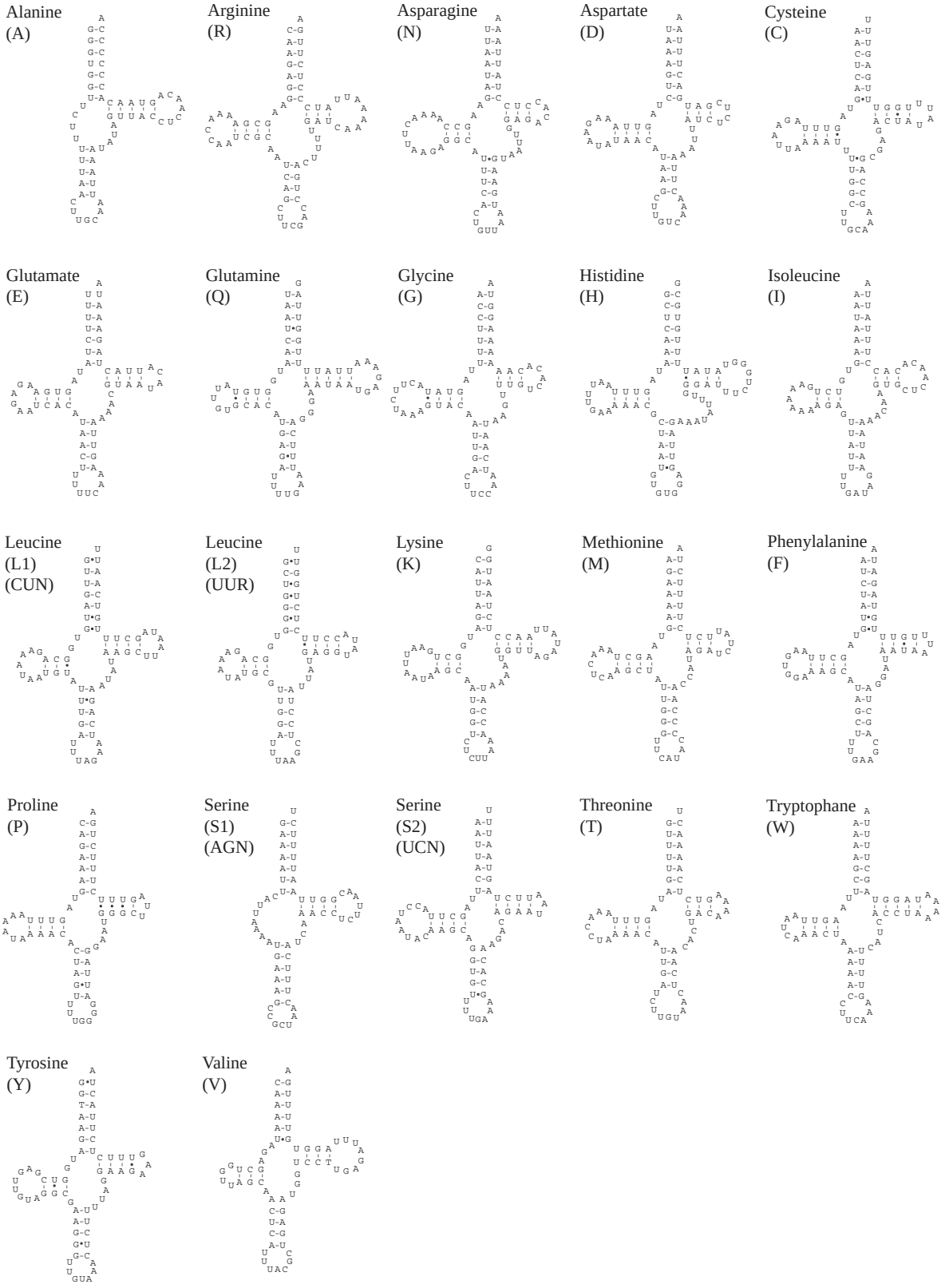
[1] Hwang et al. 2001, [2] Simon et al. 1994, [3] Kambhampati and Smith 1995

Supplementary file 3.2. Genome organisation of *Damon diadema* (14786 bp).

Gene	Strand	Position		Length (nuc.)	GC-skew	Start-codon	Stop-codon	Intergenic nucleotides
cox1	+	1 -	1539	1539	-0.235	ATC	TAA	+5
cox2	+	1545-	2220	676	-0.379	ATG	T	+3
trnK	+	2224-	2293	70				+1
trnD	+	2295-	2358	64				0
atp8	+	2359-	2514	156	-0.765	ATC	TAA	-7
atp6	+	2508-	3178	671	-0.471	ATG	TA	0
cox3	+	3179-	3962	784	-0.336	ATG	T	0
trnG	+	3963-	4029	67				0
nad3	+	4030-	4371	342	-0.486	ATT	TAA	0
trnA	+	4372-	4429	58				+4
trnR	+	4434-	4500	67				+3
trnN	+	4504-	4570	67				0
trnS-AGN	+	4571-	4630	60				0
trnE	+	4631-	4695	65				-2
trnF	-	4694-	4758	65				0
nad5	-	4759-	6460	1702	0.620	ATG	T	-11
trnH	-	6450-	6526	77				-15
nad4	-	6512-	7858	1347	0.645	ATA	TAA	-11
nad4L	-	7848-	8159	312	0.696	ATC	TAG	-8
trnT	+	8152-	8215	64				-1
trnP	-	8215-	8278	64				+1
nad6	+	8280-	8729	450	-0.636	ATG	TAA	+3
cob	+	8733-	9835	1103	-0.383	ATG	TA	0
trnS-UCN	+	9836-	9902	67				+23
nad1	-	9926-	10837	912	0.619	ATG	TAA	0
trnL-UUR	-	10838-	10900	63				+6
trnL-CUN	-	10907-	10972	66				-7
rrnL	-	10966-	12259	1294	0.551			0
trnV	-	12260-	12322	63				0
rrnS	-	12323-	13072	750	0.490			+364
non-coding		13073-	13436	364				0
trnI	+	13437-	13502	66				0
trnQ	-	13503-	13571	69				+2
trnM	+	13574-	13638	65				0
nad2	+	13639-	14601	963	-0.545	ATC	TAA	-2
trnW	+	14600-	14665	66				-1
trnC	-	14665-	14730	66				0
trnY	-	14731-	8	64				-8

Supplementary file 3.3. Genome organisation of *Phrynus sp.* (14764 bp).

Gene	Strand	Position	Length (nuc.)	GC-skew	Start-codon	Stop-codon	Intergenic nucleotides
cox1	+	1 - 1536	1536	-0.192	TTA	TAA	+3
cox2	+	1540- 2212	673	-0.362	ATG	T	-3
trnK	+	2210- 2275	66				+3
trnD	+	2279- 2345	67				0
atp8	+	2346- 2501	156	-0.628	ATC	TAA	-7
atp6	+	2495- 3163	669	-0.475	ATG	TAA	+6
cox3	+	3170- 3955	786	-0.338	ATG	TAA	-2
trnG	+	3954- 4020	67				0
nad3	+	4021- 4366	346	-0.528	ATT	T	0
trnA	+	4367- 4423	57				+2
trnR	+	4426- 4491	66				+2
trnN	+	4494- 4557	64				+13
trnS-AGN	+	4571- 4627	57				0
trnE	+	4628- 4694	67				-1
trnF	-	4694- 4755	62				0
nad5	-	4756- 6433	1678	0.569	ATG	T	+3
trnH	-	6437- 6499	63				-1
nad4	-	6499- 7839	1341	0.559	ATG	TAA	-20
nad4L	-	7820- 8119	300	0.786	ATG	TAG	+8
trnT	+	8128- 8190	63				-1
trnP	-	8190- 8254	65				+2
nad6	+	8257- 8712	456	-0.592	ATT	TAA	-1
cob	+	8712- 9821	1110	-0.471	ATG	TAA	-1
trnS-UCN	+	9821- 9884	64				+17
nad1	-	9902- 10810	909	0.647	ATG	TAA	+1
trnL-UUR	-	10812- 10882	71				+2
trnL-CUN	-	10885- 10948	64				-7
rrnL	-	10942- 12159	1218	0.570			+21
trnV	-	12181- 12246	66				+6
rrnS	-	12253- 13038	786	0.543			+359
non-coding		13039- 13397	359				0
trnI	+	13398- 13461	64				+3
trnQ	-	13465- 13530	66				+7
trnM	+	13538- 13601	64				+1
nad2	+	13603- 14580	978	-0.487	ATG	TAA	-2
trnW	+	14579- 14641	63				+1
trnC	-	14643- 14706	64				0
trnY	-	14707- 5	63				-5



Supplementary file 3.4. Plots of inferred secondary structure of tRNAs from the mitochondrial genome of *Damon diadema*. Dots illustrate the base pairing of the pyrimidine base uracil with the purine base guanine.

Supplementary file 4.1 a: Genome organisation of *Cryptocellus narino* (14554 bp).

Gene	*Strand*	*Position*		*Length (nuc.)*	*CG-skew*	*Start-codon*	*Stop-codon*	*Intergenic nucleotides*
cox1	+	1-	1548	1548	0,254	CTA	TAA	-1
cox2	+	1548-	2219	672	0,411	ATG	TAA	+1
trnK	+	2221-	2290	70				-2
trnD	+	2289-	2350	62				0
atp8	+	2351-	2500	150	0,6	GTA	TAA	-7
atp6	+	2494-	3165	672	0,458	ATG	TAA	-1
cox3	+	3165-	3947	783	0,328	ATG	TAA	+1
trnG	+	3949-	4011	63				+6
nad3	+	4018-	4348	331	0,49	ATC	T	-1
trnA	+	4348-	4415	68				+2
trnR	+	4418-	4479	62				-1
trnN	+	4479-	4542	64				0
trnS-AGN	+	4543-	4603	61				+1
trnE	+	4605-	4667	63				+8
trnF	-	4676-	4735	60				0
nad5	-	4736-	6422	1687	-0,605	ATT	T	0
trnH	-	6423-	6487	65				+7
nad4	-	6495-	7806	1312	-0,546	ATG	T	+1
nad4L	-	7808-	8080	273	-0,696	TTG	TAG	+2
trnT	+	8083-	8149	67				-1
trnP	-	8150-	8210	61				+5
nad6	+	8216-	8650	435	0,669	ATA	TAA	0
cob	+	8650-	9753	1104	0,48	ATG	TAA	+5
trnS-UCN	+	9759-	9821	63				-10
nad1	-	9812-	10702	891	-0,528	ATT	TAA	+9
trnL- UUR	-	10712-	10773	62				+2
trnL- CUN	-	10776 -	10838	63				*
rrnL	-	10839-	12025	1187	-0,44			*
trnV	-	12026-	12088	63				*
rrnS	-	12089-	12828	740	-0,361			*
non-coding		12829-	13232	404				*
trnI	+	13233-	13295	63				+4
trnQ	-	13300-	13366	67				+9
trnM	+	13376-	13437	62				-18
nad2	+	13420-	14385	966	0,345	ATG	TAA	+2
trnY	-	14388-	14449	62				0
trnW	+	14450-	14511	62				-8
trnC	-	14504-	10	61				-10

* Borders determined according to adjacent gene boundaries

Supplementary file 4.1 b: Genome organisation of *Pseudocellus gertschi* (14477 bp).

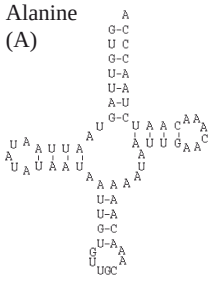
Gene	*Strand*	*Position*		*Length (nuc.)*	*CG-skew*	*Start-codon*	*Stop-codon*	*Intergenic nucleotides*
cox1	+	1-	1536	1536	0,233	TTA	TAA	-1
cox2	+	1536-	2206	671	0,290	ATG	TA	0
trnK	+	2207-	2277	71				-2
trnD	+	2276-	2339	64				0
atp8	+	2340-	2492	153	0,545	ATA	TAA	-7
atp6	+	2486-	3157	672	0,472	ATG	TAA	-1
cox3	+	3157-	3940	784	0,24	ATG	T	-3
trnG	+	3938-	3999	62				0
nad3	+	4000-	4333	334	0,551	ATA	T	0
trnA	+	4334-	4399	68				-1
trnR	+	4399-	4461	63				0
trnN	+	4462-	4524	63				0
trnS-AGN	+	4525-	4582	58				-1
trnE	+	4582-	4644	63				-2
trnF	-	4643-	4702	60				0
nad5	-	4703-	6386	1684	-0,536	ATG	T	0
trnH	-	6387-	6448	62				+8
nad4	-	6457-	7771	1315	-0,567	ATG	T	-13
nad4L	-	7759-	8055	297	-0,612	GTT	TAG	-16
trnT	+	8040-	8103	64				0
trnP	-	8104-	8165	62				+10
nad6	+	8176-	8610	435	0,625	ATA	TAA	+2
cob	+	8613-	9717	1105	0,468	ATA	T	0
trnS-UCN	+	9718-	9783	66				+4
nad1	-	9788-	10673	886	-0,582	GTT	TAT	+9
trnL- UUR	-	10683-	10745	63				-2
trnL- CUN	-	10744-	10807	64				*
rrnL	-	10808-	11969	1162	-0,509			*
trnV	-	11970-	12026	57				*
rrnS	-	12027-	12769	743	-0,467			*
non-coding		12770-	13157	388				*
trnI	+	13158-	13220	63				+2
trnQ	-	13223-	13287	65				-4
trnM	+	13284-	13350	67				0
nad2	+	13351-	14301	951	0,391	ATT	TAA	+1
trnW	+	14303-	14364	62				+8
trnC	-	14357-	14417	61				+4
trnY	-	14422-	5	61				-5

* Borders determined according to adjacent gene boundaries

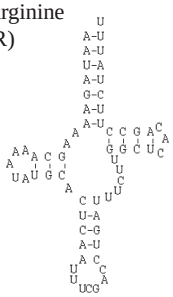
Supplementary file 4.2: Inferred secondary structures of tRNAs from the mitochondrial genome of *Cryptocellus narino* (a) and *Pseudocellus gertschi* (b). The grey shaded parts in the secondary structures of the leucine genes are marking termination signals for *rrnL* transcription or modifications of this termination signal. Dots illustrate the base pairing of the pyrimidine base uracil with the purine base guanine.

a) *Cryptocellus narino*

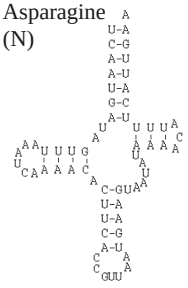
Alanine (A)



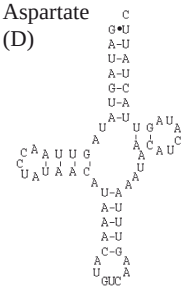
Arginine (R)



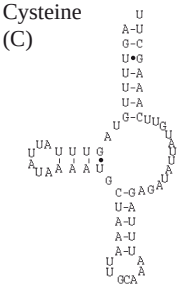
Asparagine (N)



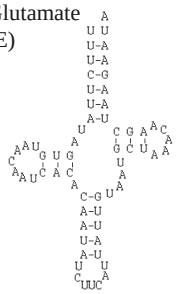
Aspartate (D)



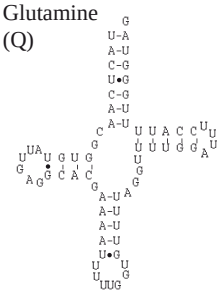
Cysteine (C)



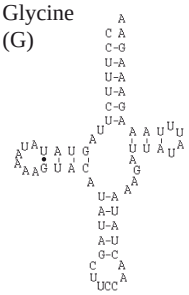
Glutamate (E)



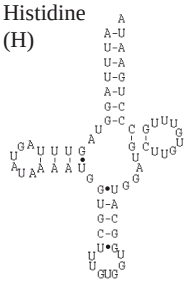
Glutamine (Q)



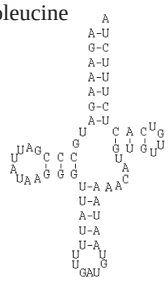
Glycine (G)



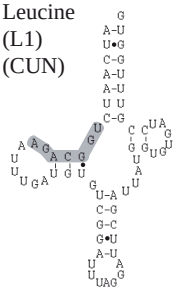
Histidine (H)



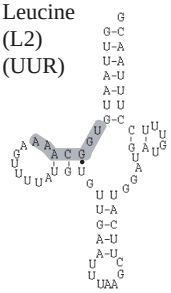
Isoleucine (I)



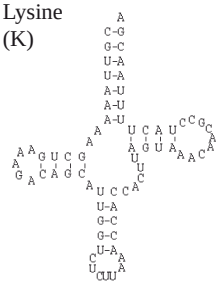
Leucine (L1) (CUN)



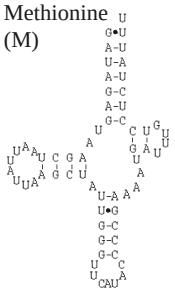
Leucine (L2) (UUR)



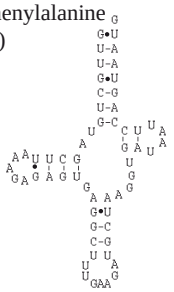
Lysine (K)



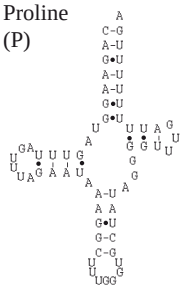
Methionine (M)



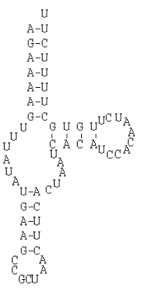
Phenylalanine (F)



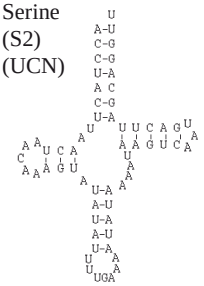
Proline (P)



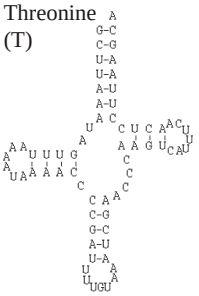
Serine (S1) (AGN)



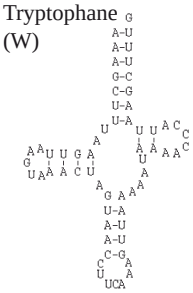
Serine (S2) (UCN)



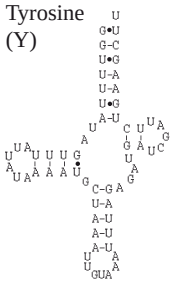
Threonine (T)



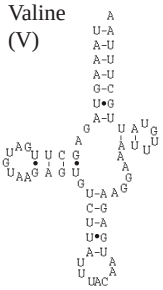
Tryptophane (W)



Tyrosine (Y)

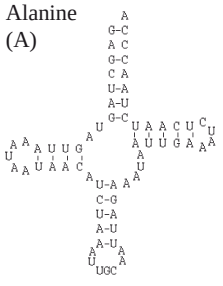


Valine (V)

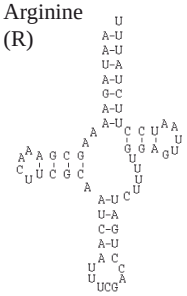


b) *Pseudocellus gertschi*

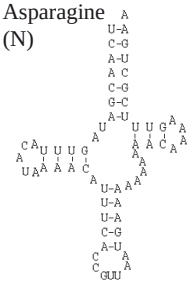
Alanine (A)



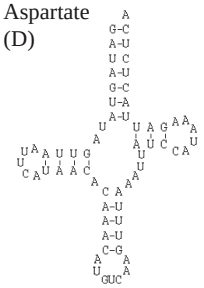
Arginine (R)



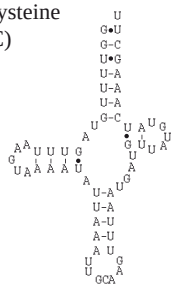
Asparagine (N)



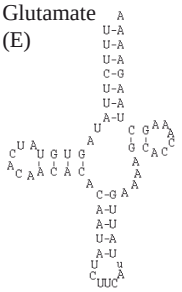
Aspartate (D)



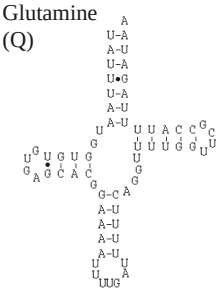
Cysteine (C)



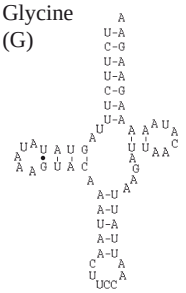
Glutamate (E)



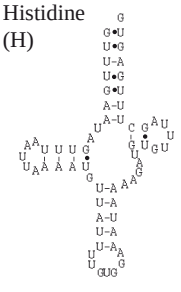
Glutamine (Q)



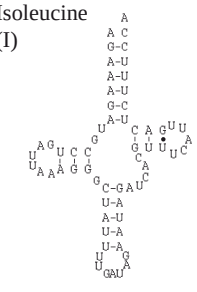
Glycine (G)



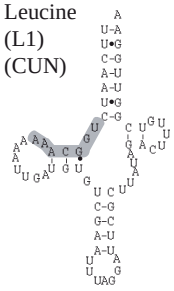
Histidine (H)



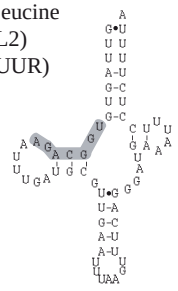
Isoleucine (I)



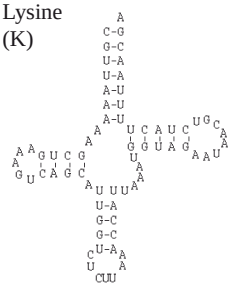
Leucine (L1) (CUN)



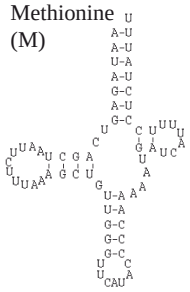
Leucine (L2) (UUR)



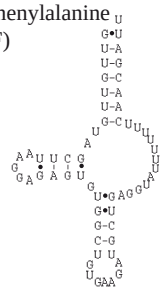
Lysine (K)



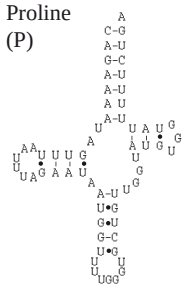
Methionine (M)



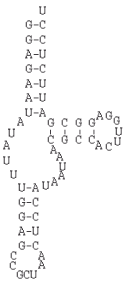
Phenylalanine (F)



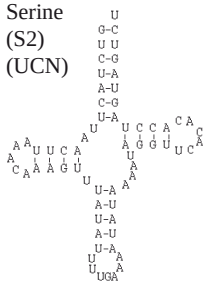
Proline (P)



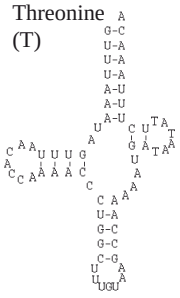
Serine (S1) (AGN)



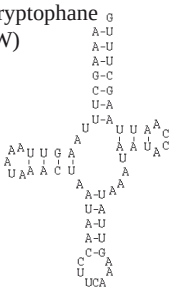
Serine (S2) (UCN)



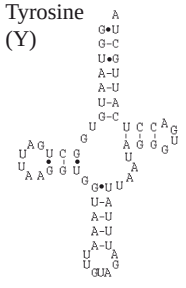
Threonine (T)



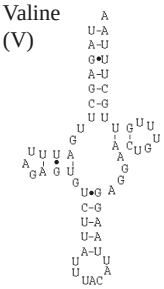
Tryptophane (W)



Tyrosine (Y)



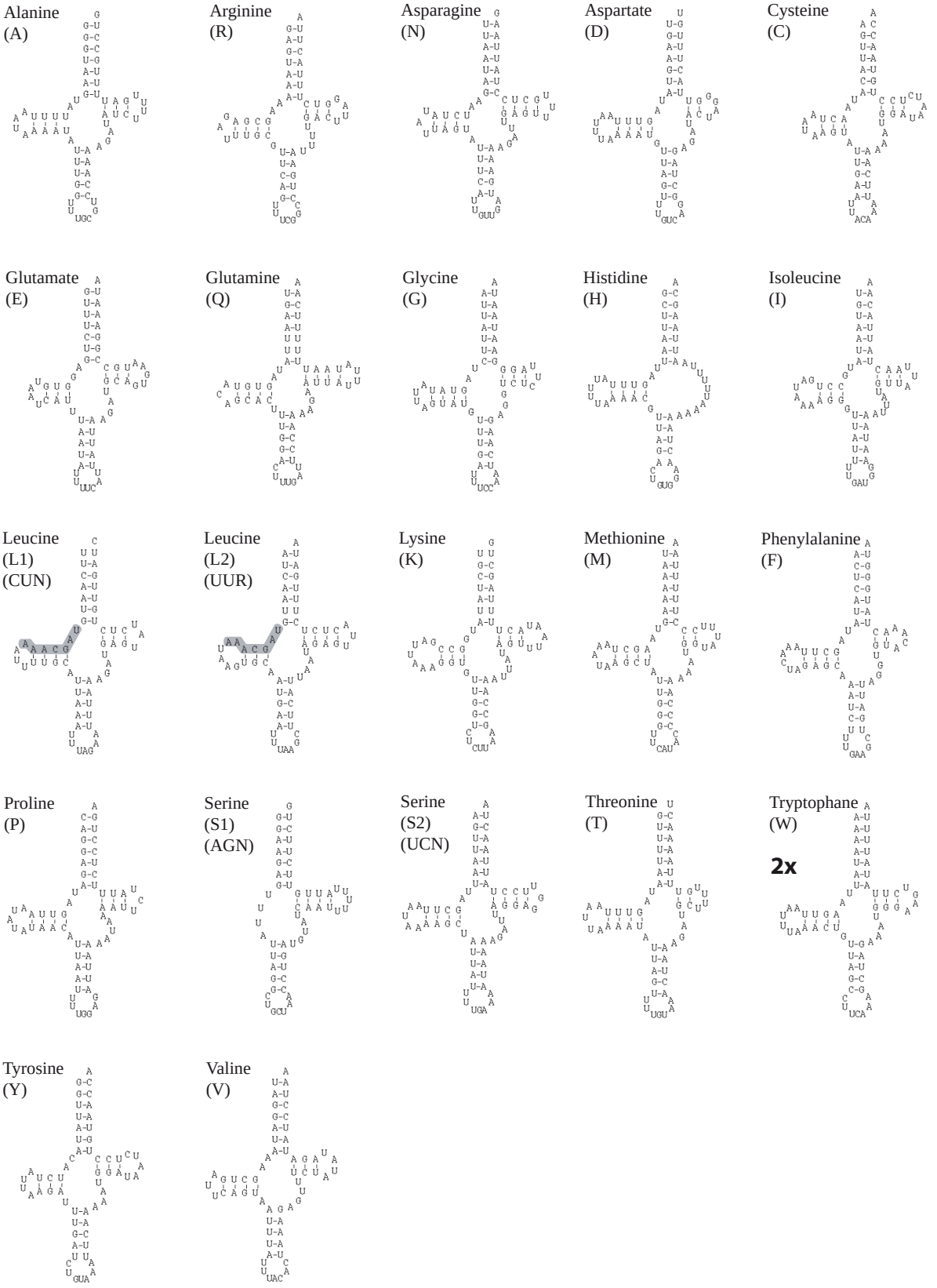
Valine (V)



Supplementary file 6.1: Genome organisation of *Hadogenes bicolor* (15083 bp).

Gene	*Strand*	*Position*		*Length (nuc.)*	*CG-skew*	*Start-codon*	*Stop-codon*	*Intergenic nucleotides*
cox1	+	1-	1539	1539	-0,326	ATA	TAA	+3
cox2	+	1543-	2212	670	-0,444	ATG	T	0
trnK	+	2213-	2276	64				-2
trnD	+	2275-	2335	61				0
atp8	+	2336-	2491	156	-0,689	ATT	TAG	-7
atp6	+	2485-	3150	666	-0,414	ATG	TAA	0
cox3	+	3151-	3933	783	-0,462	ATG	TAA	-1
trnG	+	3933-	3991	59				0
nad3	+	3992-	4331	340	-0,534	ATG	T	0
trnA	+	4332-	4392	61				0
trnR	+	4393-	4452	60				-3
trnN	+	4450-	4511	62				0
trnS1-AGN	+	4512-	4565	54				-1
trnE	+	4565-	4624	60				-2
trnF	-	4623-	4683	61				0
nad5	-	4684-	6367	1684	0,422	ATT	T	0
trnH	-	6368-	6426	59				0
nad4	-	6427-	7756	1330	0,313	ATG	T	-1
nad4L	-	7756-	8046	291	0,014	ATG	TAA	+2
trnT	+	8049-	8108	60				-1
trnP	-	8108-	8169	62				+2
nad6	+	8172-	8607	436	-0,391	ATT	T	0
cob	+	8608-	9729	1122	-0,125	ATG	TAG	-10
trnS2-UCN	+	9720-	9783	64				+18
nad1	-	9802-	10725	924	-0,066	ATA	TAA	-3
trnL2-UUR	-	10723-	10784	62				-3
trnL1-CUN	-	10782-	10841	60				0
rrnL	-	10842-	12003	1162	-0,232			*
trnV	-	12004-	12063	60				*
rrnS	-	12064-	12854	791	-0,271			*
non-coding		12855-	13579	725				*
trnI	+	13580-	13640	61				0
trnQ	-	13641-	13702	62				+1
trnM	+	13704-	13764	61				0
nad2	+	13765-	14733	969	-0,308	GTT	TAA	-2
trnW^{1}	+	14732-	14794	63				0
trnC	-	14795-	14854	60				*
non-coding		14855-	14966	112				*
trnW^{2}	+	14967-	15029	63				0
trnY	-	15030-	6	60				-6

* Borders determined according to adjacent gene boundaries



Supplementary file 6.2: Putative secondary structures of tRNAs from the mitochondrial genome of *Hadogenes bicolor*. Except for tRNA-H and tRNA-S1(AGN) all tRNAs can be folded into the usual cloverleaf secondary structure. From tRNA-W two copies are present (2x). The grey shaded parts in the secondary structures of both leucine genes are marking modified termination signals for *rrnL* transcription.

Supplementary file 6.3: PCR primers used in this study for amplification of mitochondrial fragments of *Hadogenes bicolor* (further primers according to: [1] Hwang et al. 2001, [2] Simon et al. 1994, [3] Kambhampati & Smith 1995).

Primer	Sequence (5`- 3`)	Ann. temp. (°C)
Art-HPK16Saa [1]	ATGCTACCTTTGCACRGTCAAGATACYGCGGC	65
Art-HPK16Sbb [1]	CTTATCGAYAAAAAAGWTTGCGACCTCGATGTTG	65
Art-HPK16SA [2]	CGCCTGTTTATCAAAAACAT	53
Art-HPK16SB [3]	CCGGTTGAACTCAGATCA	51
Hb-Cox1F	CATTTGTTTTGGTTTTTTGG	53
Hb-Cox1R	TCAGAATAACGACGCGG	53
Hb-16S-12S	GATTATCTCTCTTATTTGGTCCTTTC	57
Hb-12S-16S	AAGCTGATTCAGTAATTTATTTACAC	55
Hb-CrF	CTAAAAGGGAAGAAGATGGG	54
Hb-CrR	CAATAAATGAAAAGGAAGAGAAAC	55

Supplementary file 71: Genome organisation of *Liphistius erewan* (a), *Phyxioschema suthepium* (b), *Pholcus phalangioides* (c).

a) *Liphistius erewan* (14797 bp)

Gene	***Strand***	***Position***		***Length (nuc.)***	***CG-skew***	***Start-codon***	***Stop-codon***	***Intergenic nucleotides***
cox1	+	1-	1533	1533	0,196	CTG	TAA	-1
cox2	+	1533-	2202	670	0,361	ATG	T	-3
trnK	+	2200-	2263	64				-2
trnD	+	2262-	2317	56				0
atp8	+	2318-	2467	150	0,629	ATC	TAA	-7
atp6	+	2461-	3123	663	0,577	ATG	TAA	0
cox3	+	3124-	3907	784	0,288	TTG	T	0
trnG	+	3908-	3960	53				+15
nad3	+	3976-	4296	321	0,339	ATT	TAG	-2
trnA	+	4295-	4355	61				-1
trnR	+	4355-	4412	58				0
trnN	+	4413-	4470	58				0
trnS-AGN	+	4471-	4519	49				0
trnE	+	4520-	4577	58				-12
trnF	-	4566-	4619	54				0
nad5	-	4620-	6261	1642	-0,386	ATT	T	0
trnH	-	6262-	6319	58				-7
nad4	-	6313-	7623	1311	-0,375	ATG	TAA	-1
nad4L	-	7623-	7898	276	-0,481	ATG	TAA	+2
trnT	+	7901-	7956	56				0
trnP	-	7957-	8016	60				+2
nad6	+	8019-	8456	438	0,649	ATA	TAA	-1
cob	+	8456-	9583	1128	0,323	ATG	TAA	-2
trnS-UCN	+	9582-	9640	59				-10
nad1	-	9631-	10554	924	-0,475	ATA	TAG	-3
trnL-UUR	-	10552-	10610	59				0
trnL-CUN	-	10611-	10670	60				*
rrnL	-	10671-	11812	1142	-0,391			*
trnV	-	11813-	11867	55				*
rrnS	-	11868-	12557	690	-0,286			*
non-coding		12558-	12895	338				*
trnI	+	12896-	12954	59				+2
trnQ	-	12957-	13017	61				+2
trnM	+	13020-	13080	61				+1
nad2	+	13082-	14035	954	0,522	ATG	TAA	-2
trnW	+	14034-	14089	56				-1
trnC	-	14089-	14141	53				0
trnY	-	14142-	2	58				-2

b) *Phyxioschema suthepium* (13931 bp)

Gene	***Strand***	***Position***		***Length (nuc.)***	***CG-skew***	***Start-codon***	***Stop-codon***	***Intergenic nucleotides***
cox1	+	1-	1536	1536	-0,38	TTA	TAG	0
cox2	+	1537-	2202	666	-0,543	TTG	TAG	+3
trnK	+	2206-	2252	48				-7
trnD	+	2247-	2295	49				-3
atp8	+	2293-	2448	156	-0,72	ATT	TAA	-7
atp6	+	2442-	3110	669	-0,593	ATG	TAG	0
cox3	+	3111-	3894	784	-0,479	TTG	T	0
trnG	+	3895-	3949	55				-8
nad3	+	3942-	4280	339	-0,66	ATT	TAG	-9
trnL2-UUR	-	4272-	4323	52				-7
trnN	+	4317-	4372	56				-12
trnA	+	4361-	4420	60				-5
trnS1-AGN	+	4416-	4466	51				-4
trnR	+	4463-	4510	48				-6
trnE	+	4505-	4554	50				-12
trnF	-	4543-	4595	53				-8
nad5	-	4588-	6231	1644	0,51	ATT	TAT	-8
trnH	-	6224-	6276	53				-3
nad4	-	6269-	7549	1276	0,606	ATT	T	-5
nad4L	-	7545-	7814	270	0,41	ATC	TAT	-3
trnP	-	7812-	7860	49				0
nad6	+	7861-	8286	426	-0,635	TTG	TAT	-1
cob	+	8286-	9422	1137	-0,437	TTG	TAA	-11
trnS2-UCN	+	9412-	9472	61				0
trnT	+	9473-	9534	62				-17
nad1	-	9518-	10423	906	0,383	ATT	TAA	-8
trnL1-CUN	-	10416-	10471	56				*
rrnL	-	10472-	11536	1065	0,211			*
trnV	-	11537-	11590	54				*
rrnS	-	11591-	12252	662	0,143			*
trnI	-	12253-	12318	66				+24
trnQ	-	12343-	12407	65				*
non-coding		12408-	12794	387				*
trnM	+	12795-	12853	59				-9
nad2	+	12845-	13771	927	-0,789	ATT	TAG	+18
trnW	+	13790-	13840	51				-16
trnY	-	13825-	13876	52				+15
trnC	-	13892-	7	47				-7

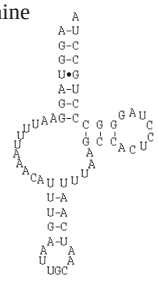
c) *Pholcus phalangioides* (14459 bp)

Gene	***Strand***	***Position***	***Length (nuc.)***	***CG-skew***	***Start-codon***	***Stop-codon***	***Intergenic nucleotides***
cox1	+	1 - 1536	1536	-0,304	ATA	TAG	-1
cox2	+	1536 - 2175	640	-0,375	GTG	T	+12
trnK	+	2188 - 2237	50				-6
trnD	+	2232 - 2300	69				-22
atp8	+	2279 - 2428	150	-0,412	ATT	TAA	-7
atp6	+	2422 - 3084	663	-0,327	GTG	TAA	+3
cox3	+	3088 - 3876	789	-0,345	ATG	TAA	-5
trnG	+	3872 - 3920	49				-4
nad3	+	3917 - 4249	333	-0,657	ATT	TAG	-5
trnL2-UUR	-	4245 - 4298	54				+67
trnN	+	4366 - 4416	51				-11
trnA	+	4406 - 4456	51				-5
trnS1-AGN	+	4452 - 4503	52				-13
trnR	+	4491 - 4540	50				-6
trnE	+	4535 - 4684	50				-21
trnF	-	4564 - 4622	59				-12
nad5	-	4611 - 6225	1615	0,484	ATA	T	-8
trnH	-	6218 - 6274	57				-14
nad4	-	6261 - 7535	1275	0,531	ATG	TAA	-1
nad4L	-	7535 - 7789	255	0,449	ATA	TAA	-8
trnP	-	7782 - 7837	56				0
nad6	+	7838 - 8266	429	-0,621	ATG	TAA	+3
cob	+	8270 - 9388	1119	-0,396	ATG	TAG	-9
trnS2-UCN	+	9380 - 9441	62				-1
trnT	+	9441 - 9501	61				-15
nad1	-	9487 - 10380	894	0,336	ATT	TAG	-5
trnL1-CUN	-	10376 - 10430	55				*
rrnL	-	10431 - 11435	1005	0,176			*
trnV	-	11436 - 11485	50				*
rrnS	-	11486 - 12175	690	0,162			*
trnI	+	12176 - 12229	54				-14
trnQ	-	12216 - 12277	62				*
non-coding		12278 - 13353	1076				*
trnM	+	13454 - 13422	69				-19
nad2	+	13404 - 14336	933	-0,479	GTG	TAA	0
trnW	+	14337 - 14387	51				-19
trnY	-	14369 - 14422	54				-5
trnC	-	14418 - 6	58				-6

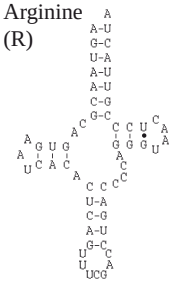
Supplementary file 7.2: Inferred secondary structures of tRNAs from the mitochondrial genomes of *Liphistius erewan* (a), *Phyxioschema suthepium* (b) and *Pholcus phalangioides* (c). In all three genomes these putative secondary structures of the detected tRNAs differ in several cases from the usual cloverleaf secondary structure. (b) also shows the putative secondary structure of the detected pseudogene of *trnI* in the genome of *P. suthepium*. The grey shaded parts in the secondary structures of the leucine genes are marking termination signals for *rrnL* transcription or modifications of this termination signal. Dots illustrate the base pairing of the pyrimidine base uracil with the purine base guanine.

a) *Liphistius erewan*

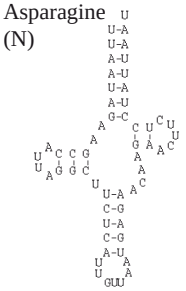
Alanine (A)



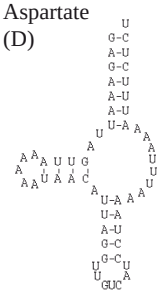
Arginine (R)



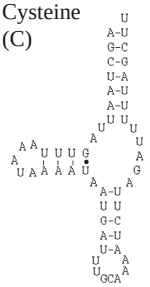
Asparagine (N)



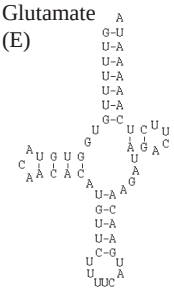
Aspartate (D)



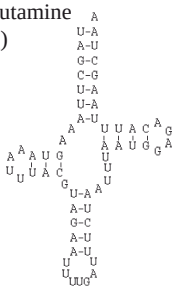
Cysteine (C)



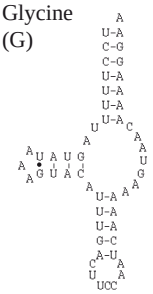
Glutamate (E)



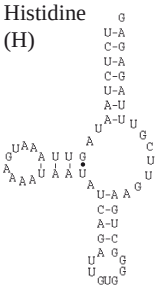
Glutamine (Q)



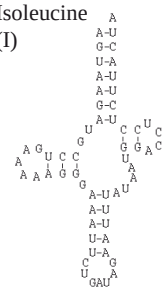
Glycine (G)



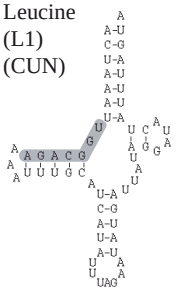
Histidine (H)



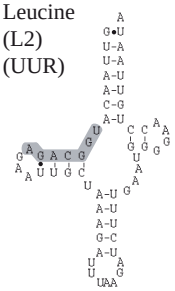
Isoleucine (I)



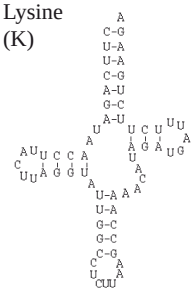
Leucine (L1) (CUN)



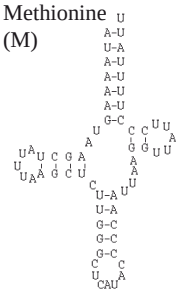
Leucine (L2) (UUR)



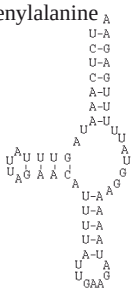
Lysine (K)



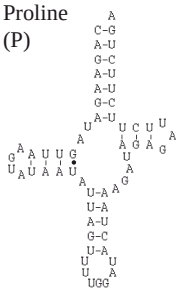
Methionine (M)



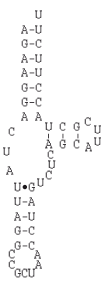
Phenylalanine (F)



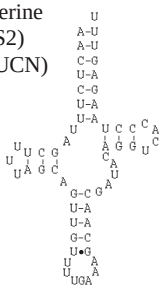
Proline (P)



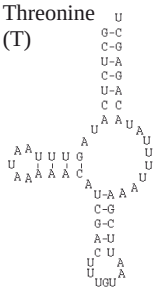
Serine (S1) (AGN)



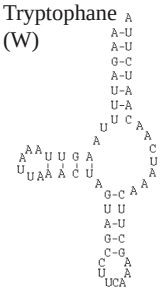
Serine (S2) (UCN)



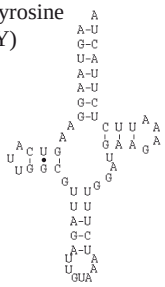
Threonine (T)



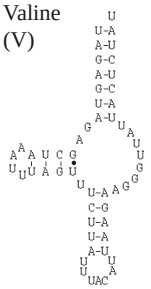
Tryptophane (W)



Tyrosine (Y)

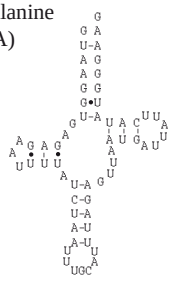


Valine (V)



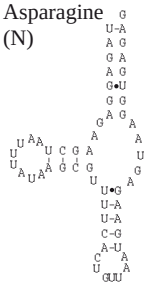
b) *Phyxioschema suthepium*

Alanine (A)



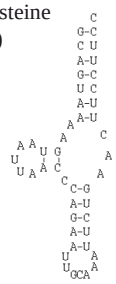
Arginine (R)

Asparagine (N)

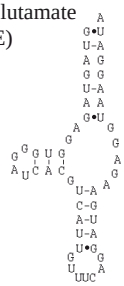


Aspartate (D)

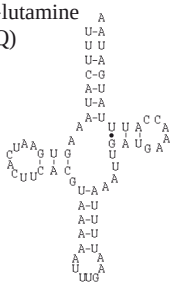
Cysteine (C)



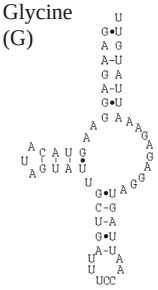
Glutamate (E)



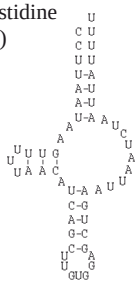
Glutamine (Q)



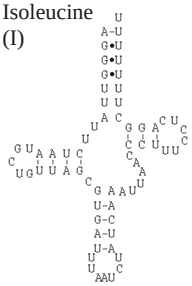
Glycine (G)



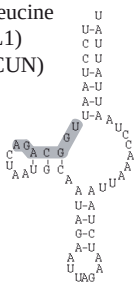
Histidine (H)



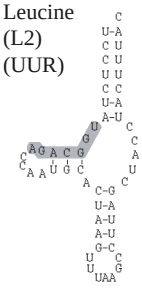
Isoleucine (I)



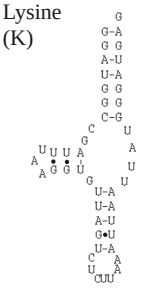
Leucine (L1) (CUN)



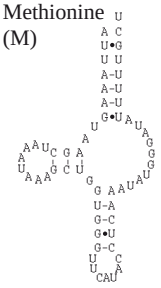
Leucine (L2) (UUR)



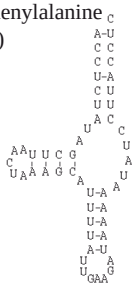
Lysine (K)



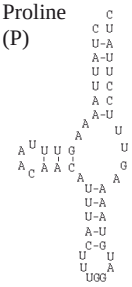
Methionine (M)



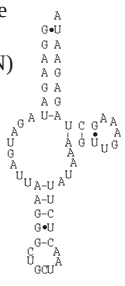
Phenylalanine (F)



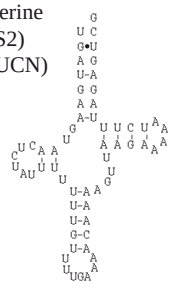
Proline (P)



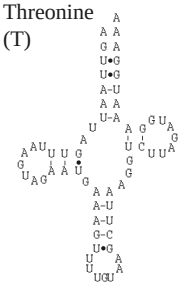
Serine (S1) (AGN)



Serine (S2) (UCN)

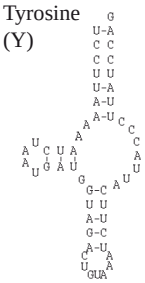


Threonine (T)

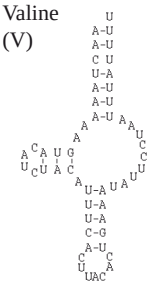


Tryptophane (W)

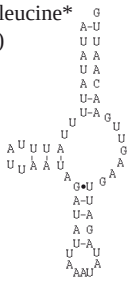
Tyrosine (Y)



Valine (V)

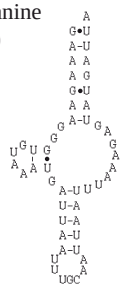


Isoleucine* (I*)

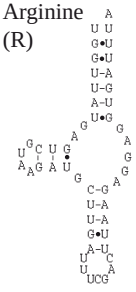


c) *Pholcus phalangioides*

Alanine (A)

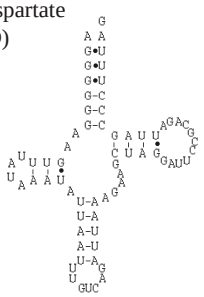


Arginine (R)

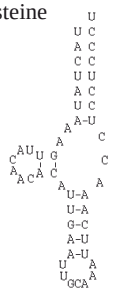


Asparagine (N)

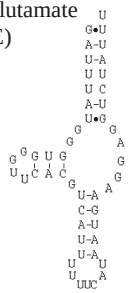
Aspartate (D)



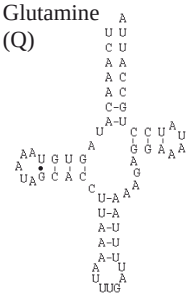
Cysteine (C)



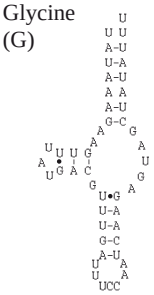
Glutamate (E)



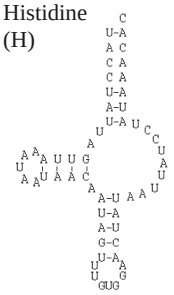
Glutamine (Q)



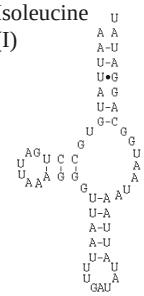
Glycine (G)



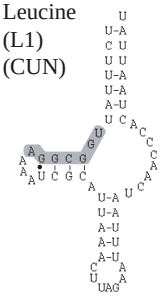
Histidine (H)



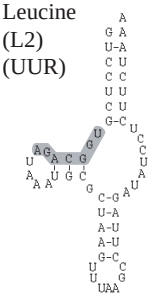
Isoleucine (I)



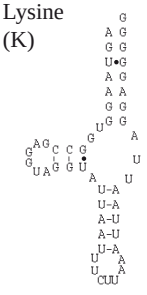
Leucine (L1) (CUN)



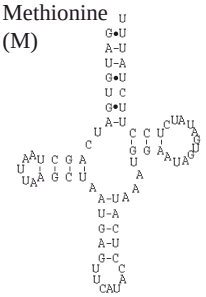
Leucine (L2) (UUR)



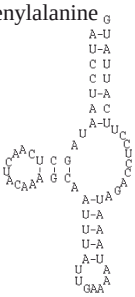
Lysine (K)



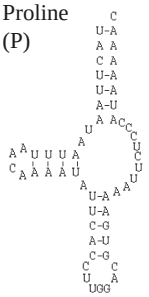
Methionine (M)



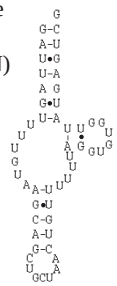
Phenylalanine (F)



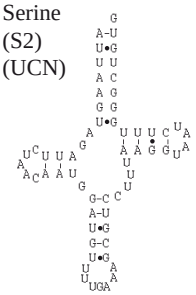
Proline (P)



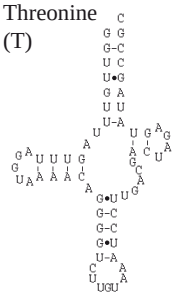
Serine (S1) (AGN)



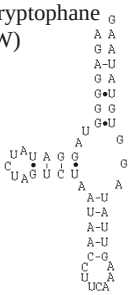
Serine (S2) (UCN)



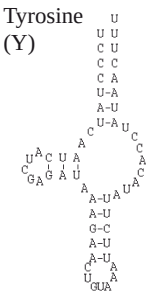
Threonine (T)



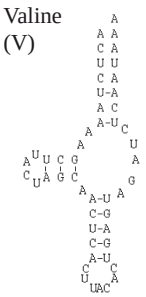
Tryptophane (W)



Tyrosine (Y)



Valine (V)



Supplementary file 7.3: PCR primers used in this study for amplification of mitochondrial fragments of *Liphistius erewan*[1], *Phyxioschema suthepium*[2], and *Pholcus phalangioides*[3].

Primer	Sequence (5`- 3`)	Ann. temp. (°C)
11-15[1]	ATAGGGGCTGTTTTTGC	51
15-11[1]	GCTGTTTGGGGGATAGAG	52
15-38[1]	CATCCTTACTACCTTTTCCC	51
38-15[1]	GATCCCAATGTTTCATACTTC	51
38-4[1]	TGCTACTTACCTTATTTTCCC	52
4-38[1]	GCCACCGATTCTTCCC	53
10-43[2]	TGCCTGTATGAGGTTCTTTG	53
43-10[2]	CATTCGCATTTTCAGGCATTC	60
44-8[2]	AGATAACTTTCTTCTTGGGG	51
8-44[2]	AGTTGAATATAATCATCGCAG	50
9-26[3]	GACATCGTTCTTCATGATACTTAC	53
26-9[3]	TAACCACCATTTACAACCTACG	55
28-9[3]	TGAAAGTCTGGCTGTGATTGAGG	62
9-28[3]	TCATAACCAAAGCATGAGCAGTTAC	61