MOLECULAR MARKERS OF THE MITOCHONDRIAL GENOMES OF ISOPODA AND IMPLICATIONS ON THE PHYLOGENY OF PERACARIDA (CRUSTACEA: MALACOSTRACA)

D i s s e r t a t i o n

Zur Erlangung des Grades

Doktor der Naturwissenschaften (Dr. rer. nat.)

Institut für Evolutionsbiologie und Zooökologie

Mathematisch-Naturwissenschaftliche Fakultät

Rheinische Friedrich-Wilhelms-Universität

Fabian Kilpert

aus Lübeck

Bonn, Dezember 2009

Diese Arbeit wurde mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Rheinischen Friedrich-Wilhelms-Universität Bonn angefertigt.

Erstgutachter: Prof. Dr. Thomas Bartolomaeus
Zweitgutachter: Privat-Dozent Dr. Lars Podsiadlowski

Tag der mündlichen Prüfung: 19. April 2010

Abgabe: Dezember 2009
Erscheinungsjahr: 2010

Eidesstattliche Erklärung

Hiermit erkläre ich, dass ich die vorliegende Arbeit ohne unzulässige Hilfe Dritter und ohne Benutzung 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.

Bonn, Dezember 2009

Fabian Kilpert

The dissertation is based on following publications:

Chapter 2: Kilpert F, Podsiadlowski L: 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 2006, 7:241

Chapter 3: Kilpert F, Podsiadlowski L: The Australian fresh water isopod (Eophreatoicidea: Isopoda) allows insights into the early mitogenomic evolution of isopods. Comp. Biochem. Phys. D; doi: 10.1016/j.cbd.2009.09.003

Chapter 4: Kilpert F, Held C, Podsiadlowski L: Rearrangements of the mitochondrial genome of Isopoda and implications on the phylogeny of Peracarida. Gen. Biol. Evol. – *submitted*.

Chapter 5: Kilpert F, Podsiadlowski L: The mitochondrial genome of the Japanese skeleton shrimp *Caprella mutica* (Amphipoda: Caprellidea) reveals a unique gene order and shared apomorphic translocations with Gammaridea. Mitochon. DNA – *submitted*.

Contributions of the listed authors:

Chapter 2: FK did the majority of laboratory work. LP was the initiator and the supervisor of the work. Manuscript writing was done by FK and LP in equal parts.

Chapter 3: FK did the majority of laboratory work and manuscript writing. LP was supervisor and helped in writing the manuscript.

Chapter 4: FK did the majority of laboratory work and manuscript writing. CH provided the *Glyptonotus* specimen and helped in manuscript writing. LP was supervisor and helped in writing the manuscript.

Chapter 5: FK did the majority of laboratory work and manuscript writing. TB was co-supervisor. LP was supervisor and helped in writing the manuscript.

Contributions of further persons are mentioned in the acknowledgement section of respective chapters.

Contents

1. General introduction

1.1 The mitochondrion

The mitochondrion is a specialized, independent compartment and of particular importance to the eukaryotic cell. A double membrane encloses the organelle and allows it to process several vital metabolic pathways separate from the cytoplasm (Brand, 1997; Seyffert, 2003). In this regard it also takes a key role in a number of crucial cell processes, e.g. aging (Wei, 1998), apoptosis (Kroemer *et al.*, 1998), diseases (Graeber and Muller, 1998). The mitochondrion is mainly well-known for being the site of oxidative phosphorylation. Electrons are transferred through a series of protein complexes (electron transport chain), located in the inner mitochondrial membrane, to an electron acceptor, which is in most animals oxygen (O_2). The energy released during that redox reaction is used to generate a potential of protons across the inner mitochondrial membrane, which is on its part used as an energy source for the enzyme ATP synthase to phosphorylate adenosine diphosphate (ADP) to adenonsine triphosphate (ATP). The molecule ATP serves as a general energy supply for the cell. For this reason, at least one or several mitochondria are present in almost every respiring animal cell, dependent on the required amount of energy. It is conspicuous that mitochondria reproduce by binary fission only and are generally inherited by the female germ line (maternal inheritance). Mitochondria also contain their own genome, referred to as the mitochondrial DNA (mtDNA), which encodes protein subunits of the electron transport chain. The mtDNA also uses a derived genetic code to specify amino acids (Osawa *et al.*, 1990; Osawa *et al.*, 1992).

The commonly accepted endosymbiont theory (Sagan, 1967) supposes that mitochondria are derived from prokaryotes (alpha-proteobacteria), which established a symbiotic relationship with the primitive eukaryotic cell. Eukaryotes bearing mitochondria are assumed to exist at least since the Palaeoprotozoic (2,500 to 1,600 mya) (Knoll *et al.*, 2006), the age when the oxygen level in the earth's atmosphere significantly rose by photosynthesis. Since then the mitochondrial (mt) genome has evolved alongside with the nuclear genome. Consequently, the evolutionary history should be reflected in the mtDNA of successive organisms (Saccone *et al.*, 1999).

1.2 The mitochondrial genome

The genomes of alpha-proteobacteria (e.g. of *Rickettsia prowazekii*) revealed a high similarity to the mtDNA of eukaryotes — a strong indication for their common origin. Early mitochondria certainly possessed the complete genetic information of their bacterial ancestors. In the course of gradual transformation to a cell organelle, however, the gene content was significantly reduced. Some genes were simply lost, as they were dispensable in an endosymbiontic environment, but most of the genes were successively transferred to the nuclear genome (Adams and Palmer, 2003; Martin *et al.*, 2001). This general trend is well traced by the declining number of mt genes in several protists, rhodophytes, and chlorophytes (Seyffert, 2003).

In mt genomes of bilaterian animals (Bilateria) the gene content has been reduced to 37 genes (2 rRNA genes, 13 protein subunit genes, 22 tRNA genes). The encoded polypeptides are all subunits of the protein complexes, which are part of the mitochondrial oxidative phosphorylation system (Taanman, 1999). The rRNAs are part of the mt protein synthesizing machinery. Apart from genes, also one non-coding region is found in the mtDNA. Bearing controlling elements for replication and transcription, it is called the mitochondrial control region (CR). Even though mitochondria may maintain their own system of DNA replication, transcription, mRNA processing and protein translation, they strongly rely on proteins and RNAs from the cytoplasm as well.

The mtDNA usually has a size of 15-20 kb and is organized as a ring-shaped double-helix. The larger mt genomes that are known up to now, only result from duplications of parts of the genome and not from variations of the gene content. Gene losses, in most cases of tRNA genes, were occasionally reported from certain lineages, e.g. nematodes, cnidarians, bivalves (Boore, 1999), and chaetognathes (Helfenbein *et al.*, 2004). The mt nucleotide sequence is generally evolving faster than in sequences from the nuclear genome.

The highly economical organization of the mt genome is striking: Unlike in nuclear DNA no introns and nearly no non-coding intergenic sequences exist except for a few bases. Instead, small gene overlaps are not uncommon. In cases, where all genes are distributed on the same DNA strand, genes are transcribed as usual. In most species, however, genes are located on both strands with opposite orientations. It has been studied for mammalian mt genomes representatively that in these cases the transcription

follows the ‘polycistron model’. Starting from two functionally independent transcription initiation sites, each strand is transcribed as a large polycistron, which is enzymatically processed to specific messages in subsequent steps (Boore, 1999; Taanman, 1999).

The mitochondrial replication process is also based on two initiation sites. But this time both sites are spatially separated. The second initiation site is located about two-thirds downstream of the mt genome. The result is known as the asymmetrical replication process of mtDNA (Clayton, 1982). When the unidirectional replication is started for one strand, the replication of the complementary strand is retarded and does not start until the replication loop unveiled the second initiation site. For this reason the second waiting strand remains single stranded for longer than the first, which leads to varied mutational constraints. This is probably the reason for the frequently observed asymmetric distribution of complementary nucleotides between both strands (Saccone *et al.*, 1999). That asymmetry can be expressed in GC- and AT-skews, whereby the GC-skew is the more significant identifier. As the strand bias is strongly correlated to the orientation of the replication process (leading/lagging strand), a reversed bias indicates an inversion of the control region. This finding also has to be considered in phylogenetic analyses, as a reversed bias consequently leads to a higher base substitution rate in the concerned taxa. Long branch attraction (LBA) (Felsenstein, 1978), the clustering of unrelated taxa due to many homoplasious substitutions, is certainly a resulting problem (Hassanin *et al.*, 2005).

1.3 Characters of ‘genome morphology’

Many studies on population genetics, phylogeography, and phylogeny used animal mt genomes as a primary source of data, due to a number of particularly suitable features like their small size, uniparental inheritance, lack of recombination, conserved gene content, exclusive presence of orthologous genes, and relatively constant gene order (Boore, 1999; Lavrov, 2007). While analyses mainly focused on nucleotide sequences, mostly of single genes (e.g. *cox1*, *rrnL*, *rrnS*), in the beginning, the use of non-sequence characters became more apparent just recently. They were named characters of ‘genome morphology’, since they are of the same descriptive type as morphological characters,

but furthermore also comprise all kinds of additional structural genomic features (Dowton *et al.*, 2002). Changes of gene arrangements are certainly best-known. But also changes of compositional strand bias (mentioned above), gene content, genetic code, mt genome topology, and secondary structure modifications turned out to be valuable resources after an increasing number of complete mt genomes were made available for comparison. These characters can also be combined together with morphological characters in data matrices.

Characters of 'genome morphology' may provide more reliable evidence on deeper phylogenetic relationships than comparative molecular data (Dowton *et al.*, 2002). Molecular sequences are affected by random mutations accumulated over time, which can blur the existent phylogenetic signal. One must bear in mind that there are only four alternate character states for each base position (A, C, G, T), making homoplasious mutations common. Evolutionary divergences of a few hundred of million years might also lie beyond the scope of sequence analyses, as they are based on mathematical and biochemical models, which become increasingly inexact in this vast timeframe (Penny *et al.*, 2001).

Mitochondrial gene rearrangements deserve particular mention, as they are the most promising of these different types of structural mt characters. They are naturally investigated in almost every study covering multi-gene fragments of the entire mt genome of a new species. One reason for their reliability certainly is the general rareness of gene order changes. Due to the absence of considerable non-coding sections in the mtDNA, gene translocation events are very likely to corrupt vital genes — which are in most cases fatal for the respective cell/organism. Thus, gene rearrangements are only possible within narrow limits, certainly the reason why re-organizations in mt genomes are observed less frequently than in the nuclear genome (Dowton *et al.*, 2002).

An impressive example of constancy of mt gene arrangements is known from arthropods: A variety of crustacean and insect species (e.g. Crustacea: *Daphnia pulex*, Insecta: *Drosophila yakuba*) exhibit exactly the same derived mt gene order (pancrustacean ground pattern), which differs from the ancestral arthropod condition only by a single translocated *trnL* gene (Dowton *et al.*, 2002; Kilpert and Podsiadlowski, 2006). This finding not only suggests a sister group relationship of crustaceans and insects, but also proves the flawless conservation of the mt gene arrangement in taxa which probably diverged more than 500 million years ago (Pisani, 2009).

However, if changes in gene order occur, they are assumed to be neutral to selection. The transcription mechanism predicts that all mt genes are transcribed as a polycistron, meaning that their actual position in the mtDNA is not important. Indeed, markedly different arrangements of mt genome are known from divergent taxa (e.g. invertebrates and vertebrates), suggesting that a complete reshuffling of gene order is not threatening the mitochondrial efficiency in any way.

Due to its complexity it is unlikely the same gene order arises convergent from independent reshaping. Dowton et al. (2002) calculated the chance of two mt genomes sharing the identical derived position of a single gene due to convergent evolution with only 1/2664, under the precondition of equal probability for all gene translocation events. At least tRNA genes are known for an accelerated mobility with respect to other mt genes (Saccone *et al.*, 1999; Saccone *et al.*, 2002); and genes within clusters of tRNA genes were found to swap positions more frequently (Dowton and Austin, 1999). This underlines the importance to consider gene translocation mechanisms a well, since they have the potential to raise the chance of homoplasy.

1.4 Mechanisms of gene translocation

The initial concept of mt gene translocation was developed for genes which were translocated to a new position on the same strand. The duplication random loss model (Figure 1.1a) implies that the newly synthesized strand can slip back to an already replicated part of the template strand (slipped strand mispairing) during the ongoing replication process. When the DNA polymerase now resumes the extension, parts of the genome are synthesized again. As a result, duplicates of genes exist on the new strand. Due to the high selection pressure to keep the mt genome small, microdeletions accumulate by chance in supernumerary genes copies, allowing only one of each orthologous gene to stay functional. The deletion of genes over time either reconstructs the original gene order or gives rise to a new gene arrangement, where single genes appear to have changed their relative positions (Boore, 2000; Dowton *et al.*, 2002). Inaccurate termination of replication was also suggested to cause duplication of genes (not depicted); in this case the synthesis of the new strand continues after the complete replication of the mitogenomic ring (Boore *et al.*, 1998)

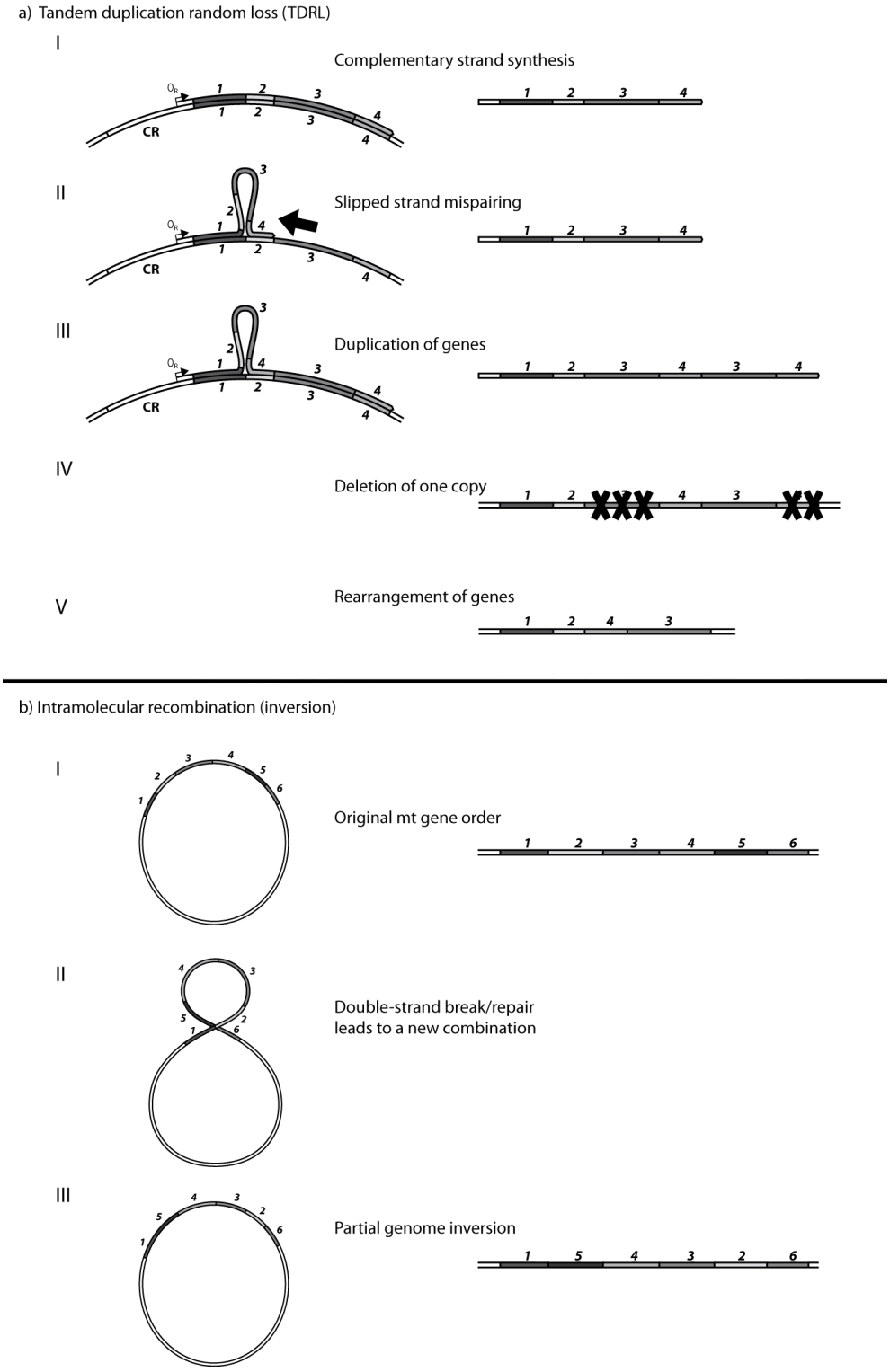


Figure 1.1: Mechanisms of gene translocation. a) Tandem duplication random loss (Dowton *et al.*, 2002). I) Starting at the replication origin (O_R) the complementary strand is synthesized. II) Due to chance homology the new strand slips back and allows base pairings of different genes. III) Complementary strand synthesis continues; genes *3* and *4* exist twice. IV) microdeletions accumulate over time. One redundant gene copy gets lost, V) resulting in a new gene order. b) Inversion as a result of intramolecular recombination. I) Initial gene order in the mt genome. II) A twist of the mtDNA ring causes a strand crossing. The double-strand break/repair integrates an inverted sequence. III) Genes 5 to 2 are inversed.

The remolding of tRNA genes can be regarded as an extension of the duplication random loss model. Subsequent to a duplication of a tRNA gene mutations occur in the anticodon and change the gene identity. A random drop of one supernumerary gene follows as described. This kind of event can be indicated by a high similarity in nucleotide sequences of original gene and gene copy. In addition, both genes are often located next to each other. A remolding from trnL(UUR) to trnL(CUN) seems to be not unusual, as it originated in several gastropods and crustaceans as result of convergent evolution (Rawlings *et al.*, 2003).

Long-range translocations of genes are probably not caused by a duplication random loss event. The original gene order would have to be restored for all but the translocated genes, a process which becomes more unlikely with increasing distance of translocations. Gene inversions can also not be explained by the common duplication random loss model. These events probably result from intramolecular recombination. This mechanism not only rearranges parts of the genome, but also inverts them at the same time (Figue 1.1b). There is so far no conclusive proof whether duplication random loss or intramolecular recombination is dominating mt gene rearrangements.

The lack of substantial non-coding sequences in the mt genome makes mt gene order changes generally difficult. The only exception is the mt control region, which bears some regulatory elements but no genes. For this reason it might be a favorite insertion area for translocations.

As mitochondria are only maternal inherited, apart from some exceptions, only those gene rearrangements are passed, which occur during the oogenesis (Dowton *et al.*, 2002).

1.5 Characterization of Isopoda

The Isopoda (Malacostraca: Peracarida) are a very species-rich suborder of peracarid crustaceans. More than 10,100 isopod species are currently known (Schotte *et al.*, 2009), reflecting the amazing ecological diversity of the isopod group. Just for comparison, for Crustacea as a whole the estimate number of described species was stated approximately 52,000 (Martin and Davis, 2001). Like other crustaceans isopods were initially marine species. In the course of time isopods found a world-wide

distribution. Nowadays they not only occupy all kinds of aquatic habitats, comprising the oceans from the littoral to the deep sea, but are also found in various fresh water environments including subterranean waters. The largest isopod subgroup is formed by terrestrial isopods (Oniscidea, >5,300 species), though. These isopods are largely independent form open water and are able to inhabit even arid regions (Schmidt, 2008). The direct development of eggs to juveniles takes place in a ventral brood pouch (marsupium) of the female, a main autapomorphy of Peracarida. The maternal brood care enabled by the marsupium is considered as a prerequisite for the later colonization of land by oniscidean isopods (Westheide and Rieger, 2007). Most isopod species are free living, but parasitic or symbiotic species are known as well. The type of feeding is diverse, since predators as well as carrion eaters, sediment feeders, and filterers are known. The typical isopod size is 1 to 5 centimeters, but extremes from 300 micrometers (Microcerberidae) to nearly 50 centimeters (*Bathynomus*) do also exist. The oval, dorsoventrally flattened habitus, that is usually associated with isopods, is certainly an adaptation to the living on the ground. Nevertheless, there are alterations of body shape, e.g. Phreatoicidea are laterally thinned and more reminiscent of amphipod species. Parasitic isopods have undergone great changes (e.g. Bopyridae), sometimes even lack segmentation, and show as adults only little similarity to isopods at all.

Isopods, belonging to the greater group of malacostracan crustaceans, share the typical subdivision of the thorax in segments with walking legs (peraeopods) and segments with lamellar legs for swimming and breathing (pleopods). In Isopoda there are seven pairs of peraeopods and five pairs of pleopods. The name Isopoda is a reference to the high similarity of the peraeopods, meaning “similar” (iso) and “foot” (pod) (Schotte *et al.*, 2009).

1.6 Phylogeny of Isopoda

Apart from serious efforts of phylogeneticists, the Isopoda persist to be a group with uncertain phylogenetic relationships (Wilson, 2009). A general phylogeny seems to have settled (Brusca and Wilson, 1991; Wägele, 1989), but some issues still remain ambiguous, due to high degree of diversification and the large number of isopod species. This applies above all to derived groups of isopods, like those which were

formerly comprised as “Flabellifera” (Brandt and Poore, 2003), but also to the earliest divergences within the Isopoda. The specialized fresh water-inhabiting Phreatoicidea are placed as the sister group to all other isopods, followed by the Asellota as the next taxon in recent studies. However, there are first indications of Phreatoicidea and Asellota being sister groups (Wilson, 1999). Most recent molecular 18S data show no resolution for the major isopod divergences or places the Phreatoicidea on a higher level in the isopod tree (Wilson, 2009). Other relationships of major isopod taxa have been raised, too, e.g. the Asellota being the earliest branching taxon within isopods prior to the Oniscidea (Schmalfuss, 1989), or a sister group relationship of Oniscidea and Valvifera (Tabacaru and Danielopol, 1999).

The placement of the Isopoda within the Peracarida is also uncertain. Nine major subclades are currently recognized for the various subgroups of Peracarida (Martin and Davis, 2001), but their relationships remain subject to continuing controversial discussions, as well as the membership of different species to these groups. A brief overview, summarizing eight alternative hypotheses of peracarid phylogeny from the last decades, can be looked up in Spears et al. (2005). The greatly varying position of the Isopoda in these peracarid trees presented by different authors is an obvious problem. Some authors suggest a basal branching of the Isopoda, whereas others favor a later origin. Hence, different putative sister groups are proposed, most notably Amphipoda and Tanaidacea. 18S and combined 18S/morphological analyses indicate, however, that the closest relative to isopods might also be Cumacea (Wilson, 2009). Spears et al. (2005) emphasize that a lot of the persisting difficulties in determining the peracarid phylogeny can be ascribed to the very limited list of synapomorphic peracarid characters. Some of the features do not even occur in all peracarid suborders, and are partially found in non-peracarid crustaceans as well.

1.7 Thematic complexes

The following chapters are intended for an in-depth analysis of additional mitogenomic characters, also referred to as characters of 'genome morphology', of isopods and related peracarid species. For this purpose a number of thematic complexes and questions will be addressed within this thesis:

(1) What kinds of characters of 'genome morphology' actually exist in the examined mt genomes? Are these characters valuable for phylogenetic inference? Do they occur frequently enough to establish a phylogeny apart from sequence data? (Chapters 2 to 6)

(2) Are there any rearrangement hotspots in the mt genome? Or is an equal distribution of translocated genes over the entire genome observed? Areas of increased gene order variability may be of less phylogenetic value, as they might be subject to homoplasious change. (Chapters 2 to 4)

(3) What is the scenario of 'genome morphology' changes within Isopoda? How can the observed succession of character changes be most parsimoniously explained? A broad comparison of species should not only allow tracing back character evolution, but should also enable inferences on the underlying mechanisms. What is the putative gene order in the isopod ancestor?
(Chapter 4)

(4) How reliable are characters of 'genome morphology'? Is there any conflict in the data, e.g. are there specific examples of homoplasious events? (Chapter 2 and 4)

(5) What is the basal split in isopod phylogeny? Phreatoicidea, Asellota or a common clade of both taxa are worth being considered to take the most basal position relative to the other isopods. (Chapters 3 and 4)

(6) What is the sister group of the Isopoda? (Chapters 5 and 6)

2. The complete mitochondrial genome of the common sea slater, *Ligia oceanica* (Crustacea, Isopoda) bears a novel gene order and unusual control region features

2.1 Abstract

Background: Sequence data and other characters from mitochondrial genomes (gene translocations, secondary structure of RNA molecules) are useful in phylogenetic studies among metazoan animals from population to phylum level. Moreover, the comparison of complete mitochondrial sequences gives valuable information about the evolution of small genomes, e.g. about different mechanisms of gene translocation, gene duplication and gene loss, or concerning nucleotide frequency biases.

The Peracarida (gammarids, isopods, etc.) comprise about 21,000 species of crustaceans, living in many environments from deep sea floor to arid terrestrial habitats. *Ligia oceanica* is a terrestrial isopod living at rocky seashores of the european North Sea and Atlantic coastlines.

Results: The study reveals the first complete mitochondrial DNA sequence from a peracarid crustacean. The mitochondrial genome of *Ligia oceanica* is a circular double-stranded DNA molecule, with a size of 15,289 bp. It shows several changes in mitochondrial gene order compared to other crustacean species. An overview about mitochondrial gene order of all crustacean taxa yet sequenced is also presented. The largest non-coding part (the putative mitochondrial control region) of the mitochondrial genome of *Ligia oceanica* is unexpectedly not AT-rich compared to the remainder of the genome. It bears two repeat regions (4x 10bp and 3x 64bp), and a GC-rich hairpin-like secondary structure. Some of the transfer RNAs show secondary structures which derive from the usual cloverleaf pattern. While some tRNA genes are putative targets for RNA editing, *trnR* could not be localized at all.

Conclusions: Gene order is not conserved among Peracarida, not even among isopods. The two isopod species *Ligia oceanica* and *Idotea baltica* show a similarly derived gene order, compared to the arthropod ground pattern and to the amphipod *Parhyale*

hawaiiensis, suggesting that most of the translocation events were already present the last common ancestor of these isopods. Beyond that, the positions of three tRNA genes differ in the two isopod species. Strand bias in nucleotide frequency is reversed in both isopod species compared to other Malacostraca. This is probably due to a reversal of the replication origin, which is further supported by the fact that the hairpin structure typically found in the control region shows a reversed orientation in the isopod species, compared to other crustaceans.

2.2 Background

The metazoan mitochondrial genome is a circular double stranded DNA molecule of about 12-20 kb length. Due to the important role of mitochondria for cell metabolism its gene content is highly conserved and typically contains the same 37 genes: 13 protein-coding genes, two ribosomal genes and 22 transfer RNA genes (Wolstenholme, 1992). In addition one A+T-rich non-coding part is present which contains essential regulatory elements for transcription and replication. It is therefore referred to as the mitochondrial control region (Wolstenholme, 1992). The organization of the mtDNA is compact with very little non-coding sequences between genes, even gene overlaps by a few nucleotides are commonly found (especially at the boundaries between *nad4/nad4L* and *atp6/atp8*). As a result the gene order of mitochondrial genomes is relatively stable because rearrangements are likely to disrupt genes. Thus changes in gene order are relatively rare, whereas tRNA genes more frequently change their position than larger protein-coding and rRNA genes (Boore, 1999). Mollusca (Akasaki *et al.*, 2006; Boore *et al.*, 2004; Dreyer and Steiner, 2004; Knudsen *et al.*, 2006), Brachiopoda (Endo *et al.*, 2005; Helfenbein *et al.*, 2001; Noguchi *et al.*, 2000; Stechmann and Schlegel, 1999) and Nematoda (He *et al.*, 2005; Hu *et al.*, 2003; Keddie *et al.*, 1998; Lavrov and Brown, 2001) represent phyla where a lot of rearrangements of mitochondrial genomes were reported, whereas in Chordata only few changes in gene order were found (Boore, 1999). Among arthropods a lot of species have retained the arthropod ground pattern (or a slight modification in Hexapods and Crustacea), while some taxa show frequent genome rearrangements, e.g. Myriapoda (Lavrov *et al.*, 2000; Lavrov *et al.*, 2002;

Negrisolo *et al.*, 2004), Hymenoptera (Dowton *et al.*, 2003; Dowton and Austin, 1999), Acari (Black and Roehrdanz, 1998; Shao *et al.*, 2005), and Araneae (Masta and Boore, 2004; Qiu *et al.*, 2005). Peracarid crustaceans seem to represent another example, as two partially sequenced mitochondrial genomes (Cook *et al.*, 2005; Podsiadlowski and Bartolomaeus, 2006) exhibit strong differences between each other and from the arthropod ground pattern.

Mitochondrial genomes offer a broad range of characters to study phylogenetic relationships of animal taxa. Besides nucleotide and amino acid sequences, tRNA secondary structures (Macey *et al.*, 2000), deviations from the universal genetic code (Castresana *et al.*, 1998; Telford *et al.*, 2000), as well as changes in the mitochondrial gene order (Boore *et al.*, 1998; Boore *et al.*, 1995) are successfully used as characters in phylogenetic inference. Especially the changes in gene order prove as extremely reliable phylogenetic characters because the probability that homoplastic translocations occur in closely related taxa is very low. Dowton et al. (Dowton *et al.*, 2002) calculated a chance of 1/2664 for a single gene translocation event occuring independently in two mitochondrial genomes (starting from the same gene order in both). However, this probability could be underestimated according to yet unidentified constraints on modes of gene rearrangements and should be handled with care.

With about 21,000 known species peracarids comprise approximately one third of all crustacean species so far described. Within Crustacea the isopods form the largest subtaxon (10,000 species). Isopods show an amazing ecological diversity and morphological flexibility. They are common around the globe, their habitats range from deep sea plains over freshwater wells to terrestrial, even arid deserts. Although to date 30 complete mitochondrial genomes from crustaceans are available - thereof 16 from malacostracan species - a complete mitochondrial sequence of a peracarid species is still missing. Recent sequencing efforts with the amphipod *Parhyale hawaiiensis* (Cook *et al.*, 2005) and the isopod *Idotea baltica* (Podsiadlowski and Bartolomaeus, 2006) produced almost complete genome records, lacking only the control region and some of the tRNA genes. Here we present the first complete sequence of a peracarid mitochondrial genome. *Ligia oceanica* (Isopoda: Oniscidea) is a terrestrial species living on rocky seashore habitats. It is found from Norway to Iceland in the north, around the british islands and the north sea coasts south to northern Spain and Portugal. We discuss changes in gene order compared to other crustacean taxa and give an

overview about genome rearrangements in Crustacea. In addition we compare nucleotide composition of isopod mitochondrial genes and tRNA secondary structure, and describe in detail uncommon features of the mitochondrial control region from *Ligia oceanica*.

2.3 Results and Discussion

Genome organization

The complete mitochondrial genome sequence of *L. oceanica* has an overall length of 15,289 bp [GenBank:DQ442914]. Successfully accomplished PCRs have proven a circular organization of the molecule (Figure 2.1). Although this is the general state of metazoan mitochondrial DNA, it is mentioned here, because there is evidence for a linear organization in a related species, *Armadillidum vulgare* (Isopoda: Oniscidea) (Raimond *et al.*, 1999). All 13 protein coding subunits which are usually found in metazoan mitochondrial genomes are present, as well the two rRNA subunits (Table 2.1). In contrast only 21 tRNA genes instead of the typical number of 22 were identified (see below). In addition one major non-coding sequence was detected, which presumably contains the origin of replication and regulatory elements for transcription (mitochondrial control region). There are small gene overlaps at 14 gene borders. The largest has a length of 15 nucleotides (between *nad2* and *trnC*). Some small non-coding sequences exist which occur quite often in arthropod mitochondrial genomes. The largest extends up to 52 nucleotides and is located between *trnT* and *nad5*.

Protein-coding genes

The A+T content of the protein coding genes of the *L. oceanica* mitochondrial genome is with 60.1% (A=28.6; C=16.7%; G=23.2%; T=31.5%) at the lower end observed for malacostracan species. The values range from a 60.0% minimum given by *Cherax destructor* (Miller *et al.*, 2004) to a 69.3% maximum by *Penaeus monodon* (Wilson *et al.*, 2000). Whereas the majority of the 13 protein coding genes show usual start codons

for mtDNA, two genes begin with exceptional codons (Table 2.1). The gene *atp8* probably starts with codon GTG. Although there is an ATA codon nine bp downstream from this start codon, alignments with *atp8* amino acid sequences from other arthropods suggest the presence of more amino acids in the starting region. GTG is probably also in use as start codon in mitochondrial genes from *Idotea baltica* (*nad1* and *cox2*) (Podsiadlowski and Bartolomaeus, 2006). The second gene with an apparently exceptional codon is *cox1*, which starts with ACG. Although this seems to be unusual in metazoan mitochondrial genomes, almost all other malacostracan crustaceans studied so far have this start codon for *cox1*. The only known exception concerns the crab *Portunus trituberculatus* (Yamauchi *et al.*, 2003). Two of the protein coding genes

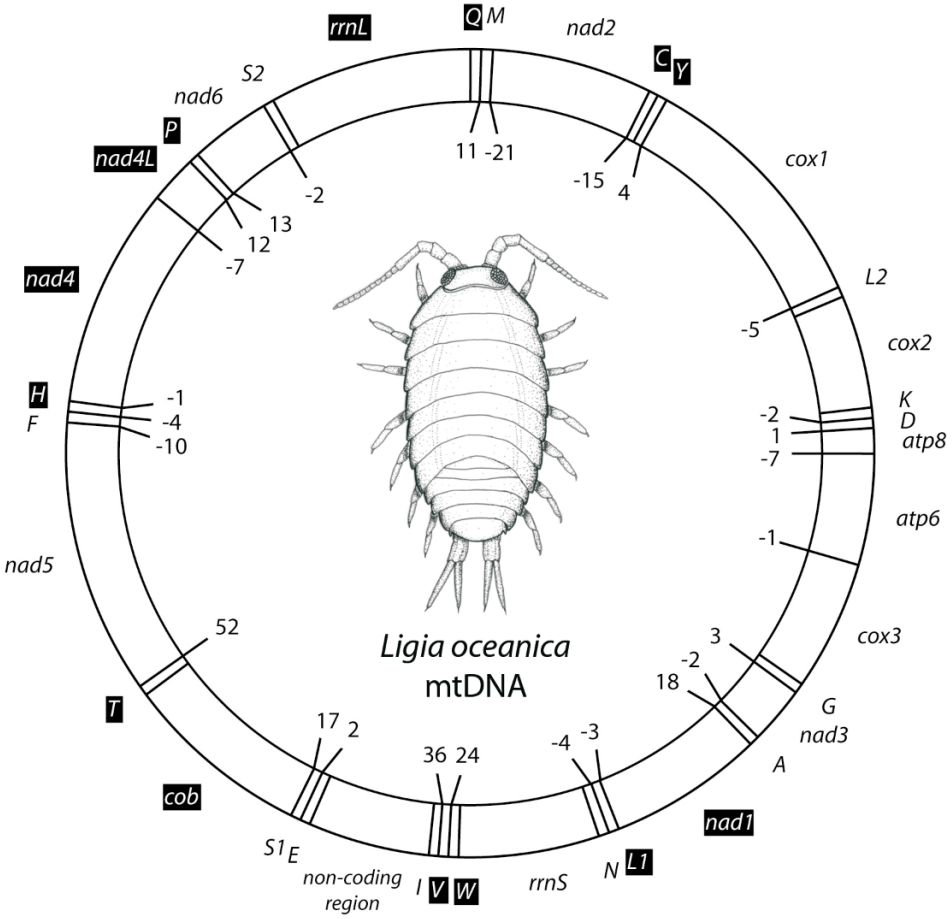


Figure 2.1. Map of the mitochondrial genome of *Ligia oceanica*. Transfer-RNAs are represented by their one-letter amino acid code. Inverted (white on black) legends indicate genes located on (-)strand. Numbers specify the length of non-coding sequences or the extent of gene overlaps (negative values), respectively.

Table 2.1. Gene content of the mitochondrial genome of *Ligia oceanica*.

Gene	Strand	GenBank position no.	Size (nts)	Size (aa)	Start codon	Stop codon	Intergenic nucleotides
trnQ	-	1-63	63				11
trnM	+	75-140	66				-21
nad2	+	120-1142	1023	340	ATG	TAA	-15
trnC	-	1128-1181	54				0
trnY	-	1182-1243	62				4
cox1	+	1248-2786	1539	512	ACG	TAA	-5
trnL2-UUR	+	2782-2843	62				0
cox2	+	2844-3527	684	227	ATA	TAG	0
trnK	+	3528-3590	63				-2
trnD	+	3589-3648	60				1
atp8	+	3650-3805	156	51	GTG	TAA	-7
atp6	+	3799-4470	672	223	ATG	TAA	-1
cox3	+	4470-5256	787	262	ATG	T	0
trnG	+	5257-5317	61				3
nad3	+	5321-5671	351	116	ATT	TAA	-2
trnA	+	5670-5729	60				18
nad1	-	5748-6691	944	314	ATA	TA	-3
trnL1-CUN	-	6689-6750	62				-4
trnN	+	6747-6810	64				*
rrnS	+	6811-7660	850				*
trnW	-	7661-7726	66				24
trnV	-	7751-7806	56				36
trnI	+	7843-7899	57				*
non-coding region		7900-8636	737				*
trnE	+	8637-8696	60				2
trnS1-AGY	+	8699-8760	62				17
cob	-	8778-9911	1134	377	ATA	TAA	0
trnT	-	9912-9971	60				52
nad5	+	10024-11688	1665	554	ATT	TAG	-10
trnF	+	11679-11741	63				-4
trnH	-	11738-11800	63				-1
nad4	-	11800-13134	1335	444	ATG	TAA	-7
nad4L	-	13128-13418	291	96	ATA	TAA	12
trnP	-	13431-13492	62				13
nad6	+	13506-13994	489	162	ATT	TAG	-2
trnS2-UCN	+	13993-14055	63				*
rrnL	-	14056-15289	1234				*

* Gene borders defined by borders to adjacent genes

show truncated stop codons. The gene for *nad1* terminates with TA whereas *cox3* bears a single thymine at its end. This is a well known phenomenon in the mitochondrial genome and is frequently reported for several species. The stop codons are very likely completed by post-transcriptional polyadenylation, so that each transcript finally obtains a functional UAA terminal codon (Ojala *et al.*, 1981).

In most arthropods there is a strand specific bias in nucleotide frequencies (Hassanin *et al.*, 2005; Hassanin, 2006). In detail the (+)strand contains more cytosine and adenine, while the (-)strand consequently is more rich in guanine and thymine. Some taxa show a reversal in that strand bias, among them the isopod *Idotea baltica* (Podsiadlowski and Bartolomaeus, 2006). Strand bias is best reflected in GC skew (Hassanin *et al.*, 2005; Perna and Kocher, 1995) of mitochondrial genes (Table 2.2). In *Ligia oceanica*, as well as in *Idotea baltica*, GC skew is positive in (+)strand encoded genes, while it is negative in (-)strand encoded genes. This is in contrast to most other malacostracan crustaceans and is probably due to an inversion of the mitochondrial control region, or at least the replication origin (Hassanin *et al.*, 2005). Further evidence comes from sequence analysis of the control region (see below).

The effective number of codons (ENC) is a statistic describing how far codon usage in protein-coding genes departs from the equal usage of all synonymous codons (Wright, 1990). Its range lies between 20 (when only one codon is used for each amino acid) and 62 (when all synonymous codons are equally in use). The latter departs from the usual value of 61 for nuclear genes as in invertebrate mitochondrial genomes 62 codons are in use (instead being a stop codon, UGA codes for tryptophane in the invertebrate mitochondrial code). The ENC of all published crustacean mitochondrial genomes was determined for all genes (Figure 2.2), except *nad4L* and *atp8*, because these genes are too short (less than 100 codons) to get proper results. A positive correlation with G+C content in third codon positions was revealed ($r2 = 0.3381$; $p<0.01$). There is no obvious difference seen between malacostracan and other crustaceans. Genes from the two isopod species (*Ligia oceanica* and *Idotea baltica*) are of higher G+C content and therefore show a higher than average number of effective codons. For numbers of effective codons for individual species and genes, as well as GenBank accession numbers see Supplementary file 2.1.

Transfer RNAs

We identified 21 out of normally found 22 transfer-RNA genes in the mitochondrial genome of *Ligia oceanica*. Despite extensive efforts to find secondary structures in non-coding regions the gene *trnR* was not found in the mitochondrial genomes sequence. By

all means tRNA-Arg is essential for maintaining translation of mitochondrial gene products, so it has to be either imported into the mitochondrion, or its gene exists in the mitochondrial genome, but is subject to extensive RNA editing and therefore not identifiable by now.

Transfer-RNA genes are spread over the entire genome and are located on both strands (Figure 2.1, Table 2.1). 14 of them were identified using tRNAscan-SE 1.21 (Lowe and Eddy, 1997). The other seven tRNA genes (*trnD, trnC, trnE*, *trnI*, *trnF*, *trnS1*, *trnW*) were found by eye inspection of otherwise non-coding regions. Some of the putative secondary structures derive from the usual cloverleaf pattern (Figure 2.3): tRNA-Cys and tRNA-Ser(AGY) lack the DHU-arm. The loss of this arm in tRNA-Ser(AGY) was also observed in many other arthropod species, among malacostracan crustaceans *Pseudocarcinas gigas* and *Macrobrachium rosenbergi* (Miller *et al.*, 2005), *Euphausia superba* (Machida *et al.*, 2004b), *Cherax destructor* (Miller *et al.*, 2004), *Penaeus monodon* (Wilson *et al.*, 2000), and *Portunus trituberculatus* (Yamauchi *et al.*, 2003). In contrast to that, the derived structure of tRNA-Cys seems to be unique among malacostracan species studied so far. Transfer-RNA-Val and tRNA-Ile miss the TΨC-arm. Again these features are not seen in other malacostracan crustaceans.

A misplaced adenine was recorded in the anticodon loop of tRNA-Val. Its existence has been proven by repeated sequencing of different PCR-products. To assure the functionality of this gene a correctional RNA editing must be presumed in which a single nucleotide is removed. Similar post-transcriptional events with insertion and deletion of single nucleotides are known from the mitochondrial mRNAs of trypanosomes (Benne, 1994), tRNA editing was demonstrated in the centipede *Lithobius forficatus* (Lavrov *et al.*, 2000). In addition several mismatches are found in tRNA stems, most of them in the acceptor stem (Figure 2.3: tRNA-Gln, tRNA-Ile, tRNA-Leu1, tRNA-Leu2, tRNA-Pro, tRNA-Val, and in the anticodon stem (Figure 2.3: tRNA-Ala, tRNA-Asp, tRNA-Thr, tRNA-Tyr). Such mismatches were also reported from other animal mitochondrial tRNAs and are probably further subjects to RNA editing (Lavrov *et al.*, 2000; Masta and Boore, 2004; Yokobori and Paabo, 1995).

Table 2.2. Nucleotide frequencies, AT content, AT- and GC-skew for mitochondrial genes of *L. oceanica*. GC-skew from *Idotea baltica* genes for comparison. GC-skews from genes coding on (-)strand are shown in bold numbers.

Gene (strand)	Nucleotide frequency A	C	G	T	%AT	AT skew	GC skew *L.oceanica*	GC skew *I.baltica*
atp6 (+)	0.244	0.205	0.222	0.329	57.3	-0.148	0.038	0,012
atp8 (+)	0.288	0.157	0.157	0.399	68.6	-0.162	0.000	0,100
cox1 (+)	0.236	0.190	0.224	0.350	58.6	-0.195	0.083	0,122
cox2 (+)	0.253	0.183	0.228	0.336	58.9	-0.142	0.111	0,169
cox3 (+)	0.219	0.212	0.234	0.336	55.4	-0.211	0.048	0,029
cob (-)	0.250	0.248	0.153	0.348	59.9	-0.164	-0.235	-0,212
nad1 (-)	0.236	0.205	0.213	0.346	58.3	-0.188	0.020	-0,090
nad2 (+)	0.269	0.149	0.236	0.347	61.6	-0.127	0.226	0,299
nad3 (+)	0.219	0.194	0.268	0.319	53.9	-0.185	0.161	0,320
nad4 (-)	0.273	0.229	0.142	0.357	63.0	-0.134	-0.235	-0,247
nad4L (-)	0.261	0.196	0.155	0.388	65.0	-0.196	-0.118	-0,204
nad5 (+)	0.275	0.127	0.255	0.343	61.7	-0.111	0.335	0,246
nad6 (+)	0.268	0.125	0.233	0.374	64.2	-0.166	0.303	0,240
rrnL (-)	0.345	0.186	0.163	0.306	65.2	0.060	-0.065	-0,103
rrnS (+)	0.324	0.187	0.212	0.278	60.1	0.076	0.062	0,098
CR	0.240	0.231	0.212	0.318	55.8	-0.139	-0.043	n.d.
total (+)	0.292	0.169	0.222	0.317	60.9	-0.041	0.136	n.d.

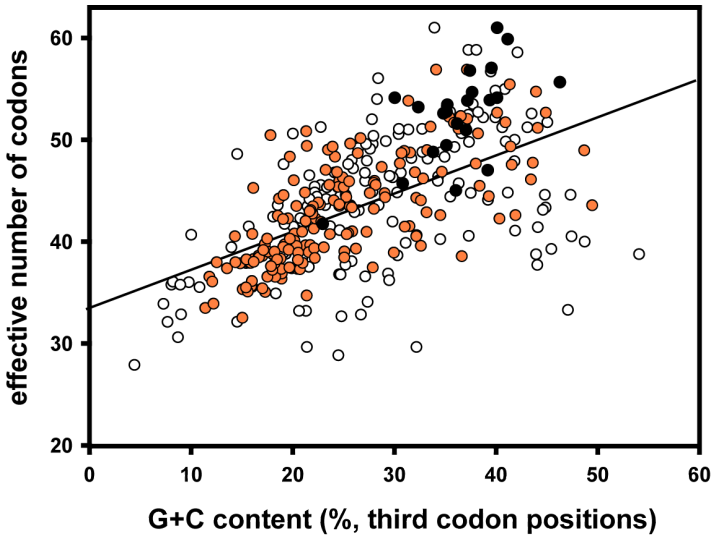


Figure 2.2. Effective number of codons versus G+C content in third codon position in crustacean mitochondrial genes. All species with complete mitochondrial genome entries are included (for a species list, GenBank accession numbers and single values see Supplementary file 2.1). For each species eleven mitochondrial protein-coding genes were evaluated and plotted (all except *nad4L* and *atp8*, which contain less than 100 codons). Black dots: genes from Isopoda; orange dots: genes from all other Malacostraca; white dots: genes from Crustacea excl. Malacostraca. Regression line with $r^2 = 0.3381$; $p<0.01$.

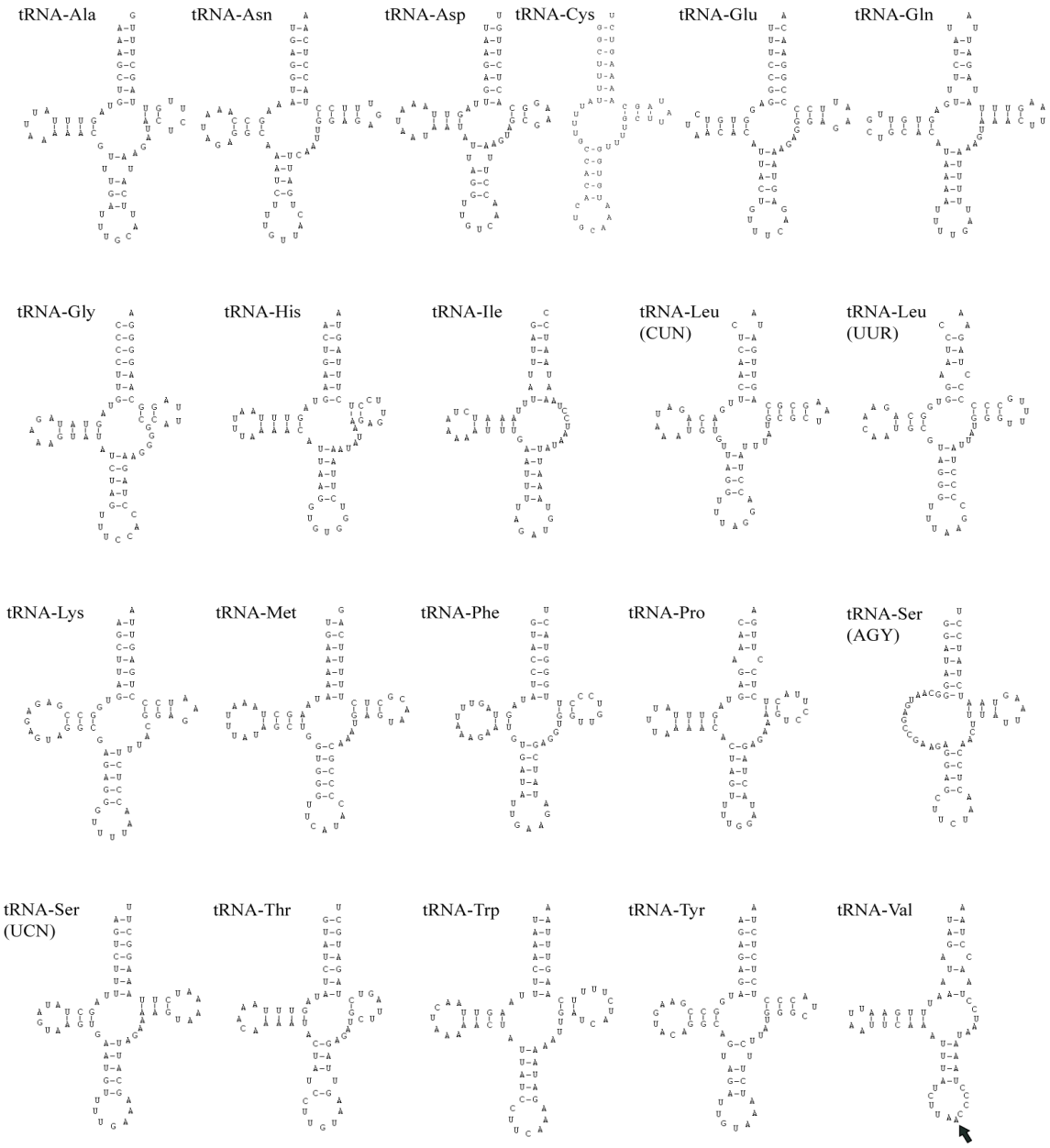


Figure 2.3. Plots of the mitochondrial tRNAs found in *Ligia oceanica*. An additional nucleotide, probably deleted by RNA editing, is found in the anticodon loop of *trnV* (arrow).

Control region and repetitive sequences

There is one major non-coding region of 737 bp length located between *trnI* and *trnE*. It is assumed to be the mitochondrial control region. At its boundary to *trnI* it contains two sections with repetitive sequences (Figure 2.4). The first consists of a series of four completely matching sequences of 10 bp each and extends into *trnI*. The second section is formed by a consecutive triplicate 64 bp segment. No similarities of these sequences

to any other mitochondrial gene could be identified. Only a few other mitochondrial genomes were shown to contain any repeat region: rabbit mitochondrial genomes show repeated 153 bp motifs in their mitochondrial control region, varying in copy number between different individuals or tissues (Casane *et al.*, 1994); the highly aberrant mitochondrial genome of the brachiopod *Lingula anatina* posesses ten different unassigned repeated elements ranging in size between 28 bp and 1092 bp and in copy number between 2 and 11 (Endo *et al.*, 2005). Also some insects show tandem repetitions in mitochondrial DNA (Zhang, X and Hewitt, 1997).

Contrary to expectations the A+T content in the control region (55.8%) is lower than in other parts of the genome (protein coding genes: 60.1%). In contrast most other arthropods have an A+T-rich control region. While the repeat region is A+T-rich (70.3%), a 65 bp region near the 3`-end of the control region has an A+T content of only 14.1% (Figure 2.4). That region is putatively folded into a hairpin-like structure with a stem consisting of 19 paired nucleotides (two mismatches) and a loop consisting of 11 nucleotides (Figure 2.5). This hairpin-like structure highly resembles stem-loop structures known from insect mitochondrial control regions which have stems ranging

Figure 2.4. Sequence of *trnI* and the major non-coding region (control region) of *Ligia oceanica*. The region contains two sections with tandem repeats (4x 10bp, 3x 64bp) and a GC-rich region containing the putative hairpin structure (see also Figure 2.5).

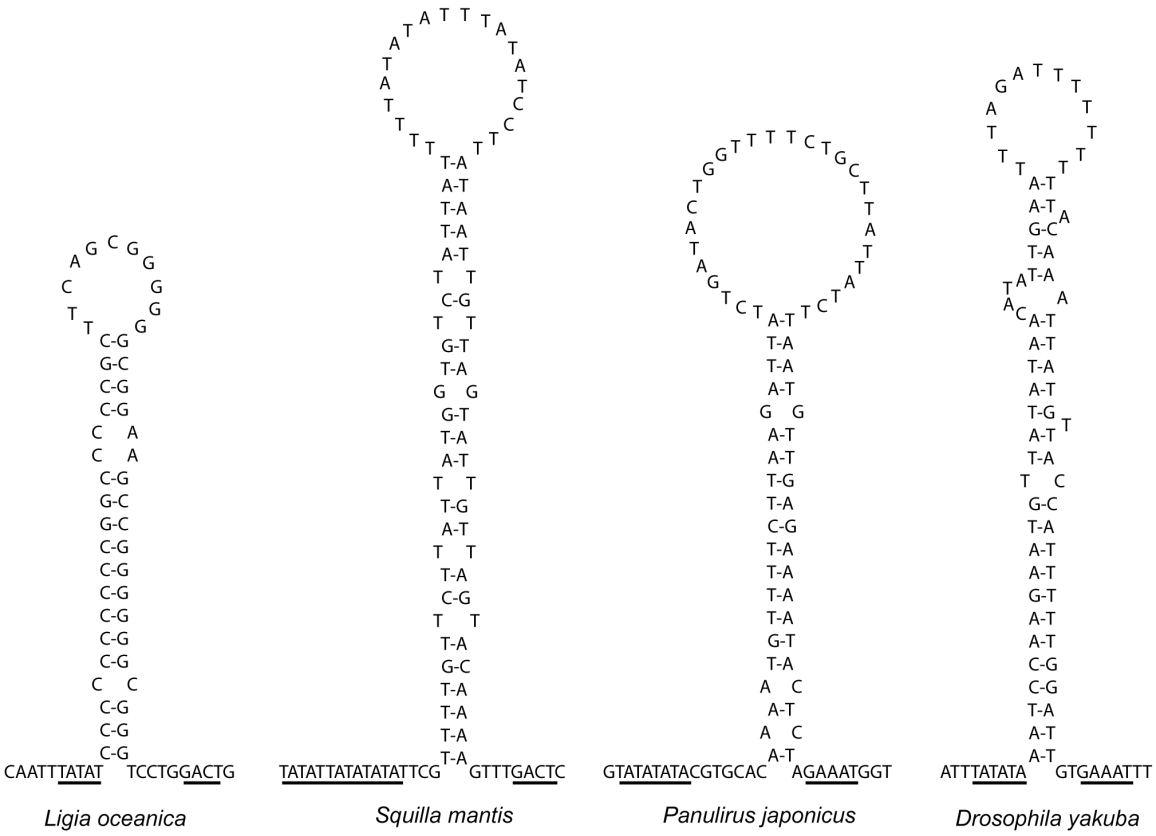


Figure 2.5. Hairpin structures in the mitochondrial control regions of *Ligia oceanica* (Isopoda), *Squilla mantis* (Stomatopoda), *Panulirus japonicus* (Decapoda) and *Drosophila yakuba* (Hexapoda). Conserved motifs in 5´- and 3`-flanking sequences are underlined. *Drosophila yakuba* structure according to (Monforte *et al.*, 1993), the other structures were deduced from GenBank entries.

between 15-30 bp and loops of about 9-15 nucleotides (Zhang *et al.*, 1995). Similar stem-loop structures were found in other crustacean species, like the mantis shrimp *Squilla mantis* and the spiny lobster *Panulirus japonicus* (Fig 5). The flanking sequences around the stem region show conserved motifs: 5`-flanking sequences show a TATA element, while 3`-flanking sequence contains a GACT in *Ligia* and *Squilla*, while the GAAAT motif typical for insects is found in *Panulirus*. It is assumed that these structures are of functional importance in conjunction with the origin of replication (Zhang *et al.*, 1995). In *Ligia* the flanking motifs are found in opposite direction and strand compared to that of *Squilla* and *Panulirus*. This fact gives direct evidence for an inversion of the control region in isopods (in addition to the reversed strand bias of nucleotide frequency mentioned above).

An overview about gene translocations in Crustacea

Crustacean systematics is far from being settled. While Malacostraca seems to be a well defined clade, the interrelationships between crustacean subtaxa is under debate and even monophyly of Crustacea is doubtful, with respect to the position of Hexapoda, which are probably the next relatives to a crustacean subtaxon (Malacostraca, Branchiopoda or Copepoda) (Cook *et al.*, 2005; Mallatt and Giribet, 2006; Nardi *et al.*, 2001; Regier *et al.*, 2005; Wilson *et al.*, 2000). Mitochondrial genome rearrangements may serve as phylogenetic markers which support sistergroup relationships among Crustacea. From 35 species of Crustcea complete or almost complete mitochondrial genome sequences are recorded in GenBank. Gene order is not conserved among these taxa: only 13 species show no changes compared to the pancrustacean ground pattern (Figure 2.6 and 2.8: Pancrustacea ground pattern). Transfer-RNA genes are more often translocated than other genes, probably because of their small size.

Crustacea and Hexapoda (united as Pancrustacea (Boore *et al.*, 1998; Friedrich and Tautz, 1995; Shultz and Regier, 2000) or Tetraconata (Richter, 2002)) share the same ground pattern in mitochondrial gene order (Boore *et al.*, 1998). It differs from the euarthropod ground pattern (Staton *et al.*, 1997) by the position of one tRNA gene: *trnL2* is located between *cox1* and *cox2*, whereas in Chelicerata, Myriapoda and Onychophora *trnL2* is located between *nad1* and *rrnl*, adjacent to *trnL1* (Boore *et al.*, 1998; Boore *et al.*, 1995). Among other data, mitochondrial gene translocations have shown that the enigmatic Remipedia and Pentastomida definitely belong to Pancrustacea, as they show the above mentioned translocation of *trnL2* (Lavrov *et al.*, 2004). Only three crustacean species do not show this character: the cephalocarid *Hutchinsoniella macrantha* (Lavrov *et al.*, 2004), where *trnL2* probably is secondarily translocated to another position (Figure 2.6, No. 7 from *Hutchinsoniella*), and the two copepod species *Tigriopus japonicus* (Machida *et al.*, 2002) and *Lepeophtheirus salmonis* (Tjensvoll *et al.*, 2005), which underwent a complete shuffling of the mitochondrial genome.

Three species (belonging to Cephalocarida, Branchiura and Pentatomida) share a translocation of *trnK* to a position between *trnR* and *trnN* (Lavrov *et al.*, 2004). Among these, the tongue worm *Armillifer armillatus* and the fish louse *Argulus americanus* share one further translocation (*trnQ*), together with mtDNA sequence analysis

supporting a close relationship between Pentastomida and Branchiura (Lavrov *et al.*, 2004; Regier *et al.*, 2005). That was already discussed according to sperm morphology (Storch and Jamieson, 1992; Wingstrand, 1972) and 18S molecular sequence data (Abele *et al.*, 1989). *trnK* is also translocated in all other taxa refered to as members of "Maxillopoda", a systematic unit only weakly based on morphological characters. However, Ostracoda, Copepoda and Cirripedia each show different positions for *trnK* compared to the above mentioned taxa, so there is no good reason to take this as a homology. Because of contrary results from morphological and sequence based analyses (Giribet *et al.*, 2005; Regier *et al.*, 2005) it is also questionable to unite the Cephalocarida with Branchiura and Pentastomida to one clade, solely based on the common translocation of *trnK* (No. 1 for *Hutchinsoniella* / *Argulus* / *Armillifer* in Figure 2.6 and Figure 2.7).

The three species of Cirripedia (Lavrov *et al.*, 2004), [GenBank:NC_006293; GenBank:NC_008974] share several translocations of tRNA genes (*trnA*, *trnE*, *trnP*). Another series of events is difficult to reconstruct: *trnC* and *trnY* are translocated in all three species to a position between *trnS2* and *nad1*, but in different order and on different strands (referred to as No. 6 from Cirripedia in Figure 2.6 and 2.7). In addition one species shows a triplication of *trnC* (No. 7 from *Pollicipes polymerus* in Figure 2.6 and 2.7). Two further differences are reported in *Megabalanus volcano*: an inversion of a block of five genes (No. 7) and the probable translocation of *trnK* and *trnQ* to a position between *trnY* and *trnC* (No. 8 in Figure 2.6 and 2.7). An alternative explanation is that *trnK/trnQ* were primarily translocated to that position seen in *Megabalanus* and secondarily translocated to the position seen in *Pollicipes* and *Tetraclita*. With data from these three species alone, it is not possible to reconstruct a ground pattern of mitochondrial gene order of Cirripedia with respect to the position of *trnK*, *trnQ*, *trnY* and *trnC*.

A lot of further translocation events are recorded only in single species, making them useless in phylogenetic analysis of the actual data set. Large genome rearrangements involving also protein-coding genes are seen in the branchiuran *Argulus americanus* (Lavrov *et al.*, 2004), and in the ostracod *Vargula hilgendorfi* (Ogoh and Ohmiya, 2004), both accompanied by a duplication of the control region. But the highest degree of genome rearrangement was found in the two copepod species (Machida *et al.*, 2002; Tjensvoll *et al.*, 2005), where a complete reshuffling of the mitochondrial genomes has

led to a gene order with almost no similarities between the two species and to other crustaceans. Partial genomes from two other copepod species revealed even more rearrangements (Machida *et al.*, 2004a). In contrast three species of Branchiopoda have retained the pancrustacean ground pattern. The fourth species, *Artemia franciscana* shows two tRNA gene translocations (*trnI, trnQ*) (Valverde *et al.*, 1994).

Among Malacostraca 10 from 19 species have retained the pancrustacean ground pattern: six mantis shrimps (Stomatopoda) and four members of Decapoda (Figure 2.8 and 2.9). Among Decapoda independent translocation events changed gene order in *Pagurus longicarpus* (Hickerson and Cunningham, 2000), *Cherax destructor* (Miller *et al.*, 2004), and Brachyura. The four species of Brachyura share a translocation of *trnH*, the freshwater crabs *Geothelphusa dehaani* (Segawa and Aotsuka, 2005) and *Eriocheir japonica* (Sun *et al.*, 2005) show further translocations. In the Euphausiacea *Euphausia superba* a swap between *trnL1* and *trnL2* seems to have happened, probably preceded by a gene duplication (Machida *et al.*, 2004b).

Besides the two isopod species (*Ligia oceanica*, *Idotea baltica*), only one other peracarid mitochondrial sequence, from the amphipod *Parhyale hawaiiensis*, was published before (Cook *et al.*, 2005). From six gene translocations that must be assumed to get the gene order of *Parhyale hawaiiensis*, none is shared with the mitochondrial genomes of isopods (Figure 2.8 and 2.9). Only *trnI* is translocated in both taxa, but as its new position is different in *Parhyale hawaiiensis* and *Ligia oceanica* (and not known in *Idotea baltica*), there is no reason to presume that a translocation had already happened in their common ancestor. Instead we assume an independent translocation of this gene in isopods and amphipods. This implies, that the ground pattern of gene order in Peracarida must be identical to that of Pancrustacea.

Gene translocations in isopods

A comparison of the complete mitochondrial genome of *L. oceanica* with the ancestral state of pancrustacea (Boore *et al.*, 1998) demonstrates several changes in gene order (Figure 2.8). All in all 11 genes (*cob*, *nad1*, *nad5*, *rrnS*, *trnI*, *trnL1*, *trnF*, *trnS1*, *trnT*, *trnW*, *trnV*) and the control region are found in other relative positions than reported in other malacostracan crustaceans. By reason of parsimony we assume that these

positional changes were due to nine gene translocations (Figure 2.8, 2.9; No. 1-9) and a translocation of the major non-coding region (NCR). The genes *nad5* and *trnF*, as well as *nad1* and *trnL1* retained their adjacent positions, so that they probably were translocated as a block. The other genes were most likely repositioned by single translocation events. Five translocations led also to inversion of genes to the complementary strand: *cob*, *trnT* and *trnW* changed from the (+)strand to the (-)strand, whereas *rrnS* and the block of *nad*5 and *trnF* were inverted from (-)strand to (+)strand.

From the 11 genes being translocated in *Ligia oceanica*, seven are also found in the same new positions in the mitochondrial genome of *Idotea baltica* (Figure 2.8) (Podsiadlowski and Bartolomaeus, 2006). In addition the mitochondrial control region, *trnW*, *trnI* and *trnS1* are not found in their original position in *Idotea baltica*, but will probably be found in the region not sequenced yet - between *rrnS* and *cob*, similar to *Ligia oceanica*. Of all genes translocated in *Ligia oceanica* only *trnV* is found in its original position in *Idotea baltica*. So of nine gene translocation events supposed for *Ligia oceanica*, eight must already have happended in the common ancestor of both species (Figure 2.8, 2.9; No. 1-8). The derived gene order is probably the ground pattern for an isopod subtaxon comprised of Oniscidea and Valvifera (and probably more taxa). Translocation of *trnV* probably happened after the split of the oniscidean and valviferan lineages (Figure 2.9, No. 9). In contrast, translocation of *trnN* probably happened in the lineage leading to *Idotea baltica* (Figure 2.9, No.11). It is located in original position in *Ligia oceanica*, but missing in that position in *Idotea baltica*. We do not know about the fate of *trnR* in *Ligia oceanica*, but in *Idotea baltica* it was also subject to a translocation (Figure 2.9, No. 10).

It is noticeable that in *Ligia oceanica* all translocated genes found their new position in a segment comprising about one third of the complete genome (between *trnA* and *trnH*). This area bears a cluster of tRNA genes in the ancestral gene order of arthropods. It seems to be a "hot spot" of genome rearrangements in arthropods (Black and Roehrdanz, 1998; Dowton *et al.*, 2003).

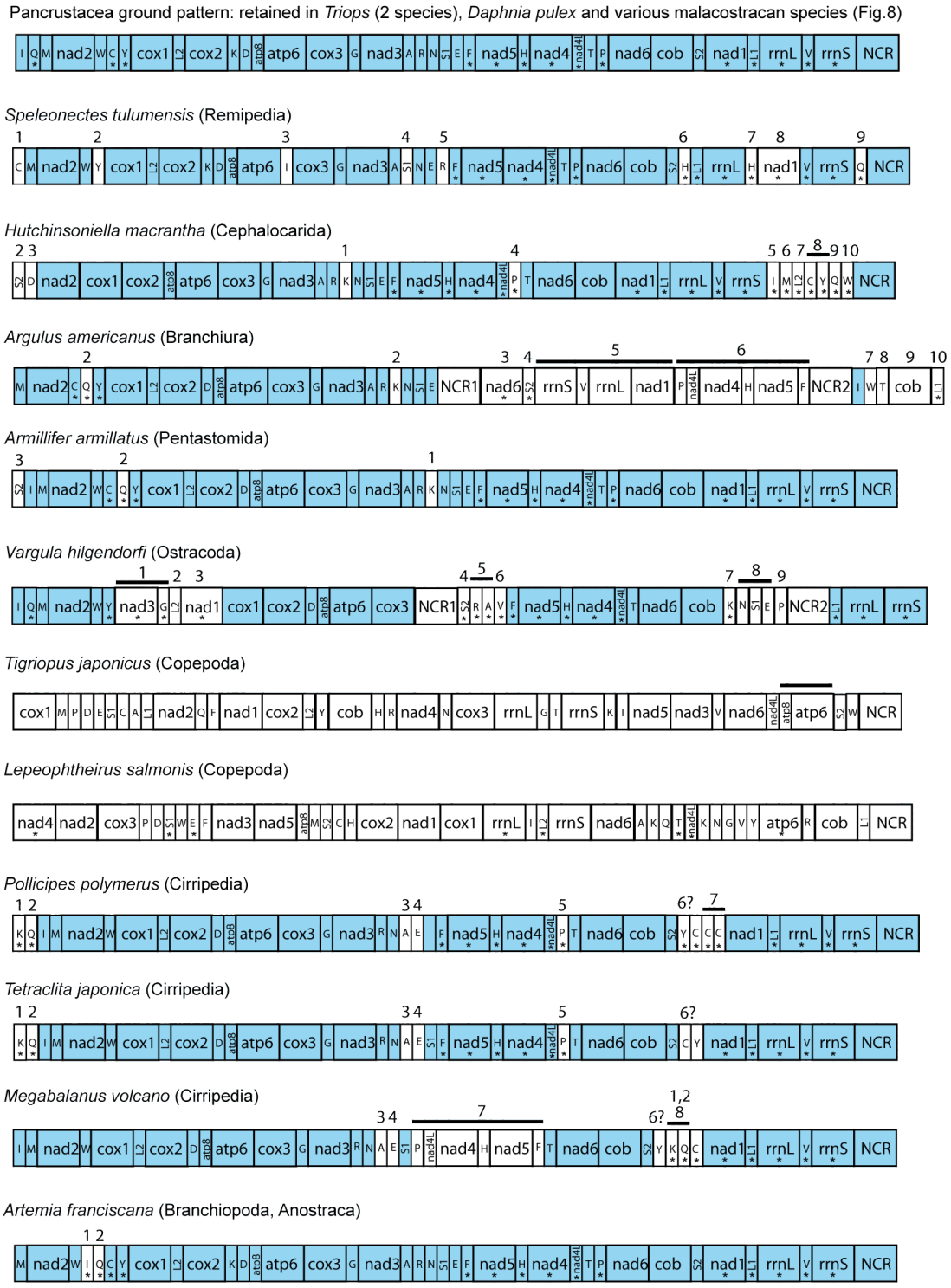


Figure 2.6. Gene order of mitochondrial genomes from Crustacea (excl. Malacostraca). All species with complete mitochondrial genomes are listed. White colours indicate changes compared to the pancrustacean ground pattern. Numbers refer to different gene translocation events (compare Figure 2.7), horizontal lines combine adjacent genes, which were probably subject to a joint translocation. NCR = major non-coding region, the putative mitochondrial control region. Asterisks indicate genes located on (-)-strand. For GenBank accession numbers see Supplementary file 2.1.

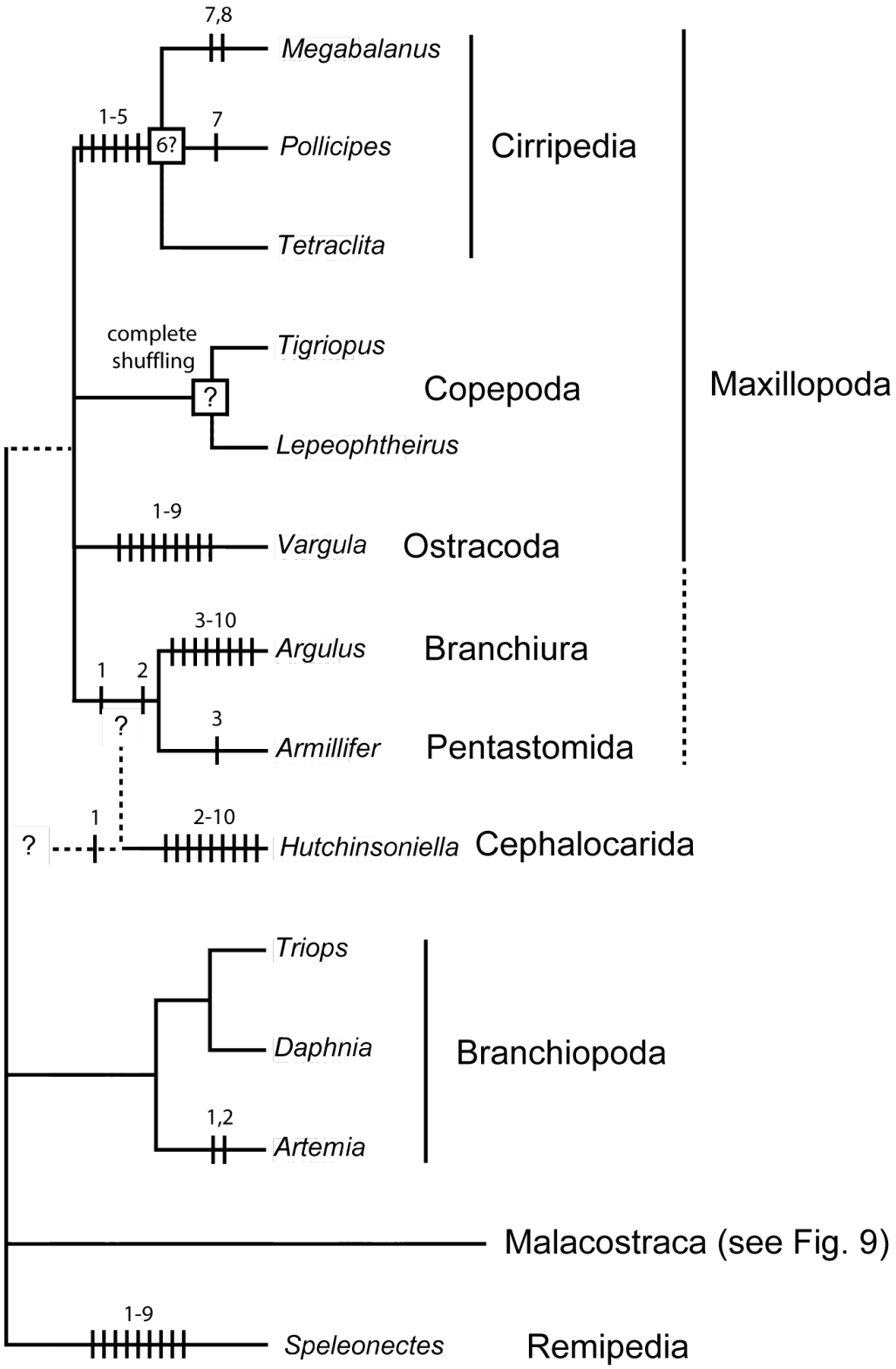


Figure 2.7. Phylogenetic tree of Crustacea. Only those species with GenBank entries for complete mitochondrial genomes are included. Hypotheses of gene translocation events are mapped to the corresponding clades (numbers correspond to those in Figure 2.6). Due to the uncertain homology, translocations of the major non-coding region were not considered.

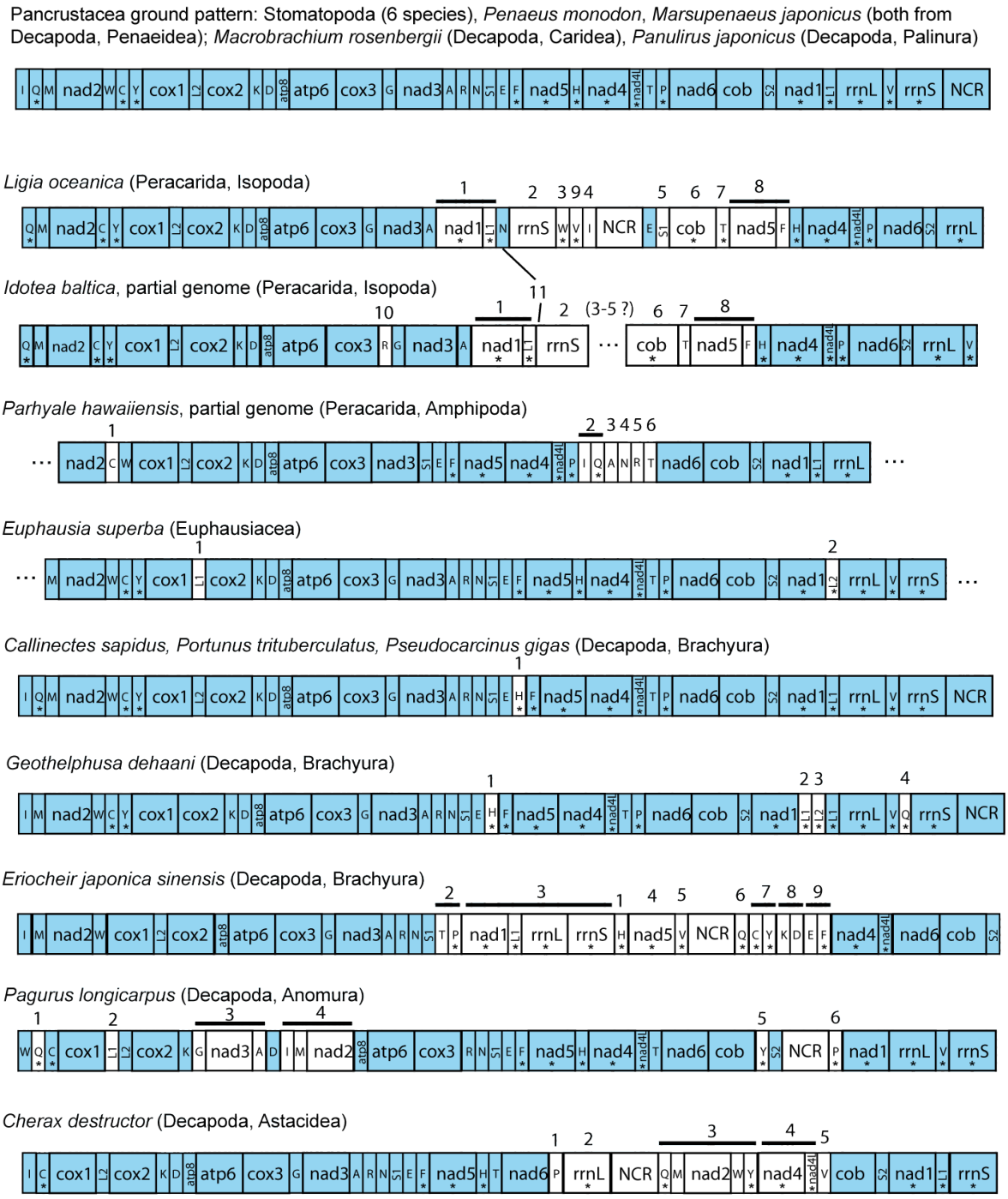


Figure 2.8. Gene order of mitochondrial genomes from malacostracan crustaceans. All species with complete or almost complete mitochondrial genomes are listed. White colours indicate changes compared to the pancrustacean ground pattern. Numbers refer to different gene translocation events (compare Figure 2.9), horizontal lines combine adjacent genes, which were probably subject to a joint translocation. NCR = major non-coding region, the putative mitochondrial control region. Asterisks indicate genes located on (-)-strand. *Idotea baltica* [GenBank:DQ442915]; *Parhyale hawaiiensis* [GenBank:AY639937]; *Euphausia superba* [GenBank: AB084378]; for all other GenBank accession numbers see Supplementary file 2.1.

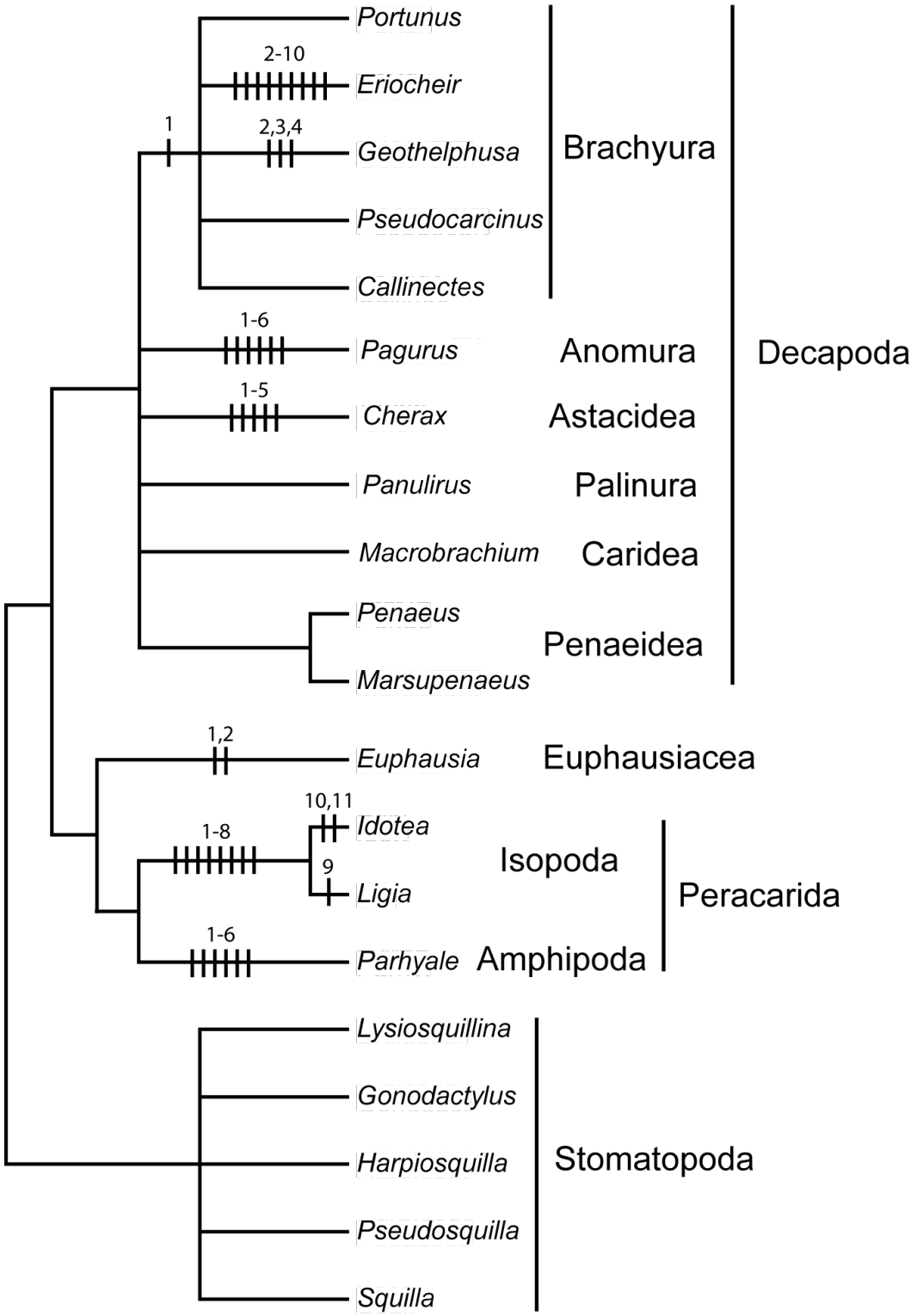


Figure 2.9. Phylogenetic tree of Malacostraca. Only those species with GenBank entries for complete or almost complete mitochondrial genomes are included. Hypotheses of gene translocation events are mapped to the corresponding clades (numbers correspond to those in Figure 2.8). Due to the uncertain homology, translocations of the major non-coding region were not considered. Phylogeny according to (Richter and Scholtz, 2001).

2.4 Conclusions

The first complete mitochondrial genome sequence of a peracarid arthropod, the isopod *Ligia oceanica*, shows the usual compact and circular organization known from other Metazoa. Gene order is not conserved among peracarids and even not among isopods. In *Ligia oceanica* 11 genes plus the control region have changed their relative positions in comparison to the pancrustacean ground pattern, implying to be the result of nine gene translocation events. No gene translocation is shared with the amphipod *Parhyale hawaiiensis*, whereas eight gene translocations were probably already present in the common ancestor of *Ligia oceanica* and another isopod, *Idotea baltica*. Both isopod mitochondrial genomes differ by the position of three tRNA genes (*trnR*, *trnV*, *trnN*) and share an inverted strand bias of nucleotide frequencies compared to other malacostracan crustaceans. Reason for this is probably an inversion of the replication origin. This is confirmed by the fact that the typical hairpin-like secondary structure commonly found in mitochondrial control regions is found in opposite orientation compared to other crustacean species.

A broad survey of mitochondrial gene rearrangements in Crustacea reveals a great variation of gene order. Characters derived from gene order (= gene translocations, inversions or duplications) do not solve overall phylogenetic relationships between major crustacean subtaxa. However, they will probably be helpful in analyses of internal phylogeny of some of these subtaxa (Cirripedia, Brachyura, Peracarida), when more data will be provided.

2.5 Methods

Sample and DNA extraction

A specimen of *Ligia oceanica* originally collected at the coast of the North Sea island Helgoland (Germany) and preserved in 99% ethanol was utilized for the DNA extraction process. 2-3 pleopods were applied to the DNeasy Tissue Kit (Qiagen,

Hilden, Germany) following the manufacturer's protocol to receive the total genomic DNA.

PCR primers

The first partial mitogenomic sequences were obtained by using two insect based primer pairs N4 + 16S2 and CytB + N4(87) (Roehrdanz *et al.*, 2002). Additional intragenetic parts of *cox1*, *cox3*, *nd4*, *nad5*, *rrnL* and *rrnS* were determined by using six crustacean primer pairs (Podsiadlowski and Bartolomaeus, 2005; Podsiadlowski and Bartolomaeus, 2006). The remaining gaps were closed by making use of specific primer pairs designed according to the sequences of the aforementioned genes (for primer sequences see Supplementary file 2.2). Larger PCR products were sequenced by primer walking strategy. To abbreviate this longsome process two longer PCR products were sequenced with the primer pairs S1-S19 from a set of primers which was successfully applied to decapod crustaceans (Yamauchi *et al.*, 2004). All primers were purchased from Metabion (München, Germany).

PCR and purification of PCR products

The PCRs were performed with an Eppendorf Mastercycler or Eppendorf Mastercycler Gradient. The cycling was set up with an initial denaturation step at 94°C for 2 minutes, followed by 40 cycles comprising denaturation at 94°C for 30 seconds, annealing at 45-52°C (primer specific) for 1 minute and elongation at 72° for 2-5 minutes depending on the expected length of the PCR product. The process was completed with a final elongation at 68°C for 2 minutes. The reaction volume amounted 50µl containing 1µl dNTP mix (Eppendorf), 0.25µl HotMaster*Taq* DNA polymerase (5U/µl; Eppendorf), 5µl HotMaster*Taq* buffer (Eppendorf), 1µl primer mix (10µM each), 1µl DNA template and 41.75µl sterilized distilled water (Eppendorf). The PCR products were separated with a 1% TBE agarose gel, stained with ethidium bromide and inspected subsequently under UV transillumination.

For the purification of the PCR Products the QIAquick PCR Purification Kit (Qiagen) as well as the Blue Matrix DNA Purification Kit (Eurx) were used. Abiding to the

manufacturers protocols both kits produced equivalent yields. All PCR products were stored at -20°C until sequencing was performed.

Cloning and transformation

In one case a single PCR fragment, due to low sequence signal quality, ranging from *rrnS* to *cob*, which contained the major non-coding region, was sequenced after cloning with the pGEM-T Easy Vector System (Promega). We followed the manufacturers protocol with the exception that half volumes (5µl) were used for the cloning reaction. For the transformation *Escherichia coli* XL-gold (Stratagene) were used. Colonies that contained recombinant plasmids were identified with selection plates (LB/ampicillin/IPTG/X-Gal). To verify insertion of the PCR product, few cells were applied to a Colony PCR using the vector primers M13F and M13R. The reaction volume amounted 20µl with aforementioned PCR ingredients added proportionally. The colony-cycling consisted of an initial denaturation step (10 minutes, 95°C) followed by 25 cycles of denaturation (30 seconds, 94°C), annealing (30 seconds, 46°C) and elongation (4 minutes, 68°C). The colony PCR was closed by a final elongation (3 minutes, 68°C). The result was inspected with an agarose gel under UV transillumination. Positive tested colonies were proliferated in a LB/ampicillin Medium. Subsequently the plasmids were extracted with the Quantum Prep Kit (Bio Rad) and finally stored at -20°C.

Sequencing and sequence analysis

Cycle sequencing reactions were performed with the CEQ DTCS Quick Start Kit (Beckman Coulter) following the manufacturers protocols. The same primers and thermocyclers were used as in PCRs. The temperature profile included 30 cycles comprising denaturation at 94°C for 20 seconds, annealing at 45-52°C (primer specific) for 20 seconds and elongation at 60°C for 4 minutes. Plasmids were preheated additionally before the sequencing reaction (96°C for 1 minute). The separation was executed by a CEQ 8000 cappillary sequencer (Beckman Coulter) and analyzed with the appendant CEQ software (software version: 5.0.360, instrument version: 6.0.2).

Gene annotation and sequence analysis

The alignment of the fragments to complete the whole mitochondrial DNA sequence was done in BioEdit 7.0.5.2 (Hall, 1999). Each partial sequence was ascertained twice at least to prevent sequencing faults. Ambiguous base pairs were validated manually referring to chromatograms. Gene identification was determined by BLAST search on GenBank databases (2006) and by comparison to the mitochondrial Genome of *Drosophila yakuba* (NC001322). Boundaries of the protein coding genes were determined with a multiple alignment of other crustacean amino acid sequences. It was assumed that they were specified by the first start and stop codons in frame. Transfer RNA genes were determined with tRNAscan-SE 1.21 (Lowe and Eddy, 1997) or by eye inspection for anti-codon sequences and secondary structures in regions between identified genes. Hairpin structures in non-coding regions were also identified by eye inspection. The control region and RNA genes were assumed to extend to adjacent genes, due to the lack of resources for a better determination of their boundaries. Nucleotide frequencies of protein coding and RNA genes were calculated with the DAMBE software package (Xia and Xie, 2001), effective number of codons was determined according to (Wright, 1990) with INCA 1.20 (Supek and Vlahovicek, 2004). The complete genome sequence is submitted to NCBI GenBank [GenBank:DQ442914].

2.6 Abbreviations

A, adenine; *atp6* and *8*, genes encoding ATPase subunit 6 and 8; bp, base pairs; *cox1-3*, genes encoding cytochrome oxidase subnunits I-III; *cob*, gene encoding cytochrome b; C, cytosine; ENC, effective number of codons; G, guanine; mtDNA, mitochondrial DNA; *nad1-6* and *nad4L*, genes encoding NADH dehydrogenase subunits 1-6 and 4L; nt, nucleotide(s); NCR, non-coding region; PCR, polymerase chain reaction; rRNA, ribosomal RNA; *rrnl*, large rRNA subunit (gene); *rrnS*, small rRNA subunit (gene); T, thymine; tRNA, transfer RNA; *trnX* (where X is replaced by one letter amino acid code), tRNA gene.

2.7 Authors` contributions

F. Kilpert did the majority of the laboratory work and the primary sequence analysis; L. Podsiadlowski was the initiator and supervisor of this work. Final analyses of data, discussion of results and drawing of the manuscript was done by both authors in equal shares.

2.8 Additional material

Supplementary file 2.1
Effective numbers of codons used in mitochondrial protein-coding genes of various crustacean taxa. These numbers are the data base for figure 2. [http://www.biomedcentral.com/content/supplementary/1471-2164-7-241-S1.pdf]

Supplementary file 2.2
PCR primers used to amplify mitochondrial gene fragments from *Ligia oceanica*. [http://www.biomedcentral.com/content/supplementary/1471-2164-7-241-S2.pdf]

The files can also be found in the appendix of this thesis (chapter 11).

2.9 Acknowledgements

The authors thank Prof. Thomas Bartolomaeus for his kind support during our work in his group. We also thank Christine Bergmann (FU Berlin), Dipl.biol. Achim Meyer (Universität Mainz), and Dr. Christoph Bleidorn (Universität Potsdam) for technical support during this study, and Prof. Heinz-Dieter Franke (Biologische Anstalt Helgoland) for providing us with isopods. We also thank two anonymous reviewers for valuable suggestions to improve the manuscript. This study was in part supported by grants from the German science foundation (DFG): Ba 1520/10-1 (to LP) and Po 765/2-1 (to LP).

3. The Australian fresh water isopod (Phreatoicidea: Isopoda) allows insights into the early mitogenomic evolution of isopods

3.1 Abstract

The complete mitochondrial (mt) genome sequence of the Australian fresh water isopod *Eophreatoicus* sp. 14 has been determined. The new species is a member of the taxon Phreatoicidea, a clade of particular interest, as it is often regarded as the sister group to all other Isopoda. Although the overall genome organisation of *Eophreatoicus* sp. 14 conforms to the typical state of Metazoa—it is a circular ring of DNA hosting the usual 37 genes and one major non-coding region—it bears a number of derived characters that fall within the scope of "genome morphology". Earlier studies have indicated that the isopod mitochondrial gene order is not as conserved as that of other crustaceans. Indeed, the mt genome of *Eophreatoicus* sp. 14 shows an inversion of seven genes (including *cox1*), which is as far as we know unique. Even more interesting is the derived arrangement of *nad1*, *trnL(CUN)*, *rrnS*, control region, *cob*, *trnT*, *nad5* and *trnF* that is shared by nearly all available isopod mt genomes. A striking feature is the close proximity of the rearranged genes to the mt control region. Inferable gene translocation events are, however, more suitable to trace the evolution of mt genomes. Genes like *nad1*/*trnL(CUN)* and *nad5/trnF*, which retained their adjacent position after being rearranged, were most likely translocated together. A very good example for the need to understand the mechanisms of translocations is the remolding of *trnL(UUR)* to *trnL(CUN)*. Both tRNA genes are adjacent and have a high sequence similarity, probably the result of a gene duplication and subsequent anticodon mutation. Modified secondary structures were found in three tRNAs of *Eophreatoicus* sp. 14, which are all characterized by the loss of the DHU-arm. This is common to crustaceans for tRNA Serine(AGY), while the arm-loss in tRNA Cysteine within Malacostraca is only shared by other isopods. Modification of the third tRNA, Isoleucine, is not known from any other related species. Nucleotide frequencies of genes have been found to be indirectly correlated to the orientation of the mitochondrial replication process. In *Eophreatoicus*

sp. 14 and in other Isopoda the associated nucleotide bias is inversed to the state of other Malacostraca. This is a strong indication for an inversion of the control region that most likely evolved in the isopod ancestor.

3.2 Introduction

Mitochondria are regarded as relicts of prokaryotic endosymbionts that were incorporated into the early eukaryotic cell. Endosymbiont and host have developed since then a remarkably relationship, so close that it also blurred the autonomy of their initially distinct genomes. As a result most of the mitochondrial proteins are now encoded in the nuclear genome and must be imported into the mitochondria for operation. In bilaterian animals only a few genes are retained in the mitochondrial genome (Boore 1999; Wolstenholme 1992). They comprise parts of the protein-synthesis machinery (2 rRNAs, 22 tRNAs), and subunits of the respiratory chain enzyme complexes located in the inner membrane of mitochondria (13 genes).

Maternal inheritance, clear orthology of its genes and the absence of introns made the mitochondrial genome a favourite choice for a lot of applications in molecular ecology and phylogenetics. In addition, it was often assumed that only stabilizing selection acts on mitochondrial genes, so all sequence variation is neutral to selection. Despite several flaws and exceptions to these assumptions, mitochondrial genomes are still a powerful source of sequence information utilized in a wide range of phylogenetic studies on population, species and higher taxonomic levels (Rubinoff and Holland 2005). Besides nucleotide or amino acid sequences, other genomic characters were used for phylogenetic inference, too. Most prominently are examples utilizing gene order data: e.g. Crustacea and Hexapoda both show a derived condition in the position of *trnL(UUR)*, which is a good apomorphy for combining these two taxa to Pancrustacea, excluding myriapods and chelicerates (Boore et al 1998). Other good examples are the sister group relationship of Pentastomida and Branchiura, which is also supported by tRNA translocations (Lavrov et al 2004), or the highly similar gene order of Myzostomida and Annelida (Bleidorn et al 2007). In a broad comparison of mitochondrial gene order among Crustacea we identified some taxa with higher

variability of gene order, e.g. Copepoda, Cirripedia, Brachyura, and Isopoda (Kilpert and Podsiadlowski 2006).

Isopoda are a highly diverse and species-rich group of crustaceans, living in marine habitats from deep sea to the coast, as well as freshwater and terrestrial habitats. In addition, ecto- and endoparasitic species occur (Brusca and Wilson 1991; Wägele 1989). A broad variety of physiological studies were conducted with isopod examples, e.g. adaptations in conjunction with the water-land transition (Wright and Ting 2006).

The number of isopod mitochondrial genomes being available at the moment is relatively small: Only one complete mitochondrial genome is published, that of *Ligia oceanica* (suborder Oniscidea) (Kilpert and Podsiadlowski 2006), and another two almost complete sequences are published of *Idotea balthica* (suborder Valvifera) (Podsiadlowski and Bartolomaeus 2006), and *Armadillidium vulgare* (suborder Oniscidea) (Marcade et al 2007). Here we provide the mitochondrial genome of *Eophreatoicus* sp. 14, the first complete mt genome of a member of the isopod suborder Phreatoicidea and the second complete sequence record for Isopoda. Gene order varies among the published isopod sequences. Although isopod phylogeny is far from being well resolved (Wilson 2009), Phreatoicidea are often regarded as the sister group to all other Isopoda (Brusca and Wilson 1991; Wägele 1989). Therefore, a comparison of mt genome data of a phreatoicidean species to other isopod species allows important insights into the ground pattern and mt genome evolution in Isopoda. Phreatoicidea are freshwater inhabitants with a disjunct southern continent distribution.

3.3 Materials and Methods

Sample and DNA Extraction

Specimens of *Eophreatoicus* sp. 14 were collected (December 2004) in a fresh water lake near Cannon Hill/Hawk Dreaming Stockyard Creek, Northern Territory, Australia (Wilson et al 2009). They were preserved in 99% Ethanol until DNA extraction (May 2006). Total genomic DNA was isolated from embryos, which a single brooding female carried in her ventral brooding pouch (marsupium). Due to the small size of the

embryos three of them were used for an extraction with the DNeasy Tissue Kit (Qiagen, Germany). All in all four extractions were needed to amplify the entire mt genome in overlapping PCR fragments.

PCR Primers

A brief summary of the procedure used to obtain the mt genome of *Eophreatoicus* sp. 14 is given in Additional File 3.1. It shows all PCR products needed for a complete coverage together with the primer combinations used for amplification. A complete list of primers, including the ones applied for primer walking, is provided in Additional File 3.2. The primers were designed with spacious overlappings (usually >100 bp) in mind, which facilitated the creation of the contig from obtained DNA fragments at a later time. All primers were purchased from Metabion (Germany).

First parts of the mitochondrial genome were amplified by intragenic crustacean primer pairs (Crust-Xf/r) (Podsiadlowski and Bartolomaeus 2005), which were useful in previous studies of isopod mt genomes (Kilpert and Podsiadlowski 2006; Podsiadlowski and Bartolomaeus 2006). Two additional primers (Isop-16sf, Isop-16sr) were designed from an alignment of unpublished isopod sequences. With results becoming available from sequencing, further primers were designed on the basis of obtained *Eophreatoicus* sp. 14 sequences and were tested in various combinations. Only four primer combinations yielded PCR products, so together with initial fragments the complete mt genome was covered by eight PCR products, with overlaps between neighbouring fragments of at least 100 bp. Primer walking strategy was used to sequence large PCR products (>1 kb). In a few cases it was possible to reduce the overall time for sequencing larger PCR products by using primers of a primer set that was originally designed for decapod crustaceans for secondary PCR or direct sequencing (Yamauchi et al 2004).

DNA Amplification and Purification of PCR products

Fragmnents of the mitochondrial genome were amplified by standard PCRs that were subsequently sequenced. Eppendorf Mastercycler and Eppendorf Mastercycler Gradient

machines (Eppendorf, Hamburg, Germany) were used for PCR. Short range and medium range PCRs up to 3 kb were performed in 50 µl reaction volumes consisting of 1 µl dNTP mix (10 mM each, Eppendorf), 0.25 µl Taq DNA polymerase (5 U/µl; 5Prime, Hamburg, Germany), 5 µl 10x Taq buffer advanced, 1 µl primer mix (10 µM each; Metabion, München, Germany), 41.75 µl sterilized distilled water (Bio Mol grade, 5Prime) and 1 µl DNA template. The thermal cycling protocol reads as follows: Initial denaturation (94°C, 2 minutes) followed by 40 cycles of denaturation (94°C, 30 seconds), annealing (45-55°C, 45 seconds) and extension (68°C, 2-5 minutes). The cycling was closed by a final extension step (70°C, 4 minutes).

All other PCRs that were expected to yield large fragments (> 3 kb)—like the one defined by Isop-16sr/Eo-7298—were amplified using a long range PCR kit (TaKaRa LA Taq; Takara Bio Inc., distributed by MoBiTec, Göttingen, Germany). We used half volumes, so that a single reaction volume of 25µl consisted of 2.5µl 10x Takara LA PCR buffer, 4µl Takara dNTP mixture (2.5mM each), 0.25µl Takara LA Taq polymerase (5 units/µl), 16.75µl sterile distilled water (Bio Mol grade, 5Prime), 0.5µl primer mixture (10 µM each; Metabion) and 1µl DNA template. The corresponding thermal cycler protocol started with an initial denaturation step (94°C, 1 minute) followed by 30 cycles of denaturation (98°C, 15 seconds), annealing (58°C, 45 seconds) and extension (68°C, 12 minutes), and finally ended in another extension (72°C, 10 minutes).

The PCR products were examined by agarose gel electrophoresis (1% agarose in TBE, stained with ethidium bromide) under UV transillumination. PCR clean-up was done by silica-membrane spin columns (NucleoSpin Extract II, Macherey-Nagel). PCR products were subsequently sequenced or stored at -20°C.

Sequencing and sequence analysis

For the most part of the sequencing we used a CEQ 8000 capillary sequencer (Beckman Coulter, Krefeld, Germany), provided together with a CEQ software package (instrument version: 6.0.2, software version: 5.0.360) for operation and analysis. Sequencing reaction was performed according to the manufacturer's instructions (CEQ DTCS Quick Start Kit; Beckman Coulter). The profile for the thermal cyclers was set

up with 30 cycles, each comprising three steps: denaturation (94°, 20 seconds), annealing (45-55°C, 20 seconds) and extension (60°C, 4 minutes). The reactions were cleaned up for sequencing by ethanol precipitation (three washing steps with subsequent vacuum drying). A part of the mt geome spanning from *rrnL* to *nad2* was processed by a commercial sequencing service (AGOWA GmbH, Berlin, Germany) using a 3730xl DNA Analyzer (ABI). Every part of the mt genome is covered at least twice to allow the detection of sequencing errors. Ambiguous parts were carefully inspected referring to the chromatograms and repeatedly sequenced if necessary.

Gene annotation and sequence analysis

Overlapping fragments obtained from sequencing were aligned and annotated in BioEdit 7.0.9.0 (Hall 1999). The complete annotated mt genome sequence of *Eophreatoicus* sp. 14 was submitted to NCBI GenBank (FJ790313). Transfer RNAs were automatically identified by two computer programs, tRNAscan-SE 1.21 (Lowe and Eddy 1997) and ARWEN 1.2.3.c (Laslett and Canback 2008), and manually reviewed by visual inspection of the conforming genomic regions. A reliable identification of tRNA genes is generally not trivial. The most common approach currently used is to search for base pairings that conform to a typical tRNA cloverleaf structure. Both programs show good detection rates on these typical tRNAs, but perform different on modified tRNAs. As far as we can say from experience, some tRNAs are identified by only one of the two programs (see Additional File 3.3). Very few mismatches can break the close limit of a used software algorithms and prevent the proper detection of a tRNA. A combination of two advanced programs ensures therefore the confidence of having exhaustively searched the mt genome for tRNAs. In addition, remaining non-coding parts of the mt genome were checked manually for further tRNAs.

ARWEN performed better on tRNAs that lack an arm of the secondary structure (tRNA-Cys, tRNA-Ile). tRNAscan-SE, on the other hand, offers a variety of modifiable settings, like the option for the ‘Nematode Mito’ model, which alternatively extends the detection threshold. This may decrease the quality of the proposed secondary structures, and should therefore be regarded rather an indication for a tRNA that requires manual

inspection. In the present case we manually improved the pairings of the secondary structure of tRNA-Phe.

Protein-coding genes and RNA subunit genes were identified by BLAST search of GenBank databases and by comparison to the mitochondrial genome of *Drosophila yakuba* (NC001322). Borders of protein-coding genes were specified by aligning them to multiple amino acid sequences of other crustaceans. The protein-coding genes are assumed to be located within the first start and stop codon in frame. A localization of that kind is not possible for RNA genes and the major non-coding region. Therefore they are considered to be bounded by neighbouring genes. Nucleotide frequencies of the protein coding genes, rRNA genes as well as the total genome were determined with DAMBE 4.2.13 (Xia and Xie 2001).

3.4 Results and discussion

Genome organisation

The complete mt genome of *Eophreatoicus* sp. 14 (FJ790313) has a size of 14,994 bp (Figure 3.1, Table 3.1). It is most likely organized in a single circular molecule of DNA. This topology is the result of an alignment of sequences from eight overlapping PCR products (Additional File 3.2). The sequence annotation revealed the standard set of 37 mt genes (13 protein-coding genes, 22 tRNA genes, 2 rRNA genes typical for animal mt genomes (Wolstenholme 1992). Therefore, not a single gene is missing on the ring. This is different to the only other complete isopod mt genome of *Ligia oceanica* (Kilpert and Podsiadlowski 2006), where the *trnR* gene could not be identified. Indications that the mt genome may be splitted into smaller mini-circles (Shao et al 2009) or that it may be organized in linear fragments or dimers like in *Armadillidium vulgare* (Raimond et al 1999) have not been found. The mt genome of *Ligia oceanica* has been demonstrated to be a circular molecule, too. Thus, the non-circular mt genome of *Armadillidium vulgare* (Marcade et al 2007) is assumed to have evolved within an isopod subtaxon containing *Armadillidium*, but not *Ligia* and *Eophreatoicus*.

Like most other bilaterian mt genomes, the *Eophreatoicus* mt genome bears one major non-coding region, referred to as the mitochondrial control region. The remaining genome is very economically organized: Only few unassigned nucleotides (up to 8 bp) exist between individual genes (Table 3.1). The longest non-coding regions (31 and 48 bp) that have been found are too short to carry additional genes.

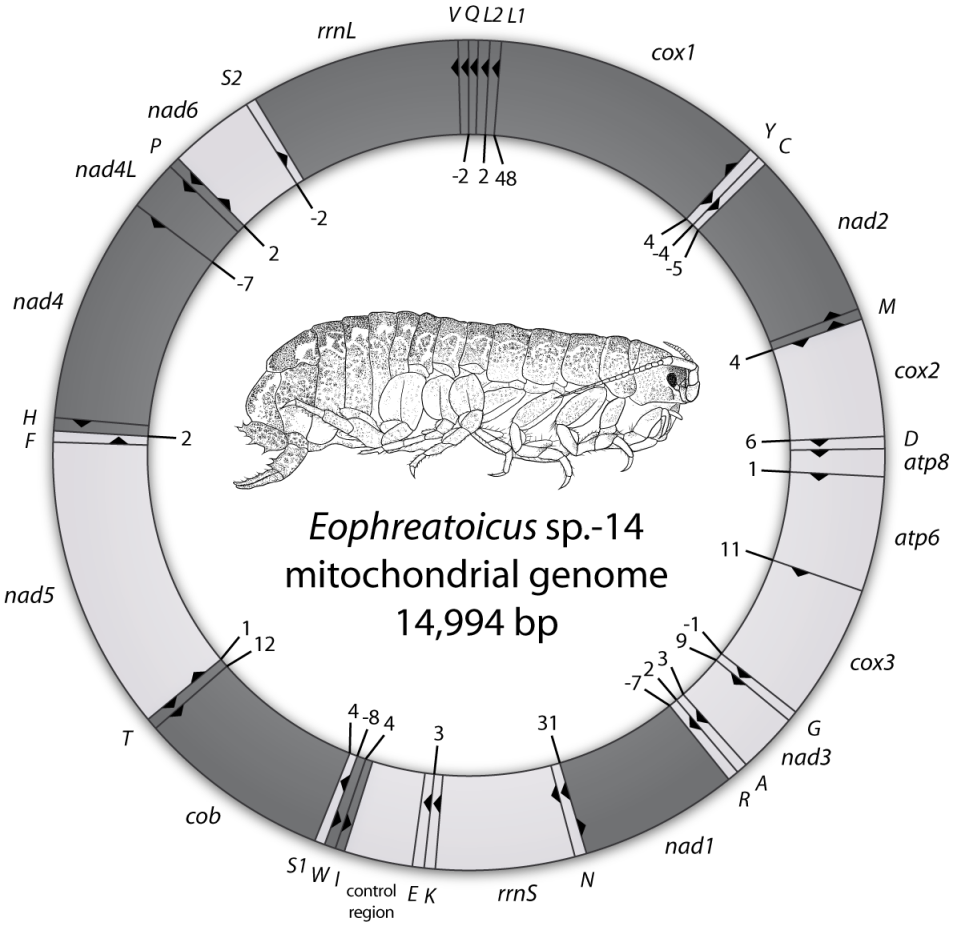


Figure 3.1. Map of the mitochondrial genome of *Eophreatoicus* sp. 14. Transfer-RNAs are represented by their one-letter amino acid code. Numbers specify the length of non-coding sequences or the extent of gene overlaps (negative values), respectively. Arrows pointing clockwise indicate (+)strand genes; counter-clockwise arrows indicate (-)strand genes. The latter are shaded dark additionally.

Table 3.1. Gene content of the complete mitochondrial genome of Eophreatoicus sp. 14. The overall length of the whole, circular DNA molecule is 14,994 bp.

Gene	Strand	GenBank position no.	Size (nts)	Size (aa)	Start codon	Stop codon	Intergenic nucleotides
trnQ	-	1-62	62				0
trnL(UUR)	-	63-123	61				2
trnL(CUN)	-	126-190	65				48
cox1	-	239-1775	1537	512	ACG	T	4
trnY	+	1780-1841	62				-4
trnC	+	1838-1896	59				-5
nad2	-	1892-2890	999	333	ATA	TAG	0
trnM	-	2891-2953	63				4
cox2	+	2958-3644	687	228	ATA	TAA	6
trnD	+	3651-3713	63				0
atp8	+	3714-3875	162	53	TTG	TAA	1
atp6	+	3877-4545	669	222	ATG	TAA	11
cox3	+	4557-5342	786	261	ATT	TAA	-1
trnG	+	5342-5404	63				9
nad3	+	5414-5758	345	114	ATT	TAA	3
trnA	+	5762-5823	62				2
trnR	+	5826-5889	64				-7
nad1	-	5883-6815	933	310	ATA	TAA	31
trnN	+	6847-6909	63				*
rrnS	+	6910-7693	784				*
trnK	+	7694-7757	64				3
trnE	+	7761-7824	64				*
major-ncod		7825-8225	401				*
trnI	-	8226-8279	54				4
trnW	-	8284-8349	66				-8
trnS(AGY)	+	8342-8405	64				4
cob	-	8410-9534	1125	374	ATT	TAA	12
trnT	-	9547-9606	60				1
nad5	+	9608-11323	1716	571	TTG	TAG	0
trnF	+	11324-11384	61				2
trnH	-	11387-11451	65				0
nad4	-	11452-12790	1339	446	ATG	T	-7
nad4L	-	12784-13083	300	99	ATC	TAA	0
trnP	-	13084-13148	65				2
nad6	+	13151-13648	498	165	TTG	TAG	-2
trnS(UCN)	+	13647-13709	63				*
rrnL	-	13710-14933	1224				*
trnV	-	14934-2	63				-2

* Gene borders determined by borders of adjacent genes

Transfer RNAs

All 22 tRNA genes that are usually present in mt genomes of Bilateria were identified. They are distributed throughout the genome and can be found on both strands (Table 3.1). The putative secondary structures of all identified tRNAs are plotted in Figure 3.2.

The great majority of tRNAs (19 of 22) features a common t-shaped or clover leaf secondary structure.

In tRNA-Cys and tRNA-Ile this pattern is modified: The DHU-arm is totally absent. Loss of a tRNA arm has been reported from many malacostracan crustaceans and was also found in isopods. In *Ligia oceanica,* the only other isopod with published tRNA secondary structures, the DHU-arm is similarly missing in tRNA-Cys (Kilpert and Podsiadlowski 2006). This feature might be a putative autapomorphy of the Isopoda, since it is not known from other malacostracan crustaceans. An arm of tRNA-Ile is lacking in *Ligia oceanica* as well, but there it is the TΨC-arm. Apparently, tRNA-Ile is less conserved within isopod species than other tRNAs.

Transfer RNA-Ser(AGY) was recognised by software (ARWEN) to have a DHU-arm that shows only one base pair in the stem structure. It is questionable if it is sufficient to form an effective stem at all, as DHU-stem is usually formed by at least three parings. It seems more probable that tRNA-Ser(AGY) from *Eophreatoicus* sp. 14 lacks the DHU-arm as reported from *Ligia oceanica* and many other malacostracan crustaceans, e.g. *Pseudocarcinus gigas* (Miller et al 2005), *Euphausia superba* (Machida et al 2004), *Cherax destructor* (Miller et al 2004), *Penaeus monodon* (Wilson et al 2000), and *Portunus trituberculatus* (Yamauchi et al 2003).

Protein-coding genes

Nucleotide frequencies of all mitochondrial genes of *Eophreatoicus* sp. 14 are listed in Table 3.2. We determined an overall AT content of 69.6% for the mitochondrial genome of *Eophreatoicus* sp. 14. Up to now, the highest value reported from a malacostracan crustacean was 69.3%, found in *Penaeus monodon* (Wilson et al 2000). The lowest AT content of 60.0% is known from *Cherax destructor* (Miller et al 2004). The high AT content of *Eophreatoicus* sp. 14 contrasts to the relatively low AT content the other complete isopod mt genome from *Ligia oceanica* (60.1%).

A bias of nucleotide frequencies is generally found in the mt genome of arthropods (Hassanin 2006): On the (+)strand cytosine is more frequent than guanine (negative GC skew), whereas the (-)strand contains more guanine than cytosine (positive GC skew).

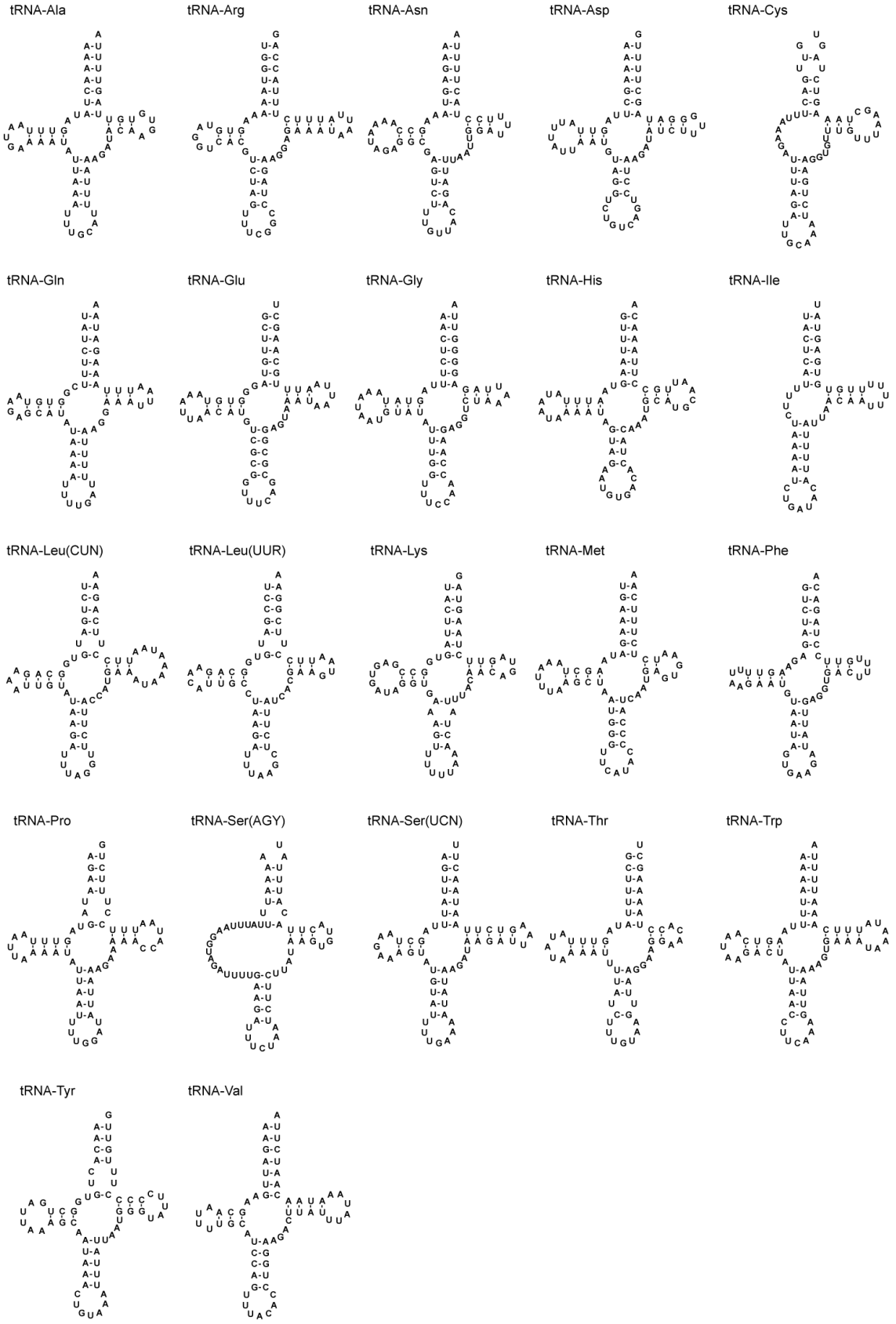


Figure 3.2. Transfer RNA secondary structure plots of *Eophreatoicus* sp. 14. Most tRNAs feature a standard clover-leaf structure. Exceptions: DHU-arm is non-existent in tRNA-Cys, tRNA-Ile, tRNA-Ser(AGY).

Table 3.2. Nucleotide frequencies, AT content, AT- and GC-skew for mitochondrial genes of *Eophreatoicus* sp. 14. Sequences form genes located on (-)strand were inverted prior to calculation. GC-skews from genes coding on (-)strand are shown in bold numbers. GC-skews from *Ligia oceanica* and *Idotea balthica* listed for comparison.

Gene (strand)	Nucleotide frequencies				% AT	AT skew	GC skew	GC skew	GC skew
	A	C	G	T				*L. oceanica*	*I. baltica*
atp6 (+)	0.262	0.109	0.194	0.435	69.7	-0.249	0.281	0.038	0.012
atp8 (+)	0.284	0.099	0.161	0.457	74.1	-0.233	0.238	0.000	0.100
cob (-)	0.335	0.213	0.115	0.337	67.2	-0.003	**-0.301**	**-0.235**	**-0.212**
*cox1 (-)**	0.323	0.187	0.156	0.333	65.7	-0.015	**-0.091**	0.083	0.122
cox2 (+)	0.272	0.130	0.188	0.411	68.3	-0.203	0.184	0.111	0.169
cox3 (+)	0.238	0.116	0.207	0.439	67.7	-0.297	0.283	0.048	0.029
nad1 (-)	0.337	0.182	0.128	0.354	69.0	-0.025	**-0.177**	**0.020**	**-0.090**
*nad2 (-)**	0.357	0.159	0.111	0.372	73.0	-0.021	**-0.178**	0.226	0.299
nad3 (+)	0.232	0.081	0.209	0.478	71.0	-0.347	0.440	0.161	0.320
nad4 (-)	0.354	0.189	0.114	0.343	69.7	0.016	**-0.246**	**-0.235**	**-0.247**
nad4L (-)	0.357	0.177	0.117	0.350	70.7	0.009	**-0.204**	-0.118	-0.204
nad5 (+)	0.254	0.093	0.210	0.444	69.8	-0.272	0.387	0.335	0.246
nad6 (+)	0.235	0.058	0.231	0.476	71.1	-0.339	0.597	0.303	0.240
rrnL(-)	0.427	0.147	0.121	0.305	73.2	0.167	**-0.098**	**-0.065**	**-0.103**
rrnS(+)	0.329	0.100	0.184	0.388	71.7	-0.082	0.297	0.062	0.098
major non-coding region	0.354	0.107	0.307	0.232	58.6	0.209	0.482	-0.043	-
total mt genome (+)	0.312	0.114	0.190	0.384	69.6	-0.104	0.251	0.136	-

* *cox1* and *nad2* are on opposite strand in *Eophreatoicus* compared to other isopods

The skew is probably caused by the mitochondrial replication process, which brings about different mutation rates for guanine and cytosine nucleotides of each strand, dependent on the direction of replication. Therefore, the GC skew allows inference on the orientation of the origins of replication, which are located in the mitochondrial control region (Hassanin et al 2005).

Eophreatoicus sp. 14 has a positive GC skew for all genes on the (+)strand, and consequently a negative GC skew on the (-)strand—the reversed condition to what has been reported from other malacostracan crustaceans (Hassanin 2006). This indicates a reversal of the origins of replication (Hassanin et al 2005). The other isopods *Ligia oceanica* and *Idotea balthica* show the same condition as *Eophreatoicus*, so the reversal might have happened at the base of Isopoda, or even earlier. Mt genome data is still missing from Cumacea and Tanaidacea, probably the next relatives to Isopoda. Two genes of *Eophreatoicus* sp. 14 (*cox1*, *nad2*) differ in their GC skews from that of the other isopods. This is certainly due to their location on the opposite strand, caused by an additional inversion of the genomic segment containing these genes (discussed below).

In many cases gene boundaries of protein coding genes could be easily identified, as they conform to the ordinary start codons (ATG, ATA, ATT) and stop codons (TAA, TAG) of arthropod mt genomes (Table 3.1). Nevertheless, some genes show variations of these codons: The gene *cox1* begins with ACG. Although uncommon in usual metazoan genomes, it is the standard start codon for *cox1* found in malacostracan crustaceans (Additional File 3.4), first noticed in *Penaeus monodon* (Wilson et al 2000). The stop codon of *cox1*, however, is abbreviated and consists of a single T nucleotide only. The same applies to *nad4*. Incomplete stop codons have been reported for mitochondrial genes previously. It has been suggested that they are completed by post-transcriptional polyadenylation to functional UAA termination codons (Ojala et al 1981).

The genes *atp8*, *nad5* and *nad6* are assumed to start with the codon TTG, whereas *nad5* most likely starts with ATC. The number of exceptional start and stop codons is closely linked to the strategy of identification. Currently there is no other practical way than aligning genes under study to other genes of related species—in this study to multiple species of Malacostraca. Where a common start codon was not available at the expected positions, we chose a chemically similar codon instead. Alternatively, conventional start codons would not only mean to assume a resizing of genes, but also large non-coding gaps and extended gene overlappings. TTG as a start codon was reported also from other invertebrates, although by now not from an arthropod. ATC as a start codon has already been found in the honey bee *Apis mellifera* (Crozier and Crozier 1993).

Gene translocations

The mitochondrial genome of *Eophreatoicus* sp. 14 bears significant changes in gene arrangement compared to the gene order of other crustaceans and the hypothetical pancrustacean ground pattern (Kilpert and Podsiadlowski 2006). We inferred gene translocation events (Figure 3.3a) from differences among isopod mitochondrial gene orders and the pancrustacean ground pattern (Figure 3.3b). In *Eophreatoicus* sp. 14 translocated genes exist in two regions of the mitochondrial genome. The first is located between the conserved genes *trnQ* and *cox2*; it comprises the following genes: *trnL(UUR)*, *trnL(CUN)*, *cox1*, *trnY*, *trnC*, *nad2* and *trnM*. This gene series certainly is the result of an inversion of a single multi-gene fragment, ranging from *trnL(UUR)* to

trnM (Figure 3.3b, No. 11). GC-skews of *cox1* and *nad2* are negative for both protein coding genes, thus reflecting the partial inversion of that mitochondrial region (Table 3.2). A similar arrangement is not known from other Isopoda or other Peracarida. Therefore, we must assume the origin of the new gene order in the lineage leading to *Eophreatoicus* somewhere between the split of Phreatoicidea from the remaining isopods and the origin of *Eophreatoicus* sp-14. The inverted part was further modified by an additional change in gene order. In *Eophreatoicus* sp. 14 the tRNA gene *trnL(CUN)* is now located in a position between *trnL(UUR)* and *cox1* (Figure 3.3b, No. 12). Next to *nad1*, where it was usually located, only a non-coding sequence is

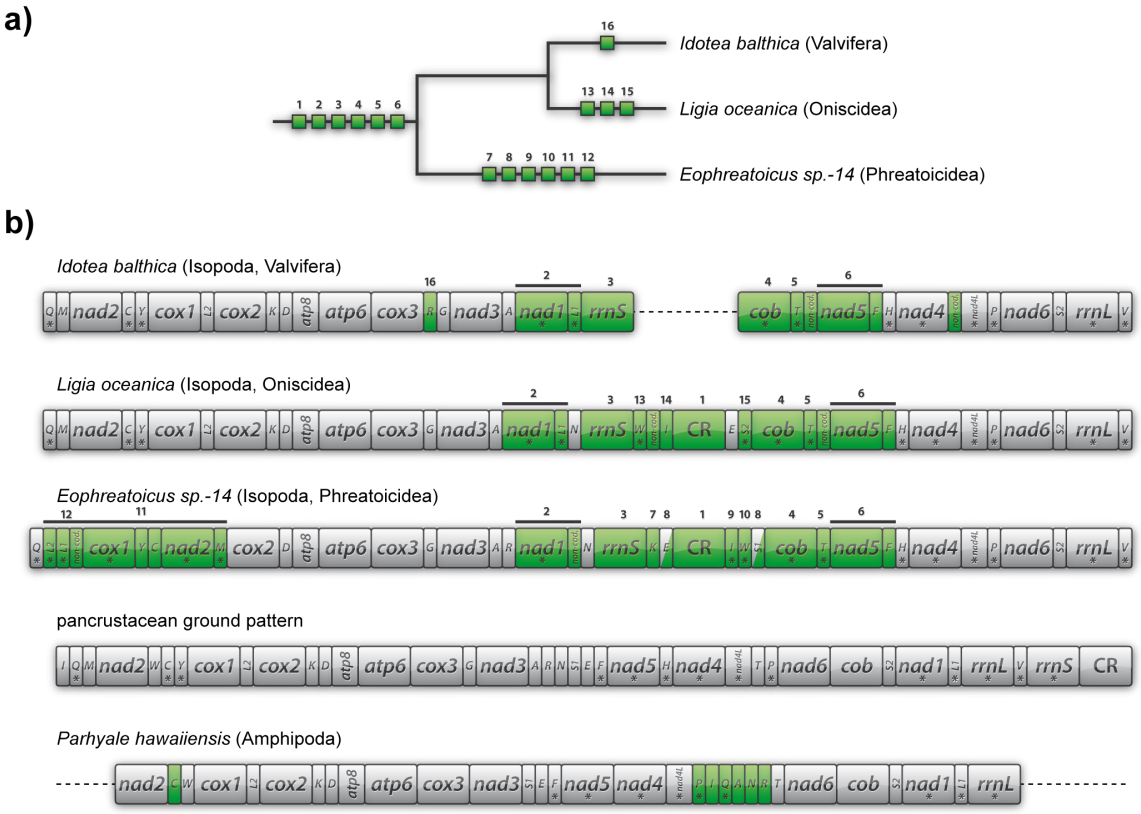


Figure 3.3. Gene arrangement of the mitochondrial genome of *Eophreatoicus* sp. 14 compared to other mt genomes of Peracarida. a) Translocation events for three Isopoda species (*Eophreatoicus* sp. 14, *Ligia oceanica*, *Idotea balthica*) mapped on a simplified isopod tree. Numbers indicate inferred translocation events. b) Comparison of gene arrangements of *Eophreatoicus* sp. 14 to the pancrustacean ground pattern and to the other available Isopoda. Changes to the ground pattern are highlighted in green. Numbers indicate translocation events. Combined translocation of adjacent genes was assumed for event 2, 6, and 11. Only one translocation (8) of the two genes *trnE* and *trnS1* must be assumed for reasons of parsimony. Asterisks mark (-)strand genes.

detectable. On closer inspection the probable mechanism of this translocation event becomes apparent: A comparison of both neighbouring *trnL* genes reveals a very high similarity in gene sequence (77.3%) and as a consequence also in secondary structure (Figure 3.2). It is very likely that, after *trnL(UUR)* was duplicated, one tRNA gene changed its identity to *trnL(CUN*) by random point mutation of the anticodon triplet. After that the original tRNA gene degraded, so that the *trnL(CUN)* gene is now placed next to *trnL(UUR).* All in all, the translocation of *trnL(CUN*) is a perfect example for a identity change of a tRNA gene. Similar remolding events of *trnL(UUR)* genes to *trnL(CUN)* were already observed in decapod crustaceans (Rawlings et al 2003).

The rest of the repositioned genes can be found in a second part of the mitochondrial genome located between *trnR* and *trnH*. This "hot spot" of gene rearrangement is placed around the control region, which was likewise an object of translocation. The altered arrangement reads as follows: *nad1*, NCR, *rrnS*, *trnK*, *trnE*, control region, *trnI*, *trnW*, *trnS(AGY)*, *cob*, *trnT*, *nad5*, and *trnF*. Under the most parsimonious explanation and in comparison with other isopods (*L. oceanica, I. baltica*) we assume a translocation of the control region (Figure 3.3b, No. 1) and nine gene translocations (Figure 3.3b, No. 2-10).

All isopod species studied so far show a similar derived gene order in this region. Especially the arrangement of *nad1*, *trnL(CUN)*, *rrnS*, control region, *cob*, *trnT*, *nad5* and *trnF* seems to be typical for the Isopoda (Figure 3.3, No. 1-6). As the most parsimonious explanation of gene order change in this region, we assume six translocation events. Tandem-duplication/random loss models offer no better way to reach the new gene order, as there is a mixture of inversions and genome shuffling. Some of these genes were likely translocated together. E.g. *nad1* and *trnL(CUN)*, as well as *nad5* and *trnF* have retained same relative position to each other in the derived isopod gene order from the pancrustacean ground-pattern, so only one event is assumed in translocation of these neighbouring genes. In *Eophreatoicus* the relative position of *trnL(CUN)* is further modified by tRNA duplication and identity change, as discussed above. As well *trnK* seems to be translocated in Eophreatoicus independently from other isopods.

It is quite possible that the control region was translocated (or duplicated) first and that the other genes followed afterwards. Beyond doubt, the control region plays an important role in the replication process, bearing the origins of replication and

transcription. But bearing no genes, it is probably more tolerant to insertions than any other part of the mitochondrial genome (as well shown by its highly variable sequence between species). Therefore, the control region or some of its non-functional parts are possibly the preferable destination for a gene translocation without disrupting a functional gene.

At the moment complete mitochondrial sequences from Peracarida apart from isopod species do not exist. Only a few partial genomes have been published, among them the near-complete mitochondrial genome of the amphipod *Parhyale hawaiiensis* (Amphipoda) (Cook et al 2005) (Figure 3.3b). Differences in gene order to the pancrustacean ground pattern are restricted to translocations of seven tRNA genes. None of these events can be considered to have a common origin with any of the discussed rearrangements found in isopods. Other partial mt genomes comprise only few genes and do not share any gene translocations with Isopoda as well.

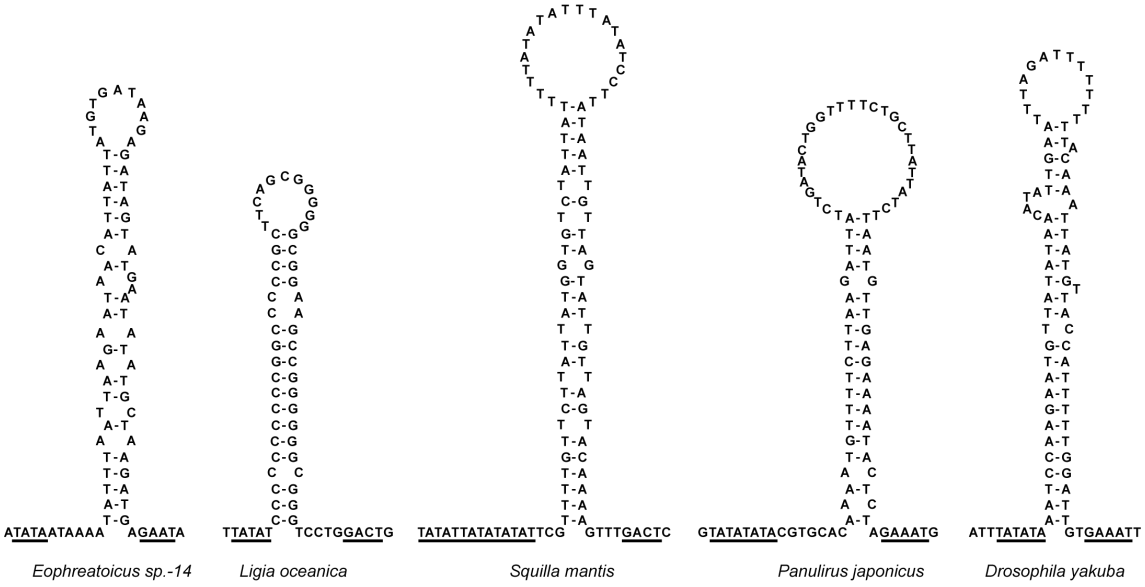


Figure 3.4. Hairpin structure in the mitochondrial control region of *Eophreatoicus* sp. 14, and comparison to hairpin structures of *Ligia oceanica* (Isopoda), *Squilla mantis* (Stomatopoda), *Panulirus japonicus* (Decapoda) and *Drosophila yakuba* (Hexapoda). The depicted sequence corresponds to 8130-8206 bp of the *Eophreatoicus* mt genome. A conserved TATA 5'-motif is flanking the hairpin structure; an elongation by repetitions of this motif is not unusual. On the 3'-side a GAAT motif is present in *Eophreatoicus*, GACT in *Ligia* and *Squilla*, GAAAT in *Panulirus* and *Drosophila*.

Control region

The largest non-coding region (401 bp) is regarded as the mitochondrial control region, probably containing the origins of replication and regulatory elements for transcription. It seems to be relatively short in length, compared to the control region of *L. oceanica*, which has a length of 737 bp. But unlike in *L. oceanica* there are no repetitive sequences (Kilpert and Podsiadlowski 2006) in the control region of *Eophreatoicus* sp. 14. Repetitive mitochondrial sequences are very rarely reported so far, e.g. from insect species (Zhang, X and Hewitt 1997), and show a high variability in length as well as in number of repetitions (Endo et al 2005). Deduced by its repetitive sequences the remaining length of the control region of *L. oceanica* would amount 450 bp, roughly the scale of *Eophreatoicus* sp. 14. It is unlikely that the smaller total size is also accompanied by a loss of function.

In the major non-coding region we determined a hairpin structure comparable to the hairpin of other crustaceans and insects (Figure 3.4). The hairpin is thought to be involved in the replication process (Zhang et al 1995). In both isopod species it is located at a similar position at the end of the control region, suggesting the same orientation in both species, a fact corresponding to the similar nucleotide bias among isopods.

3.5 Conclusions

The Phreatoicidea are often regarded as the sister group to the remaining Isopoda (Brusca and Wilson 1991; Wägele 1989). In this study the mt genome of *Eophreatoicus* sp. 14, a member of Phreatoicidea, shows a unique inversion including multiple genes from (*trnL-UUR* to *trnM*), but also corresponds in large sections to the typical derived condition that is shared by other isopods (*Idotea baltica*, *Ligia oceanica*). The common rearrangements involve multiple translocations of genes and of the mitochondrial control region (*nad1*, *trnL(CUN)*, *rrnS*, control region, *cob*, *trnT*, *nad5* and *trnF*, Figure 3.3b, No. 1-6). This “isopod gene arrangement” might represent the putative ground pattern of Isopoda (Figure 3.3a, No. 1-6). However, there is a chance that the comparison to other mitochondrial genomes of Peracarida, in particular the closely

related Cumacea and Tanaidacea, will reveal some conformity in gene order. These commonly rearranged genes would serve as perfect apomorphic characters to constitute peracarid subtaxa, and possibly could solve the open question of the sister taxon to Isopoda. Recent molecular (Spears et al 2005) and combined studies (Wilson 2009) failed to answer this question without unequivocally.

It is also possible, by comparison to the crustacean ground pattern and to other isopod mitochondrial genomes, to make assumptions on the relative sequence of gene translocations. At least for some genes we can reconstruct the sequence of events (e.g. *trnK*, relocation region from *trnL-UUR* to *trnM*). Furthermore, there is evidence by nucleotide frequencies that the origins of replication and therefore the control region was inverted in comparison with other malacostracans, including Amphipoda.

3.6 Acknowledgements

The authors sincerely thank Dr. George D.F. Wilson (Australian Museum) for providing the specimen of *Eophreatoicus* sp. 14, his work on Phreatoicideans is granted by the Australian Biological Resources Survey Grant, 204-59 (GDFW and CL Humphrey). GDFW also assisted with editing the manuscript. The credit for collecting the specimen goes to A. Cameron, S. Atkins and J. Hanley. We also thank Prof. Dr. Thomas Bartolomaeus for his kind support during all stages of work. This study was supported by a German Science Foundation (DFG) grant to LP (Po 765/2-1,2) and a NaFöG Berlin grant to FK.

3.7Additional material

Additional Files 3.1 to 3.4 are provided with the online version.

Files can also be found in the appendix of this thesis (chapter 11).

4. Rearrangements of the mitochondrial genome of Isopoda and implications on the phylogeny of Peracarida

4.1 Abstract

In this study, we analyse the evolutionary dynamics and phylogenetic implications of gene order rearrangements in five newly sequenced mitochondrial (mt) genomes and four published mt genomes of isopod crustaceans. The sequence coverage is nearly complete, only the control region could not be recovered in some species. Mitochondrial gene order in isopods seems to be more plastic than in other crustacean lineages making all known nine mt gene orders different, especially the asellote *Janira* is characterized by many autapomorphies. The following inferred ancestral isopod mt gene order exists slightly modified in modern isopods: *nad1*, *tnrL1*, *rrnS*, control region, *trnS1*, *cob*, *trnT*, *nad5*, *trnF*.

We consider the inferred gene translocation events leading to gene rearrangements as valuable characters in phylogenetic analyses. In this first study covering major isopod lineages potential apomorphies were identified, e.g. a shared relative position of *trnR* in Valvifera. We also report one of the first findings of homoplasy in mitochondrial gene order, namely a shared relative position of *trnV* in unrelated isopod lineages. In addition to increased taxon sampling secondary structure, modification in tRNAs and GC-skew inversion may be potentially fruitful subjects for future mt genome studies in a phylogenetic context.

4.2 Introduction

The number and identity of the genes encoded in the mitochondrial (mt) genome is highly conserved in bilaterian animals, while the order and orientation in which they appear on the circular doublestrand DNA molecule is not. Historically, our knowledge of evolutionary change of the mt genome has been dominated by analyses of nucleotide sequences of single genes. As more and more complete mt genomes became known, however, it became apparent that in addition to the accumulation of point mutations another kind of mutation is shaping the mt genome: the translocation of one or several entire genes to a new relative position. The molecular mechanisms causing these gene translocations are still only incompletely understood, but the increasing availability of entire mt genomes suggests that (1) translocations do not occur randomly across the mt genome but typically respect borders between genes, thus leaving the genes functional, and (2) translocations of mt genes may not be rare exceptions but rather an important factor in the evolution of the mt genome. Although there is no doubt that mt gene arrangements are happening less frequently than single nucleotide substitutions in the primary sequence of the mt genome, there are increasing reports from different taxa where changes of gene order appear more often than previously assumed. Examples of a higher degree of gene order divergences are found in Mollusca, Brachiopoda and Nematoda, and within Arthropoda from Myriapoda, Hymenoptera, Acari, and Araneae (see references in Kilpert and Podsiadlowski (2006)). Recent studies also reported extensive mt rearrangements from Gastropoda (Grande, Templado, and Zardoya 2008), Neogastropoda (Cunha, Grande, and Zardoya 2009), Bivalvia (Wu et al. 2009), Symphyla (Gai et al. 2008), Hymenoptera (Oliveira et al. 2008), and also from the crustacean suborders Copepoda (Ki, Park, and Lee 2009) and Decapoda (Liu and Cui 2009). From Peracarida modified gene orders are known from Amphipoda (Bauza-Ribot et al. 2009; Cook, Yue, and Akam 2005) and, of course, Isopoda (Kilpert and Podsiadlowski 2006; Kilpert and Podsiadlowski 2009; Marcade et al. 2007; Podsiadlowski and Bartolomaeus 2006).

The frequent translocations of genes within the mt genome can at least in part be attributed to the fact that the mt genome is characterised by a low degree of structural complexity. This reduces the probability of a gene becoming dysfunctional in a new location as it would be the case if the function depended on extensive molecular

interactions with regulatory elements elsewhere in the genome. The regulation of the mt genes is very simple, involving only start and stop codons, compared to the sophisticated regulation and mutual dependencies that are typical for many genes located in the cell nucleus. In most cases the mt genome is a single doublestrand DNA molecule of intermediate size (12-20 kb) with a circular topology. It typically contains 37 genes (22 tRNA genes, 13 protein coding genes, 2 rRNA genes) and one control region (CR). Other important characters of the mtDNA are the strict orthology of genes, maternal inheritance and the absent (or very low) recombination (Gissi, Iannelli, and Pesole 2008). The mt genome possesses no introns and almost no non-coding sequence. In contrast, some genes even overlap by a few nucleotides.

In addition to being subject of investigations themselves, translocation events of mt genes are increasingly recognised as valuable characters in phylogenetic studies of higher taxa. Mutations in the primary nucleotide sequence information have been used for some time already, but the homology of individual characters is difficult to ascertain due to multiple substitutions of single sites over long evolutionary times and the small size of the molecular alphabet (four bases: A, C, G, T). In contrast, the lower rate of translocation events, seems to be less affected by saturation. The higher complexity of gene translocation characters comprising stretches of several hundreds of nucleotides makes the evaluation of character states (relative position of mt genes) less error-prone in comparison to using nucleotide substitutions as phylogenetic markers, due to a reduced importance of multiple substitutions (saturation) and insecurity regarding positional homology (e.g. requiring multi-sequence alignments).

The relatively small number of known mt genomes sets a limit to the gene rearrangements available as potential synapomorphies. An impressive example for the informative value of this kind of characters is given by the tRNA gene *trnL(UUR)*, which was translocated from a position from between *trnL(CUN)* and *nad1* to a derived position between *cox1* and *cox2* in crustaceans and hexapods, thus providing support for the distinctness of the taxon Pancrustacea from myriapods and chelicerates (Boore, Lavrov, and Brown 1998).

Mitochondrial gene rearrangements are the most prominent characters of ‘genome morphology’ in a phylogenetic context, but additional features may be valuable as well: e.g. absence/presence of specific repetitive sequences, reversal of nucleotide biases,

modified RNA secondary structures, deviations from the circular organisation of mt genomes, changes of the mitochondrial genetic code. These characters can be coded in an absence/presence matrix and therefore can be easily analysed in combination with morphological characters.

Many open questions remain about the phylogeny of arthropods, due to convergent evolution of morphological characters and saturation of nucleotide or protein sequences (Dowton, Castro, and Austin 2002). This is particularly problematic for recovering the early evolutionary history of this highly diverse group, which is probably older than 400 million years. Especially for these deep divergences the analysis of higher order mutation events such as gene rearrangements holds a lot of promise, at least for some splits.

Despite recent progress in understanding the overall phylogenetic relationships of Isopoda (Brusca and Wilson 1991; Wägele 1989), many open questions remain in isopod phylogeny (Wilson 2009). Especially the divergence of higher isopods ("Flabellifera sensu lato") is unclear, as well as the earliest isopod divergences. Here, the fossil record and the morphological data tend to favour the Phreatoicidea as the sister group of all other Isopoda, followed by the Asellota. Recently, 18S rRNA analyses failed to support a basal position of Phreatoicidea and split up Asellota in several unrelated suborders (Wilson 2009). A sistergroup relationship of Phreatoicidea and Asellota is also conceivable (Wilson 1999). Because of ambiguities of morphological and 18S data and an inconclusive fossil record in isopod phylogeny, this study evaluates mt genome data as another set of phylogenetic informative characters.

This study compares nine nearly complete isopod mt genomes including five new sequences. The ancestral isopod mt gene order was determined and compared to the inferred pancrustacean ground pattern. We projected the inferred gene rearrangements onto molecular trees estimated from a concatenation of all mt protein-coding and rRNA genes of the isopods in study. In addition to estimating the evolutionary plasticity of the isopod mitochondrial genome with regard to gene translocations, we investigated other complex properties of the isopod mt genome (strand-specific nucleotide bias; start and stop codon usage; inferred secondary structure of transfer RNAs) and inferred their value in a phylogenetic context.

4.3 Material and Methods

General Approach

Total genomic DNA was extracted from preserved specimens (see paragraph: DNA Extraction). Starting with non species-specific primers (see: Primers) initial parts of each mt genome were amplified (see: PCR Amplification) and sequenced (see: DNA Sequencing) by primer walking strategy: new primers were designed based on these sequences, which then served for another cycle of amplification and sequencing. These steps were repeated several times until no new fragments could be obtained. All overlapping fragments from each species were integrated to a single contig sequence and annotated afterwards (see: Data Assembly). Finally, the annotated partial mt genomes were utilised for phylogenetic analyses (see: Phylogenetic Analyses).

DNA Extraction

All specimens that were used for DNA extraction were stored in Ethanol (96 to 99 %). Depending on the size of the animals one or several legs were used for DNA extraction with the DNeasy Tissue Kit (Qiagen, Germany) to isolate the total genomic DNA according to the recommendations of the manufacturer. The examined specimens of *Eurydice pulchra* (Cymothoidea) and *Janira maculosa* (Asellota) were sampled from the rocky coastline of Roscoff (France). *Sphaeroma serratum* (Sphaeromatidae) was collected at the coast of the island Helgoland (Germany) in the North Sea. *Glyptonotus* cf. *antarcticus* (Valvifera) was dredged from the Antarctic shelf. *Asellus aquaticus* (Asellota) was collected near Bremen (Germany), and the terrestrial isopod *Armadillidium vulgare* (Oniscidea) was picked up in the backyard of the FU Berlin zoological institute (Germany).

Primers

With no species-specific primers available first PCR amplifications were attempted with the combinations of the primers N4, 16S2, CytB and N4(87), which already have been proven useful on insects (Roehrdanz, Degrugillier, and Black 2002). In addition,

crustacean-specific primers were used, which amplify intragenic parts of *cox1*, *cox3*, *nad4*, *nad5*, *rrnL*, and *rrnS* (Podsiadlowski and Bartolomaeus 2005). For Glyptonotus we were successful with invertebrate primers (Folmer et al. 1994). All these primers were tried in different combinations, even those which require gene rearrangements. Another set of crustacean specific primers were used for nested PCRs and sequencing with the intention to accelerate the sequencing process of long PCR fragments (Yamauchi et al. 2004). Detailed lists of primes that worked best for PCR and sequencing are provided in the supplementary material (Supplementary files 4.1a: *Eurydice pulchra*, 4.1b: *Sphaeroma serratum*, 4.1c: *Glyptonotus* cf. *antarcticus*, 4.1d: *Armadillidium vulgare*, 4.1e: *Asellus aquaticus*, 4.1f: *Janira maculosa*). Further primers were designed on the basis of obtained sequences and are therefore species specific. All primers were purchased from Metabion (Munich, Germany). The whole process of primer selection and design for PCR and sequencing was exemplarily detailed for another isopod species, *Eophreatoicus* sp. 14 (Kilpert and Podsiadlowski 2009).

PCR Amplification

The PCR amplifications were carried out in 50 µl reaction volumes, comprising 1 µl dNTP mix (10 mM; Eppendorf, Hamburg, Germany), 0.25 µl Taq DNA polymerase (5 U/µl; 5Prime, Hamburg, Germany), 5 µl 10x Taq buffer advanced (5Prime), 1 µl primer mix (10 µM each, Metabion), 1 µl DNA template and 41.7 µl sterile distilled water (Eppendorf). All reactions were executed on thermal cyclers (Eppendorf Mastercycler, Eppendorf Mastercycler Gradient, Peqlab Primus 25 advanced): Initial denaturation (94°C, 2 min), followed by 40 cycles, each comprising denaturation (94°C, 30 sec), annealing (primer specific at 45-55°C, 1 min) and elongation (68°C, 2-5 min), followed by a final elongation step (68°C, 2 min). The quality of PCR products was validated by electrophoresis in 1% TBE ethidium bromide stained agarose gel.

Individual long-range PCRs conditions: A 5181 bp (*cox1* to *rrnS*) fragment of the *Eurydice* mt genome was amplified by the Ep-7504/Ep-12685 primer combination; in *Sphaeroma* the primers Ss-175/Ss-5244 yielded a 5069 bp (*cob* to *rrnL*) fragment; in *Asellus* Aa-7773/Aa-13378 amplified 5605 bp (*cox1* to *rrnS*). These larger PCR products were received by using the more robust TAKARA Taq polymerase (Takara Bio Inc., distributed by MoBiTec, Göttingen, Germany), in total volumes of 25 µl (2.5

µl 10x LA PCR buffer II (Mg^{2+} plus), 4 µl dNTP mix (2.5 mM), 0.5µl primer mix (10µM each), 0.25 µl Takara La *Taq*, 1 µl DNA template, 16.75 µl sterile distilled water (Bio Mol grade, 5Prime)). The related PCR machine profile was set up with an initial denaturation step (94°C, 1 min) followed by 35 cycles, each with a denaturation step (98°C, 10 sec), annealing (60°C, 1 min) and extension (68°C, 10 min). The amplification was completed with a final extension (68°C, 4 min).

PCR products were purified with the NucleoSpin Extract II (Machery-Nagel, Dueren, Germany) kit, as well as the QIAquick PCR Purification Kit (Qiagen, Hilden, Germany) with comparable results, and finally stored at -20°C.

DNA Sequencing and Contig Assembly

The great majority of PCR products were sequenced in our laboratory on a CEQ 8000 capillary sequencer (Beckman Coulter, software version: 5.0.360, instrument version: 6.0.2). The setup of the machine as well as the initial cycle sequencing step were executed according to the manufacturer. The CEQ DCTS Quick Start Kit (Beckman Coulter, Krefeld, Germany) was used to set up a single 10 µl reaction volume; it included 5 µl of purified PCR product, 4 µl DCTS master mix (Beckman Coulter), and 1 µl primer (10 mM). The temperature cycling profile conducted 30 cycles, each comprising denaturation (94° C, 20 sec), annealing (primer specific at 45-55° C, 20 sec), and extension (60° C, 2 min). The reaction cleanup was done by ethanol precipitation (three washing steps with subsequent vacuum drying). Primers and thermocyclers were the same as those used for PCR amplifications. Some sequencing was outsourced to AGOWA (Berlin, Germany), which is using a 3730xl DNA Analyzer (ABI).

Overlapping sequence reads were collapsed to a contig for each species using BioEdit 7.0.9.0 (Hall 1999). If necessary, sequencing reactions were repeated until every part of the sequence was represented by at least two sequences to track down sequencing errors. Protein coding and rRNA genes were identified by GenBank BLAST search and by comparison (bl2seq) to the *Drosophila yakuba* mt genome (NC001322). The latter was also used for the relative location of genes. A multiple alignment of crustacean amino acid sequences served as a reference to specify the start and stop codon of protein

coding genes. Where no canonical start codon could be located, we suggest a nearby codon with similar biochemical characteristics. For ribosomal RNA genes and for the major non-coding region there is currently no practicable way to determine the exact borders. Therefore, they are assumed to extend up to adjacent genes.

In recent years the detection of tRNA genes, which is based on computational predictions of secondary structures, has noticeably improved. We used two software tools, tRNAscan 1.21 (Lowe and Eddy 1997) and ARWEN 1.2.3.c (Laslett and Canback 2008). Using both, a more reliable and reproducible identification is possible. The algorithms of the programs perform quite differently on unusual tRNAs, which deviate from the canonical clover-leaf secondary structure (e.g. lack an arm). These tRNAs are generally the hardest to identify, so that the combination of both programs should increase the detection rate and enable an evaluation of the found tRNA genes. In addition to the automated detection a manual search for tRNAs was performed in remaining non-coding areas of the sequence. The method of tRNA identification is documented for each tRNA in each species (Supplementary files 4.2a-f).We also made a revision of the already published isopod mt genomes using this approach, which led to an improved tRNA annotation (Supplementary file 4.3).

DAMBE 4.2.13 (Xia and Xie 2001) was used to calculate nucleotide frequencies of protein coding and RNA genes. For calculations of tRNA similarity BioEdit was used on aligned DNA sequences (ClustalW). The determination of the AT content was done in BioEdit as well. The AT- and GC skew was calculated according to the formula proposed by Perna and Kocher (1995): AT skew = (A-T)/(A+T); GC skew = (G-C)/(G+C), where A, T, C, G are the four bases.

Phylogenetic Analyses of Mt Gene Sequences

In addition to the nine available isopod species, we chose nine further crustacean sequences as outgroup: *Cherax destructor* (NC_011243), *Euphausia superba* (AB084378), *Hutchinsoniella macracantha* (NC_005937), *Metacrangonyx longipes* (NC_013032), *Parhyale hawaiensis* (AY639937), *Pollicipes polymerus* (NC_005936), *Portunus trituberculatus* (NC_005037), *Squilla mantis* (NC_006081), *Vargula hilgendorfii* (NC_005306). Because computational time requirements of Bayesian and

maximum likelihood analyses depend strongly on the number of taxa, the total number of species was restricted to 18. All analyses were performed with alignments from all 13 protein coding genes. An amino acid alignment was built for each gene including data from all 18 species using Muscle 3.6 (Edgar 2004). They were then concatenated to a single alignment (3873 aa). The Bayesian analysis was performed with MrBayes 3.1.2 (Huelsenbeck et al. 2001; Ronquist and Huelsenbeck 2003) assuming a fixed amino acid model (mtRev: prset aamodelpr=fixed(mtrev)), with proportional evolutionary rate (prset ratepr=variable), gamma-shaped rate variation (lset rates=gamma), five rate categories (lset ngammacat=5), and unlinked model parameters (unlink shape=(all)) for 13 single-gene partitions. One million generations (ngen=1000000) were run with a sample frequency of 1000 (samplefreq=1000), and 2 x 4 search chains (nchains=4). After inspection of the graphical output the first 100 trees were omitted as burnin (sumt burnin=100).

Treefinder (version of October 2008) (Jobb, von Haeseler, and Strimmer 2004) was used for maximum likelihood analysis. Again, the prepared amino acid alignment was used with optimization of model parameters for the single genes (=13 partitions). Preceding to the actual analysis the substitution model was tested with Treefinder's 'Analysis -> Propose Model' function. The choice was between three user defined models: mtArt, mtMam, mtRev. The mtArt model was found most appropriate with optimization of amino acid frequency parameters, gamma distribution, optimization of heterogeneity, and 5 rate categories: (mtArt[,Optimum]:G[Optimum]:5). The actual bootstrap analysis was run with 100 pseudoreplicates according to the suggested parameters.

4.4 Results

Partial Genomes

The partial genome nucleotide sequences that were determined in the course of this study are available in GenBank (*Eurydice pulchra*: GU130253, *Sphaeroma serratum*: GU130256, *Glyptonotus* cf. *antarcticus*: GU130254, *Armadillidium vulgare*:

GU130251, *Asellus aquaticus*: GU130252, *Janira maculosa*: GU130255). It was not possible to obtain the area of the mt control region for any of these species. However, the mt genomes can be considered as 'nearly complete', as the comparison to already known isopod mt genomes of *Ligia oceanica* (Kilpert and Podsiadlowski 2006) and *Eophreatoicus* sp. 14 (Kilpert and Podsiadlowski 2009) suggests. These species feature a circular mt genome topology and a similar gene order, which make it very likely that the missing part of the isopod genomes in the present study is limited to the control region and few adjacent tRNA genes only. An exception is the mt genome of *Janira maculosa*, which could not be sequenced to a similar extent, so that also some of the protein coding genes are missing in the provided mt sequence (see next paragraph) (Figure 4.1).

Individual mt Genome Characterizations

We determined 13,055 bp of the *Eurydice pulchra* mt genome. In this part the annotation revealed 31 genes (Supplementary file 4.4a), including all 13 protein coding genes, both rRNA subunit genes, and 16 tRNA genes (Supplementary file 4.5a). As already mentioned the correct annotation of tRNA genes is crucial for analysis. Therefore, the method of tRNA identification is also referred (Supplementary file 4.2a). If one assumes that mt genome usually holds 22 tRNA genes, then 6 of them (*trnE, trnF, trnI, trnL(CUN), trnR, trnS(AGY)*) are probably located in the part near to the control region, which unfortunately could not be sequenced. A compact overall mt genome organization was found in all examined isopod species. Like in other mt genomes, the mitochondrial genes are densly packed, with few non-coding sections and frequent gene overlaps. In *Eurydice pulchra* two non-coding sections of 60 bp between *trnW* and *cox2* and 91 bp between *nad3* and *trnA* are the exceptions (Supplementary file 4.4a). The largest gene overlap (26 bp) was determined at the border of *trnH* and *nad4*.

The start codon GTG was specified for the *cox2* and *atp8,* due to the absence of a traditional start codon. The gene *nad4L* is probably initiated by ATC. Abbreviated stop codons, only consisting of a single T, terminate *cox2* and *nad3*. Peculiar start codons and abbreviated stop codons of all isopod species will be addressed in detail in the discussion part of this study.

The sequence of the *Sphaeroma serratum* mt genome is 13,467 bp in length and features 30 genes (Supplementary file 4.4b). All 13 protein coding genes are available, also the two rRNA genes, and 15 genes coding for tRNAs (Supplementary file 4.5b). In addition, the method of tRNA identification is listed (Supplementary file 4.2b). One non-coding section of 40 bp exists between *nad2* and *trnL2*; the largest gene overlap is located between *nad5* and *trnF*. Three unusual initiation codons were found: ACT (*cox1*), ATC (*nad6*), GTG (*nad3*). Incomplete stop codons, only consisting of a single T, were found in *cox1*, *cox3*, *nad3*, and *nad4*.

In the partial mt genome (13,809 bp) of *Glyptonotus* cf. *antarcticus* 33 gene were identified (Supplementary file 4.4c). All of the usual 13 protein coding genes and two rRNA genes are present, just as 18 tRNA genes (Supplementary file 4.5c, Supplementary file 4.2c). The mt genome does not feature any important non-coding sections. The genes *trnR* and *trnG* share 13 bp, which is the largest gene overlap in this mt genome. Unusual start codons were found in three genes: ACG (*cox1*), GTG (*cox2*), TTG (*atp8*). A single T terminal nucleotide for termination is present in *nad4*, TA is supposed to end *nad5*.

The partial genome of *Armadillidium vulgare* has a size of 12,890 bp (Supplementary file 4.4d). All protein coding genes and rRNA genes are present, likewise 13 tRNA genes (Supplementary file 4.5d, Supplementary file 4.2d). The longest non-coding part amounts 55 bp (located between *cox1* and *cox2*). Relatively large gene overlaps (up to 26 bp) are more frequent than in other examined isopod species. The only protein coding gene which shows an unusual start codon is GTT (*atp8*). In *cox1* the only indication for a termination signal is a single T nucleotide.

We decided to provide this sequence, although a similar partial mt genome of a specimen from another *A. vulgare* population was already published by other authors (Marcade et al. 2007). The comparison of both *Armadillidium vulgare* mt genomes showed a very high sequence similarity of 97.2%. The differences are limited to single nucleotide polymorphisms (SNP). This clearly indicates their membership to the same species, even though to different populations. For the analyses of this study we rely on the data of the specimen collected in Berlin (Germany). The identified gene order of *Armadillidium vulgare* differs compared to (Marcade et al. 2007), due to an annotation update (Supplementary file 4.3).

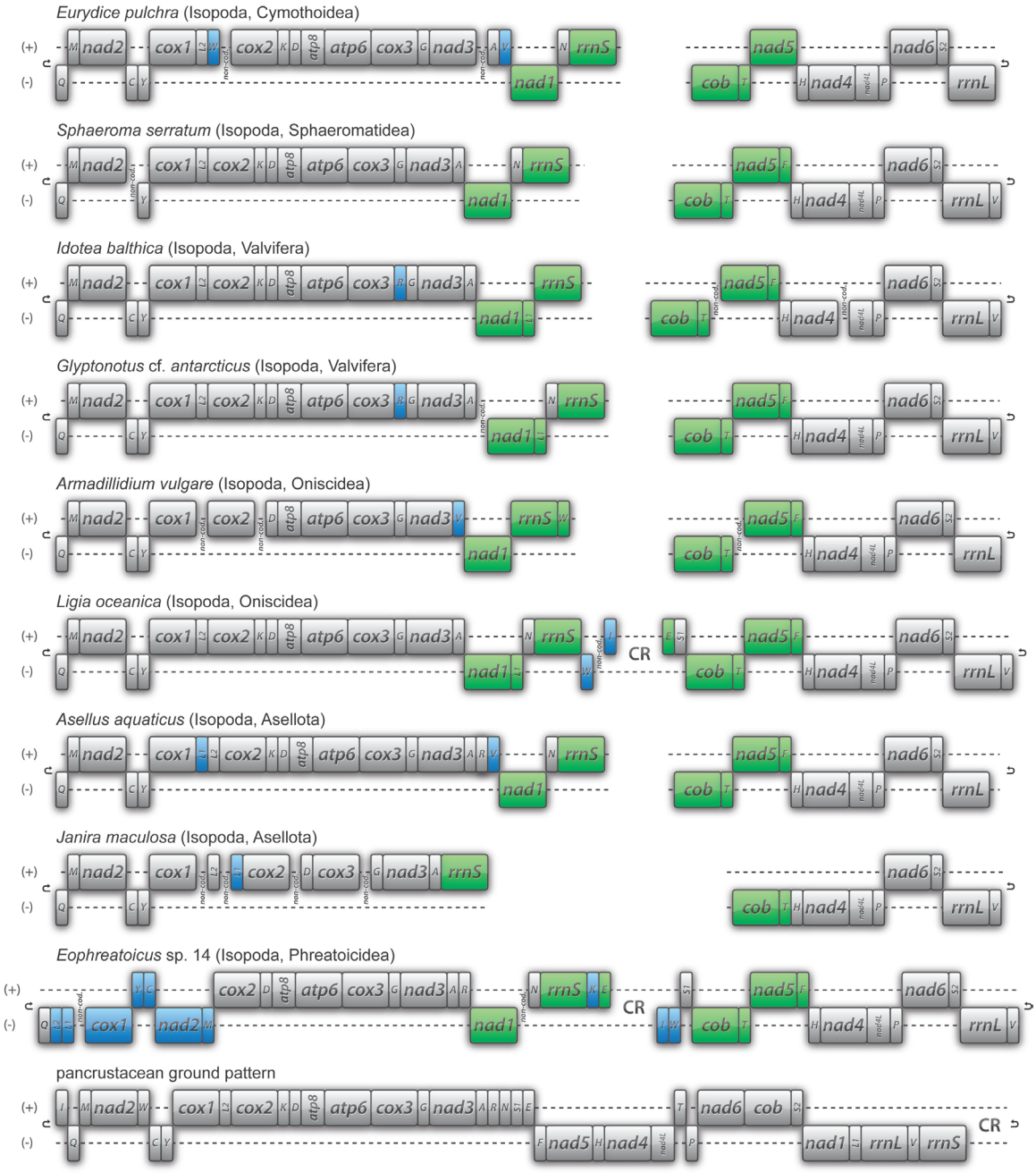


Figure 4.1. Comparison of mitochondrial gene arrangements of nine species of Isopoda. In addition the pancrustacean ground pattern, the putative ancestral state, is provided. The mt genomes of *Ligia oceanica* and *Eophreatoicus* sp. 14 are completely available; the other mt genomes are incomplete and are missing the area of the control region (CR). Coloured genes mark translocated genes in comparison to the ground pattern. Green colour indicates genes, which share a derived position in isopods. Uniquely derived gene positions of individual species are depicted in blue.

Asellus aquaticus is another species from which we were able to obtain a partial mt genome sequence (13,639 bp, Supplementary file 4.4e). Like most isopod mt genomes it also features 13 protein coding genes and 2 rRNA genes. 18 tRNA were identified as well (Supplementary file 4.5e, Supplementary file 4.2e). The largest part not being utilized for gene coding is only 16 bp in length (between *nad1* and *trnN*); the largest gene overlap amounts 15 bp (of *nad2* and *trnC*). There are a number of uncommon start codons: ACA (*nad4*), ACC (*cox1*), and CTG (*atp8*). The *nad4* gene shows only a single T instead of a fully formed stop codon in the DNA sequence.

The partial mt sequence of *Janira maculosa* is considerably shorter than the other mitogenomes presented here (9,871 bp). Although the fragment shares the same genes at its ends with the other obtains partial genomes (*cob* to *rrnS*) (Supplementary file 4.4f), several protein coding genes are missing (*apt6*, *atp8*, *nad1*, *nad5*). This is in clear contrast to the other isopod mt genomes, where both rRNA genes and all 13 protein coding genes were found in this segment. In addition only 14 tRNA genes could be identified in *Janira maculosa* (Supplementary file 4.5f, Supplementary file 4.2f). A relatively large non-coding section (215 bp) exists in this mt genome between the two *trnL* genes. A maximum of 27 bp is shared by *trnV* and *trnQ*. One exceptional start codon was detected, ACG in *cox1*. The gene also shows an incomplete termination codon (T).

Four other complete or almost complete mt genomes of isopod species have been published up to now: *Idotea balthica* (incomplete, Podsiadlowski and Bartolomaeus 2006), *Ligia oceanica* (complete, Kilpert and Podsiadlowski 2006), *Armadillidium vulgare* (incomplete, Marcade et al. 2007), and *Eophreatoicus* sp. 14 (complete, Kilpert and Podsiadlowski 2009). The sequences of these mt genomes were validated by the same software tools used to annotate the newly sequenced species of this study. This update to older annotations ensures the comparability to the species of the current analysis. The improved mt genome annotations of *Armadillidium vulgare* and *Ligia oceanica* are provided in Supplementary file 4.3).

Nucleotide Composition

Nucleotide frequencies, AT content, AT and GC skew vary considerably among the species in this study (Supplementary files 4.6a-f). To facilitate a comparison of the available isopod mt genomes, we completed this list with the respective data for *Idotea balthica* (Supplementary file 4.6g), which was not included in a previous publication (Podsiadlowski and Bartolomaeus 2006).

Large differences were observed for the AT content of different isopod species (all protein coding genes, all rRNA genes, total): *E. pulchra*: 55.7%, 54.9%, 55.9%; *S. serratum*: 52.2%, 53.0%, 54.4%, *G.* cf. *antarcticus*: 65.5%, 67.4%, 65.4%, *A. vulgare*: 71.7%, 72.1%, 71.4%; *A. aquaticus*: 61.0%, 67.7%, 62.0%; *J. maculosa*: 70.9%, 69.4%, 71.2%; *I. balthica*: 60.5%, 61.6%, 61.0%; *L. oceanica*: 60.5%, 62.6%, 60.9%; *Eophreatoicus* sp. 14: 69.8%, 72.5%, 69.6%.

When AT and GC skews are compared, it becomes evident that a decisive nucleotide bias is only existing in GC skews (Supplementary file 4.6a-g). Most species isopod species show a clearly positive GC skew for +strand genes and a negative one for –strand genes. For example in *E. pulchra* the GC skew for the +strand gene *nad3* is 0.4, whereas the –strand gene *nad4* has a GC skew of -0.15. The corresponding AT skews are both negative: -0.2 and -0.22, respectively. A similar condition can be found in other genes as well. Therefore, the GC skew has to be preferred to detect a bias in nucleotide frequencies. No clear bias exists in *G.* cf. *antarcticus* and *J. maculosa*, where the GC skews are nearly 0. However, a noticeable bias was detected for *A. aquaticus*, where GC skews are negative for +strand genes and positive for –strand genes. This is in clear contrast to the majority of examined species of Isopoda!

Table 4.1 gives an overview over the relative molecular weight of the +strand (compared to the -strand), AT content, and GC skew in Crustacea. It illustrates the tendency of isopod species for a positive GC skew on the +strand (except Asellota), while the other crustaceans generally show a negative GC skew (except *Hutchinsoniella macracantha*). The situation is similar with the molecular weight of the DNA strands, which also differs between isopod species (except Asellota) and the other crustaceans in this comparison (except *Hutchinsoniella macracantha*).

Phylogenetic Analyses of Amino Acid Sequences

The resulting trees from Bayesian inference (MrBayes) and maximum likelihood analyses (Treefinder), both based on concatenated alignments of 13 protein coding genes (3873 aa), show a very similar topology (see Figure 4.2 for the best tree of ML analysis). The Isopoda were always recovered as monophyletic. The relationships of isopod suborders are also identical. *Eophreatoicus* sp. 14 is sister to all other isopods, which are split into monophyletic Asellota, comprising *Asellus aquaticus* and *Janira maculosa*, and monophyletic Scutocoxifera (among others Oniscidea, Valviferea, Cymothoidea, Spaeromatidea) (Dreyer and Waegele 2002). Not recovered was the monophyly of the terrestrial Oniscidea, as *Armadillidium vulgare* and *Ligia oceanica* are paraphyletic in both trees. Well supported are the Valvifera, here represented by *Glyptonotus* cf. *antarcticus* and *Idotea balthica*. They form the sister group to the cymothoidean species *Eurydice pulchra* and the sphaeromatidean species *Sphaeroma serratum*. A monophylum of isopod and amphipod species, which would represent the Peracarida in this analysis, was not recovered. Instead, the cirripede *Pollicipes polymerus* is placed with the amphipod species. For this reason there is also no support for Malacostraca.

Mitochondrial Gene Translocations

Figure 4.1 shows a comparison of all nine currently available isopod mt genomes, including four previously published mt genomes. In addition the putative ancestral state, the pancrustacean ground pattern, is provided (Boore, Lavrov, and Brown 1998; Kilpert and Podsiadlowski 2006; Lavrov, Brown, and Boore 2004).

The comparison revealed three noticeable facts: First, all of the nine isopod mt genomes differ in their gene order but most differences are limited to the position of one or a few tRNA genes. Second, the gene order of isopod mt genomes can be clearly distinguished from the pancrustacean ground pattern. Inferred changes involve tRNA genes, protein coding genes, rRNA genes, and the mitochondrial control region (CR). Third, the isopods share a derived order of genes, although there are modifications in individual species. Using the similarities common to most isopods it was possible to infer a most

Table 4.1: Comparison of crustacean +strand AT contents and GC skews.

Species	Taxon	Acc. number	Strand	AT content (%)	GC skew
Ligia oceanica	Isopoda, Oniscidea	NC_008412	H	60.85	0.134
Armadillidium vulgare	Isopoda, Oniscidea	GU130251	H	71.36*	0.175*
Glyptonotus cf. *antarcticus*	Isopoda, Valvifera	GU130254	H	65.38*	0.038*
Idotea balthica	Isopoda, Valvifera	DQ442915	H	60.97*	0.163*
Sphaeroma serratum	Isopoda, Sphaeromatidea	GU130256	H	54.40*	0.219*
Asellus aquaticus	Isopoda, Asellota	GU130252	L	61.97*	-0.122*
Janira maculosa	Isopoda, Asellota	GU130255	L	71.23*	-0.026*
Eophreatoicus sp. 14	Isopoda, Phreatoicidea	FJ790313	H	69.57	0.250
Eurydice pulchra	Isopoda, Cymothoidea	GU130253	H	55.86*	0.198*
Parhyale hawaiensis	Amphipoda, Gammaridea	AY639937	L	73.66*	-0.121*
Metacrangonyx longipes	Amphipoda, Gammaridea	NC_013032	L	76.03	-0.035
Squilla mantis	Stomatopoda	NC_006081	L	70.17	-0.130
Portunus trituberculatus	Decapoda	NC_005037	L	70.22	-0.241
Cherax destructor	Decapoda	NC_011243	L	62.43	-0.280
Euphausia superba	Euphausiacea	AB084378	L	67.77*	-0.143*
Hutchinsoniella macracantha	Cephalocarida	NC_005937	H	70.99	0.313
Vargula hilgendorfii	Ostracoda, Myodocopa	NC_005306	L	61.61	-0.355
Pollicipes polymerus	Maxillopoda, Cirripedia	NC_005936	L	67.04	-0.134

* incomplete mt genome

H: heavy strand

L: light strand

parsimonious hypothesis for the isopod ground pattern of mt gene order (Figure 4.3),comprising the unique arrangement of *nad1*, *trnL1*, *rrnS*, CR, *trnS1*, *cob*, *trnT*, *nad5*, and *trnF*. Direct evidence for the position of CR is only available for *Ligia oceanica* and *Eophreatoicus* sp. 14, because in all other isopods the segments containing the control region could not be sequenced. But as there is no large non-coding segment between *rrnL* and *nad2*, it is obvious that the CR was translocated in these mt genomes as well and not retained in its ancestral position.

A derived relative position of *trnV* is shared in three isopod species. It is located upstream of *nad1* in *Eurydice pulchra*, *Armadillidium vulgare* and *Asellus aquaticus*. The other six isopods, however, have *trnV* next to *rrnL*, which corresponds to the pancrustacean ground pattern. Also for another gene, *trnR*, a commonly derived position was detected: In *Idotea balthica* and *Glyptonotus* cf. *antarcticus* it is located in-between *cox3* and *trnG*, this position is a potential apomorphy for the isopod subtaxon Valvifera. In contrast, the ancestral position of *trnR* next to *trnA* was retained in

Eophreatoicus sp. 14 and *Asellus aquaticus*. For the other isopods *trnR* could not be identified in the sequenced parts of the mt genomes.

Furthermore, there are several tRNA genes, which were translocated in addition to the already mentioned isopod rearrangements (e.g. *trnW* in *Eurydice pulchra*). Whether these tRNA gene translocations are unique to individual species only or characterize taxa on a lower taxonomic level remains to be seen when more mt genomes of isopods become avaiable. For other genes (e.g. *trnC* in *Sphaeroma serratum*) it is only certain by comparison that they are no longer located in an expected position. The definite new position of the gene remains unclear for now.

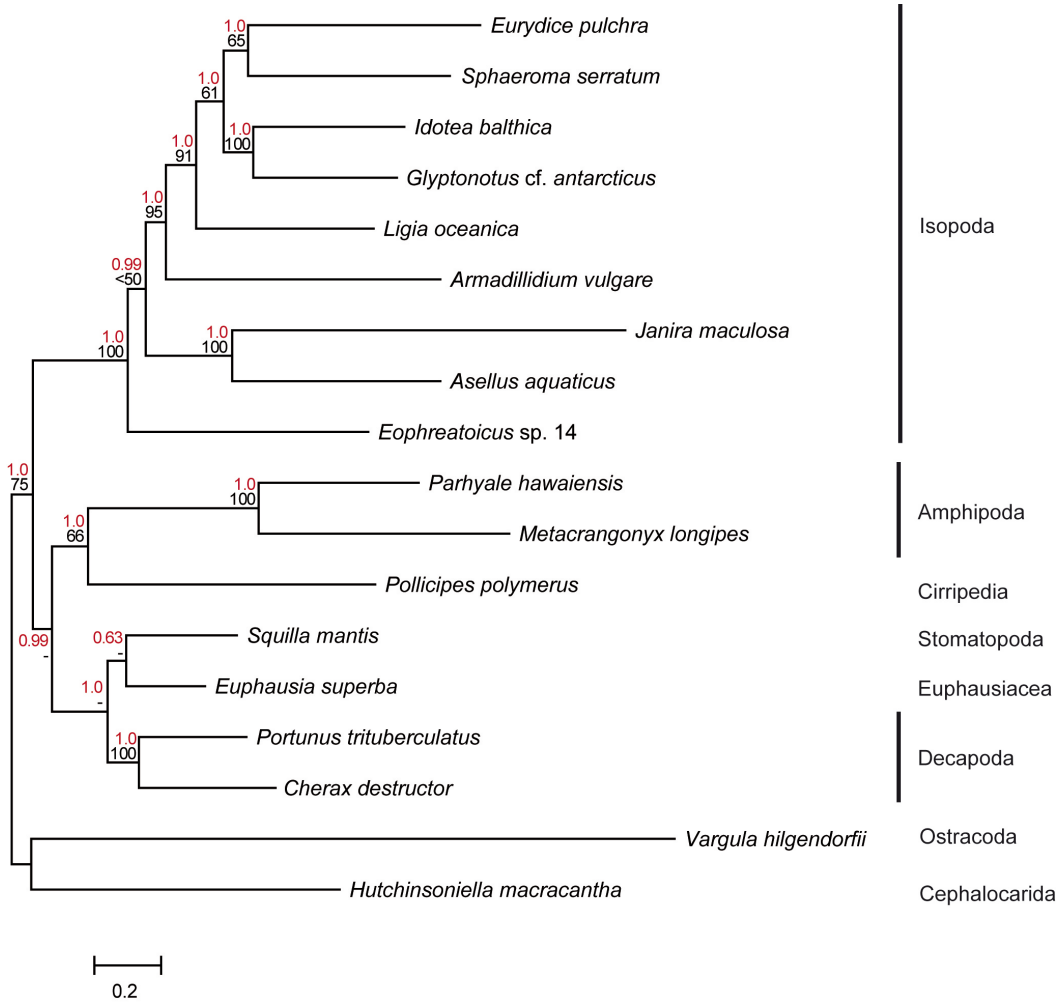


Figure 4.2. Phylogenetic analyses of isopod mt genomes and out group species. Majority-rule consensus tree from Treefinder analysis (mtART+G). Numbers above branches indicate (from top to bottom): Bayesian posterior probabilities (red) and bootstrap percentages (black) of the ML analysis. Scale bar below the tree indicates evolutionary distances (substitutions per site).

It is conspicuous that the overall distribution of rearrangements is unequally distributed throughout the isopod mt genome. All major translocations, which are shared among isopods, found a destination in a well defined area between *trnA* (or *trnR* for *Asellus aquaticus* and *Eophreatoicus* sp. 14) and *trnH*. For *Eophreatoicus* sp. 14 and *Ligia oceanica* it is likely that the CR is located in the centre of this isopod ”rearrangement hotspot”. The surrounding tRNA genes are characterized by an increased variability. Both species differ in gene content, gene arrangement, as well as in the orientation of genes. The area of the control region (between *rrnS* and *cob*) is in *Eophreatoicus* sp. 14: *trnK*(+), *trnE*(+), CR, *trnI*(-), *trnW*(-), *trnS1*(+); and in *Ligia oceanica*: *trnW*(-), *trnI*(+), CR, *trnE*(+), *trnS2*(-). This part of the mt genome is currently not available for other isopods.

The specified inversion (*trnL2*, *trnL1*, *cox1*, *trnY*, *trnC*, *nad2*, *trnM*) of a part of the *Eophreatoicus sp. 14* mt genome (Kilpert and Podsiadlowski 2009) was not found in any other isopod mt genome in this study. However, a similarly derived gene position of *trnL1* (inbetween *cox1* and *trnL2*) was detected in the asellotan species *Asellus aquaticus*. The sequence similarity of both *trnL* genes of *Asellus aquaticus* is 59.7%. In the other asellotan species, *Janira maculosa, trnL2* is also located in proximity to *trnL1*, but on the other side and separated by a non-coding region of 215 bp (gene order: cox1, *trnL2*, non-coding, *trnL1*). Here, the sequence similarity of *trnL* genes is 49.2%.

The mt gene arrangement of *Janira maculosa* surely is an exception in this comparison of isopod mt genomes. A segment comparable to that sequenced for the other species lacks tRNA genes and even four protein coding genes (*atp6*, *atp8*, *nad1*, *nad5*). This resulted in a significant shorter mtDNA fragment (ranging from *rrnS* to *cob)* than in other isopods (see Supplementary file 4.4). In addition, four noticeable non-coding sections were detected, in contrast to the compact organization of usual mt genomes. *Janira maculosa* also shows a conspicuous location of *trnL1* (nearby *trnL2*). Despite all these unique features, indicating more extensive rearrangements than in other isopods, large sections are in conformity to the typical isopod mt genome. In fact, the area ranging from *nad4L* to *trnQ* is characterized in isopods by the absence of several genes (*trnT*, *cob*, *nad1*, *trnL1*, *rrnS*, CR, *trnI*) compared to the ground pattern. For this important part, being the source of many translocation events typical for isopods, the gene order of *Janira maculosa* is very similar to that of other isopods.

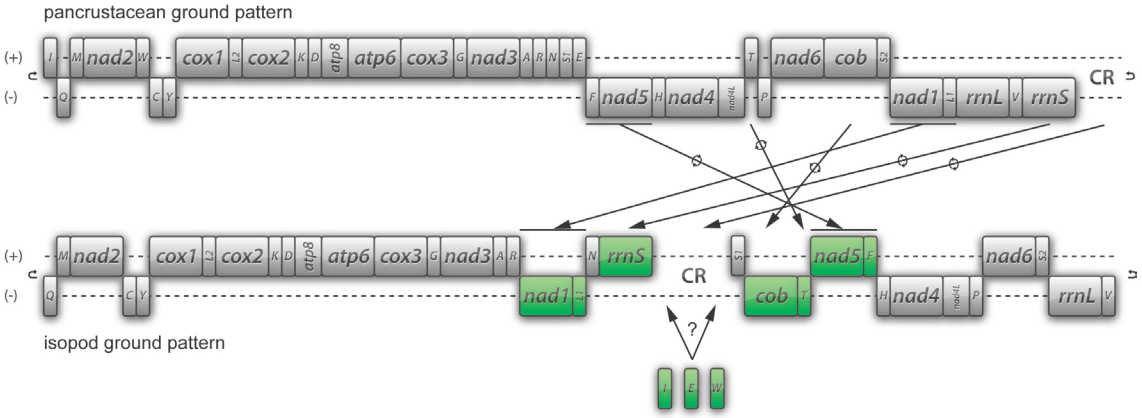


Figure 4.3. Isopod mitochondrial genome ground pattern. The mt gene arrangement given by the isopod ground pattern putatively evolved from the pancrustacean ground pattern by several translocation events (arrow lines). All translocated genes are coloured in green. For parsimony reasons the concatenated genes *nad1*/*trnL1* were probably translocated in one step. This applies to *nad5/trnF* as well, but in combination with a strand switch. Single gene translocations involved *cob*, *rrnS*, and *trnT*; here again strands were changed. The control region (CR) is also found in a new area of the isopod mt genome. The isopod mitochondrial ground pattern must remain preliminary, because too little is known about the tRNA genes next to the CR. This area probably contains the genes *trnI*, *trnE*, *trnS1*, and *trnW*.

4.5 Discussion

Mitochondrial Gene Translocations in Isopod Species

The comparison of the nine available isopod mt genomes (Figure 4.1) leads to identification of a number of genes sharing a derived relative position in isopods compared to other arthropods. In addition, the mt control region seems to be translocated in isopods. Together, they prominently characterize the isopod mt genome by a typical arrangement of *nad1*, *trnL1*, *rrnS*, CR, *trnS1*, *cob*, *trnT*, *nad5*, and *trnF*. But also a less apparent area ranging from *nad4* to *trnQ* has a unique gene order in isopods, as this region lost a number of the above mentioned genes. So the gene order of isopods differs significantly from the pancrustacean ground pattern (Boore, Lavrov, and Brown

1998; Kilpert and Podsiadlowski 2006), the hypothetical mt gene arrangement, which evolved in a common ancestor of Crustacea and Hexapoda and is actually present in a wide range of different species (e.g. *Drosophila yakuba* (Hexapoda, Diptera), *Daphnia pulex* (Branchiopoda, Cladocera), *Penaeus monodon* (Malacostraca, Decapoda). From the malacostracan clade Peracarida, which also includes the Isopoda, only two other mt genome sequences are currently available, from the amphipods *Parhyale hawaiensis* (Cook, Yue, and Akam 2005) and *Metacrangonyx longipes* (Bauza-Ribot et al. 2009). Their gene orders both exhibit differences to the pancrustacean ground pattern, but different ones than isopods, so there is no evidence for common gene translocation events for Isopoda and Amphipoda. Therefore, the pancrustacean ground pattern represents the most reasonable mt gene order of the common ancestor of amphipods and isopods.

Figure 4.3 shows a minimum number of postulated translocation events, which are needed to explain the change from the pancrustacean ground pattern into the putative mt gene arrangement of the common isopod ancestor, inferred from the observed shared features of nine isopod mt genomes (Figure 4.1). Six major translocations most likely happened before the Isopoda split up into different lineages (Figure 4.3): *nad1* and *trnL1* were probably translocated together, due to their retained adjacent position in the new location. This is also true for *nad5* and *trnF*, but accompanied by an inversion. Single genes which were also translocated and inverted are *rrnS*, *cob*, and *trnT*. Likewise, the CR was translocated and probably inverted, suggested by nucleotide frequency bias (see below). A decision about translocation of the tRNA genes (*trnI*, *trnE*, *trnS1*, *trnW*) surrounding the CR is not possible at the moment; data from more isopod species are needed here. *trnK* is obviously no candidate for a location near the CR (as seen in *Eophreatoicus* sp. 14) in the isopod ground pattern, because in most other isopods it is found between *cox2* and *trnD*, which is a plesiomorphic character state found also in the pancrustacean ground pattern. Features of the isopod ground pattern are found in nearly all isopods included in this study, although the exact gene order is modified in single species. The strongest deviation from the supposedly ancestral mt gene order is found in *Janira maculosa* (Figure 4.1). Here non-coding sequences are unusually frequent; protein coding genes are missing in expected positions. In this respect it is the most unusual isopod mt genome examined so far. Nevertheless, it can be clearly assigned to the isopod clade, due to the typical gene order

of the area ranging from *nad4L* to *trnQ*. Another unusual major mt arrangement is an inversion covering the region from *trnM* to *cox1* found in *Eophreatoicus* sp. 14. This unique feature might become relevant for further phylogenetic analyses of this clade when data from more species become available.

A potential apomorphy of the Valvifera is the derived position of *trnR* (between *cox3* and *trnG*) found in *Idotea balthica* and *Glyptonotus* cf. *antarcticus*. The *trnR* gene never appears in this position in other isopods but rather in a plesiomorphic position (next to *trnA*) in the mt genomes of *Asellus aquaticus* and *Eophreatoicus* sp. 14, which supports basal placement of Phreatoicidea and Asellota within the Isopoda (Wilson 2009). The absence of *trnR* in the other asellote, *Janira maculosa*, is most likely due to the heavily modified nature of this mt genome as mentioned above.

Another noticeable gene order change is the occurrence of *trnL1* next to *trnL2*, which is shared between phreatoicidean and asellotan species but not by other isopods. Thus, with the exception of the above mentioned inversion of *Eophreatoicus*, the gene order is the same in *Asellus aquaticus* and *Eophreatoicus* sp. 14. Again the other asellotan species, *Janira maculosa* shows further modifications and non-coding sequences in the particular area. The three species show a strong similarity of both *trnL* genes in gene sequence as well as in secondary structures (Supplementary file 4.5): 77.3% in *Eophreatoicus* sp. 14, 59.7% in *Asellus aquaticus*, and 49.2% in *Janira maculosa*. These high similarities indicate a translocation mechanism which involved a duplication of the already existing *trnL2* gene, followed by an identity change to *trnL1*, while the original *trnL1* gene degraded. Such a series of events was already suggested for a few cases of mt gene order change (Rawlings, Collins, and Bieler 2003). It is probably a rare event, in this case weakly supporting the hypothesis of a common clade including asellotan and phreatoicidean species.

Control Region Inversion

Most animal species show a clear strand asymmetry in their mt nucleotide composition (Hassanin, Leger, and Deutsch 2005) such as their ratio of guanine to cytosine nucleotides within a single strand of DNA (the so-called GC skew value). Similarly, an AT skew is sometimes also present, but it is often less distinctive, like in the species of

this study. It is assumed that these biases are caused by mutational constraints, which affect each strand to a varying degree, due to an asymmetric replication process, described by the strand-displacement model (Bogenhagen and Clayton 2003; Clayton 1982): When replication begins, both strands become temporarily single-stranded. Then a unidirectional complementation to a double-stand is immediately started on one strand, while the complementation in the other direction is postponed, probably due to a wide separation of the replication origins for each strand. This extra time of one strand staying single-stranded and being particularly prone to mutational nucleotide deamination of cytosine (and adenine), then leads to the shift in the GC skew (and AT skew). A reversal in the direction of the replication therefore induces a reversal of the mutational exposure as well, which will be reflected in a reversed GC skew after some time.

In crustaceans the GC skew usually is negative for genes located on the +strand (light strand) and positive for genes on the –strand (heavy strand) (Hassanin 2006). The only isopod in which this ancestral state was detected is the asellotan species *Asellus aquaticus* (Table 4.1; Supplementary file 4.6e). On the contrary most other isopods show a clearly reversed GC skew (Table 4.1; Supplementary file 4.6a-d, 4.6f+g). We therefore assume that the origin of replication, which is a prominent part of the mt control region, is oriented in the opposite direction in *Eophreatoicus* sp. 14 (Phreatoicidea), *Ligia oceanica* (Oniscidea), *Armadillidium vulgare* (Oniscidea), *Idotea balthica* (Valvifera), *Sphaeroma serratum* (Sphaeromatidea), and *Eurydice pulchra* (Cymothoidea). No clear bias was detected for the valviferan *Glyptonotus* cf. *antarcticus* and the asellotan *Janira maculosa*. The reason for the taxon-specific differences of these biases is unclear. It could be caused by an alteration of the replication process, which is less asymmetrical, or by an additional reversal a short time ago. In *Janira maculosa* additional changes in gene order indicate intensive rearrangements, which might also include another reversal of the control region and probably blurred a previous bias. In this regard, *Janira maculosa* is a special case.

The fact that a clearly reversed bias was detected in all examined major isopod taxa except Asellota suggests a common origin of this state for parsimony reasons. This could be the case, if the clade which includes *Asellus aquaticus* would branch off first in isopod phylogeny. The reversal of the origin of replication would be a shared autapomorphy of the branch including all other isopods.

The fossil record (Schram 1970) and previous studies of isopod phylogeny favor the Phreatoicidea as the sister clade of the remaining isopods (Brusca and Wilson 1991; Wägele 1989), a position which is also supported by the molecular analyses of this study. However, this causes a conflict in explaining the reversed isopod GC skew, which is present in *Eophreatoicus sp. 14* and in the majority of isopods, but not in asellotan species. Considering Phreatoicidea as sistergroup to all other isopods, at least two reversals of the orientation of the mt control region must be assumed: Either a) at the base of the whole Isopoda clade and a re-reversal for some Asellota (including *Asellus aquaticus*); or b) two independent reversals for the Phreatoicidea and the Scutocoxifera (a clade of all remaining isopods, except Asellota). Although similar independent reversals of the GC skews were reported from other arthropod clades before (Hassanin 2006), they are not assumed to be common events. Thus, we take the differences in GC-skew to be a faithful indicator of reversals of control region orientation inside the mt genomes, which lends some support for the hypothesis that Asellota might have branched off earlier than Phreatoicidea. A recent combined analysis based on morphological and 18S data showed similar results (Wilson 2009) and placed the Asellota and not the Phreatoicidea as sister group to all other isopods. It has to be tested by further analyses including GC skews of more isopods, particularly asellotan species.

The detection of a general reversed GC skew in some species should be considered also in phylogenetic analyses, as the compositional heterogeneity among taxa may cause higher substitution rates (Hassanin, Leger, and Deutsch 2005), promoting problems such as long branch attraction (LBA) (Felsenstein 1978).

Non-canonical Start and Stop Codons

Some of the protein coding genes begin with unconventional start codons in relation to the commonly used invertebrate mitochondrial code. Nevertheless, alternative initiation codons are known from literature, e.g. ATC in honey bee *Apis mellifera* (Crozier and Crozier 1993), GTG in the chiton *Katharina tunicate* (Boore and Brown 1994), or ACG in the shrimp *Penaeus monodon* (Wilson et al. 2000). The reason for postulating the existence of new codons is the absence of canonical start or stop codons at the beginning or end of a protein coding gene. A multiple alignment of genes from other

species allows to estimate the range that usually holds the start codons. In the case that an appropriate start codon is unavailable, a new potential start codon is suggested, which is in a comparable relative position like in the other related species. This seems more reasonable than to assume a significant expansion or reduction of a vital gene, which, however, cannot be ruled out completely.

In general, ATG is the most frequent initial codon in isopod mt genomes, followed by the alternatives ATT and ATA. Another prevalent initial codon is ACG for the *cox1* gene, which has been repeatedly reported from malacostracan crustaceans, the first time for *Penaeus monodon* (Wilson et al. 2000). Isopods, which likewise show an ACG start codon in *cox1* are *Glyptonotus* cf. *antarcticus* (this study), *Janira maculosa* (this study), *Ligia oceanica* (Kilpert and Podsiadlowski 2006), *Idotea balthica* (Podsiadlowski and Bartolomaeus 2006), *Eophreatoicus* sp. 14 (Kilpert and Podsiadlowski 2009). The start codon of *cox1* is ACT in *Sphaeroma serratum*, and ACC in *Asellus aquaticus*. These are probably equivalent start codons to of ACG.

The GTG initiation codon might not be too unusual as well. It was not only detected in the chiton mentioned above, but also in *Ligia oceanic*a in the *atp8* gene. In *Eurydice pulchra* we assume GTG to initiate *atp8* and *cox2*, in *Glyptonotus* cf. *antarcticus* it initiates *cox2*, too. Another GTG finding is in *nad3* of *Sphaeroma serratum*.

Generally, a lot of different non-canonical codons were found for the extremely short gene *atp8*: GTG as mentioned above, CTG in *Asellus aquaticus*, TTG in *Glyptonotus* cf. *antarcticus*, GTT in *Armadillidium vulgare*. In addition there are usual start codons like ATT in *Spharoma serratum*. As the taxon sampling increases it becomes clearer that the invertebrate mitochondrial code may be less widely applicable than previously assumed, at least for start codons of some genes.

In every isopod species included in this study incomplete termination codons have been found, at least for some of the mt protein coding genes (Supplementary file 4.4a-f). Like described for the initiation codons, no appropriate codon was found in a reasonable area. To ensure the proper termination of the gene it is assumed that the stop codon is not completely coded in the DNA. Rather it is created by polyadenylation of the mRNA subsequent to the transcription (Anderson et al. 1981; Ojala, Montoya, and Attardi 1981). Therefore, only shortened stop codons like TA or T exist in the genome sequence, which are finally completed to a functional UAA codon able to terminate the

following translation process. In isopods incomplete terminal codons are particularly prevalent in the genes *cox1* (in *S. serratum, A. vulgare, J. maculosa*,) and *nad4* (in *S. serratum, G.* cf. *antarcticus, A. aquaticus, E. sp. 14*).

tRNA Content and Secondary Structure Modifications

The complete mt genome of *Eophreatoicus* sp. 14 demonstrates that the typical set of 22 tRNA genes is present in isopod mt genomes as well (Kilpert and Podsiadlowski 2009). However, the gene *trnR*, specific for the transfer of arginine, was not be identified in the complete *Ligia oceanica* mt genome (Kilpert and Podsiadlowski 2006). Mitochondrial genes are unlikely to be dropped without replacement, as every gene usually is needed to maintain all the essential biochemical functions (e.g. oxidative phosphorylation), which the mitochondrion provides to the cell (Boore 1999). Since arginine is encoded in almost every mt protein coding gene of the species, the specific tRNA must be either imported from the cytoplasm or the corresponding gene (*trnR*) is modified beyond recognition in the mtDNA, due to altered nucleotides, which are probably subject to RNA editing (Brennicke, Marchfelder, and Binder 1999; Stuart et al. 2005).

Due to the incomplete character of the available sequences from the other examined isopod species, not all of the tRNA genes were identified yet. A summary of the set of tRNA genes present in each taxon is provided in Supplementary file 4.7. The missing tRNA genes and the CR are probably located in the unknown part of the mt genomes, as suggested by the examples of *Eophreatoicus* sp. 14 and *Ligia oceanica*. These species host up to five tRNA genes surrounding CR. This corresponds to the expectations for *Asellus aquaticus*, *Glyptonous* cf. *antarcticus*, and *Idotea balthica*, where 4, respectively 5 for the latter, of the complete set of 22 tRNAs are missing. *Sphaeroma serratum* (7) and *Eurydice pulchra* (6) even show a higher number of missing of tRNAs in the known parts of their mtDNA sequence. Data from the control region are needed to elucidate the fate of these genes. The extraordinarily high number of absent genes in *Janira maculosa* (8) can be explained by greater changes compared to other isopods and the significantly less complete coverage of its mt genome. A maximum of tRNA genes is lacking in *Armadillidium vulgare* (9) despite our annotation update, which already increased the number of determined genes (Supplementary file 4.3). Although the overall mt gene order seems to correspond to the typical isopod arrangement, a

linearization of the mtDNA molecule was determined for this genome (Marcade et al. 2007; Raimond et al. 1999). If the missing tRNA genes are located at the outer ends of the linear molecule is not clear.

We consider the risk of having overlooked single tRNA genes as relatively low. Generally, in isopod mt genomes the remaining non-coding regions are short, leaving no place for additional tRNA genes. All larger non-coding regions, which still occur in some species, were thoroughly checked using different software and manual inspection. If tRNA genes remained undetected in these regions, they must be strongly modified.

Some of the identified tRNAs show modifications to the usual clover-leaf secondary structure (Supplementary files 4.5 and 4.7). Most frequent is the loss of the DHU-arm (left arm). It is missing in tRNA-Cys in nearly all examined isopods, with the exception of *Eurydice pulchra*, where the TΨC-arm (right arm) is missing and *Sphaeroma serratum*, where tRNA-Cys could not be identified (Supplementary file 4.7). A loss of the DHU-arm in tRNA-Cys is so far not known from other malacostracan crustaceans and is a potential apomorphic character of isopods. In tRNA-Ser(AGY) of *Eophreatoicus* sp. 14 and *Ligia oceanica* the left arm is also commonly missing. But here it corresponds to the usual state of this tRNA in Bilateria (Haen et al. 2007) and was also reported from various species of Malacostraca (Kilpert and Podsiadlowski 2006). In other isopod species the according tRNA gene was not recovered by the partial genomes, but is most likely present in the unsequenced mt area next to the mt control region. A loss of the TΨC-arm (right arm) was detected in other tRNAs of several isopod species but no common pattern is detectable (Supplementary file 4.7). The highest number of derived tRNA secondary structures is found in *Armadillidium vulgare*. In addition to loss of arms *Armadillidium* also features an enlarged variable loop in tRNA-Asp and only few base parings in the acceptor arm of tRNA-His. The frequent tRNA modifications in this mt genome are another indication for an overall modified condition of the mt genome in this species. The determined tRNA modifications might emerge to be particularly useful in isopod phylogeny on a lower taxonomic level, e.g. comparing different *Armadillidium* species.

Isopod Phylogeny Inferred from Sequence Analysis

The results from Bayesian inference (BI) and Maximum likelihood (ML) analyses (Figure 4.2) demonstrate some of the problems that sequence based analyses are facing in crustacean phylogeny. Resolution and support of isopod suborders is quite good with both methods. The tree topologies confirm the hypothesis of Phreatoicidea being the sister group of all other isopods, followed by successively branching Asellota, Oniscidea, Valvifera, and Sphaeromatidea+Cymothoidea. The only commonly accepted isopod clade that is not supported by our analysis is Oniscidea, which appears as a paraphylum here. In other analyses monophyly of the group is beyond doubt, as it comprises all terrestrial isopods (Brusca and Wilson 1991; Wägele 1989). Outside the Isopoda the results have to be regarded less suitable, as the Peracarida, here represented by Isopoda and Amphipoda, were not recovered as monophyletic in any of our analyses. Neither do the trees support Malacostraca. This may be due to the limited taxon sampling for the outgroup. But as the focus of this study is on the phylogeny of Isopoda, we consider the inadequacies in resolving peracarid and malacostracan relationships acceptable. The isopod branch of the sequence analysis was used to map structural characters of the mt genomes (see next paragraph; Figure 4.4). This way it serves as our isopod phylogeny hypothesis and allows evaluation of the putative apomorphic character states. It should also help to check genome morphology for homoplasious character states.

Homoplasies in Gene Order Re-Arrangments

Mt gene rearrangements are generally considered as very reliable not only because of rareness and selective neutrality of gene order, but also because of the high number of potential gene orders, which make convergently derived gene arrangements unlikely (Boore et al. 1995). Thus, reports about homoplasy of gene order are virtually inexistent; the only example for parallel evolution of mt genome organization is known from birds (Mindell, Sorenson, and Dimcheff 1998), but here only the relative position of genes towards the control region is different. In isopod species we have now detected strong indications for convergently derived structural characters. The gene *trnV* is located in most species in the ancestral relative position between *rrnL* and *trnQ* (Figure 4.1). But three species from different main isopod suborders share a derived relative

position of *trnV* close to *nad1*: *Eurydice pulchra* (Cymothoidea), *Armadillidium vulgare* (Oniscidea), and *Asellus aquaticus* (Asellota). This derived gene arrangement cannot be reconciled with any reasonable phylogeny of Isopoda and must therefore be considered as homoplasious (Figure 4.4).

Another strong indication for a convergently derived gene order was found in *Asellus aquaticus* and *Eophreatoicus* sp. 14, which also show a specific gene translocation mechanism: *trnL1* probably is a remolded duplicate of *trnL2*, suggested by sequence similarity and a tandem arrangement of the two genes in both species. A modified gene order caused by this mechanism is probably more likely and occurs more frequently than usual gene translocations caused by duplication-random-loss. However, the shared position of *trnL1* could also indicate a monophyly of a combined asellotan and phreatoicidean clade, as suggested by (Wilson 1999). This hypothesis cannot be ruled out currently, due to the open discussion about the first splits in isopod phylogeny.

Most changes of gene order and of other characters of genome morphology are limited to single species (Figure 4.4) such as the multigene inversion (*trnM* to *trnL2*) in *Eophreatoicus* sp. 14 and the high number of rearranged genes in *Armadillidium vulgare* and *Janira maculosa*. But modified structural characters are also present in the mt genome of the other species. They emphasize that the mitochondrial gene order is less conserved inside the Isopoda than what is known from other major crustacean taxa. We are confident that a lot of these characters provide valuable information in further studies of isopod phylogeny. They may be especially helpful for studying lower level relationships, e.g. of isopod suborders and families.

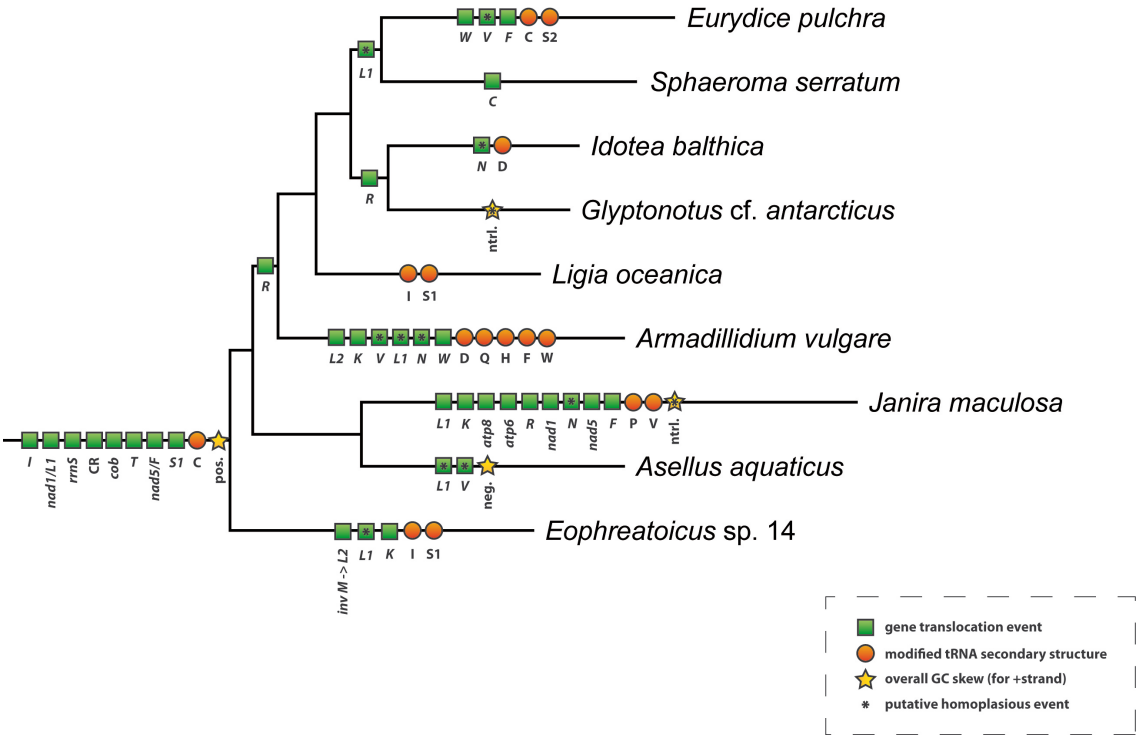


Figure 4.4. Apomorphic characters of genome morphology. Gene translocation events, tRNA secondary structure modifications and changes of GC skew mapped on the consensus tree from Bayesian and maximum likelihood analyses (Figure 4.3).

4.6 Conclusions

The taxon sampling of mt genomes from Isopoda was extended by five partial mt genomes (*Eurydice pulchra*, *Sphaeroma serratum*, *Glyptonotus* cf. *antarcticus*, *Asellus aquaticus*, *Janira maculosa*). With the exception of *Janira maculosa*, which features a higher degree of mt gene rearrangements, these mt genomes can be regarded as nearly complete, missing only a small proportion surrounding the control region. In addition, we provide new data from *Armadillidium vulgare*, which benefits from an update of genome annotation. The broad comparison of all nine currently available isopod mt genomes reveals a greater variability of gene order and other structural characters than generally assumed for crustaceans. It can be confirmed that a number of derived genes are typically shared by isopod species of all major suborders. They are attributable to an arrangement of *nad1*/*trnL1*, *rrnS*, CR, *cob*, *trnT*, *nad5*/*trnF*, *trnS1*, which probably had its origin in a common isopod ancestor and is part of the isopod ground pattern. The

plesiomorph relative position of *trnR* was retained in the phreatoicidean species and in one of the asellotan species only. At the base of the Scutocoxifera *trnR* was translocated. Another derived position of *trnR* is shared by the two valviferan species. Although gene order is generally considered as a reliable phylogenetic character, we found in some cases evidence of homoplasy: The derived location of *trnV* is shared by three unrelated isopod species; some species show a congruent loss of *trnN*, but with no proof of the new location; a translocation of *trnL1*, probably caused by a duplication/remolding mechanism, might also be evolved in parallel in phreatoicidean and asellotan species. However, the early splits in the isopod tree are still unclear. The phylogenetic analyses (ML, Bayes) based on amino acid sequences of all 13 protein coding genes favor the Phreatoicidea as the sister group of all other isopoda, comprised of the sistergroups Asellota and Scutocoxifera. A further structural character of mitogenomes, the nucleotide bias (GC-skew), was found inversed in most isopod species (except *Asellus aquaticus*) in comparison to other crustaceans. This character is able to indicate a very complex modification of the mt genome, the reversal of the replication origin, and probably is of special phylogenetic value. tRNA secondary structure modifications, found in nearly all isopod species may be helpful for deducing evolutionary relationships of lower level isopod groups.

4.7 Acknowledgments

The authors wish to thank Dr. Christoph Bleidorn (Univ. Leipzig) for collecting specimens from all over the world, in this case from Roscoff (France). We thank Prof. Dr. Heinz-Dieter Franke (Biologische Anstalt Helgoland) and Achim Meyer (Univ. Mainz) for providing us with specimens. Christine Bergmann (FU Berlin) is acknowledged for technical assistance. This study was supported by a NaFöG Berlin grant to FK and by research grants from the German science foundation (DFG)Po 765/2-1,2 (LP) and Ba 1520/10-1,2 (TB).

4.8 Additional material

Supplementary files 4.1-4.7 can be found in the appendix of this thesis (chapter 11).

5. The mitochondrial genome of the Japanese skeleton shrimp *Caprella mutica* (Amphipoda: Caprellidea) reveals a unique gene order and shared apomorphic translocations with Gammaridea

5.1 Abstract

This study presents the complete mitochondrial genome of the amphipod *Caprella mutica*, an east-asian species, which recently invaded coastal regions of North America, Europe and New Zealand. It is the first complete sequence of a member of the amphipod subclade Caprellidea. The mitochondrial genome has a total length of 15,427 bp and is organized in a circular doublestrand molecule. All 37 mitochondrial genes are present, including the common set of 22 tRNAs. Particularly noticeable is a duplication of the control region. The additional control region is located between *nad6* and *cob* and is almost identical to the original one. The most extensive changes in gene order affect *nad5* and a block consisting of *trnH, nad4, nad4L,* and *trnP*—all were inserted near the original control region. The gene *nad5* is also inverted. Furthermore, the comparison to the pancrustacean ground pattern reveals additional changes of individual tRNA genes. Some of these changes are also shared by *Metacrangonxy longipes* and by *Parhyale hawaiensis*. These arrangements were found only in amphipods and might be considered as apomorphic character states of Amphipoda. In all three species there is good evidence that *trnG* originated from a rare duplication/remolding event of the adjacent *trnW* gene. In addition to these common features all of the three available amphipod mt genome sequences bear unique rearrangements. *Caprella mutica*, however, shows the most extensive rearrangement in comparison to the pancrustacean ground pattern.

5.2 Introduction

The mitochondrial (mt) DNA has become an important source of information for a variety of studies on populations, phylogeography, and phylogeny of animals in the last years. This may be due to certain advantages of these small extranuclear genomes (~16 kb), which are among others: exclusive maternal inheritance, clear orthology of genes, and the lack of introns. (Avise 2000; Boore 1999; Lavrov 2007; Wolstenholme 1992). In addition, the nucleotide sequence shows a higher rate of evolution compared to the nuclear genome, whereas the sequence variation itself is generally assumed to be neutral to selection. A circular double-strand organization and a slightly different genetic code are additional special features. Although there are exceptions to all of this rules, mtDNA is still a valuable marker for population genetics and phylogenetics (Rubinoff and Holland 2005).

The arrangement of genes shows a very economical organization, leaving almost no non-coding sequence between adjacent genes, while small gene overlaps occur quite frequently. During the evolution of endosymbiotic bacteria to mitochondria, most of the genes were lost or transferred to the nucleus. In bilaterian animals a standard set of only 37 genes (2 rRNA genes, 13 protein-coding genes, 22 tRNA genes) and one control region (CR) is retained in the mitochondrial genome (Boore 1999). However, exceptions are known, e.g. from the mt genomes of Chaetognatha, which lack all but one tRNA, as well as the protein coding genes *atp6* and *atp8* (Helfenbein et al. 2004; Papillon et al. 2004). Generally, mt genomes feature a relative conserved gene order. As there is virtually a limited amount of non-coding or non-vital sequence, rearrangements are likely to disrupt other genes. Identical gene arrangements can be retained for a long time, as found in distantly related species, e.g. among vertebrates hagfishes and mammals have the same mt gene order.

The pancrustacean ground pattern is similarly present in many different species from Crustacea and Hexapoda (Kilpert and Podsiadlowski 2006). It differs from the arthropod ground pattern in the relative position or a single gene (*trnL2*) (Boore and Brown 2000; Boore, Lavrov, and Brown 1998). In Peracarida this ancestral state was modified in different ways. Isopoda show other changes in gene order (Kilpert and Podsiadlowski 2009) than the two available mt genomes from gammarid amphipods *Parhyale hawaiensis* (Cook, Yue, and Akam 2005) and *Metacrangonxy longipes*

(Bauza-Ribot et al. 2009). None of the detected rearrangements seems to be common to both, Amphipoda and Isopoda. Mt genomes from other peracarid taxa are so far unavailable.

Here we present the complete mt genome of *Caprella mutica*. It is the first complete dataset of a member of the Caprellidea, a suborder of Amphipoda (Crustacea, Malacostraca, Peracarida) with more than 300 classified species (8 suborders; Laubitz 1993; Martin and Davis 2001). The phylogeny of Caprellidea is not settled jet; historically *Caprella mutica* is belonging to the Caprellidae (skeleton shrimps). Species of this group can be found in benthic marine enviroments. Most of them use other organisms as substrata like hydroids, sponges, and seaweed. *Caprella mutica* is indigenous to sub-boreal waters of north-east asia, but has recently gained attention as one of the most rapidly invading marine species, meanwhile found at the coasts of North America, Europe and New Zealand (Ashton et al. 2007).

5.3 Material and Methods

Sample and DNA extraction

A single specimen of *Caprella mutica* preserved in 99% ethanol was used for DNA extraction. Only legs (peraeopods) were applied to the DNeasy Tissue kit (Qiagen, Hilden, Germany) following the manufacturer's protocol for total genomic DNA extraction. The specimen was collected at the coast of the North Sea island Helgoland in spring 2006.

mtDNA amplification and processing

We successfully amplified the almost complete mt genome of *Caprella mutica* by a single long range PCR. The whole procedure was inspired by a paper describing one-step PCR amplification of complete mitogenomes in arthropods (Hwang et al. 2001). We used a combination of the primers Isop-16sf and Isopd-16sr, which already proved their usefulness for long range PCRs of an isopod species (Kilpert and Podsiadlowski

2009). The generated PCR product had a size of about 15 kb, indicated by gel electrophoresis. It was amplified using the TaKaRa LA Taq kit (Takara Bio Inc., distributed by MoBiTec, Göttingen, Germany). The reaction volumes amounted 25µl, each consisting of 2.5 µl 10× Takara LA PCR buffer, 4 µl Takara dNTP mixture (2.5 mM each), 0.25 µl Takara LA Taq polymerase (5 units/µl), 16.75 µl sterile distilled water (Bio Mol grade, 5Prime), 0.5 µl primer mixture (10 µM each; Metabion, München,Germany) and 1 µl DNA template. The thermal cycling protocol was setup with an initial denaturation step (94 °C, 1 min) followed by 30 cycles each comprising denaturation (98 °C, 15 s), annealing (58 °C, 45 s) and extension (68 °C, 14 min). The cycling was completed by a final extension (72 °C, 10 min). Eppendorf Mastercycler and an Eppendorf Mastercycler Gradient (Eppendorf, Hamburg, Germany) were used for all PCRs of this study. Clean-up of PCR products was accomplished by silica-membrane spin columns (NucleoSpin Extract II, Macherey-Nagel, Dueren, Germany); Storage temperature for purified PCR products was always -20° C.

The sequencing of this large PCR product generally followed the primer walking strategy. A large number of additional primers helped to advance the sequencing process: We used decapod, peracarid and isopod primers we built in the past based on conserved segments from single-gene alignments (Kilpert and Podsiadlowski 2009; Podsiadlowski and Bartolomaeus 2005). The insect based primers 16S2, N4, and N4(87) were proven to be useful as well (Roehrdanz, Degrugillier, and Black 2002), likewise several primers of a set of primers, which was originally designed for decapods crustaceans (Yamauchi et al. 2004). The complete list of primers successfully used is provided in Additional File 5.1.

The two areas that turned out to contain control regions were hard to sequence and were verified by nested PCRs with the initial long-range PCR product as template. The primers pair Cm-6441/Cm-7613 (1173 bp) and Cm-464/Cm-14669 (1224 bp) were used for these amplifications.

Sequencing

Sequencing was performed on a CEQ 8000 capillary sequencer (Beckman Coulter, Krefeld, Germany), provided with the CEQ software package (instrument version:

6.0.2, software version: 5.0.360) for operation and analysis. For sequencing reactions we used the CEQ DTCS Quick Start Kit (Beckman Coulter) according to the manufacturer's protocol. The thermal cycling comprised 30 cycles (20 sec at 94° C, 20 sec at 45-58° C (primer specific), 4 min at 60° C). The reactions were cleaned-up by three ethanol washing/precipitation steps, subsequent vacuum drying and finally solved in sample loading solution. After we had complete sequencing of the mt genome of *Caprella mutica* by primer walking, we got the opportunity to verify the results by next generation sequencing. Thus, the PCR product from the first long-range amplification was introduced to the commercial 454 service provided by AGOWA (Berlin, Germany). We identified four contig sequences (from 1209 single reads) which covered almost the entire mt genome: A (457 reads, covering bp 22-6605), B (100 reads, covering bp 6606-7290), C (169 reads, covering 7291-9704), D (383 reads, covering bp 10021-14764).

Gene annotation and sequence analysis

BioEdit 7.0.9.0 (Hall 1999) was used to align the complete mt genome from single DNA sequences. The complete mtDNA sequence of *Caprella mutica* was deposited at GenBank (NCBI, GU130250). Primers were designed to allow a fragment overlap of at least 50 bp. Every part of the mt genome was covered by at least two sequences to allow a detection of sequencing errors. In addition, there are four large contigs from 454 sequencing covering 93% of the complete genome. Therefore, the coverage of two independent sequencing methods ensures a high sequence quality. Two computer programs, tRNAscan-SE 1.21 (Lowe and Eddy 1997) and ARWEN 1.2.3.c (Laslett and Canback 2008) were used to identify tRNAs by their secondary structure. Non-coding areas were manually checked for tRNA genes as well. Because identification of tRNA genes is anything but trivial and single genes are easily overlooked, we spent utmost care on this topic (see Additional File 5.2). Other genes were identified by BLAST search on GenBank databases and by comparison to the mitochondrial genome of *Drosophila yakuba* (NC_001322). Gene borders of protein-coding genes were determined by comparison to alignments of mitochondrial genes from several crustacean species. Gene boundaries were assumed to be the first start and stop codons in frame. rRNA genes are assumed to extend up to adjacent genes, due to the lack of

better resources to determine gene boundaries. The DAMBE 5.1.1 software package (Xia and Xie 2001) was used to calculate nucleotide frequencies.

5.4 Results and Discussion

Genome organization

The mitochondrial genome of *Caprella mutica* is a circular doublestrand molecule of 15,427 bp length (Figure 5.1, GenBank: GU130250). The ring shape topology is assumed due to the results of the sequencing of overlapping PCR fragments. All 37 genes (2 rRNA genes, 13 protein-coding genes, and probably 22 tRNA genes), which are usually present in metazoan mt genomes were found. The genes are encoded on both strands and often overlap by a few nucleotides (Table 5.1); non-coding sequences are rare. Unusual, however, is the existence of a two major non-coding regions (control regions) instead of one. The highly similar nucleotide sequences suggest these were subject to a recent duplication event, because there is no significant amount of substitutions detectable. Normally the major non-coding regions of mt genomes are among the most variable regions, often useful markers in population genetics. . The relative positions of the major non-coding regions in the mt genome suggest that CR2 between *trnP* and *trnI* is the original one, as it is found in a relative position comparable to that of other Crustceans, while and CR2 between *nad6* and *trnC* seems to be the duplicate. Besides this duplication a number of gene translocation events also happened (see below).

Strand bias

It has been shown that the frequency of complementary nucleotides, most notably of G and C, differs between both strands in mtDNA (Perna and Kocher 1995). In malacostracan crustaceans there is regularly more C than G in +strand genes (negative GC skew) and more G than C in –strand genes (positive GC skew) (Hassanin 2006). This skew is probably caused by the asymmetrical replication process of the mt genome, whereby each mtDNA strand is exposed to different mutational pressure,

dependent from the direction of replication. It is therefore an indicator for the orientation of the control region (Hassanin, Leger, and Deutsch 2005). *Caprella mutica* keeps the usual arthropod GC skews (negative GC skew for +strand genes, positive GC skew for –strand genes) (Additional File 5.3). The gammarid amphipods *Metacrangonyx longipes* (Bauza-Ribot et al. 2009) and *Parhyale hawaiensis* (Cook, Yue, and Akam 2005) share these regular GC skews. This is an important difference to the mt genomes of Isopoda, which show a reversed GC skew (Kilpert and Podsiadlowski 2006; Kilpert and Podsiadlowski 2009; Podsiadlowski and Bartolomaeus 2006).

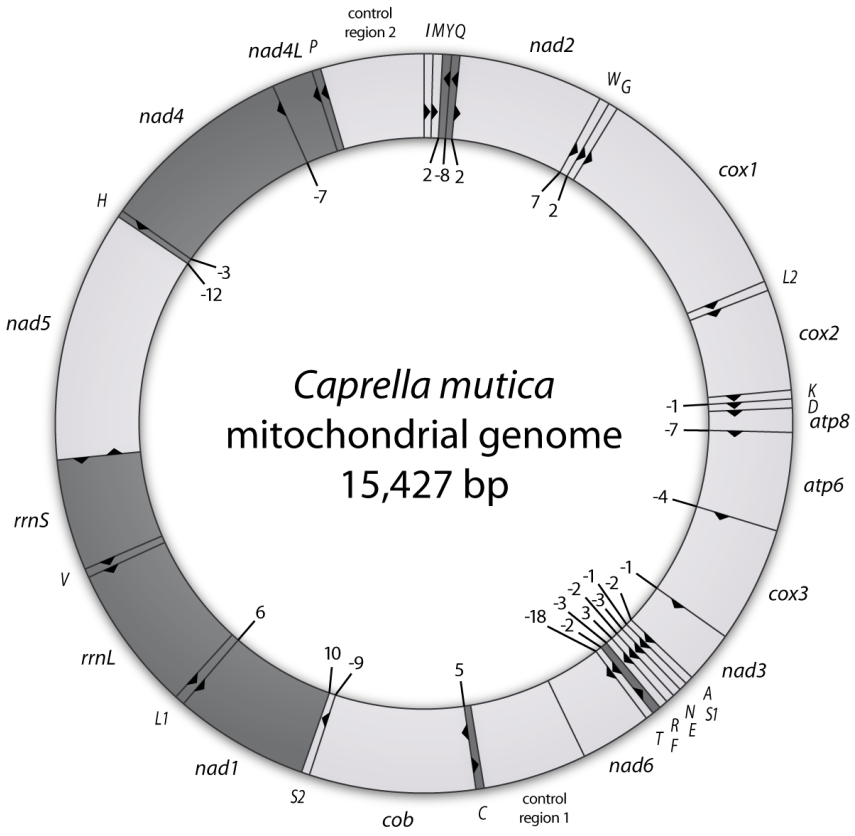


Figure 5.1. Map of the complete mitochondrial genome of *Caprella mutica* (Amphipoda: Caprellidae). Transfer-RNAs are represented by their one-letter amino acid code. Numbers specify the length of non-coding sequences (positive values) or the extent of gene overlaps (negative values). Arrows pointing clockwise indicate +strand genes; counter-clockwise arrows indicate -strand genes. The latter are shaded dark additionally.

Table 5.1. Gene content of the complete mitochondrial genome of *Caprella mutica*. The mtDNA is a circular molecule.

Gene	Strand	GenBank position no.	Size (nts)	Size (aa)	Start codon	Stop codon	Intergenic nucleotides
trnI	+	1-62	62				0
trnM	+	63-124	62				2
trnY	-	127-186	60				-8
trnQ	-	179-242	64				2
nad2	+	245-1222	978	325	ATT	TAA	7
trnW	+	1230-1289	60				2
trnG	+	1292-1353	62				0
cox1	+	1354-2890	1537	512	ATT	T	0
trnL(UUR)	+	2891-2951	61				0
cox2	+	2952-3633	682	227	ATA	T	0
trnK	+	3634-3696	63				-1
trnD	+	3696-3756	61				0
atp8	+	3757-3915	159	52	ATA	TAA	-7
apt6	+	3909-4580	672	223	ATG	TAA	-4
cox3	+	4577-5365	789	262	ATA	TAA	-1
nad3	+	5365-5715	351	116	ATG	TAA	-2
trnA	+	5714-5774	61				-1
trnS(AGN)	+	5774-5829	56				-3
trnN	+	5827-5889	63				-2
trnE	+	5885-5946	59				3
trnR	+	5950-6009	60				-3
trnF	-	6007-6066	60				-2
trnT	+	6065-6123	59				-18
nad6	+	6106-6606	501	166	ATT	TAA	*
non-cod. 1		6607-7311	705				*
trnC	-	7312-7367	56				5
cob	+	7373-8488	1116	371	ATG	TAA	-9
trnS(UCN)	+	8480-8544	65				10
nad1	-	8555-9457	903	300	ATA	TAA	6
trnL(CUN)	-	9464-9526	63				*
rrnL	-	9527-10516	990				*
trnV	-	10517-10574	58				*
rrnS	-	10575-11326	752				*
nad5	+	11327-13027	1701	566	ATA	TAA	-12
trnH	-	13016-13075	60				-3
nad4	-	13073-14390	1318	439	ATG	T	-7
nad4L	-	14384-14668	285	94	TTG	TAA	0
trnP	-	14669-14728	60				*
non-cod. 2		14729-15427	699				*

* Gene borders determined by borders of adjacent genes

Protein-coding genes

The A+T content of the protein-coding genes amounts 67.7% (total genome: 68.0%; Table 5.2), which is within the usual 60-70% range of malacostracan mt genomes. The regular start codons ATA, ATT, and ATG can be found at all but one protein-coding

gene; only *nad4L* is supposed to start with a TTG (Table 5.1). This start codon is not regularly found in arthropods, but was at least suggested as start codon in an isopod species (Kilpert and Podsiadlowski 2009). Most genes uniformly end by the stop codon TAA. However, the three genes *cox1*, *cox2*, and *nad4* show abbreviated stop codons, which consist of a single T. It is generally thought that these shortened stop codons are completed by post-transcriptional polyadenylation to a functional UAA in the transcript (Ojala, Montoya, and Attardi 1981). Abbreviated stop codons are often reported from mt genomes of various species, including the amphipods *Parhyale hawaiensis* (Cook, Yue, and Akam 2005) and *Metacrangonyx longipes* (Bauza-Ribot et al. 2009).

Table 5.2. Nucleotide frequencies, AT content, AT- and GC-skew for mitochondrial genes of *Caprella mutica*.

Gene(strand)	Nucleotide frequency				%AT	AT skew	GC skew
	A	C	G	T			
atp6(+)	0.298	0.223	0.121	0.359	65.6	-0.09	-0.30
atp8(+)	0.453	0.151	0.057	0.340	79.2	0.14	-0.45
cob(+)	0.280	0.217	0.135	0.368	64.8	-0.14	-0.23
cox1(+)	0.260	0.201	0.176	0.362	62.3	-0.16	-0.07
cox2(+)	0.318	0.199	0.145	0.337	65.5	-0.03	-0.16
cox3(+)	0.278	0.197	0.165	0.361	63.9	-0.13	-0.09
nad1(-)	0.233	0.145	0.207	0.415	64.8	-0.28	**0.18**
nad2(+)	0.305	0.186	0.112	0.398	70.3	-0.13	-0.25
nad3(+)	0.288	0.168	0.134	0.410	69.8	-0.18	-0.11
nad4(-)	0.250	0.131	0.208	0.411	66.1	-0.24	**0.23**
nad4L(-)	0.246	0.098	0.165	0.491	73.7	-0.33	**0.25**
nad5(+)	0.276	0.202	0.145	0.377	65.3	-0.15	-0.17
nad6(+)	0.321	0.198	0.114	0.367	68.9	-0.07	-0.27
prot. cod. total	0.293	0.178	0.145	0.384	67.7	-0.14	-0.11
rrnL(-)	0.332	0.116	0.171	0.381	71.3	-0.07	**0.19**
rrnS(-)	0.358	0.112	0.153	0.378	73.5	-0.03	**0.16**
rRNA total	0.345	0.114	0.162	0.379	72.4	-0.05	**0.17**
non-cod. 1	0.355	0.136	0.112	0.397	75.2	-0.06	-0.10
non-cod. 2	0.342	0.142	0.119	0.398	74.0	-0.08	-0.09
total mt genome	0.332	0.188	0.133	0.348	68.0	-0.02	-0.17

rRNA genes

The size of *rrnL* (16S), located on the negative strand, has a length of only 990 bp. It is therefore significantly shorter than available *rrnL* genes from related peracarid species: *Metacrangonyx longipes* (Amphipoda): 1137 bp, *Eophreatoicus* sp.-14 (Isopoda): 1224, *Ligia oceanica* (Isopoda): 1234 bp, *Idotea balthica* (Isopoda): 1216 bp. Although a

number of nucleotides are missing at the 3'-end of the gene, as shown by an alignment of these *rrnL* genes, a general loss of function is not probable.

Transfer RNAs

The regular set of 22 tRNA genes was identified in the mt genome of *Caprella mutica* (Table 5.1). With the exception of the two isoaccepting *trnS* genes, which were found by manual inspection of remaining non-coding regions, all tRNA genes were detected by the utilized computer software (Additional File 5.2). The secondary structure plots reveal that most tRNAs feature a usual t-shape or clover leaf secondary structure (Figure 5.2). Deviations from this pattern were found in tRNA-Cys, tRNA-Glu, and tRNA-Thr, where the TΨC-arm is missing. tRNA-Ser(AGN) on the other hand has no DHU-Arm. An elongation of the anticodon loop was found in tRNA-Ser(UCN). Aberrant secondary structures are also known from other related amphipod or peracarid species: In isopods tRNA-Cys is lacking the DHU-arm as well (Kilpert and Podsiadlowski 2009). The same arm shows very bad parings in tRNA-Ser(AGN) of *Metacrangonyx longipes* (Amphipoda) (Bauza-Ribot et al. 2009) and is also missing in isopod species. This is similar with the TΨC-arm of tRNA-Thr, which is missing in *Metacrangonyx longipes* (Amphipoda) and is most likely also absent in *Ligia oceanica* (Isopoda).

There is evidence for a remolding event of two adjacent tRNA genes. The high sequence similarity of *trnW* and *trnG* (74.2 %) probably resulted from a duplication of *trnW* followed by several nucleotide mutations, which changed the identity to a functional *trnG* gene. Similar remolding events were reported only from isoaccepting *trnL* genes so far (Kilpert and Podsiadlowski 2009; Rawlings, Collins, and Bieler 2003). A complete identity change, however, requires not only mutations in the anticodon, but also in the accepting aminoacetyl tRNA synthetase . This kind of remolding is probably less frequent, due to the higher degree of complexity.

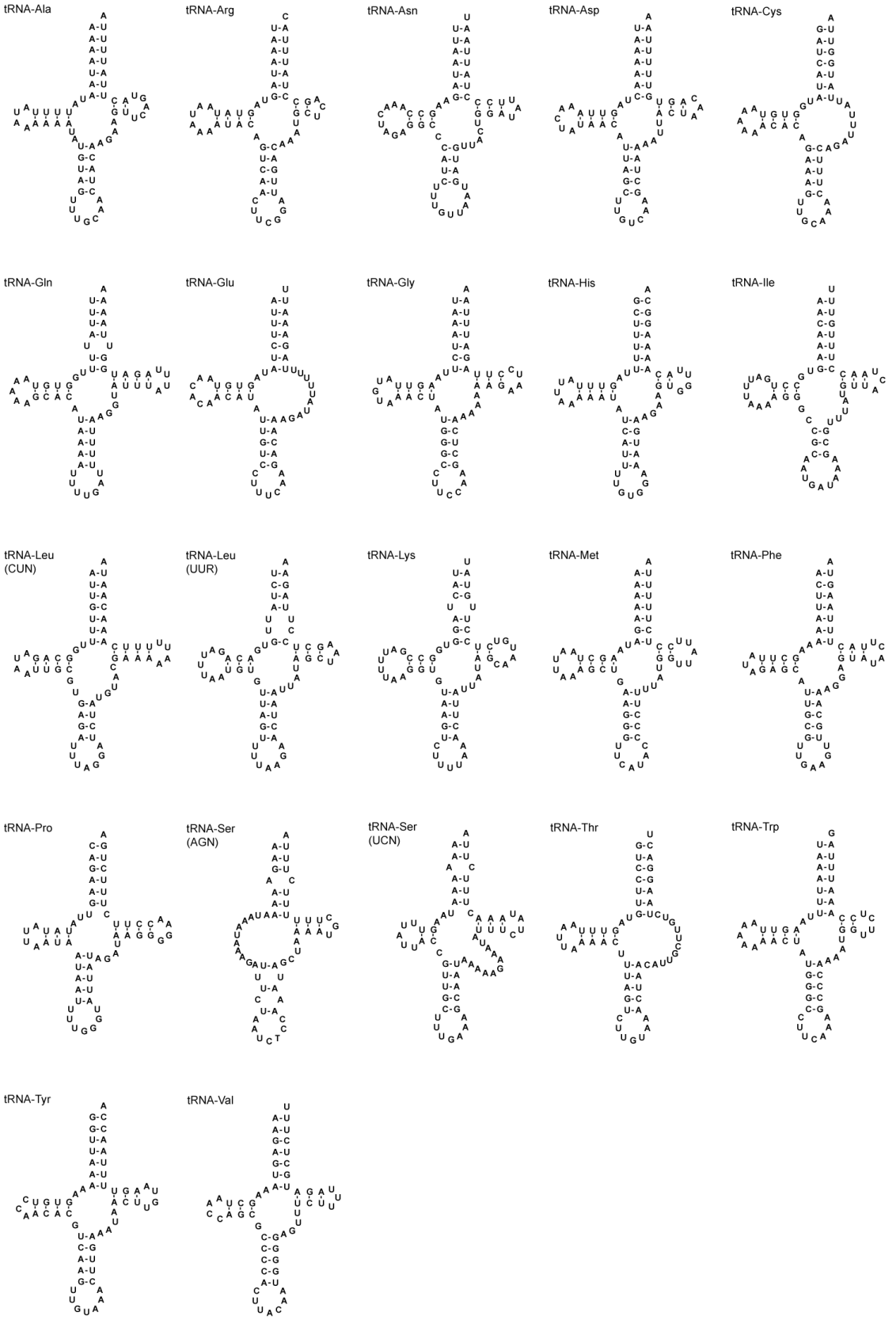


Figure 5.2. Secondary structure plots of tRNAs of *Caprella mutica*. Most tRNAs feature a typical clover-leaf structure. Exceptions: DHU-arm missing in tRNA-Cys, tRNA-Glu, and tRNA-Thr. TΨC-arm missing in tRNA-Ser(AGN).

Control region duplication and repeats

The mt genome of *Caprella mutica* features two control regions. As the relative positions suggest (see next paragraph), CR1 (705 bp; located between *nad6* and *trnC*) most likely is a copy of CR2 (699 bp; located between *trnP* and *trnI*). The nucleotide sequence is practically identical (99.1%), except for additional sequence at the 3'-end of CR1 (41) and the 5'-end of CR2 (35 bp; Additional File 5.3). The occurrence of another control region is quite uncommon in mt genomes, as usually one control region bears everything needed for replication and transcription. It seems likely that the duplicates had no time to accumulate a significant amount of mutations or deletions. Therefore, it seems reasonable to assume very recent duplication event.. Further studies will show if a duplication of the control region also occurs in other caprellid amphipods. Other completely sequenced mt genomes from Peracarida are only known from Isopoda. The analyzed species show only one control region, although it is likely that the derived relative position of the isopod control region is the result of a duplication/random-loss event (Kilpert and Podsiadlowski 2006; Kilpert and Podsiadlowski 2009), which is generally assumed as the basic mechanism of gene rearrangements (Dowton, Castro, and Austin 2002). Duplications of the control region are occasionally reported, mainly from vertebrates (Amer and Kumazawa 2005; Eberhard, Wright, and Bermingham 2001; Gibb et al. 2007). Within the control region there is a 69 bp sequence, which—except for one nucleotide—is inversely repeated starting 56 bp downstream (Additional File 5.3). Together they may form a huge hairpin with a stem of about 70 bp length and a loop of 56 bp. However, typical control region hairpins are significantly smaller (Kilpert and Podsiadlowski 2006). A part of the inverse repeat sequence (28 bp) is present another time in the control region, but with no counterpart to form a second hairpin.

Mt genome rearrangements

The mt genome of *Caprella mutica* differs significantly from the gene order of the pancrustacean ground pattern and also from the other available amphipod species (Figure 5.3). The pancrustacean ground pattern (Lavrov, Boore, and Brown 2000) is still realized in many crustacean and hexapod species (Kilpert and Podsiadlowski 2006). A striking difference of *Caprella mutica* is the already mentioned duplication of the

control region. The comparison to the ground pattern indicates that the position of CR2 in *Caprella mutica* corresponds to the control region of the ground pattern. CR1 probably is a duplication of CR2 that was inserted between *nad6* and *cob*. The other striking difference is a derived relative position of *nad5* and a block of *trnH*, *nad4*, *nad4L* and *trnP*. Furthermore, *nad5* must have switched strands in comparison to the pancrustacean ground pattern. It is conceivable that the whole area ranging from *nad5* to *trnP* was translocated as a whole and *nad5* was inverted in place afterwards. Generally, the large number of potential gene arrangements makes it very unlikely that genes stay adjacent when moved by independent translocation events (Dowton, Castro, and Austin 2002). For *nad5* a separate translocation followed by an inversion is the most parsimonious explanation. Generally, the best founded mechanistic explanation for mt gene translocations is the traditional duplication-random-loss model (Moritz, Dowling, and Brown 1987). However, large scale duplications followed by random losses, is not the first choice to parsimoniously generate the observed gene order of *nad5* and the *trnH*/*nad4*/*nad4L*/*trnP* block. Intramitochondrial recombination may be more likely to move a block of genes around the mt genome (Dowton, Castro, and Austin 2002).

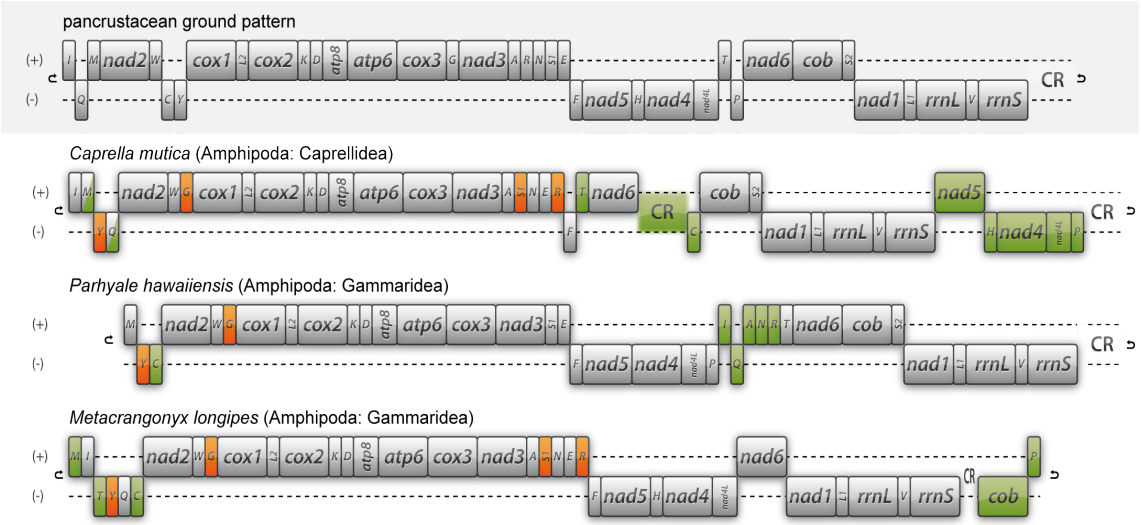


Figure 5.3. Gene arrangements of Amphipoda. *Caprella mutica* (this study) is compared to *Parhyale hawaiensis* (FM957525, FM957526, Bauza-Ribot, Jaume, Juan, and Pons 2009; AY639937, Cook, Yue, and Akam 2005), and *Metacrangonyx longipes* (NC_013032, Bauza-Ribot, Jaume, Juan, and Pons 2009). Colored genes indicate changes in comparison to the ground pattern of Pancrustacea (ancestral state). Green colored genes signal changes in individual species only; orange colored genes share derived positions.

A clear reference about the mechanism that caused the derived position of *trnG* can be found in all available amphipod mt genomes. In *Caprella mutica* high sequence similarity of *trnW* and *trnG* (74.2 %) suggest that *trnG* probably emerged from a duplication of *trnW*. The pancrustacean ground pattern shows that *trnW* stayed in its original position, while the position of *trnG* is new. The same high similarity of both tRNA genes is incidentally present *Metacrangonyx longipes* (76,6%; NC_013032) and *Parhyale hawaiensis* (67,2%; AY639937) as well. For the latter, a verification of tRNAs with ARWEN lead to an update of the annotation: *trnC* (in AY639937) is in fact *trnW;* and *trnW* is *trnG*. Therefore, all amphipods, including *Parhyale hawaiensis*, share the derived position of *trnG*, which apparently is the result from a duplication-random-loss event that was followed by a remolding of the tRNA. For isoaccepting *trnL* genes it seems to happen every now and then at least in some lineages (Rawlings, Collins, and Bieler 2003), but shifts into different tRNAs, which also require changes of the acceptor stem, seem to occur much more rarely (Lavrov and Lang 2005). The derived location of *trnG* is an unequivocal apomorphic character of the Amphipoda, as it is no longer found between *cox3* and *nad3*. In isopods, the only other peracarids with available mt genomes, *trnG* was not translocated (Kilpert and Podsiadlowski 2006; Kilpert and Podsiadlowski 2009).

Other genes with derived relative positions in *Caprella mutica* are *trnY* (beside *trnQ*) and *trnS1* (between *trnA* and *trnN*), and *trnR* (between *trnE* and *trnR)* (Figure 5.3). These genes are also translocated in *Metacrangonxy longipes* as well. In *Parhyale hawaiensis* the relative postions of *trnY* and *trnR* are also derived, but in comparison to the other two amphipods, additional changes are required to explain the situation. The region between the control region and *nad2*, which contains multiple tRNA genes, shows an increased variability in all amphipods. Even the number of tRNA genes varies significantly: *Caprella mutica* (4 tRNA genes), *Parhyale hawaiensis* (3 tRNA genes), *Metacrangonxy longipes* (7 tRNA genes). However, there is higher similarity in gene order between the two gammarid amphipods (*Metacrangonxy longipes, Parhyale hawaiensis)* than between these and to the caprellid amphipod *Caprella mutica*. In *Parhyale hawaiensis* changes are limited to rearranged tRNAs, mainly located in an area between *trnP* and *trnT*. In *Metacrangonxy longipes* most significant rearrangements can be found between the CR and *nad2* (even involving *cob*). But most significant changes were found in *Caprella mutica*, which not only feature derived

tRNA gene orders but also a duplication of the CR and of protein-coding genes (*nad4*, *nad4L* and *nad5*).

5.5 Acknowledgements

The authors wish to thank Prof. Thomas Bartolomaeus for kindly supporting all phases of this study at the FU Berlin as well as at the Univ. of Bonn. We are grateful to Prof. Heinz-Dieter Franke (Biologische Anstalt Helgoland) and Achim Meyer (Univ. Mainz) for sending specimens of *Caprella mutica*. We also thank Maria Orland (Uni v. Bonn) for technical assistance in the molecular laboratory. This study was supported by a NaFöG Berlin grant to FK and by a German science foundation (DFG) grant (Po 765/2-1) to LP.

5.6 Declaration of interest

The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper. F. Kilpert did the majority of laboratory work and of manuscript writing. L. Podsiadlowski was supervisor and editor of the manuscript

5.7 Additional material

Additional Files 5.1 to 5.3 are provided with the online version.

The files can be found in the appendix of this thesis (chapter 11).

6. Insights from the mitochondrial genome of *Leucon nasica* (Cumacea) – Implications for peracarid phylogeny

6.1 General remarks and background

This chapter presents first results from the ongoing sequencing process of the mitochondrial genome of *Leucon nasica*. About half of the mt genome has been determined so far. Therefore, our data represents the most extensive dataset for a cumacean species right now. On Genbank there are only single mt gene sequences (*cox1, rrnS*) available of different cumacean species, whereby the majority of them is far from complete. Although the sequencing of the *Leucon nasica* mt genome is not finished yet, the already determined gene order is meaningful enough to allow serious inferences on peracarid phylogeny and the relationship of Cumacea to Amphipoda and Isopoda.

Cumacean species inhabit marine benthic sediments all over the world. They are particularly frequent in the deep sea, but are also found in the littoral zone. The normal size of these small peracarids is about 5-10 mm. The habitus is characterized by a balloon-shaped carapax, which regularly covers the first 3-6 thoracomers, and a long and thin pleon. The Cumacea comprise about thousand cumacean species, currently divided in eight subclades (Martin and Davis, 2001).

6.2 Material and methods

As the sequencing of *Leucon nasica* is work in progress, as mentioned above, it might be appropriate to briefly refer to the materials and methods sections of the previous chapters (Chapter 2-5). The mt genome of *Leucon nasica* was gained pretty much by the same techniques and protocols like the mt genomes of the isopods and the amphipod

species. Detailed information on each process can be looked up there; exceptions and particularities for *Leucon nasica* are listed here.

The specimen of *Leucon nasica* (8 mm length) was collected in Bergen (Norway) and preserved in 99% ethanol. We used a complete pleon for DNA extraction with the DNeasy Tissue kit. It was possible to amplify the whole sequence, which is available by now, with primer pair L12167-16S and H4244-CO3 (Yamauchi *et al.*, 2004) in a long range PCR. This fragment was then sequenced by primer walking using species specific primers (available in upcoming publication). For sequencing we relied on a commercial sequencing service (AGOWA, Berlin, Germany). Annotation and analysis were accomplished to the proven standards as before.

6.3 Results and discussion

A part of the mt genome of *Leucon nasica* with a size of 6216 bp was sequenced. The annotation revealed one rRNA gene (*rrnL*), six protein-coding genes (*atp6, atp8, cox1, cox2, cox3, nad2*), and seven tRNA genes (*trnD, trnK, trnL1, trnL2, trnM, trnQ, trnV*) in the linear fragment (Table 6.1). All tRNAs examined so far show a regular t-shape (clover-leaf) secondary structure (not shown). An unusually large non-coding sequence of 134 bp was detected between *trnM* and *nad2*. Apart from that only small non-coding regions (up to 3 bp) exist between genes as well as small gene overlaps (up to 7 bp), which are commonly found in mt genomes. Exceptional start codons were detected for the three protein coding genes: *atp8* (TTG), *cox1* (GCG), *cox2* (GTT). Usual start codons do not exist near the start of the genes being preset by multiple sequence alignments of other crustaceans. This indicates that exceptional start codons also occur in cumacean mt genomes, similar to those in amphipods and isopods. An incomplete stop codon terminates *atp6*. These abbreviated stop codons (Ojala *et al.*, 1981) have been reported from every peracarid mt genome. The partial mt genome generally features more C than G nucleotides for the +strand (negative GC-skew) and an inverse relationship for the –strand (positive GC-skew). The overall +strand GC skew of the sequenced fragment is -0.08. This slightly negative value corresponds to the generally negative GC skew of crustacean species, which was maintained in amphipods, but not

in most isopods. As the GC skew allows conclusions on the orientation of the control region (Hassanin *et al.*, 2005), there is currently no indication for an inversion in peracarids before the origin of isopods.

The gene order of the sequenced fragment, ranging from *rrnL* to *cox3*, shows some significant differences in comparison to the pancrustacean ground pattern (putative ancestral state) (Figure 6.1). The loss of a section containing *rrnS*, CR, and *trnI* is certainly most noticeable among them. As a result, *trnV* is located adjacent to *trnQ* in *Leucon nasica*, which is the exactly the same state as assumed for the ground pattern of isopods! A translocation of an rRNA gene or even the control region is anything but frequent. The commonly shared gene order therefore is a strong indication for a shared origin of this character in isopod and cumacean species. Convergence in derived gene order, especially of multiple genes, is unlikely, due to the enormous number of alternative gene orders (Dowton *et al.*, 2002).

Although the new relative position of *rrnS*, CR, and *trnI* are unknown by now, as they are probably located in the by now unsequenced part of the mt genome, the cumacean species *Leucon nasica* seems to share one of the most distinctive isopod mt gene rearrangements. This shared apomorphic character (synapomorphy) qualifies the Cumacea as a potential isopod sister group in the Peracarida tree. Amphipod species retained the original position of the control region of the pancrustacean ground pattern. Some species even show additional genes between *trnV* and *trnQ* (Chapter 5). Therefore, Amphipoda are certainly not the sister group of the Isopoda. For Tanaidacea there is no data available as it has still not yet been possible to sequence multigene fragments of the mt genome.

An interesting difference in *Leucon nasica* is the loss of tRNAs gene between *nad2* and *cox1*. Usually this position always holds multiple tRNA genes. In the pancrustacean ground pattern there are three tRNA genes (*trnW, trnC, trnY*) of which two remained in isopods (*trnC, trnY*). The Amphipoda show here always *trnW* and *trnG*, in which *trnG* probably originated from a duplication/remolding of *trnW* (Chapter 5). The fact that no tRNA genes are located between *nad2* and *cox1* therefore is an exception among Peracarida. It could be another good synapomorphic character for Cumacea, if it turns out that tRNA genes are similarly absent in other species of the group.

Table 6.1. Gene content of the partial mt genome of *Leucon nasica* (Cumacea)

Gene	Strand	GenBank position no.	Size (nts)	Size (aa)	Start codon	Stop codon	Intergenic nucleotides
rrnL[a]	-	1-1017[a]	1017[a]				*
trnV	-	1018-1080	63				-3
trnQ	-	1078-1139	62				-2
trnM	+	1138-1201	64				134
nad2	+	1336-2304	969	322	ATT	TAA	1
cox1	+	2306-3841	1536	511	GCG	TAA	3
trnD	+	3845-3904	60				0
trnL(UUR)	+	3905-3964	60				3
trnL(CUN)	+	3968-4027	60				0
cox2	+	4028-4693	666	221	GTT	TAA	1
trnK	+	4695-4757	63				0
atp8	+	4758-4907	150	49	TTG	TAA	-7
atp6	+	4901-5564	664	221	ATG	T	0
cox3[a]	+	5565-6216[a]	652[a]	[a]	ATG	[a]	

* Gene borders determined by borders of adjacent genes

[a] Incomplete gene

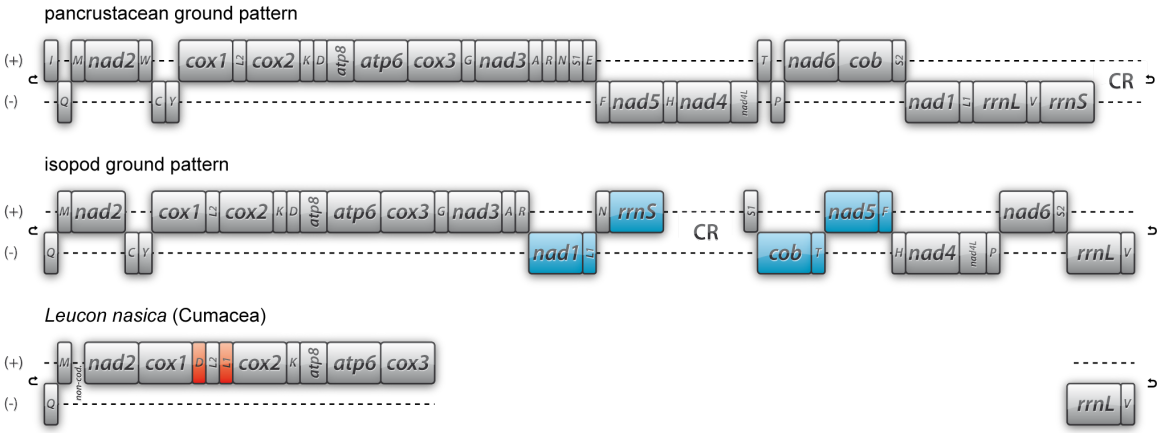


Figure 6.1. The partial mt gene arrangement of *Leucon nasica* (Cumacea) compared to the putative ancestral state (pancrustacean ground pattern) and the isopod ground pattern. rrnS, CR, and trnI were translocated in *Leucon nasica*, similar to Isopoda. Derived relative position of genes color: blue (Isopoda), red (Cumacea).

Another derived position was detected for *trnD*, which is located between *cox1* and *trnL2* in *Leucon nasica*. This relative position was never reported for any crustacean species (see chapter 2: figures 6 and 8; Kilpert and Podsiadlowski, 2006). In the pancrustacean ground pattern and also in all known amphipod and isopod species *trnD* is always located between *cox2* and *trnK*.

The gene position of *trnL1* (between *trnL2* and *cox2*) is probably attributable to a duplication/remolding event of the adjacent *trnL2* gene (Rawlings *et al.*, 2003). The sequence similarity of both *trnL* genes amounts 62.9%; therewith it is much higher than chance similarity of unrelated genes. Similar translocation events of *trnL1* caused by the same mechanism are also known from phreatoicidean and asellotan isopods (Chapters 3 and 4). However, *trnL1* is located on the other side of *trnL2* next to *cox1* in two of the isopod species. Altogether, it seems like translocation events of *trnL1* caused by duplication and remolding might be more frequent than most other translocation events. The shared position of *trnL1* in the particularly derived mt genome of *Janira maculosa* and in the cumacean *Leucon nasica* is quite likely a result of homoplasy.

6.4 Acknowledgements

Prof. Dr. Stefan Richter (Univ. Rostock) provided several species for analysis, among them *Leucon nasica*. Thank you very much.

7. Concluding discussion

Concluding remarks on the questions posed in the introduction:

(1) What kinds of characters of 'genome morphology' actually exist in the examined mt genomes? Are these characters valuable for phylogenetic inference? Do they occur frequently enough to establish a phylogeny apart from sequence data?

The comparison of isopod, amphipod, and cumacean mt genomes (Chapters 2 to 6), including new data from ten species, revealed variation in several structural and organizational characters, referred to as characters of 'genome morphology'. In Isopoda, the taxon for which there is currently most data (Chapters 2 to 4), extensive rearrangements of the mt gene order were determined. In fact, none of the nine isopod species shares the same gene order with another species. In addition to gene order there are further remarkable features of mt genomes, which may be used in a character matrix for phylogenetic analysis: a) Some species show a reversed strand bias in nucleotide frequencies, expressed in the GC skew value, in comparison to other crustaceans. b) Modifications of the typical cloverleaf tRNA secondary structure are frequently recorded, predominantly missing stems. c) Some protein-coding genes feature unusual start codons apart from the invertebrate mitochondrial code and abbreviated stop codons. d) In species where it was possible to sequence the mt control region (CR), a hairpin-like stem-loop structure has been determined. e) In one species (*Ligia oceanica*) repetitive sequences were also found in the CR. With the exception of repetitive sequences, all the above described features also occured in the mt genomes of Amphipoda (Chapter 5) and Cumacea (Chapter 6).

In the above-named cases, the mt genome serves as a model for the evolution of character states of 'genome morphology'. Obviously, only changes of character states are useful for phylogenetic inferences. A comparison with character states from the outgroup allows to distinguish between plesiomorphic and apomorphic character states. The unequivocal traceability of changes may be limited in some cases: too many changes may blur ancestral similarity, so it may become impossible to reconstruct the series of events leading to differences. Examples of extensive rearrangements, which make the mt gene order hardly useful for phylogenetic inferences, can be found e.g. in

Copepoda (Ki et al., 2009) and Brachiopoda (Helfenbein et al., 2001). In the dataset from Peracarida presented in this thesis, there is good support from gene order changes for some clades, e.g. Valvifera (synapomorphic location of *trnR*), Isopoda (several synapomorphic changes) (Chapter 4), Amphipoda (*trnY*, *trnG*) (Chapter 5). Isopods and cumaceans share derived features as well, but as long as there is only limited data available for Cumacea this result remains preliminary. Despite all these useful phylogenetic features, many derived character states from this 'genome morphology' dataset are restricted to single species. With a broader taxon sampling they might become valuable as apomorphic character states resolving relationships on a lower taxonomic level.

(2) Are there any rearrangement hotspots in the mt genome? Or is an equal distribution of translocated genes over the entire genome observed? Areas of increased gene order variability may be of less phylogenetic value, as they might be subject to homoplasious change.

In the mt genomes of Isopoda there is an eye-catching aggregation of translocated genes (Chapter 4). Many of the derived gene positions, especially of protein-coding genes, are shared among isopods. They even allow proposing a ground pattern hypothesis of the gene order of isopods. What is remarkable is that although many genes originate from different locations, most of them were inserted in a certain area with the CR in its center, which likewise was a subject to translocation. In contrast to this hotspot of rearrangements, another large part of the mt genome (*nad2* to *nad3*) always shows very little gene order variation in isopods, except an inversion found in one species (*Eophreatoicus* sp.14). It can be argued that the proximity to the CR might promote gene order changes, probably due to the presence of a large section of non-coding sequence. E.g. in the two examples where the CR was sequenced, the tRNA genes located around the CR show different relative positions. Unlike other parts of the mtDNA, such as those coding for vital genes, the CR might be less affected by a disruption of essential mitogenomic sequences caused by the insertion of translocated elements.

In contrast, in the mt genomes of amphipod species no such hotspot for gene translocations was detected. Instead, each of the three currently available amphipod mt genomes features unique gene order changes. Most of them seem to affect tRNA genes

of tRNA clusters (>4 tRNA genes) (Chapter 5). Derived gene positions of protein-coding genes were independently found in two species. As in isopods both were inserted adjacent to the CR.

In the cumacean species *Leucon nasica* a significant part of the mt genome (*rrnS*, CR, *trnI*) must have been subject to translocation (Chapter 6). Although the definite position of the translocated elements remains unclear for the moment, as only about one half of the mt genome is sequenced so far, there is good evidence for a derived gene order shared with isopods.

(3) What is the scenario of 'genome morphology' changes within Isopoda? How can the observed succession of character changes be most parsimoniously explained? What is the putative gene order in the isopod ancestor?

As stated above, the comparison of mt genomes from Isopoda revealed some variation in 'genome morphology'. However, there are not enough synapomorphic character states to infer isopod relationships based on structural mitogenomic characters alone. The determined character states were therefore mapped parsimoniously on a molecular tree (Chapter 4). This allowed the detection of apomorphies, synapomorphies, and putative homoplasies. In addition, the series of changes within certain lineages, e.g. of *trnR* or of the strand bias, became traceable. The broad comparison of the available mt genomes also resulted in a hypothesis for the putative isopod ground pattern (Chapter 4). In Figure 7.1 the tree from Chapter 4 was extended by the cumacean and amphipod species to give an overview of changes in mt 'genome morphology' in Peracarida as far as it is known yet.

Each change of a character state is caused by a specific event. A reversal of the GC-skew is probably the result of an inversion of the replication origin. For short distance gene translocations on the same strand a TDRL mechanism remains the best explanation. For single translocations of highly similar tRNA genes a TDRL mechanism with a subsequent remolding event is very likely, as seen for *trnL1* in isopod species (Chapter 4) or *trnG* in amphipod species (Chapter 5). The in-place inversion found in *Eophreatoicus* was most likely caused by an intramolecular recombination event.

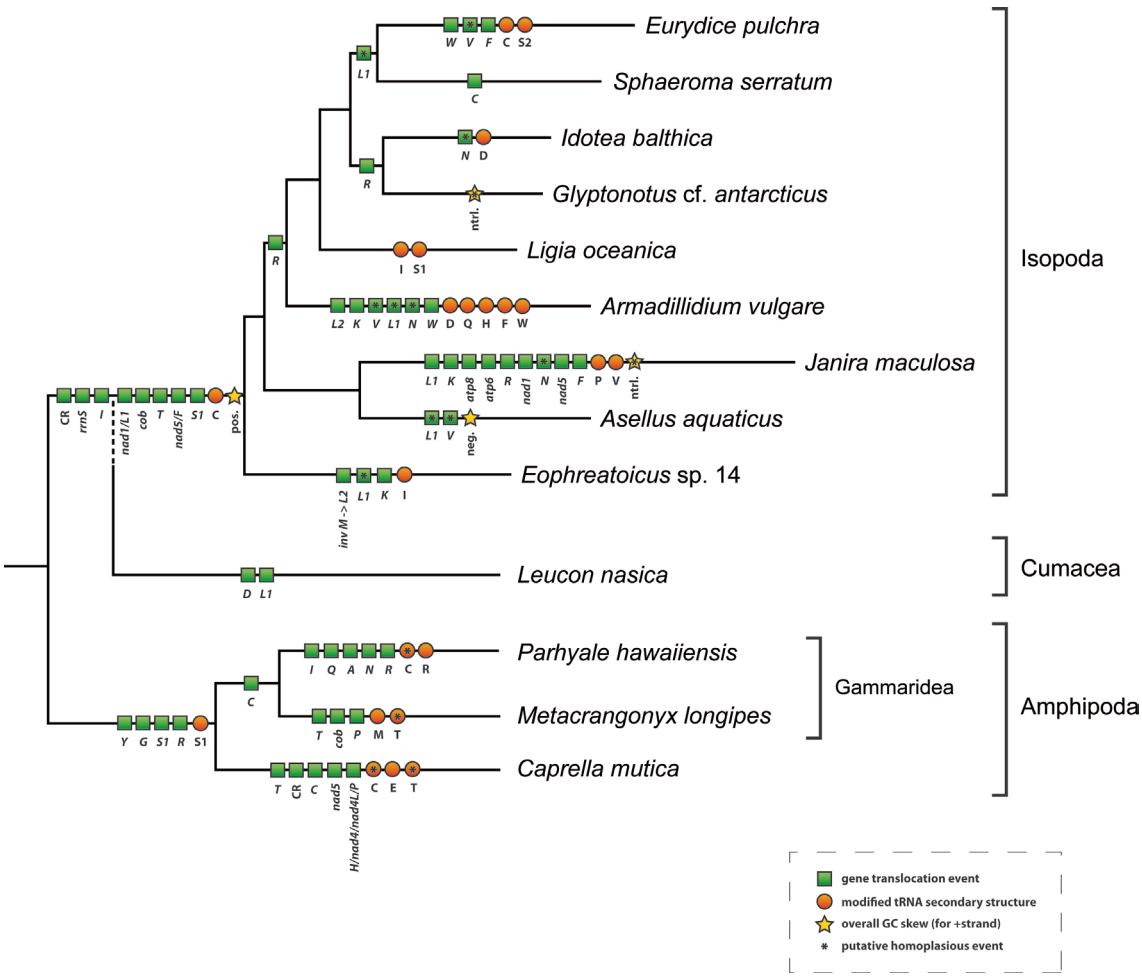


Figure 7.1. Structural characters mapped on a peracarid tree. The molecular tree of Figure 4.4 was extended by branches for cumacean and amphipod species. Similarity in gene order between the cumacean *Leucon nasica* and the isopod gene order was detected. Amphipoda show a set of derived characters different from Isopoda and Cumacea. Therefore, the ground pattern of Peracarida most probably resembles the pancrustacean ground pattern.

(4) How reliable are characters of 'genome morphology'? Is there any conflict in the data, e.g. are there specific examples of homoplasious events?

The overall reliability of characters of 'genome morphology' may be affected by homoplasious character changes. However, until now, there have been almost no reports in literature with regard to convergent evolution of mt gene order, except from birds

(Mindell et al., 1998). Homoplasious character states can be identified through conflicts in the dataset. In this study a few cases of homoplasious gene translocations were identified in isopods (Figure 7.1), which all concern tRNA genes (Chapter 4). Most striking is a shared derived location of *trnV* in three species of different isopod subtaxa, whereas the ancestral position was maintained in all other isopods! Homoplasious gene translocations of larger genes like rRNA or protein-coding genes were not detected. This finding underscores the general reliability of gene rearrangements. Convergent tRNA translocation is not likely but may happen in some cases.

Asellus aquaticus is the only isopod species that shows the same strand bias as most other crustaceans apart from Isopoda. The strand bias is an indicator for an inversion of the mitochondrial replication origin, which is generally more complex and very rare. Assuming that Phreatoicidea is the sister group of all remaining isopods, which is not undisputed, a re-reversal must have happened in the lineage leading to *Asellus aquaticus*. Otherwise two independent switches have to be assumed for Phreatoicidea and Scutocoxifera (Chapter 4). In fact, the GC skew on its own favours the Asellota as the sister group to the remaining isopods instead of the Phreatoicidea, but there is not much other evidence for this scenario.

Two cases of convergent changes in tRNA secondary structure were found in amphipod species. Whether this suffices to describe derived tRNA secondary structures as less reliable features cannot be decided yet.

(5) What is the basal split in isopod phylogeny? Phreatoicidea, Asellota or a common clade of both taxa are worth being considered to take the most basal position relative to the other isopods.

There is conflicting evidence in the mitochondrial data regarding to the earliest splits of isopod phylogeny. In phreatoicidean and asellotan species *trnR* is retained in the same position as in other arthropods, while it is shifted to other positions in the remaining isopods. The shared location of *trnL1,* derived from duplication in *Eophreatoicus* sp. 14 and *Asellus aquaticus* (but not in *Janira maculosa*), provides only weak evidence for a sister group relationship of Phreatoicidea and Asellota (Chapter 4). The reversed strand bias (GC skew) found in all isopods except Asellota favours the Asellota as the sister

group of all other isopods (Chapter 4). However, the phylogenetic analysis of mitochondrial protein-coding sequences supports a basal split between Phreatoicidea and all other isopods (Chapter 4). All these hypotheses have been raised before, based on morphological and sequence data. The data from the isopod mt genomes are somehow ambiguous on this point, but given the weakness of the two 'genome morphology' characters mentioned above, the sequence-based analysis may have more weight, and therefore Phreatoicidea is favoured as the sister group to the remaining isopods.

(6) What is the sister group of the Isopoda?

The Amphipoda were mentioned as the putative sister group to the Isopoda in some publications (Spears et al., 2005). Gene order comparison, however, shows no synapomorphies of amphipods and isopods (Figure 7.1, Chapter 5). This is significantly different for the sequenced partial mt genome of the cumacean species *Leucon nasica*, which corresponds to the mt gene order of Isopoda in one major point, the translocation of a segment spanning *rrnS*, CR and *trnI* (Chapter 6). Therefore, a sister group relationship between Amphipoda and Isopoda can clearly be rejected due to gene order characters. Currently, there is no data from Tanaidacea, another taxon which may be closely related to isopods. Despite several attempts, it was not possible to generate useful sequence information from a member of Tanaidacea. Until it will be possible to overcome the technical difficulties, the sister group to Isopoda cannot be determined by mt genome data. So far, only the isopod and cumacean mt genomes share apomorphic changes in gene order, making Cumacea a better candidate for being the sister group of Isopoda than Amphipoda.

8. Summary

The mitochondrial genome features structural and organizational characters, also referred to as characters of 'genome morphology', in addition to primary sequence data. This thesis analyzes the state, evolution, and phylogenetic value of these characters in peracarid crustaceans, with the focus on Isopoda and putative isopod sister group taxa (Amphipoda, Cumacea). Therefore, eight complete or almost complete isopod mt genomes were sequenced, as well as an entire amphipod mt genome and about half of a cumacean mt genome. The comparison revealed several significant changes of character states of the mtDNA (mitochondrial gene order, nucleotide strand bias, tRNA secondary structure etc.). These characters were mapped on a molecular tree of isopods to trace the sequence of changes and to evaluate their phylogenetic significance. It is striking that gene arrangements differ in all examined mt genomes. However, several derived gene positions are shared among isopods, which even allow inferences on the gene order of the isopod ancestor (isopod ground pattern). There are also indications that the Cumacea share main rearrangements with isopods. This is not the case for the Amphipoda, which therefore do not qualify as the isopod sister group. Although there are also indices on rare homoplasious translocation events of tRNA genes in isopods, gene order changes can be overall evaluated as phylogenetically informative characters. This applies particularly to complex rearrangements comprising protein-coding and rRNA genes, which are unlikely to emerge by convergent evolution. The overall frequency of rearrangements appears to be higher in the examined species than usually assumed for mt genomes, but is still far from saturation. The diversity in gene order should be also valuable for inferring phylogenetic relationships of closer related isopod or peracarid species. In some cases it is even possible to infer the mechanism of a translocation event, e.g. inversion by intramolecular recombination or translocation by remolding of tRNA genes. Another rare and complex character, a bias in the nucleotide composition of DNA strands, is shared by the Asellota and most other crustaceans, but is reversed in the majority of Isopoda. This finding is contrary to the position of the Phreatoicidea being the sister group to all other isopods e.g. proposed by the sequence analysis.

9. Zusammenfassung

Das mitochondriale Genom weist neben der DNA-Sequenz weitere strukturelle Merkmale auf, die als „Genom-morphologische“ Merkmale bezeichnet werden. Zur Rekonstruktion der mitochondrialen Veränderungen innerhalb der Peracariden und Bewertung des phylogenetischen Signals solcher Merkmale wurden vollständige bzw. nahezu vollständige mt-Genome von acht Isopoden-Arten sequenziert, ebenso ein weiteres vollständiges mt-Genom eines Amphipoden sowie ungefähr die Hälfte des mt-Genoms eines Vertreters der Cumacea. Im Vergleich zeigten sich mehrere deutliche Merkmalsänderungen, u.a. in der Genreihenfolge, der Basenverteilung zwischen den Strängen und der Sekundärstruktur von tRNAs. Diese Merkmale wurden auf die Ergebnisse der molekularen Stammbaumanalyse aufgetragen, um die Veränderungen nachvollziehbar zu machen und ihre Relevanz für phylogenetische Analysen festzustellen. Auffällig ist, dass alle untersuchten mt-Genome eine andere Genreihenfolge aufweisen. Dennoch gibt es einige abgeleitete Gen-Positionen, die übereinstimmend in vielen Isopoden vorhanden sind und die Rückschlüsse auf das Grundmuster der Isopoden erlauben. Ebenso gibt es Hinweise auf Gemeinsamkeiten in der Genanordnung mit den Cumacea, nicht jedoch mit den Amphipoda, die daher als Schwestergruppe der Isopoda ausscheiden. Obwohl es einige seltene Hinweise auf Homoplasien bei tRNA-Translokationen gibt, können Veränderungen der Genreihenfolge insgesamt als phylogenetisch wertvolle Merkmale angesehen werden. Insbesondere komplexe Veränderungen unter Beteiligung mehrerer Gene dürften kaum durch konvergente Evolution entstanden sein. Insgesamt scheinen Genreihenfolgeveränderungen bei den Peracariden etwas häufiger vorzukommen als bei anderen Crustacea, dennoch sind sie meist noch gut nachzuvollziehen. In einigen Fällen sind auch Rückschlüsse auf die Mechanismen möglich, die den Translokationen zugrunde liegen. So wurden z. B. Inversionen wahrscheinlich durch intramolekulare Rekombination verursacht und einige tRNAs durch Identitätsänderungen neu angeordnet. Ein anderes, selten beobachtetes, komplexes Merkmal ist die Umkehrung der Verhältnisse der Nukleotidfrequenzen für die einzelnen Stränge. Anders als die übrigen Asseln zeigen die Asellota dieselben Verhältnisse wie die meisten übrigen Crustacea. Dieses Merkmal steht somit im Widerspruch zu der sonst oft vertretenen Hypothese, wonach die Phreatoicidea die Schwestergruppe aller übrigen Isopoden sind.

10. Collected references

Abele, L. G., W. Kim and B. E. Felgenhauer. 1989. Molecular Evidence for Inclusion of the Phylum Pentastomida in the Crustacea. Mol. Biol. Evol. 6:685-691.

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

Akasaki, T., M. Nikaido, K. Tsuchiya, S. Segawa, M. Hasegawa and N. Okada. 2006. Extensive mitochondrial gene arrangements in coleoid Cephalopoda and their phylogenetic implications. Mol. Phylogenet. Evol. 38:648-658.

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

Anderson, S., A. T. Bankier, B. G. Barrell, M. H. de Bruijn, A. R. Coulson, J. Drouin, I. C. Eperon, D. P. Nierlich, B. A. Roe, F. Sanger, P. H. Schreier, A. J. Smith, R. Staden and I. G. Young. 1981. Sequence and organization of the human mitochondrial genome. Nature 290:457-465.

Ashton, G. V., K. J. Willis, E. J. Cook and M. Burrows. 2007. Distribution of the introduced amphipod, *Caprella mutica* Schurin, 1935 (Amphipoda : Caprellida : Caprellidae) on the west coast of Scotland and a review of its global distribution. Hydrobiologia 590:31-41.

Avise, J. C. 2000. Phylogeography: The History and Formation of Species. Harvard University Press viii+447 pp.

Bauza-Ribot, M. M., D. Jaume, C. Juan and J. Pons. 2009. The complete mitochondrial genome of the subterranean crustacean *Metacrangonyx longipes* (Amphipoda): a unique gene order and extremely short control region. Mitochondrial DNA 20:88-99.

Benne, R. 1994. RNA editing in trypanosomes. Eur. J. Biochem. 221:9-23.

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

Bleidorn, C., I. Eeckhaut, L. Podsiadlowski, N. Schult, D. McHugh, K. M. Halanych, M. C. Milinkovitch and R. Tiedemann. 2007. Mitochondrial genome and nuclear sequence data support myzostomida as part of the annelid radiation. Mol. Biol. Evol. 24:1690-1701.

Bogenhagen, D. F. and D. A. Clayton. 2003. The mitochondrial DNA replication bubble has not burst. Trends in Biochemical Sciences 28:357-360.

Boore, J. L. and W. M. Brown. 1994. Complete DNA sequence of the mitochondrial genome of the black chiton, *Katharina tunicata*. Genetics 138:423-443.

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

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

Boore, J. L. 2000. The Duplication/Random Loss Model for Gene Rearrangement Exemplified by Mitochondrial Genomes of Deuterostome Animals. *Comparative Genomics.* Kluwer Academic Publishers 1:133-147.

Boore, J. L. and W. M. Brown. 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., M. Medina and L. A. Rosenberg. 2004. Complete sequences of the highly rearranged molluscan mitochondrial genomes of the Scaphopod *Graptacme eborea* and the bivalve *Mytilus edulis*. Mol. Biol. Evol. 21:1492-1503.

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

Brand, M. D. 1997. Regulation analysis of energy metabolism. J. Exp. Biol. 200:193-202.

Brandt, A. and G. C. B. Poore. 2003. Higher classification of the flabelliferan and related Isopoda based on a reappraisal of relationships. Invertebrate Systematics 17:893-923.

Brennicke, A., A. Marchfelder and S. Binder. 1999. RNA editing. FEMS Microbiol. Rev. 23:297-316.

Brusca, R. C. and G. D. F. Wilson. 1991. A phylogenetic analysis of the Isopoda with some classificatory recommendations. Memoirs of the Queensland Museum 31:143-204.

Casane, D., N. Dennebouy, R. H. de, J. C. Mounolou and M. Monnerot. 1994. Genetic analysis of systematic mitochondrial heteroplasmy in rabbits. Genetics 138:471-480.

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

Clayton, D. A. 1982. Replication of Animal Mitochondrial-DNA. Cell 28:693-705.

Cook, C. E., Q. Yue and M. Akam. 2005. Mitochondrial genomes suggest that hexapods and crustaceans are mutually paraphyletic. Proc. Biol. Sci. 272:1295-1304.

Crozier, R. H. and Y. C. Crozier. 1993. The mitochondrial genome of the honeybee *Apis mellifera*: complete sequence and genome organization. Genetics 133:97-117.

Cunha, R. L., C. Grande and R. Zardoya. 2009. Neogastropod phylogenetic relationships based on entire mitochondrial genomes. BMC Evol. Biol. 9:210.

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

Dowton, M., L. R. Castro, S. L. Campbell, S. D. Bargon and A. D. Austin. 2003. Frequent mitochondrial gene rearrangements at the hymenopteran nad3-nad5 junction. J. Mol. Evol. 56:517-526.

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

Dreyer, H. and G. Steiner. 2004. The complete sequence and gene organization of the mitochondrial genome of the gadilid scaphopod *Siphonondentalium lobatum* (Mollusca). Mol. Phylogenet. Evol. 31:605-617.

Dreyer, H. and J. W. Waegele. 2002. The Scutocoxifera tax. nov. and the information content of nuclear ssu rDNA sequences for reconstruction of isopod phylogeny (Peracarida: Isopoda). Journal of Crustacean Biology 22:217-234.

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

Edgar, R. C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32:1792-1797.

Endo, K., Y. Noguchi, R. Ueshima and H. T. Jacobs. 2005. Novel repetitive structures, deviant protein-encoding sequences and unidentified ORFs in the mitochondrial genome of the brachiopod *Lingula anatina*. J. Mol. Evol. 61:36-53.

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

Folmer, O., M. Black, W. Hoeh, R. Lutz and R. Vrijenhoek. 1994. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. Mol. Mar Biol. Biotechnol. 3:294-299.

Friedrich, M. and D. Tautz. 1995. Ribosomal DNA phylogeny of the major extant arthropod classes and the evolution of myriapods. Nature 376:165-167.

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

Gibb, G. C., O. Kardailsky, R. T. Kimball, E. L. Braun and D. Penny. 2007. Mitochondrial genomes and avian phylogeny: complex characters and resolvability without explosive radiations. Mol. Biol. Evol. 24:269-280.

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

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

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

Grande, C., J. Templado and R. Zardoya. 2008. Evolution of gastropod mitochondrial genome arrangements. BMC Evol. Biol. 8:61.

Haen, K. M., B. F. Lang, S. A. Pomponi and D. V. Lavrov. 2007. Glass sponges and bilaterian animals share derived mitochondrial genomic features: A common ancestry or parallel evolution? Molecular Biology and Evolution 24:1518-1527.

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.

Hassanin, A., N. Leger and J. Deutsch. 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.

He, Y., J. Jones, M. Armstrong, F. Lamberti and M. Moens. 2005. The mitochondrial genome of *Xiphinema americanum sensu stricto* (Nematoda: Enoplea): considerable economization in the length and structural features of encoded genes. J. Mol. Evol. 61:819-833.

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

Helfenbein, K. G., H. M. Fourcade, R. G. Vanjani and J. L. Boore. 2004. The mitochondrial genome of *Paraspadella gotoi* is highly reduced and reveals that chaetognaths are a sister group to protostomes. Proc. Natl. Acad. Sci. U. S. A 101:10639-10643.

Hickerson, M. J. and C. W. Cunningham. 2000. Dramatic mitochondrial gene rearrangements in the hermit crab *Pagurus longicarpus* (Crustacea, Anomura). Mol. Biol. Evol. 17:639-644.

Hu, M., N. B. Chilton and R. B. Gasser. 2003. The mitochondrial genome of *Strongyloides stercoralis* (Nematoda) - idiosyncratic gene order and evolutionary implications. Int. J. Parasitol. 33:1393-1408.

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

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

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

Keddie, E. M., T. Higazi and T. R. Unnasch. 1998. The mitochondrial genome of *Onchocerca volvulus*: sequence, structure and phylogenetic analysis. Mol. Biochem. Parasitol. 95:111-127.

Ki, J. S., H. G. Park and J. S. Lee. 2009. The complete mitochondrial genome of the cyclopoid copepod *Paracyclopina nana*: a highly divergent genome with novel gene order and atypical gene numbers. Gene 435:13-22.

Kilpert, F. and L. Podsiadlowski. 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. and L. Podsiadlowski. 2009. The Australian fresh water isopod (Phreatoicidea: Isopoda) allows insights into the early mitogenomic evolution of isopods. Comp. Biochem. Phys. D DOI: 10.1016/j.cbd.2009.09.003.

Knoll, A. H., E. J. Javaux, D. Hewitt and P. Cohen. 2006. Eukaryotic organisms in Proterozoic oceans. Philos. Trans. R. Soc. Lond B Biol. Sci. 361:1023-1038.

Knudsen, B., A. B. Kohn, B. Nahir, C. S. McFadden and L. L. Moroz. 2006. Complete DNA sequence of the mitochondrial genome of the sea-slug, *Aplysia californica*: conservation of the gene order in Euthyneura. Mol. Phylogenet. Evol. 38:459-469.

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

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

Laubitz, D. R. 1993. Caprellidea (Crustacea: Amphipoda): Towards a new synthesis. Journal of Natural History 27:965-976.

Lavrov, D. V., W. M. Brown and J. L. Boore. 2000. A novel type of RNA editing occurs in the mitochondrial tRNAs of the centipede *Lithobius forficatus*. Proc. Natl. Acad. Sci. U. S. A 97:13738-13742.

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

Lavrov, D. V. and W. M. Brown. 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., J. L. Boore and W. M. Brown. 2002. Complete mtDNA sequences of two millipedes suggest a new model for mitochondrial gene rearrangements: duplication and nonrandom loss. Mol. Biol. Evol. 19:163-169.

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

Lavrov, D. V. and B. F. Lang. 2005. Transfer RNA gene recruitment in mitochondrial DNA. Trends Genet. 21:129-133.

Lavrov, D. V. 2007. Key transitions in animal evolution: a mitochondrial DNA perspective. Integrative and Comparative Biology 47:734-743.

Liu, Y. and Z. Cui. 2009. Complete mitochondrial genome of the Asian paddle crab *Charybdis japonica* (Crustacea: Decapoda: Portunidae): gene rearrangement of the marine brachyurans and phylogenetic considerations of the decapods. Mol. Biol. Rep. DOI: 10.1007/s11033-009-9773-2.

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

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

Machida, R. J., M. U. Miya, M. Nishida and S. Nishida. 2002. Complete mitochondrial DNA sequence of *Tigriopus japonicus* (Crustacea : Copepoda). Marine Biotechnology 4:406-417.

Machida, R. J., M. U. Miya, M. M. Yamauchi, M. Nishida and S. Nishida. 2004. Organization of the mitochondrial genome of Antarctic krill *Euphausia superba* (Crustacea : Malacostraca). Marine Biotechnology 6:238-250.

Machida, R. J., M. U. Miya, M. Nishida and S. Nishida. 2004. Large-scale gene rearrangements in the mitochondrial genomes of two calanoid copepods *Eucalanus bungii* and *Neocalanus cristatus* (Crustacea), with notes on new versatile primers for the srRNA and COI genes. Gene 332:71-78.

Mallatt, J. and G. Giribet. 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.

Marcade, I., R. Cordaux, V. Doublet, C. Debenest, D. Bouchon and R. Raimond. 2007. Structure and evolution of the atypical mitochondrial genome of *Armadillidium vulgare* (Isopoda, Crustacea). J. Mol. Evol. 65:651-659.

Martin, J. W. and G. E. Davis. 2001. An updated classification of the recent Crustacea. Natural History Museum of Los Angeles County Science Series i-viii.

Martin, W., M. Hoffmeister, C. Rotte and K. Henze. 2001. An overview of endosymbiotic models for the origins of eukaryotes, their ATP-producing organelles (mitochondria and hydrogenosomes), and their heterotrophic lifestyle. Biological Chemistry 382:1521-1539.

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

Miller, A. D., T. T. T. Nguyen, C. P. Burridge and C. M. Austin. 2004. Complete mitochondrial DNA sequence of the Australian freshwater crayfish, *Cherax destructor* (Crustacea : Decapoda : Parastacidae): a novel gene order revealed. Gene 331:65-72.

Miller, A. D., N. P. Murphy, C. P. Burridge and C. M. Austin. 2005. Complete mitochondrial DNA sequences of the decapod crustaceans *Pseudocarcinus gigas* (Menippidae) and *Macrobrachium rosenbergii* (Palaemonidae). Mar. Biotechnol. 7:339-349.

Mindell, D. P., M. D. Sorenson and D. E. Dimcheff. 1998. Multiple independent origins of mitochondrial gene order in birds. Proc. Natl. Acad. Sci. U. S. A 95:10693-10697.

Monforte, A., E. Barrio and A. Latorre. 1993. Characterization of the length polymorphism in the A + T-rich region of the Drosophila obscura group species. J. Mol. Evol. 36:214-223.

Moritz, C., T. E. Dowling and W. M. Brown. 1987. Evolution of Animal Mitochondrial-Dna - Relevance for Population Biology and Systematics. Annual Review of Ecology and Systematics 18:269-292.

Nardi, F., A. Carapelli, P. P. Fanciulli, R. Dallai and F. Frati. 2001. The complete mitochondrial DNA sequence of the basal hexapod *Tetrodontophora bielanensis*: evidence for heteroplasmy and tRNA translocations. Mol. Biol. Evol. 18:1293-1304.

National Center for Biotechnology Information, http://www.ncbi.nlm.nih.gov

Negrisolo, E., A. Minelli and G. Valle. 2004. Extensive gene order rearrangement in the mitochondrial genome of the centipede *Scutigera coleoptrata*. J. Mol. Evol. 58:413-423.

Noguchi, Y., K. Endo, F. Tajima and R. Ueshima. 2000. The mitochondrial genome of the brachiopod *Laqueus rubellus*. Genetics 155:245-259.

Ogoh, K. and Y. Ohmiya. 2004. Complete mitochondrial DNA sequence of the sea-firefly, *Vargula hilgendorfii* (Crustacea, Ostracoda) with duplicate control regions. Gene 327:131-139.

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

Oliveira, D. C., R. Raychoudhury, D. V. Lavrov and J. H. Werren. 2008. Rapidly evolving mitochondrial genome and directional selection in mitochondrial genes in the parasitic wasp *Nasonia* (Hymenoptera: Pteromalidae). Mol. Biol. Evol. 25:2167-2180.

Osawa, S., A. Muto, T. H. Jukes and T. Ohama. 1990. Evolutionary changes in the genetic code. Proc. Biol. Sci. 241:19-28.

Osawa, S., T. H. Jukes, K. Watanabe and A. Muto. 1992. Recent evidence for evolution of the genetic code. Microbiol. Rev. 56:229-264.

Papillon, D., Y. Perez, X. Caubit and P. Y. Le. 2004. Identification of chaetognaths as protostomes is supported by the analysis of their mitochondrial genome. Mol. Biol. Evol. 21:2122-2129.

Penny, D., B. J. McComish, M. A. Charleston and M. D. Hendy. 2001. Mathematical elegance with biochemical realism: The covarion model of molecular evolution. J. Mol. Evol. 53:711-723.

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

Pisani, D. 2009. Arthropods (Arthropoda). In Hedges,S.B. and Kumar,S. (eds), *The Timetree of Life*. Oxford University Press 251-254.

Podsiadlowski, L. and T. Bartolomaeus. 2005. Organization of the mitochondrial genome of mantis shrimp *Pseudosquilla ciliata* (Crustacea: Stomatopoda). Mar. Biotechnol. 7:618-624.

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

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

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

Rawlings, T. A., T. M. Collins and R. Bieler. 2003. Changing identities: tRNA duplication and remolding within animal mitochondrial genomes. Proc. Natl. Acad. Sci. U S A 100:15700-15705.

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

Richter, S. and G. Scholtz. 2001. Phylogenetic analysis of the Malacostraca (Crustacea). Journal of Zoological Systematics and Evolutionary Research 39:113-136.

Richter, S. 2002. The Tetraconata concept: Hexapod-crustacean relationships and the phylogeny of Crustacea. Organisms Diversity and Evolution 2:217-237.

Roehrdanz, R. L., M. E. Degrugillier and W. C. I. Black. 2002. Novel rearrangements of arthropod mitochondrial DNA detected with long-PCR: Applications to arthropod phylogeny and evolution. Mol. Biol. Evol. 19:841-849.

Ronquist, F. and J. P. Huelsenbeck. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572-1574.

Rubinoff, D. and B. S. Holland. 2005. Between two extremes: mitochondrial DNA is neither the panacea nor the nemesis of phylogenetic and taxonomic inference. Syst. Biol. 54:952-961.

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

Saccone, C., C. Gissi, A. Reyes, A. Larizza, E. Sbisa and G. Pesole. 2002. Mitochondrial DNA in metazoa: degree of freedom in a frozen event. Gene 286:3-12.

Sagan, L. 1967. On Origin of Mitosing Cells. Journal of Theoretical Biology 14:225-274.

Schmalfuss, H. 1989. Phylogenetics in Oniscidea. Monitore Zoologico Italiano (N.S.) Monografia 4:3-27.

Schmidt, C. 2008. Phylogeny of the Terrestrial Isopoda (Oniscidea): A Review. Arthropod Syst. Phyl. 66:191-226.

Schotte, M., B. F. Kensley, and S. Shilling. (1995 onwards). World list of Marine, Freshwater and Terrestrial Crustacea Isopoda. National Museum of Natural History Smithsonian Institution http://www.nmnh.si.edu/iz/isopod/.

Schram, F. R. 1970. Isopod from Pennsylvanian of Illinois. Science 169:854-855.

Segawa, R. D. and T. Aotsuka. 2005. The mitochondrial genome of the Japanese freshwater crab, *Geothelphusa dehaani* (Crustacea : Brachyura): Evidence for its evolution via gene duplication. Gene 355:28-39.

Seyffert, W. 2003. Das Chondrom - Mitochondrien-Genetik. *Lehrbuch der Genetik*. Spektrum Akademischer Verlag 2:147-184.

Shao, R., H. Mitani, S. C. Barker, M. Takahashi and M. Fukunaga. 2005. 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. F., E. F. Kirkness and S. C. Barker. 2009. The single mitochondrial chromosome typical of animals has evolved into 18 minichromosomes in the human body louse, *Pediculus humanus*. Genome Research 19:904-912.

Shultz, J. W. and J. C. Regier. 2000. Phylogenetic analysis of arthropods using two nuclear protein-encoding genes supports a crustacean + hexapod clade. Proceedings of the Royal Society Biological Sciences Series B 267:1011-1019.

Spears, T., R. W. Debry, L. G. Abele and K. Chodyla. 2005. Peracarid monophyly and interordinal phylogeny inferred from nuclear small-subunit ribosomal DNA sequences (Crustacea : Malacostraca : Peracarida). Proceedings of the Biological Society of Washington 118:117-157.

Staton, J. L., L. L. Daehler and W. M. Brown. 1997. Mitochondrial gene arrangement of the horseshoe crab *Limulus polyphemus* L.: conservation of major features among arthropod classes. Mol. Biol. Evol. 14:867-874.

Stechmann, A. and M. Schlegel. 1999. Analysis of the complete mitochondrial DNA sequence of the brachiopod *Terebratulina retusa* places Brachiopoda within the protostomes. Proceedings of the Royal Society Biological Sciences Series B 266:2043-2052.

Storch, V. and B. G. M. Jamieson. 1992. Further Spermatological Evidence for Including the Pentastomida (Tongue Worms) in the Crustacea. International Journal for Parasitology 22:95-108.

Stuart, K. D., A. Schnaufer, N. L. Ernst and A. K. Panigrahi. 2005. Complex management: RNA editing in trypanosomes. Trends Biochem. Sci. 30:97-105.

Sun, H., K. Zhou and D. Song. 2005. Mitochondrial genome of the Chinese mitten crab *Eriocheir japonica sinenesis* (Brachyura: Thoracotremata: Grapsoidea) reveals a novel gene order and two target regions of gene rearrangements. Gene 349:207-217.

Supek, F. and K. Vlahovicek. 2004. INCA: synonymous codon usage analysis and clustering by means of self-organizing map. Bioinformatics 20:2329-2330.

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

Tabacaru, I. and D. L. Danielopol. 1999. Contribution to the phylogeny of Isopoda (Crustacea). Vie et Milieu 49:163-176.

Telford, M. J., E. A. Herniou, R. B. Russell and D. T. Littlewood. 2000. Changes in mitochondrial genetic codes as phylogenetic characters: two examples from the flatworms. Proc. Natl. Acad. Sci. U. S. A 97:11359-11364.

Tjensvoll, K., K. Hodneland, F. Nilsen and A. Nylund. 2005. Genetic characterization of the mitochondrial DNA from *Lepeophtheirus salmonis* (Crustacea : Copepoda). A new gene organization revealed. Gene 353:218-230.

Valverde, J. R., B. Batuecas, C. Moratilla, R. Marco and R. Garesse. 1994. The Complete Mitochondrial-DNA Sequence of the Crustacean *Artemia franciscana*. J. Mol. Evol. 39:400-408.

Wägele, J. W. 1989. Evolution und phylogenetisches System der Isopoda. Zoologica 140:1-262.

Wei, Y. H. 1998. Oxidative stress and mitochondrial DNA mutations in human aging. Proc. Soc. Exp. Biol. Med. 217:53-63.

Wilson, G. D., C. L. Humphrey, D. J. Colgan, K. A. Gray and R. N. Johnson. 2009. Monsoon-influenced speciation patterns in a species flock of *Eophreatoicus* Nicholls (Isopoda; Crustacea). Mol. Phylogenet. Evol. 51:349-364.

Wilson, G. D. F. 1999. Some of the deep-sea fauna is ancient. Crustaceana 72:1019-1030.

Wilson, G. D. F. 2009. The Phylogenetic Position of the Isopoda in the Peracarida (Crustacea: Malacostraca). Arthropod Systematics & Phylogeny 67:159-198.

Wilson, K., V. Cahill, E. Ballment and J. Benzie. 2000. The complete sequence of the mitochondrial genome of the crustacean *Penaeus monodon*: Are malacostracan crustaceans more closely related to insects than to branchiopods? Mol. Biol. Evol. 17:863-874.

Wingstrand, K. G. 1972. Comparative spermatology of a pentastomid, *Raillietella hemidactyli*, and a branchiuran crustacean, *Argulus foliaceus*, with a discussion of pentastomid relationships. Kongelige Danske Videnskabernes Selskab Biologiske Skrifter 19:1-72.

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

Wright, F. 1990. The Effective Number of Codons Used in a Gene. Gene 87:23-29.

Wright, J. C. and K. Ting. 2006. Respiratory physiology of the Oniscidea: aerobic capacity and the significance of pleopodal lungs. Comp Biochem. Physiol A Mol. Integr. Physiol 145:235-244.

Wu, X., X. Xu, Z. Yu and X. Kong. 2009. Comparative mitogenomic analyses of three scallops (Bivalvia: Pectinidae) reveal high level variation of genomic organization and a diversity of transfer RNA gene sets. BMC Res. Notes 2:69-

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

Yamauchi, M. M., M. U. Miya and M. Nishida. 2003. Complete mitochondrial DNA sequence of the swimming crab, *Portunus trituberculatus* (Crustacea: Decapoda: Brachyura). Gene 311:129-135.

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

Yokobori, S. and S. Paabo. 1995. Transfer RNA editing in land snail mitochondria. Proc. Natl. Acad. Sci. U. S. A 92:10432-10435.

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

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

11. Appendix

The appendix contains all additional figures and tables for reference, which are part of the digital supplementary material of the respective publications (chapters 2-5). The files are usually provided by download only.

Supplementary file 2.1

Number of effective codons used in mitochondrial protein-coding genes of Crustacea.

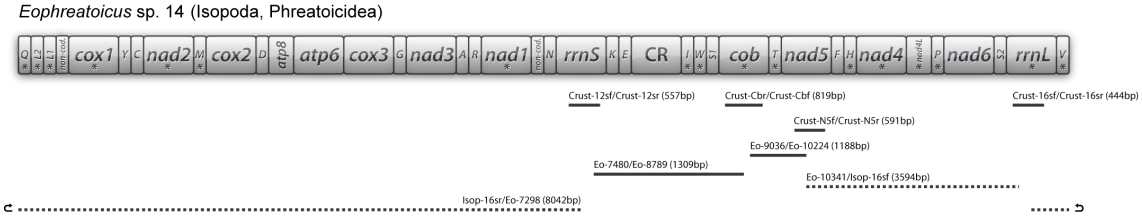
Species	Taxon	Acc.Number	*atp6*	*cox1*	*cox2*	*cox3*	*cob*	*nad1*	*nad2*	*nad3*	*nad4*	*nad5*	*nad6*
Speleonectes tulumensis	Remipedia	NC_005938	41,8	32,8	29,7	42,4	36,9	36,2	34,1	36,6	38,1	41,1	36,8
Hutchinsoniella macr.	Cephalocarida	NC_005937	32,7	34,9	33,2	34,3	37,9	41,5	28,8	32,1	45,6	45,4	39,5
Armillifer armillatus	Pentastomida	NC_005934	39,3	41,4	44,6	35,2	38,8	33,3	43,4	40,0	37,7	38,8	43,1
Argulus americanus	Branchiura	NC_005935	33,9	32,9	35,8	33,6	36,0	36,1	30,6	27,9	35,5	32,1	40,7
Vargula hilgendorfii	Ostracoda	NC_005306	50,6	49,8	51,2	49,9	47,9	44,1	44,8	40,5	41,1	44,8	49,8
Lepeophtheirus salmonis	Copepoda	NC_007215	43,6	50,6	51,1	45,6	48,1	46,4	52,7	39,9	46,0	46,3	37,7
Tigriopus japonicus	Copepoda	NC_003979	56,7	51,3	52,2	55,8	55,0	52,6	51,9	40,6	51,7	54,8	58,6
Megabalanus volcano	Cirripedia	NC_006293	44,2	42,5	47,6	41,7	48,0	37,1	44,4	37,9	43,5	44,8	49,1
Pollicipes polymerus	Cirripedia	NC_005936	50,4	44,6	49,6	50,1	48,2	44,9	46,0	43,1	48,8	42,0	52,0
Tetraclita japonica	Cirripedia	NC_008974	43,6	46,5	44,4	50,2	45,7	38,7	49,2	49,3	44,8	42,9	56,0
Artemia franciscana	Branchiopoda	NC_001620	48,0	47,2	46,8	54,4	48,8	52,2	54,0	61,0	53,3	49,9	53,8
Daphnia pulex	Branchiopoda	NC_000844	52,7	53,5	58,8	53,7	48,3	46,4	50,0	51,0	52,5	52,1	40,2
Triops cancriformis	Branchiopoda	NC_004465	46,8	41,7	45,3	49,2	42,9	38,4	41,2	48,6	42,3	38,8	51,2
Triops longicaudatus	Branchiopoda	NC_006079	41,7	38,9	43,6	40,8	44,2	37,5	46,5	50,6	41,5	41,5	36,8
Gonodactylus chiragra	Stomatopoda	NC_007442	49,0	43,5	45,4	40,8	44,8	40,6	42,9	49,6	39,0	42,0	47,3
Harpiosquilla harpax	Stomatopoda	NC_006916	39,8	40,3	40,1	39,5	47,1	44,4	37,3	43,5	39,7	38,6	32,5
Lysiosquillina maculata	Stomatopoda	NC_007443	46,8	50,6	44,1	51,1	47,7	49,4	46,1	45,8	52,3	48,8	51,7
Squilla empusa	Stomatopoda	NC_007444	46,8	43,7	37,5	41,2	48,7	43,4	41,0	34,7	39,5	38,5	44,6
Squilla mantis	Stomatopoda	NC_006081	42,0	43,0	40,3	42,0	50,2	37,6	37,9	45,3	38,4	38,2	50,9
Callinectes sapidus	Decapoda	NC_006281	43,7	41,0	46,0	46,9	39,1	40,8	43,5	43,2	36,4	41,3	40,9
Cherax destructor	Decapoda	NC_011243	55,5	51,7	52,7	51,0	54,7	44,3	49,0	47,8	52,1	51,1	43,6
Eriocheir sinensis	Decapoda	NC_006992	42,3	40,8	48,3	43,7	43,9	36,9	37,4	49,4	43,5	37,3	35,6
Geothelphusa dehaani	Decapoda	NC_007379	40,5	39,6	37,9	40,7	38,3	38,6	38,8	36,6	35,6	37,4	33,5
Macrobrachium ros.	Decapoda	NC_006880	47,6	42,3	42,6	45,1	38,6	39,6	44,5	51,2	40,5	41,5	41,5
Marsupenaeus japonicus	Decapoda	NC_007010	46,3	42,8	45,6	45,6	48,9	48,7	45,9	44,4	40,7	44,8	39,7
Pagurus longicarpus	Decapoda	NC_003058	40,8	37,9	38,0	34,2	38,3	37,4	38,3	50,5	35,1	35,3	39,5
Panulirus japonicus	Decapoda	NC_004251	49,3	47,7	56,9	56,9	46,9	52,3	51,3	53,8	49,0	46,2	45,5
Penaeus monodon	Decapoda	NC_002184	35,1	42,6	36,6	38,0	42,3	36,1	39,0	38,8	35,4	33,9	36,1
Portunus trituberculatus	Decapoda	NC_005037	42,2	39,6	39,0	40,8	38,5	37,3	39,4	44,8	37,2	38,0	35,5
Pseudocarcinus gigas	Decapoda	NC_006891	46,0	39,2	38,9	46,7	41,0	39,2	43,1	44,2	37,9	38,3	39,3
Idotea baltica	Peracarida	DQ442915	53,2	52,6	45,0	53,9	51,0	55,7	45,7	61,0	52,8	49,5	41,8
Ligia italica	Peracarida	DQ442914	56,8	53,5	54,7	59,9	51,6	57,1	53,9	54,1	54,1	48,8	47,0

Supplementary file 2.2

PCR primers used to amplify mitochondrial gene fragments from *Ligia oceanica.*

Primer	Nucleotide sequence (5'-3')	Reference
CB2H	TCCTCAAAATGATATTTGTCCTCA	Roehrdanz et al. (2002)
N4(87)	TCAGCTAATATAGCAGCTCC	Roehrdanz et al. (2002)
16S2	GCGACCTCGATGTTGGATTAA	Roehrdanz et al. (2002)
N4	GGAGCTTCAACATGAGCTTT	Roehrdanz et al. (2002)
crust-12f	CAGCAKYCGCGGTTAKAC	Podsiadlowski and Bartolomaeus (2005)
crust-12sr	ACACCTACTWTGTTACGACTTATCTC	Podsiadlowski and Bartolomaeus (2005)
crust-16sf	TGACYGTGCDAAGGTAGC	this study
crust-16sr	CCGGTCTGAACTCAYATC	Podsiadlowski and Bartolomaeus (2005)
crust-cox1f	ACTAATCACAARGAYATTGG	Podsiadlowski and Bartolomaeus (2005)
crust-cox1r	TAGTCTGAGTANCGTCGWGG	Podsiadlowski and Bartolomaeus (2005)
crust-cox3f	ATAATTCAATGATGACGAGA	Podsiadlowski and Bartolomaeus (2005)
crust-cox3r	CCAATAATWACATGWAGACC	Podsiadlowski and Bartolomaeus (2005)
crust-nd4f	TTGAGGTTAYCAGCCYG	Podsiadlowski and Bartolomaeus (2005)
crust-nd4r	ATATGAGCYACAGAAGARTAAGC	Podsiadlowski and Bartolomaeus (2005)
crust-nd5f	AGAATTCTACTAGGDTGRGATGG	Podsiadlowski and Bartolomaeus (2005)
crust-nd5r	AAAGAGCCTTAAATAAAGCATG	Podsiadlowski and Bartolomaeus (2005)
Lo-12s-f	AGGAGCAGGTGGGTTACAATC	this study
Lo-12sf-r	CTTTGGGTTTGAAGTACATAGC	this study
Lo-12sf-2	TATCTTTGAAGGATAATAGTTTTTAG	this study
Lo-12sf-3	AAATGCCTGCCTATCAAACC	this study
Lo-12sfr-2	TTACCTCAACTTGACAGATAAATGTG	this study
Lo-12sfr-3	CCTTGCTGGGTAGATTACGGTC	this study
Lo-16s-r	ATCTTAAAGGCTTACGAAATTCAG	this study
Lo-C1f-2	TTTCTTTGCATTTAGCTGGTG	this study
Lo-C1f-r	GGGGAAAGGCTATATCAGGAG	this study
Lo-C1r-2	AACCCCAATACCCCAATG	this study
Lo-C3-12	CTACTGGGAGTGTATTTTAGTCGTC	this study
Lo-C3f_r	CTACAGGGAGGTCAGATTCTAC	this study
Lo-C3f-2	GGAGTGTATTTTAGTCGTCTCTTC	this study
Lo-Cb-f	GTCCTACCACTACTCTCACACCTG	this study
Lo-Cbf-2	CCCGAATGATACTATCTATTTGCC	this study
Lo-Cbf-3	ACAGTAATGTGTTAGACACCTCCG	this study
Lo-CB-r	CAGGATATTTTTCTTGACTTTTAGG	this study
Lo-Co3-c	ATTATCTTTTTGGTATTTGATGTGG	this study
Lo-control-12s	CTTTAAAGGTTCTAAGGGTATAAGG	this study
Lo-control-cyb	AATTGGAGCCCGACCC	this study
Lo-CyB-N5	CCAGATCGACCCCAACG	this study
Lo-H4251	CAAAAAAAGGAAAAGAAGAATAGGAC	this study
Lo-L3400	TGTGGGAGAGGTTTTCAGCC	this study
Lo-L3401	ATTTGTTTGTTCTCGCCCAG	this study
Lo-L39r-2	AAGAACCATTACTACCTTTTGTATCAG	this study
Lo-L4250	CGGGGTTCAGGAGAGAGTTTAG	this study
Lo-L4251	TTTGCCCTAGAGGTTGCTGTG	this study
Lo-L5100	GAGTCACCGTTTACTATCGCCG	this study
Lo-N4f	TTATGTTAATTTTAGGATGGGGC	this study
Lo-N4f-2	TCTTACTCAAACTGACCTCAAACACC	this study
Lo-N4-N5	ATTTTAGGATGGGGCTATCAG	this study
Lo-N5-CyB	AATAAAATCTGCAGCATCGC	this study
Lo-N5f-3	TGTGTGGATTTCCGTTTATGG	this study
Lo-N5f-4	TTATGATCTAAGTGTGGTCTAGTGTTG	this study
Lo-N5-N4	TAGTTTTAGTATTCTCCTTTGTGGC	this study
Lo-N5N4-2	TAGGGCACTTGTCCATTCTTC	this study
Lo-N5N4-r	CAGAGTAGAAGAATGGACAAGTGC	this study

Supplementary file 3.1



PCR products used to cover the complete mt genome of *Eophreatoicus* sp. 14. In a first step short intragenic fragments were obtained by using general crustacean primer pairs (Crust-12sf/Crust-12sr, Crust-Cbf/Crust-Cbr, Crust-16sf/Crust-16sr, Crust-N5f/Crust-N5r.) New species specific (Eo-)primers were designed based on these sequences. They were also combined with general isopod primers (Isop-16sf, Isop-16sr). The remaining gaps were bridged by four PCR products and sequenced by primer walking strategy. For the complete list of primers see supplementary file 3.2. The largest two PCR products (dotted lines) were amplified by a long range Taq polymerase. The alignment of the overlapping fragments indicates a circular organization of the mt genome.

Supplementary file 3.2

Primers used for amplification and sequencing of mitochondrial gene fragments from *Eophreatoicus* sp. 14.

Primer	Orientation	Sequence (5'-3')	Reference
Crust-16sf	Reverse	TGACYGTGCDAAGGTAGC	Podsiadlowski & Bartolomaeus 2005
Crust-16sr	Forward	CCGGTCTGAACTCAYATC	Podsiadlowski & Bartolomaeus 2005
Crust-CBf	Reverse	CGAGATGTAAAYTAYGGSTGAC	This study
Crust-CBr	Forward	CTACGGGAGTGCACCRATYC	This study
Crust-12sf	Forward	CAGCAKYCGCGGTTAKAC	Podsiadlowski & Bartolomaeus 2005
Crust-12sr	Reverse	ACACCTACTWTGTTACGACTTATCTC	Podsiadlowski & Bartolomaeus 2005
Crust-N5f	Forward	AGAATTCTACTAGGDTGRGATGG	Podsiadlowski & Bartolomaeus 2005
Crust-N5r	Reverse	AAAGAGCCTTAAATAAAGCATG	Podsiadlowski & Bartolomaeus 2005
Crust-N4f	Reverse	TTGAGGTTAYCAGCCYG	Podsiadlowski & Bartolomaeus 2005
Crust-N4r	Forward	ATATGAGCYACAGAAGARTAAGC	Podsiadlowski & Bartolomaeus 2005
Isop-16sf	Reverse	AARAAWGATTGCGACCTCGATGTTGAATTG	This study
Isop-16sr	Forward	TATGCTACCTTAGCACAGTYAGRATACTGCGGC	This study
L39-Met	Forward	AAGCTHVTGGGCTCATACCCC	Yamauchi et al. 2004
H718-ND2	Forward	AABCCHGDGAAMGGDGGHAVHCCHCC	Yamauchi et al. 2004
L1384-CO1	Reverse	GGTCAACAAATCATAAAGATATTGG	Yamauchi et al. 2004
L1564-CO1	Reverse	ATGGTWATACCGATTWTRATTGG	Yamauchi et al. 2004
H1602-CO1	Forward	GGGAADGCTATGTCWGGGGC	Yamauchi et al. 2004
L2020S-CO1	Reverse	AATACHTCMTTCTTTGATCCWGCHGGDGGDGGDGACCC	Yamauchi et al. 2004
H2043-CO1	Forward	TAAACTTCAGGGTGACCAAAAAATCA	Yamauchi et al. 2004
H2619-CO1	Forward	GGTATWCCWGCKAGWCCTAAGAAATGTTG	Yamauchi et al. 2004
L3020-CO2	Forward	ATTTTTTTYCATGAYCATGC	Yamauchi et al. 2004
H3290CO2	Reverse	GGSATTATGTAWGAATCAAATT	Yamauchi et al. 2004
H3514-CO2	Reverse	CCACAAATTTCKGAACATTGWCCATAAAA	Yamauchi et al. 2004
L3542-CO2	Forward	GGNCAATGTTCAGAAATTTGTGG	Yamauchi et al. 2004
H4375-A6	Reverse	GCDATCATGTTDGCDGMWAGTCG	Yamauchi et al. 2004
L5170-CO3	Forward	TTGTDGCNACHGGMTTTCATGG	Yamauchi et al. 2004
L5281-CO3	Forward	CATCATKTWGGGTTTGAGGCHGCHGCWTGGTATTGGCA	Yamauchi et al. 2004
L5526-ND3	Forward	CCHTWTGAGTGTGGWTTTGATCC	Yamauchi et al. 2004
Eo-752	Reverse	TTTTTATTTACTGTAGGAGGATTGAC	This study
Eo-1326	Forward	TAAGTGAAGAGAAAAGATTGC	This study
Eo-1787	Reverse	ACAGTGTTATTTACGCAACGATG	This study
Eo-1824	Forward	GGTATTCCCCCTTTGTTGAC	This study
Eo-2400	Forward	ATTTCATTTACACCCCCATTG	This study
Eo-2568	Forward	AGAGGTAATCAGAAGTGAAAAGGTGC	This study
Eo-2677	Reverse	CTCCAAATAAGAATAACAAGGG	This study
Eo-2672	Forward	TTGGAGGAAAATAAGAGAGGC	This study
Eo-2726	Forward	TTCAGGTTTTAAGTTTCTTTTTCC	This study
Eo-2804	Reverse	AGTATGAGTAGGAATAGAACTAAACC	This study
Eo-3184	Reverse	ACCAAAATAAAAGCAGGAAGAATAG	This study
Eo-3414	Forward	CGGTTATTAGTGACTGCGGC	This study
Eo-3820	Reverse	AGCCTCACACTCCCAAATAC	This study
Eo-3900	Forward	TTTTGACCCTGCTACTGG	This study
Eo-4115	Forward	TGATAAGGAATGTAATGGG	This study
Eo-4272	Reverse	TCCCTGAGGAACTATGTG	This study
Eo-4293	Forward	TTTTATGGTTATTATTGAGAGTATTAG	This study
Eo-4390	Forward	GCTTTAATAGGAGGGTCTGC	This study
Eo-5268	Reverse	CAGCCTCAAACCCAAAATG	This study
Eo-5529	Reverse	CCACATTCAAAAGGAGACAAC	This study
Eo-5653	Forward	GGTTGAGTTTATTTCTTGGTTGTG	This study
Eo-6061	Forward	GGTCTTTCGCTTCCTCC	This study
Eo-6495	Reverse	TCTTTTTCTTTCCTGCCTGAG	This study
Eo-6865	Reverse	CTATTTTGGCTTTACTCTTTCTGC	This study
Eo-7298	Reverse	TACTACCTTTCTAAAGTCCTGGAATAAC	This study
Eo-7480	Forward	GCGTATAACTAAGTTGGGATGGG	This study

Primer	Orientation	Sequence (5'-3')	Reference
Eo-7963	Reverse	CACCTGCTTTCCACTGC	This study
Eo-8350	Forward	ATTTAAGGTAGATTTTGGAAG	This study
Eo-8375	Reverse	GAAATCTTCCAAAATCTACC	This study
Eo-8789	Reverse	ACCCAGAAAACTTTATTCCAGC	This study
Eo-9036	Forward	CTGCGAATCCTCCTCAAAGTC	This study
Eo-10224	Reverse	GGAAGCCAAGCAGAGAAAGG	This study
Eo-10341	Forward	ATTTGTTTTTATTTGTAGTAGGGGG	This study
Eo-10841	Forward	ACTAGAGTCAAAGAGGAGGATG	This study
Eo-11180	Forward	TGGAGTGAAATGTTTGGTGG	This study
Eo-11619	Forward	GCGAAAAACACTAAAAGCCTG	This study
Eo-13084	Reverse	AATCCTTCAACTAAGATCAAACC	This study
Eo-13530	Reverse	TCAACCACACTATCCAAGTAACC	This study
Eo-14120	Forward	TTTATAGGGTCTTGTCGTCTTGC	This study
Eo-14449	Forward	AGGTTTATTGTTGGTACTTTGGGG	This study
Eo-14749	Forward	TTAAATCTCCCTGATACAAAAG	This study

Supplementary file 3.3

Detection performance of utilised computer software identifying tRNA genes of *Eophreatoicus* sp. 14. Standard search with tRNAscan-SE performed with 'Invertebrate Mito' genetic code and 'Mito/Chloroplast' model. When no tRNAs were found, the less strict 'Nematode Mito' model was used instead. ARWEN was used with "Invertebrate mitochondrial genetic code" option. Results were manually checked afterwards.

tRNA	tRNAscan-SE 1.21	ARWEN 1.2.2.c	manual
tRNA-Ala	found	found	accepted
tRNA-Asn	found	found	accepted
tRNA-Asp	found	found	accepted
tRNA-Arg	found	found	accepted
tRNA-Cys	found ('Nematode Mito' model)	found	accepted
tRNA-Gln	found	found	accepted
tRNA-Glu	found	found	accepted
tRNA-Gly	found	found	accepted
tRNA-His	found	found	accepted
tRNA-Ile	not found	found	accepted
tRNA-Leu(CUN)	found	found	accepted
tRNA-Leu(UUR)	found	found	accepted
tRNA-Lys	found	found	accepted
tRNA-Met	found	found	accepted
tRNA-Phe	found ('Nematode Mito' model)	not found	improved
tRNA-Pro	found	found	accepted
tRNA-Ser(AGY)	found	found	accepted
tRNA-Ser(UCN)	found	found	accepted
tRNA-Thr	found	found	accepted
tRNA-Trp	found	found	accepted
tRNA-Tyr	found	found	accepted
tRNA-Val	found	found	accepted

Supplementary file 3.4

Start codons of *cox1* gene of crustacean species. Besides common start codons for the invertebrate mitochondrial code (ATG, ATA, ATT), alternatives exist in a number of species. It is noticeable that Malacostraca, excluding Brachyura, clearly show the ACG start codon for *cox1*. An exception is *H. rubra*, starting with CCG. Probably, the ACG start codon for *cox1* is the ancestral state of Malacostraca. Start codons are taken from published GenBank entries.

Species	Taxon	Start codon (*cox1*)
Speleonectes tulumensis	Remipedia	ATA
Hutchinsoniella macracantha	Cephalocarida	ATA
Vargula hilgendorfii	Ostracoda	ATG
Artemia franciscana	Branchiopoda	ATG
Daphnia pulex	Branchiopoda	ATT
Triops cancriformis	Branchiopoda	ATA
Triops longicaudatus	Branchiopoda	ATA
Argulus americanus	Maxillopoda	TCG
Armillifer armillatus	Maxillopoda	CTG
Lepeophtheirus salmonis	Maxillopoda	ATA
Megabalanus volcano	Maxillopoda	TTG
Pollicipes mitella	Maxillopoda	ACA
Pollicipes polymerus	Maxillopoda	TCG
Tetraclita japonica	Maxillopoda	ATC
Tigriopus californicus	Maxillopoda	ATA
Tigriopus japonicus	Maxillopoda	ATA
Cherax destructor	Malacostraca	ACG
Eophreatoicus sp.-14	Malacostraca	ACG
Euphausia superba	Malacostraca	ACG
Fenneropenaeus chinensis	Malacostraca	ACG
Gonodactylus chiragra	Malacostraca	ACG
Halocaridina rubra	Malacostraca	CCG
Harpiosquilla harpax	Malacostraca	ACG
Ligia oceanica	Malacostraca	ACG
Litopenaeus vannamei	Malacostraca	ACG
Lysiosquillina maculata	Malacostraca	ACG
Macrobrachium rosenbergii	Malacostraca	ACG
Marsupenaeus japonicus	Malacostraca	ACG
Pagurus longicarpus	Malacostraca	ACG
Panulirus japonicus	Malacostraca	ACG
Penaeus monodon	Malacostraca	ACG
Squilla empusa	Malacostraca	ACG
Squilla mantis	Malacostraca	ACG
Eriocheir sinensis	Malacostraca, Brachyura	ATG
Callinectes sapidus	Malacostraca, Brachyura	ATG
Geothelphusa dehaani	Malacostraca, Brachyura	ATG
Portunus trituberculatus	Malacostraca, Brachyura	ATG
Pseudocarcinus gigas	Malacostraca, Brachyura	ATG

Supplementary file 4.1

a) Amplification and sequencing primers used to obtain the partial mt genome of *Eurydice pulchra*.

Primer	Orientation	Sequence (5'-3')	Reference
16S2	Reverse	GCGACCTCGATGTTGGATTAA	Roehrdanz et al. 2002
CB2H	Forward	TCCTCAAAATGATATTTGTCCTCA	Roehrdanz et al. 2002
N4	Forward	GGAGCTTCAACATGAGCTTT	Roehrdanz et al. 2002
N4(87)	Reverse	TCAGCTAATATAGCAGCTCC	Roehrdanz et al. 2002
Crust-12sf	Forward	CAGCAKYCGCGGTTAKAC	Podsiadlowski & Bartolomaeus 2005
Crust-12sr	Reverse	ACACCTACTWTGTTACGACTTATCTC	Podsiadlowski & Bartolomaeus 2005
Crust-16sr	Forward	CCGGTCTGAACTCAYATC	Podsiadlowski & Bartolomaeus 2005
Crust-cox1f	Forward	ACTAATCACAARGAYATTGG	Podsiadlowski & Bartolomaeus 2005
Crust-cox1r	Reverse	TAGTCTGAGTANCGTCGWGG	Podsiadlowski & Bartolomaeus 2005
Crust-CytB-f1	Reverse	CGAGATGTAAAYTAYGGSTGAC	This study
Crust-CytB-f2	Reverse	CGAGATGTAAAYTAYGGWTGAC	This study
Isop-16sr-lrange	Forward	TATGCTACCTTAGCACAGTYAGRATACTGCGGC	This study
Pera-Co1r	Reverse	AADGCTATATCAGGAGCCCCAATTATTAAAGG	This study
HPK16saa	Forward	ATGCTACCTTTGCACRGTCAAGATACYGCGGC	Hwang et al. 2001
L329-ND2 (S5,S7)	Forward	GGWGCHGCHCCNTTWCATTTTTG	Yamauchi et al. 2004
H1368-CO1 (S5)	Reverse	TATAGAGTTCCAATRTCYTTGTGATT	Yamauchi et al. 2004
H1602-CO1 (S7)	Reverse	GGGAADGCTATGTCWGGGGC	Yamauchi et al. 2004
L3020-CO2 (S13)	Forward	ATTTTTTTYCATGAYCATGC	Yamauchi et al. 2004
H3514-CO2 (S13)	Reverse	CCACAAATTTCKGAACATTGWCCATAAAA	Yamauchi et al. 2004
L4268-A6 (S16)	Forward	GGDTGGDTTAAWMAHACWCAACA	Yamauchi et al. 2004
H4806-CO3 (S16)	Reverse	GADGHAATGAAHAGGATTATDCCTCA	Yamauchi et al. 2004
L4672-CO3 (S18, S19)	Forward	GGWCTWGTBAAATGGTTTCA	Yamauchi et al. 2004
H5244S-CO3 (S18)	Reverse	GCTTCAAATCCWAMGTGGTG	Yamauchi et al. 2004
H5252-CO3 (S19)	Reverse	CAWGCKGCVGCTTCAAATCC	Yamauchi et al. 2004
L6909-ND5 (S28)	Reverse	AAAAAGMAGGGCTTTAAAWAGWGCATG	Yamauchi et al. 2004
H7343-ND5 (S28)	Forward	TTATCWAATCGAGTGGGKGATGT	Yamauchi et al. 2004
L8321-ND4 (S32)	Forward	ATATGDGCHACAGAWGAGTAAGC	Yamauchi et al. 2004
H9302-Thr (S32)	Reverse	AAGAGGTTSTTTTGWWTTTGGTTTACAAGACC	Yamauchi et al. 2004
Ep-178	Reverse	GGGTGTTCAGATCGTTAGGGG	This study
Ep-1039	Reverse	TCTAAACCGAATCAATAAATATACCC	This study
Ep-1023	Forward	TATTGATTCGGTTTAGACCAGTATTG	This study
Ep-2250	Reverse	GCCCCCCTACCTTCATTTAC	This study
Ep-2789	Reverse	ACGGGGTTTTTCGGGTATC	This study
Ep-3161	Reverse	ATTTGATTTGGTTTTATATTTACTTTG	This study
Ep-3201	Forward	CAGCCAAAGTAACAAGAAGAGTG	This study
Ep-3229	Reverse	ATCTTCCACTCTTCTTGTTACTTTG	This study
Ep-4145	Reverse	GCTGAAGTCCTCCTCATCCTG	This study
Ep-4588	Reverse	AGCAAATAGTTCAAGTTACTTTAGGG	This study
Ep-4696	Forward	GTCTTTACCAATGAAGTTTTATAGGG	This study
Ep-5089	Reverse	ATTCCAAAAGTTTAATGTTCC	This study
Ep-5165	Reverse	ACCTAAACATTGTACCTCAGCG	This study
Ep-6588	Forward	TAAACTTCAGTCACTTCGCC	This study
Ep-7034	Forward	TTGGGGATCTTTTCGTTGC	This study
Ep-7504	Forward	GCGGGCTTATTTTACTTCTGCTAC	This study
Ep-7583	Reverse	GCAAAGTCCTTAGACCATC	This study
Ep-7954	Forward	TATTGTTTCTTCCCTTGGCTC	This study
Ep-8249	Forward	ATGGGGCAGTTTAGGGC	This study
Ep-8538	Forward	AGTTTTGCCCGCTGTG	This study
Ep-9633	Reverse	ATAGATACAGCCCCAACTCATAG	This study
Ep-10033	Reverse	AAAGCCCCTACCGAACTGG	This study
Ep-10275	Reverse	ACCTTCTATGGAAAAACGCTC	This study
Ep-10303	Reverse	ACTGCTCCTAACTCCACACACGGTC	This study
Ep-10403	Forward	TCATAGGTTGATAGAAGGAGGTC	This study
Ep-10726	Forward	TGTATTGGTGAGGAGGGGC	This study
Ep-11047	Forward	AGGCGGAGAAAGCACTG	This study

Primer	Orientation	Sequence (5'-3')	Reference
Ep-11545	Forward	AGCGATAGTCCTCCTCC	This study
Ep-12293	Reverse	ATTTCAATCCCCATCACTG	This study
Ep-12685	Reverse	TTTTGGAAGTTTACCCTCTTATAGTG	This study
Ep-12853	Forward	AAGGCAGGTCAAGGTGTAGCG	This study
Ep-12925	Forward	ATTAGATGGAGGGGCAAATGG	This study

b) Amplification and sequencing primers used to obtain the partial mt genome of *Sphaeroma serratum*.

Primer	Orientation	Sequence (5'-3')	Reference
16S2	Reverse	GCGACCTCGATGTTGGATTAA	Roehrdanz et al. 2002
Crust-12sf	Forward	CAGCAKYCGCGGTTAKAC	Podsiadlowski & Bartolomaeus 2005
Crust-12sr	Reverse	ACACCTACTWTGTTACGACTTATCTC	Podsiadlowski & Bartolomaeus 2005
Crust-16sr	Forward	CCGGTCTGAACTCAYATC	Podsiadlowski & Bartolomaeus 2005
Crust-cox1f	Forward	ACTAATCACAARGAYATTGG	Podsiadlowski & Bartolomaeus 2005
Crust-cox1r	Reverse	TAGTCTGAGTANCGTCGWGG	Podsiadlowski & Bartolomaeus 2005
Crust-cox3f	Forward	ATAATTCAATGATGACGAGA	Podsiadlowski & Bartolomaeus 2005
Crust-cox3r	Reverse	CCAATAATWACATGWAGACC	Podsiadlowski & Bartolomaeus 2005
Crust-CytB-f1	Reverse	CGAGATGTAAAYTAYGGSTGAC	This study
Crust-CytB-r	Forward	CTACGGGAGTGCACCRATYC	This study
Pera-Co1r	Reverse	AADGCTATATCAGGAGCCCCAATTATTAAAGG	This study
Pera-Co1r-seq	Reverse	GGAGCCCCAATTATTAAAGG	This study
Isop-16sf-range	Reverse	AARAAWGATTGCGACCTCGATGTTGAATTG	This study
Isop-16sr-range	Forward	TATGCTACCTTAGCACAGTYAGRATACTGCGGC	This study
L39-Met (S2)	Forward	AAGCTHVTGGGCTCATACCCC	Yamauchi et al. 2004
H718-ND2 (S2)	Reverse	AABCCHGDGAAMGGDGGHAVHCCHCC	Yamauchi et al. 2004
L329-ND2 (S5)	Forward	GGWGCHGCHCCNTTWCATTTTTG	Yamauchi et al. 2004
H1368-CO1 (S5)	Reverse	TATAGAGTTCCAATRTCYTTGTGATT	Yamauchi et al. 2004
L1384-CO1 (S8)	Forward	GGTCAACAAATCATAAAGATATTGG	Yamauchi et al. 2004
H2043-CO1 (S8)	Reverse	TAAACTTCAGGGTGACCAAAAAATCA	Yamauchi et al. 2004
H2619-CO1 (S9, S10, S11)	Reverse	GGTATWCCWGCKAGWCCTAAGAAATGTTG	Yamauchi et al. 2004
L1564-CO1 (S10)	Forward	ATGGTWATACCGATTWTRATTGG	Yamauchi et al. 2004
L2020S-CO1 (S11)	Forward	GACCCTGCGGGWGGRGGRGATCC	Yamauchi et al. 2004
L2302-CO1 (S12)	Forward	TTCCTACSGGRATTAAGATTTTTAG	Yamauchi et al. 2004
H3290-CO2 (S12)	Reverse	GGSATTATGTAWGAATCAAATT	Yamauchi et al. 2004
L3020-CO2 (S13, S14)	Forward	ATTTTTTTYCATGAYCATGC	Yamauchi et al. 2004
H3514-CO2 (S13)	Reverse	CCACAAATTTCKGAACATTGWCCATAAAA	Yamauchi et al. 2004
H4375-A6 (S14, S15)	Reverse	GCDATCATGTTDGCDGMWAGTCG	Yamauchi et al. 2004
L3542-CO2 (S15)	Forward	GGNCAATGTTCAGAAATTTGTGG	Yamauchi et al. 2004
L4672-CO3 (S18, S19)	Forward	GGWCTWGTBAAATGGTTTCA	Yamauchi et al. 2004
H5244S-CO3 (S18)	Reverse	GCTTCAAATCCWAMGTGGTG	Yamauchi et al. 2004
H5252-CO3 (S19)	Reverse	CAWGCKGCVGCTTCAAATCC	Yamauchi et al. 2004
L6369-ND5 (S26, S27)	Reverse	AGGDVWGGTAWGAATCATATWGAHCC	Yamauchi et al. 2004
H6995-ND5 (S26)	Forward	ATGATCTWAAAMGAGTWATTGC	Yamauchi et al. 2004
H7343-ND5 (S27)	Forward	TTATCWAATCGAGTGGGKGATGT	Yamauchi et al. 2004
L8519-ND4 (S33)	Forward	GGRGCCTCWACGTGGGCYTTAGG	Yamauchi et al. 2004
H9302-Thr (S33, S34)	Reverse	AAGAGGTTSTTTTGWWTTTGGTTTACAAGACC	Yamauchi et al. 2004
L8738-ND4 (S34)	Forward	CGTTCGGGTTGGTAHCCTCA	Yamauchi et al. 2004
L10061-CYB (S38)	Reverse	GGATTWTTTTTAGCKATRCATTACAC	Yamauchi et al. 2004
H10699-CYB (S38)	Forward	GCAAATAGAAAATATCATTCWGGTTG	Yamauchi et al. 2004
L12167-16S (S43)	Forward	CCTGGCTTACGCCGGTCTGAACTCAGATCATG	Yamauchi et al. 2004
H12663-16S (S43)	Reverse	CGCCTGTTTACCAAAAACAT	Yamauchi et al. 2004
L13337-12S (S48)	Reverse	YCTACTWTGYTACGACTTATCTC	Yamauchi et al. 2004
H13845-12S (S48)	Forward	GTGCCAGCAGCTGCGGTTA	Yamauchi et al. 2004
Ss-141	Reverse	TCGGGACCCCAAGAGCC	This study
Ss-175	Forward	AAATATGGATAATAACTAAACCTAGAGCG	This study
Ss-405	Reverse	TTATCGTTATAGCCGCCGC	This study

Primer	Orientation	Sequence (5'-3')	Reference
Ss-470	Forward	TCCGATGTGAAGGTAGAGGC	This study
Ss-3980	Forward	CGAAGAGAGAGAAGTATTATCATTGC	This study
Ss-4527	Reverse	TAAAAATAAGATGTACGAAAACCAC	This study
Ss-4797	Reverse	CGAGCAACAAATACAGCACG	This study
Ss-5244	Reverse	ACAAAATCCACCTAAACGTAATCAAC	This study
Ss-5259	Forward	ACCGCCCCAGTAAAACCC	This study
Ss-5295	Forward	TAAGTAGATCGTTTGAAGGG	This study
Ss-5525	Reverse	TGAAACTACTCACGGTCTGTC	This study
Ss-5560	Forward	TGCCGAGTTCTTTGACGC	This study
Ss-7206	Reverse	CCTCTTCCTCACACTCCAACG	This study
Ss-7375	Forward	TTTGTATTTTATTTTCGGGGC	This study
Ss-7673	Reverse	ACCCTCTCCTCAGCAATAATG	This study
Ss-7759	Forward	TGGTATTTTTTCGCTTCATTTAG	This study
Ss-8684	Reverse	ACACGCTGTTCCAAAACCC	This study
Ss-8970	Reverse	GAGAGACCCCGTTTTGTAGACC	This study
Ss-9347	Forward	CGTGGTCTTGCCCATACAGG	This study
Ss-9796	Forward	TCGTAAGAGTGGTTGGG	This study
Ss-10173	Forward	TGGTATTTACGGCGAGGC	This study
Ss-10712	Forward	TAGTTCAATGGTGACGAGATG	This study
Ss-11471	Reverse	TAACCAAGAACGCCAACC	This study
Ss-12310	Reverse	AGCCTCGCTATCACCC	This study
Ss-12797	Reverse	GCAGTCTGTAACAAAAACCG	This study
Ss-12861	Forward	TGAGTTCTTCAGAGAGCACCC	This study
Ss-13180	Reverse	ACTACTACCCTCTCTGCCCG	This study
Ss-13217	Forward	CCAGCACAGCCAGAGGAGAC	This study
Ss-13354	Reverse	GCACCTTGATCTGACATACTGG	This study
Ss-13420	Reverse	TCGTCCTGACTGCCCTCG	This study

c) Amplification and sequencing primers used to obtain the partial mt genome of *Glyptonotus* cf. *antarcticus*.

Primer	Orientation	Sequence (5'-3')	Reference
N4	Forward	GGAGCTTCAACATGAGCTTT	Roehrdanz et al. 2002
Crust-12sf	Forward	CAGCAKYCGCGGTTAKAC	Podsiadlowski & Bartolomaeus 2005
Crust-12sr	Reverse	ACACCTACTWTGTTACGACTTATCTC	Podsiadlowski & Bartolomaeus 2005
Crust-16sf	Reverse	TGACYGTGCDAAGGTAGC	Podsiadlowski & Bartolomaeus 2005
Crust-16sr	Forward	CCGGTCTGAACTCAYATC	Podsiadlowski & Bartolomaeus 2005
Crust-16sf2	Reverse	GCGACCTCGATGTTGGATTAA	This study
Crust-16sr2	Forward	CCGGTCTGAACTCAYGTA	This study
Crust-cox3r	Reverse	CCAATAATWACATGWAGACC	Podsiadlowski & Bartolomaeus 2005
Crust-CytB-f1	Reverse	CGAGATGTAAAYTAYGGSTGAC	This study
Crust-CytB-r	Forward	CTACGGGAGTGCACCRATYC	This study
Crust-N5r	Reverse	AAAGAGCCTTAAATAAAGCATG	Podsiadlowski & Bartolomaeus 2005
HCO2198	Reverse	TAAACTTCAGGGTGACCAAAAAATCA	Folmer t al. 1994
Pera-Co1f	Forward	GGNGCTATTACAATACTWTTGACAGACCG	This study
Pera-Co1r	Reverse	AADGCTATATCAGGAGCCCCAATTATTAAAGG	This study
L39-Met(S1,S2,S3)	Forward	AAGCTHVTGGGCTCATACCCC	Yamauchi et al. 2004
L329-ND2 (S4,S5,S6)	Forward	GGWGCHGCHCCNTTWCATTTTTG	Yamauchi et al. 2004
H718-ND2 (S1,S4)	Reverse	AABCCHGDGAAMGGDGGHAVHCCHCC	Yamauchi et al. 2004
H1368-CO1 (S2,S5)	Reverse	TATAGAGTTCCAATRTCYTTGTGATT	Yamauchi et al. 2004
L1384-CO1 (S8)	Reverse	GGTCAACAAATCATAAAGATATTGG	Yamauchi et al. 2004
H1420-CO1 (S6))	Reverse	AGTGCCWACTATWCCWGMTCA	Yamauchi et al. 2004
H1602-CO1 (S7)	Reverse	GGGAADGCTATGTCWGGGGC	Yamauchi et al. 2004
H3514-CO2	Reverse	CCACAAATTTCKGAACATTGWCCATAAAA	Yamauchi et al. 2004
L5170-CO3	Forward	TTGTDGCNACHGGMTTTCATGG	Yamauchi et al. 2004
L5526-ND3	Forward	CCHTWTGAGTGTGGWTTTGATCC	Yamauchi et al. 2004
H10104-CYB	Forward	AAGTCANCCGTAGTTTACGTCWCG	Yamauchi et al. 2004
H10699-CYB	Forward	GCAAATAGAAAATATCATTCWGGTTG	Yamauchi et al. 2004

Primer	Orientation	Sequence (5'-3')	Reference
Gly-895	Reverse	TAAAATAAATAAACCGCCTCTTG	This study
Gly-1052	Forward	AGCGGGAGTTATATGGGAG	This study
Gly-1274	Reverse	CTAAAAATAAAACCTTGTAGACCTC	This study
Gly-1275	Forward	TGGGGGTGTTGACTAATCTG	This study
Gly-1866	Forward	ATTTAGCACTGTGTGGATTTC	This study
Gly-2791	Reverse	AGGTGCTTCTTCTTTCTTATTCAG	This study
Gly-3086	Reverse	ACGGATTTTGCTCTTCTGG	This study
Gly-3648	Reverse	ACATTTATTTTAATCTTAGGCTGAGG	This study
Gly-3780	Forward	TTTGTCATCTTGCTTGAAGTATTAG	This study
Gly-4233	Forward	AAATCTAATGAAACCCATCGTG	This study
Gly-4719	Reverse	TATTTTCCCTGAGCAGTAAGTC	This study
Gly-5150	Reverse	CCAAAAATGATTACCACCTCG	This study
Gly-5191	Forward	AGTAACTTGTTCTTAATTCTTTACATAGG	This study
Gly-5595	Forward	TGCCGAGTTAATTTCTTTCTTC	This study
Gly-6534	Reverse	TAAAGATGAACAAGCCTGAACTAAG	This study
Gly-7778	Reverse	TACAAACCCTCTCCCGAGTAATAG	This study
Gly-7853	Forward	TGATATGGGAATCTTTTCTTTACAC	This study
Gly-8203	Forward	AGGAAGCCAGAAAGAAAGAGTC	This study
Gly-9243	Reverse	AATAATAGCGGGCAAGGC	This study
Gly-9415	Forward	ATGTCTGGCGGATTTCGTC	This study
Gly-9719	Forward	TGGCTGATTTAGGTGTGAG	This study
Gly-10639	Reverse	AGAGTTCTTAGGATAGCAAATAC	This study
Gly-11459	Reverse	ACATATTTGGCTATCTCCCCC	This study
Gly-12035	Reverse	ATTATTCATTCTTATTTTTTAGTCTTC	This study
Gly-12081	Forward	AACACTTTTCATGCCAATCTTATC	This study
Gly-12506	Forward	TTCATAGGAAATAGTTTGAGCCAC	This study
Gly-13015	Reverse	TTTGGCTTTACCTTGTTTGGC	This study
Gly-13423	Reverse	TCAGAGAATAGTGGGGTATCTAATCC	This study

d) Amplification and sequencing primers used to obtain the partial mt genome of *Armadillidium vulgare*.

Primer	Orientation	Sequence (5'-3')	Reference
16S2	Reverse	GCGACCTCGATGTTGGATTAA	Roehrdanz et al. 2002
CB2H	Forward	TCCTCAAAATGATATTTGTCCTCA	Roehrdanz et al. 2002
N4	Forward	GGAGCTTCAACATGAGCTTT	Roehrdanz et al. 2002
N4(87)	Reverse	TCAGCTAATATAGCAGCTCC	Roehrdanz et al. 2002
Crust-12sf	Forward	CAGCAKYCGCGGTTAKAC	Podsiadlowski & Bartolomaeus 2005
Crust-12sr	Reverse	ACACCTACTWTGTTACGACTTATCTC	Podsiadlowski & Bartolomaeus 2005
Crust-cox1f	Forward	ACTAATCACAARGAYATTGG	Podsiadlowski & Bartolomaeus 2005
Crust-cox1r	Reverse	TAGTCTGAGTANCGTCGWGG	Podsiadlowski & Bartolomaeus 2005
L39-Met (S1,S2)	Forward	AAGCTHVTGGGCTCATACCCC	Yamauchi et al. 2004
L329-ND2 (S4,S5,S7)	Forward	GGWGCHGCHCCNTTWCATTTTTG	Yamauchi et al. 2004
H718-ND2 (S1,S4)	Reverse	AABCCHGDGAAMGGDGGHAVHCCHCC	Yamauchi et al. 2004
H1368-CO1 (S2,S5)	Reverse	TATAGAGTTCCAATRTCYTTGTGATT	Yamauchi et al. 2004
L1384-CO1 (S8,S9)	Forward	GGTCAACAAATCATAAAGATATTGG	Yamauchi et al. 2004
L1564-CO1 (S10)	Forward	ATGGTWATACCGATTWTRATTGG	Yamauchi et al. 2004
H1602-CO1 (S7)	Reverse	GGGAADGCTATGTCWGGGGC	Yamauchi et al. 2004
H2043-CO1 (S8)	Reverse	TAAACTTCAGGGTGACCAAAAAATCA	Yamauchi et al. 2004
L2020S-CO1 (S11)	Forward	GACCCTGCGGGWGGRGGRGATCC	Yamauchi et al. 2004
H2619-CO1 (S9-11)	Reverse	GGTATWCCWGCKAGWCCTAAGAAATGTTG	Yamauchi et al. 2004
L3020-CO2 (S13)	Forward	ATTTTTTTYCATGAYCATGC	Yamauchi et al. 2004
H3514-CO2 (S13)	Reverse	CCACAAATTTCKGAACATTGWCCATAAAA	Yamauchi et al. 2004
L4672-CO3 (S18, S19)	Forward	GGWCTWGTBAAATGGTTTCA	Yamauchi et al. 2004
H5244S-CO3 (S18)	Reverse	GCTTCAAATCCWAMGTGGTG	Yamauchi et al. 2004
H5252-CO3 (S19)	Reverse	CAWGCKGCVGCTTCAAATCC	Yamauchi et al. 2004
L12167-16S (S43)	Forward	CCTGGCTTACGCCGGTCTGAACTCAGATCATG	Yamauchi et al. 2004
H12663-16S (S43)	Reverse	CGCCTGTTTACCAAAAACAT	Yamauchi et al. 2004
Av-320	Forward	GCTGGGTTATTTTTCCGCAC	This study

Primer	Orientation	Sequence (5'-3')	Reference
Av-439	Reverse	ATAGACTTATTCATTTATTTTCTTTCTCTC	This study
Av-871	Reverse	TCAGTAACAATGCAGAATCACC	This study
Av-1166	Forward	CTAGATTAAGGGCATTATTTGAGC	This study
Av-1616	Forward	ATGTTTCTTTAATAAGGGATGAGG	This study
Av-2618	Reverse	ATTTTGCTCTTCTGGTTTATTTTG	This study
Av-3101	Reverse	GTCTCTTCCGTTACTTGTTGC	This study
Av-3375	Forward	AAGGATAAAGTATCTGTAAATAGACC	This study
Av-4209	Reverse	TATTAAACCCTCCTAAATATACATC	This study
Av-4434	Forward	AAAATTAATACTTAAAAATTCCTAACG	This study
Av-4757	Forward	AAGGTTTTAGGGTCTTATCGTC	This study
Av-4903	Forward	CACAGTTAGGATACTGCGGCTC	This study
Av-5618	Reverse	GATAAAAAATATAAAACCCCAAAATG	This study
Av-6048	Reverse	CTACAACTCCGCCAACAATAATAC	This study
Av-6306	Reverse	CTAAGGATAAGAAACTAACTAACAAGG	This study
Av-6410	Forward	TGTGGTTACTTTATTTTATTATTTACG	This study
Av-6554	Forward	AGTTTTATTTTTTCATAAGGTTTTTGC	This study
Av-6872	Reverse	TTACTATAAAAAAAATTATTACAAAAGC	This study
Av-6995	Reverse	ATAAAGTTAAAGAAGGGGGTAAAAGTC	This study
Av-7022	Forward	AGGAGTAGGAACAGGGTGGACAG	This study
Av-7032	Reverse	TTCCTACTCCTCTTTCTAACTAATCC	This study
Av-7531	Reverse	TGGGCTCATACTACAAAACCTAAC	This study
Av-7607	Forward	TACGGGTATTAAAATCTTTAGGTG	This study
Av-7621	Reverse	ATTTTAATACCCGTAGGAACAGC	This study
Av-8140	Forward	AGTGGTTACACCCTTATCCCC	This study
Av-8377	Reverse	ATACCCTACTAAGGATAAAATTAAAAC	This study
Av-9027	Forward	AGGTTCCACAAATAGGTCCTTTAC	This study
Av-9066	Reverse	AATAAAAAAACTCACGGTAAAGGAC	This study
Av-9392	Reverse	AAAATTACAGCAACTATCTTTCTACTTC	This study
Av-9788	Reverse	GCTACTGCTACTTCTAACACCAC	This study
Av-10184	Reverse	CTCTGACCTTGGTGATAATCTTCTATG	This study
Av-10202	Reverse	AGGTCATACTCTGCCTATCTCTG	This study
Av-10357	Forward	TTGGGGTTTACTTTACAACTCTAC	This study
Av-10996	Forward	GGTGTTCTACGAATGGTGGG	This study
Av-11020	Reverse	CCCTTCCCACCATTCGTAG	This study
Av-11383	Reverse	CTGGATTTAACACAGAGTATAGCG	This study
Av-11669	Forward	TAGCTGATAGACAGCATAAAATAAAG	This study
Av-12039	Reverse	GCCTCTATGGTTTGGCTTTAC	This study
Av-12513	Reverse	AAGGTAAGATTTATCGTGGGG	This study
Av-12595	Forward	GCTTTTAGAGGAGAAGATGGG	This study
Av-12925	Reverse	ACCTCCTCTGCTATGAAAAAACAGGATGC	This study

e) Amplification and sequencing primers used to obtain the partial mt genome of *Asellus aquaticus*.

Primer	Orientation	Sequence (5'-3')	Reference
N4	Forward	GGAGCTTCAACATGAGCTTT	Roehrdanz et al. 2002
Crust-12sf	Forward	CAGCAKYCGCGGTTAKAC	Podsiadlowski & Bartolomaeus 2005
Crust-12sr	Reverse	ACACCTACTWTGTTACGACTTATCTC	Podsiadlowski & Bartolomaeus 2005
Crust-16sf	Reverse	TGACYGTGCDAAGGTAGC	Podsiadlowski & Bartolomaeus 2005
Crust-16sr	Forward	CCGGTCTGAACTCAYATC	Podsiadlowski & Bartolomaeus 2005
Crust-cox1f	Forward	ACTAATCACAARGAYATTGG	Podsiadlowski & Bartolomaeus 2005
Crust-cox1r	Reverse	TAGTCTGAGTANCGTCGWGG	Podsiadlowski & Bartolomaeus 2005
Crust-cox3r	Reverse	CCAATAATWACATGWAGACC	Podsiadlowski & Bartolomaeus 2005
Crust-ND5f	Reverse	AGAATTCTACTAGGDTGRGATGG	Podsiadlowski & Bartolomaeus 2005
L329-ND2 (S5,S6,S7)	Forward	GGWGCHGCHCCNTTWCATTTTTG	Yamauchi et al. 2004
H1368-CO1 (S5)	Reverse	TATAGAGTTCCAATRTCYTTGTGATT	Yamauchi et al. 2004
H1410-CO1 (S6)	Reverse	AGTGCCWACTATWCCWGMTCA	Yamauchi et al. 2004
H1602-CO1 (S7)	Reverse	GGGAADGCTATGTCWGGGGC	Yamauchi et al. 2004
L2302-CO1 (S12)	Forward	TTCCTACSGGRATTAAGATTTTTAG	Yamauchi et al. 2004

Primer	Orientation	Sequence (5'-3')	Reference
H3290-CO2 (S12)	Reverse	GGSATTATGTAWGAATCAAATT	Yamauchi et al. 2004
L3020-CO2 (S13)	Forward	ATTTTTTTYCATGAYCATGC	Yamauchi et al. 2004
H3514-CO2 (S13)	Reverse	CCACAAATTTCKGAACATTGWCCATAAAA	Yamauchi et al. 2004
L3542-CO2 (S15)	Forward	GGNCAATGTTCAGAAATTTGTGG	Yamauchi et al. 2004
H4375-A6 (S15)	Reverse	GCDATCATGTTDGCDGMWAGTCG	Yamauchi et al. 2004
H10315-CYB	Forward	GTGATDACHGTDGCTCCTCA	Yamauchi et al. 2004
Aa-1082	Reverse	TTATTAACCCAGCCCCATC	This study
Aa-1384	Reverse	CCAATAAGGCGATAGATAGAAGG	This study
Aa-1654	Forward	ACCTTTCTTACTACCTTTATGGC	This study
Aa-2113	Forward	TGCCCTTCACTGCCCTC	This study
Aa-3129	Reverse	TGGGCTGTTTTGTATCGC	This study
Aa-3297	Reverse	AGGTGTCTATGTAAGGATGGC	This study
Aa-3710	Reverse	TCGCTTATCCCAACTTTTATGC	This study
Aa-3885	Forward	GGAATACCCTACACTATCCCAGC	This study
Aa-4295	Reverse	AGTTGGCGTAGTGTCCTTTGTG	This study
Aa-4387	Forward	ATTTTTACTAAACTATCTTCTGCGTGG	This study
Aa-4803	Reverse	ACGGGTAGACTAAAGGTGAAAAAAG	This study
Aa-5188	Forward	TGCTGTTATCCCTAAAGTAACTTATTC	This study
Aa-5296	Reverse	TTGAGCTGGGGCGGTATTC	This study
Aa-5985	Forward	TAAATACCCCTGATACAAAAG	This study
Aa-6386	Reverse	TGTTAATATGATAATAGGTAAAAATG	This study
Aa-6830	Reverse	ATAGGTGATAAGGTAGAGAGTCCAAG	This study
Aa-7139	Forward	ACCTCTCTTTCTTATACTAGCACTCC	This study
Aa-7632	Reverse	GGGAGGGAGAAGCCAGAAAC	This study
Aa-7773	Forward	CGCTAGTTCTATTTTAGGCTCTG	This study
Aa-8270	Reverse	ATTTTAATACCTGTTGGGACTGC	This study
Aa-8324	Forward	TTATATGAGCCCTCGGGTTTG	This study
Aa-9396	Forward	CGAGTCACCCTTCCACAATC	This study
Aa-10100	Forward	CCATTCTATTACCACATAAATCTAAAC	This study
Aa-10297	Forward	ATAATAGCCCACCTTGTGCC	This study
Aa-11026	Forward	TGACCTGAGCACACCATAGAATC	This study
Aa-11352	Forward	TACACCTTTATCTATTGATGGGG	This study
Aa-11249	Reverse	ATAGTTCAGGTCCGTTTGC	This study
Aa-13084	Reverse	ATTTACTTACTTTTCTTATTTAGCAGG	This study
Aa-13307	Reverse	TTGGGTTTGTTGGGCGTATC	This study
Aa-13378	Reverse	AGTAGGGTATAGGGTGGATTGTC	This study

f) Amplification and sequencing primers used to obtain the partial mt genome of *Janira maculosa*.

Primer	Orientation	Sequence (5'-3')	Reference
16S2	Reverse	GCGACCTCGATGTTGGATTAA	Roehrdanz et al. 2002
CB2H	Forward	TCCTCAAAATGATATTTGTCCTCA	Roehrdanz et al. 2002
N4	Forward	GGAGCTTCAACATGAGCTTT	Roehrdanz et al. 2002
Crust-cox1f	Forward	ACTAATCACAARGAYATTGG	Podsiadlowski & Bartolomaeus 2005
Crust-cox1r	Reverse	TAGTCTGAGTANCGTCGWGG	Podsiadlowski & Bartolomaeus 2005
Crust-N5f	Forward	AGAATTCTACTAGGDTGRGATGG	Podsiadlowski & Bartolomaeus 2005
Crust-N5r	Reverse	AAAGAGCCTTAAATAAAGCATG	Podsiadlowski & Bartolomaeus 2005
Crust-12sf	Forward	CAGCAKYCGCGGTTAKAC	Podsiadlowski & Bartolomaeus 2005
Crust-12sr	Reverse	ACACCTACTWTGTTACGACTTATCTC	Podsiadlowski & Bartolomaeus 2005
Crust-16sr	Forward	CCGGTCTGAACTCAYATC	Podsiadlowski & Bartolomaeus 2005
Pera-Co1f	Forward	GGNGCTATTACAATACTWTTGACAGACCG	This study
L39-Met (S2)	Forward	AAGCTHVTGGGCTCATACCCC	Yamauchi et al. 2004
L329-ND2 (S5,S6,S7)	Forward	GGWGCHGCHCCNTTWCATTTTTG	Yamauchi et al. 2004
H1368-CO1 (S2,S5)	Reverse	TATAGAGTTCCAATRTCYTTGTGATT	Yamauchi et al. 2004
L1384-CO1 (S8,S9)	Forward	GGTCAACAAATCATAAAGATATTGG	Yamauchi et al. 2004
L1564-CO1 (S10)	Forward	ATGGTWATACCGATTWTRATTGG	Yamauchi et al. 2004
H1602-CO1 (S3,S7)	Reverse	GGGAADGCTATGTCWGGGGC	Yamauchi et al. 2004
H2043-CO1 (S8)	Reverse	TAAACTTCAGGGTGACCAAAAAATCA	Yamauchi et al. 2004
H2619-CO1 (S9-11)	Reverse	GGTATWCCWGCKAGWCCTAAGAAATGTTG	Yamauchi et al. 2004

Primer	Orientation	Sequence (5'-3')	Reference
Jm-91	Reverse	TTATCTCATGCTAATGGTGCTTCTC	This study
Jm-218	Reverse	TCTTTGTCTTGCTATTCAGATTGTTAC	This study
Jm-1007	Reverse	CTTTTGCTAGTTTTCTTTGTTACCG	This study
Jm-1367	Reverse	ATTATTATCTTATTTTTTATACAGTATTTGC	This study
Jm-1989	Forward	ACAAAGCAAAAAAAAGGTTC	This study
Jm-2073	Forward	ATTCTTAAATAAACTACCTTGAAAAC	This study
Jm-3017	Forward	AAAGTTTTACAGGGTCTTATCGTCTTC	This study
Jm-3717	Reverse	ATAAACACTGTAAAGGTTTCAAAATTAG	This study
Jm-4137	Reverse	TACTAAACATAAAGACCCAGTTGATTG	This study
Jm-6026	Forward	GGGACTGACAGGAGTAGTTTTAGC	This study
Jm-7118	Reverse	TGATGTTCTAACAAATGCCG	This study
Jm-7317	Forward	TGAAGATACAAAAGAGAGCC	This study
Jm-8212	Reverse	TTGTTCTAATAACCTGTGATGAGC	This study
Jm-8265	Forward	ATTTTACACGAGTCCAAGTCATAG	This study
Jm-8953	Reverse	CTCACTCATATACAGTACCCACCC	This study
Jm-9555	Reverse	TTTAGGTTTGATGGTTATCTCATTTAG	This study
Jm-9545	Forward	TCAAACCTAAAGAATGTGGCGGC	This study
Jm-9642	Forward	ACCGTCGTCTAAGTTGGCTCTGG	This study
Jm-9856	Reverse	AGAGAYTGACGGGCGATATG	This study

Supplementary file 4.2

a) Method of tRNA detection in the *Eurydice pulchra* mt genome. 16 tRNA genes were detected. Standard search with tRNAscan-SE performed with 'Invertebrate Mito' genetic code and 'Mito/Chloroplast' model. When no tRNAs were detected, the less strict 'Nematode Mito' model was used instead. ARWEN was used with 'Invertebrate mitochondrial genetic code' option. Results were manually checked afterwards. Noticeable features in bold letters.

tRNA	**tRNAscan-SE 1.21 (software)**	**ARWEN 1.2.2.c (software)**	**manual**
tRNA-Ala	detected	detected	accepted
tRNA-Asn	detected	detected	accepted
tRNA-Asp	detected ('Nematode Mito' model)	detected	accepted
tRNA-Arg	-	-	-
tRNA-Cys	not detected	detected	accepted
tRNA-Gln	not detected	detected	accepted
tRNA-Glu	-	-	-
tRNA-Gly	detected	detected	accepted
tRNA-His	not detected	not detected	detected
tRNA-Ile	-	-	-
tRNA-Leu(CUN)	-	-	-
tRNA-Leu(UUR)	detected	detected	accepted
tRNA-Lys	detected	detected	accepted
tRNA-Met	detected	detected	accepted
tRNA-Phe	-	-	-
tRNA-Pro	detected	detected	accepted
tRNA-Ser(AGY)	-	-	-
tRNA-Ser(UCN)	not detected	detected	accepted
tRNA-Thr	detected	detected	accepted
tRNA-Trp	detected	detected	accepted
tRNA-Tyr	detected	detected	accepted
tRNA-Val	detected	detected	accepted

- Genes most likely missing by incomplete mt genome sequences

b) Method of tRNA detection in the *Sphaeroma serratum* mt genome. 15 tRNA genes were detected. Standard search with tRNAscan-SE performed with 'Invertebrate Mito' genetic code and 'Mito/Chloroplast' model. When no tRNAs were detected, the less strict 'Nematode Mito' model was used instead. ARWEN was used with 'Invertebrate mitochondrial genetic code' option. Results were manually checked afterwards. Noticeable features in bold letters.

tRNA	**tRNAscan-SE 1.21 (software)**	**ARWEN 1.2.2.c (software)**	**manual**
tRNA-Ala	detected	detected	accepted
tRNA-Asn	not detected	detected	accepted
tRNA-Asp	detected	detected	accepted
tRNA-Arg	-	-	-
tRNA-Cys	-	-	-
tRNA-Gln	detected	detected	accepted
tRNA-Glu	-	-	-
tRNA-Gly	detected	detected	accepted
tRNA-His	detected	detected	accepted
tRNA-Ile	-	-	-
tRNA-Leu(CUN)	-	-	-
tRNA-Leu(UUR)	detected	detected	accepted
tRNA-Lys	detected	detected	accepted
tRNA-Met	detected	detected	accepted
tRNA-Phe	detected ('Nematode Mito' model)	detected	accepted
tRNA-Pro	detected	detected	accepted
tRNA-Ser(AGY)	-	-	-
tRNA-Ser(UCN)	detected	detected	accepted
tRNA-Thr	detected	detected	accepted
tRNA-Trp	-	-	-
tRNA-Tyr	detected	detected	accepted
tRNA-Val	detected	detected	accepted

- Genes most likely missing by incomplete mt genome sequences

c) Method of tRNA detection in the *Glyptonotus* cf. *antarcticus* mt genome (incomplete). 18 tRNA genes were detected. Standard search with tRNAscan-SE performed with 'Invertebrate Mito' genetic code and 'Mito/Chloroplast' model. When no tRNAs were detected, the less strict 'Nematode Mito' model was used instead. ARWEN was used with 'Invertebrate mitochondrial genetic code' option. Results were manually checked afterwards. Noticeable features in bold letters.

tRNA	**tRNAscan-SE 1.21 (software)**	**ARWEN 1.2.2.c (software)**	**manual**
tRNA-Ala	detected	detected	accepted
tRNA-Asn	detected	detected	accepted
tRNA-Asp	detected	detected	accepted
tRNA-Arg	detected	detected	accepted
tRNA-Cys	detected ('Nematode Mito' model)	detected	accepted
tRNA-Gln	detected ('Nematode Mito' model)	detected	accepted
tRNA-Glu	-	-	-
tRNA-Gly	detected	detected	accepted
tRNA-His	detected	detected	accepted
tRNA-Ile	-	-	-
tRNA-Leu(CUN)	detected	detected	accepted
tRNA-Leu(UUR)	detected	detected	accepted
tRNA-Lys	detected	detected	accepted
tRNA-Met	detected	detected	accepted
tRNA-Phe	not detected	detected	accepted
tRNA-Pro	detected	detected	accepted
tRNA-Ser(AGY)	-	-	-
tRNA-Ser(UCN)	detected	detected	accepted
tRNA-Thr	detected	detected	accepted
tRNA-Trp	-	-	-
tRNA-Tyr	detected	detected	accepted
tRNA-Val	detected	detected	accepted

- Genes most likely missing by incomplete mt genome sequences

d) Method of tRNA detection in the *Armadillidium vulgare* mt genome (incomplete). 13 RNA genes were detected. Standard search with tRNAscan-SE performed with 'Invertebrate Mito' genetic code and 'Mito/Chloroplast' model. When no tRNAs were detected, the less strict 'Nematode Mito' model was used instead. ARWEN was used with 'Invertebrate mitochondrial genetic code' option. Results were manually checked afterwards. Noticeable features in bold letters

tRNA	**tRNAscan-SE 1.21 (software)**	**ARWEN 1.2.2.c (software)**	**manual**
tRNA-Ala	-	-	-
tRNA-Asn	-	-	-
tRNA-Asp	detected	not detected	accepted
tRNA-Arg	-	-	-
tRNA-Cys	not detected	detected	accepted
tRNA-Gln	detected ('Nematode Mito' model)	not detected	improved
tRNA-Glu	-	-	-
tRNA-Gly	detected	detected	accepted
tRNA-His	detected ('Nematode Mito' model)	detected	accepted
tRNA-Ile	-	-	-
tRNA-Leu(CUN)	-	-	-
tRNA-Leu(UUR)	-	-	-
tRNA-Lys	-	-	-
tRNA-Met	detected	not detected	accepted
tRNA-Phe	detected ('Nematode Mito' model)	not detected	accepted
tRNA-Pro	detected	not detected	improved
tRNA-Ser(AGY)	-	-	-
tRNA-Ser(UCN)	detected	not detected	accepted
tRNA-Thr	detected ('Nematode Mito' model)	detected	accepted
tRNA-Trp	detected ('Nematode Mito' model)	detected	accepted
tRNA-Tyr	detected ('Nematode Mito' model)	detected	accepted
tRNA-Val	detected	detected	accepted

- Genes most likely missing by incomplete mt genome sequences

e) Method of tRNA detection in the *Asellus aquaticus* mt genome (incomplete). 18 tRNA genes were detected. Standard search with tRNAscan-SE performed with 'Invertebrate Mito' genetic code and 'Mito/Chloroplast' model. When no tRNAs were detected, the less strict 'Nematode Mito' model was used instead. ARWEN was used with 'Invertebrate mitochondrial genetic code' option. Results were manually checked afterwards. Noticeable features in bold letters.

tRNA	**tRNAscan-SE 1.21 (software)**	**ARWEN 1.2.2.c (software)**	**manual**
tRNA-Ala	detected	detected	accepted
tRNA-Asn	detected	detected	accepted
tRNA-Asp	detected	detected	accepted
tRNA-Arg	detected	detected	accepted
tRNA-Cys	not detected	detected	accepted
tRNA-Gln	detected	detected	accepted
tRNA-Glu	-	-	-
tRNA-Gly	detected	detected	accepted
tRNA-His	detected	detected	accepted
tRNA-Ile	-	-	-
tRNA-Leu(CUN)	detected	detected	accepted
tRNA-Leu(UUR)	detected	detected	accepted
tRNA-Lys	detected	detected	accepted
tRNA-Met	detected	detected	accepted
tRNA-Phe	detected	detected	accepted
tRNA-Pro	detected	detected	accepted
tRNA-Ser(AGY)	-	-	-
tRNA-Ser(UCN)	detected	detected	accepted
tRNA-Thr	detected	detected	accepted
tRNA-Trp	-	-	-
tRNA-Tyr	detected	detected	accepted
tRNA-Val	detected	detected	accepted

- Genes most likely missing by incomplete mt genome sequences

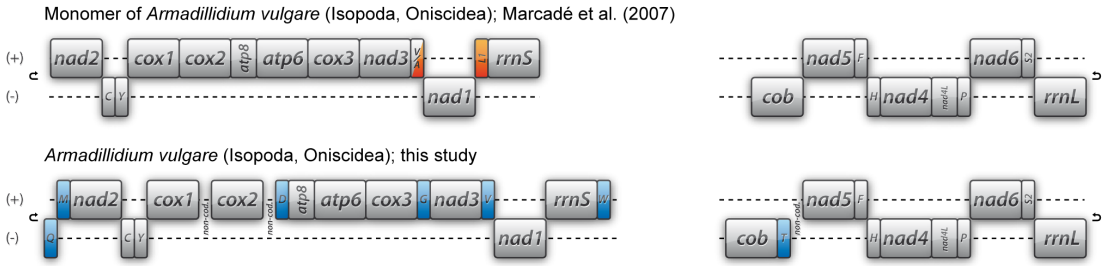
f) Method of tRNA detection in the *Janira maculosa* mt genome (incomplete). 14 tRNA genes were detected. Standard search with tRNAscan-SE performed with 'Invertebrate Mito' genetic code and 'Mito/Chloroplast' model. When no tRNAs were detected, the less strict 'Nematode Mito' model was used instead. ARWEN was used with 'Invertebrate mitochondrial genetic code' option. Results were manually checked afterwards. Noticeable features in bold letters.

tRNA	**tRNAscan-SE 1.21 (software)**	**ARWEN 1.2.2.c (software)**	**manual**
tRNA-Ala	detected	detected	accepted
tRNA-Asn	-	-	-
tRNA-Asp	detected	detected	accepted
tRNA-Arg	-	-	-
tRNA-Cys	-	detected	accepted
tRNA-Gln	detected ('Nematode Mito' model)	detected	accepted
tRNA-Glu	-	-	-
tRNA-Gly	detected	detected	accepted
tRNA-His	-	detected	accepted
tRNA-Ile	-	-	-
tRNA-Leu(CUN)	detected	detected	accepted
tRNA-Leu(UUR)	detected ('Nematode Mito' model)	detected	accepted
tRNA-Lys	-	-	-
tRNA-Met	detected ('Nematode Mito' model)	detected	accepted
tRNA-Phe	-	-	-
tRNA-Pro	detected ('Nematode Mito' model)	detected	accepted
tRNA-Ser(AGY)	-	-	-
tRNA-Ser(UCN)	detected	detected	accepted
tRNA-Thr	detected	detected	accepted
tRNA-Trp	-	-	-
tRNA-Tyr	detected	detected	accepted
tRNA-Val	detected ('Nematode Mito' model)	detected	accepted

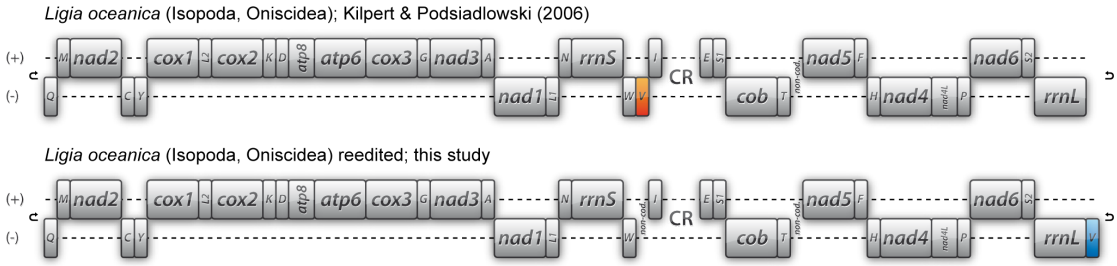
- Genes most likely missing by incomplete mt genome sequences

Supplementary file 4.3

a)



b)



Reannotations of the *Armadillidium vulgare* and the *Ligia oceanica* mt genome. The software tool ARWEN (Laslett and Canback 2008) was used in addition to tRNAscan-SE (Lowe and Eddy 1997) for an improved identification of tRNA genes. The figure confronts the old with the new annotation. Deprecated genes are colored in orange, new additions in blue. a) The presented annotation of *A. vulgare* of this study is based on our own sequence data. Seven further tRNA genes (*trnQ*, *trnM*, *trnD*, *trnG*, *trnV*, *trnW*, *trnT*) were identified by ARWEN. No indications have been found for *trnA* at the same locus of *trnV*. *trnL1* was not detected by any of the computer programs, although our sequence for this area is no different to the other published sequence. b) The new *trnV* gene identified by ARWEN is more convincing than the old one, which had an odd nucleotide in the anticodon.

Supplementary file 4.4

a) Gene content of the mitochondrial genome of *Eurydice pulchra*.

Gene	Strand	GenBank position no.	Size (nts)	Size (aa)	Start codon	Stop codon	Intergenic nucleotides
cob [a]	-	1-303 [a]	303 [a]	101 [a]	ATA	? [a]	0
trnT	-	304-363	60				1
nad5	+	365-2074	1710	569	ATT	TAG	-1
trnH	-	2074-2135	62				-26
nad4	-	2110-3462	1353	450	ATT	TAA	-13
nad4L	-	3450-3743	294	97	ATC	TAG	0
trnP	-	3744-3803	60				2
nad6	+	3806-4282	477	158	ATG	TAA	-11
trnS-UCN	+	4272-4324	53				*
rrnL	-	4325-5461	1137				*
trnQ	-	5462-5523	62				-9
trnM	+	5515-5574	60				-18
nad2	+	5557-6510	954	317	ATG	TAA	-13
trnC	-	6498-6555	58				-9
trnY	-	6547-6607	61				0
cox1	+	6608-8146	1539	512	ATG	TAA	-5
trnL-UUR	+	8142-8199	58				18
trnW	+	8218-8280	63				60
cox2	+	8341-9022	682	227	GTG	T	0
trnK	+	9023-9084	62				-2
trnD	+	9083-9141	59				0
atp8	+	9142-9294	153	50	GTG	TAA	-7
atp6	+	9288-9959	672	223	ATG	TAA	0
cox3	+	9960-10748	789	262	ATG	TAA	-1
trnG	+	10748-10807	60				0
nad3	+	10808-11153	346	115	ATA	T	91
trnA	+	11245-11304	60				-2
trnV	+	11303-11365	63				10
nad1	-	11376-12293	918	305	ATT	TAA	1
trnN	+	12295-12357	63				*
rrnS [a]	+	12358-13055 [a]	698 [a]				

* Gene borders determined by borders of adjacent genes

[a] Incomplete gene sequence

b) Gene content of the mitochondrial genome of *Sphaeroma serratum*.

Gene	Strand	GenBank position no.	Size (nts)	Size (aa)	Start codon	Stop codon	Intergenic nucleotides
cob [a]	-	1-769 [a]	769 [a]	256 [a]	ATT	? [a]	0
trnT	-	770-829	60				1
nad5	+	831-2558	1728	575	ATT	TAG	-17
trnF	+	2542-2599	58				0
trnH	-	2600-2659	60				0
nad4	-	2660-4001	1342	447	ATG	T	-7
nad4L	-	3995-4300	306	101	ATA	TAA	0
trnP	-	4301-4360	60				10
nad6	+	4371-4856	486	161	ATC	TAG	-2
trnS-UCN	+	4855-4915	61				*
rrnL	-	4916-6069	1154				*
trnV	-	6070-6131	62				-2
trnQ	-	6130-6191	62				-9
trnM	+	6183-6243	61				0
nad2	+	6244-7230	987	328	ATT	TAG	40
trnY	-	7271-7329	59				-3
cox1	+	7327-8863	1537	512	ACT	T	0
trnL-UUR	+	8864-8926	63				0
cox2	+	8927-9607	681	226	ATA	TAG	-1
trnK	+	9602-9666	60				0
trnD	+	9667-9727	61				0
atp8	+	9728-9883	156	51	ATT	TAA	-7
atp6	+	9877-10548	762	253	ATG	TAA	-1
cox3	+	10548-11334	787	262	ATG	T	0
trnG	+	11335-11395	61				0
nad3	+	11395-11743	349	116	GTG	T	-1
trnA	+	11744-11803	60				11
nad1	-	11815-12750	936	311	ATA	TAA	2
trnN	+	12753-12816	64				*
rrnS [a]	+	12817-13467 [a]	651 [a]				

* Gene borders determined by borders of adjacent genes

[a] Incomplete gene sequence

c) Gene content of the mitochondrial genome of *Glyptonotus* cf. *antarcticus*.

Gene	Strand	GenBank position no.	Size (nts)	Size (aa)	Start codon	Stop codon	Intergenic nucleotides
cob [a]	-	1-760 [a]	760 [a]	253 [a]	ATA	? [a]	0
trnT	-	761-819	59				1
nad5	+	821-2541	1721	273	ATT	TA	0
trnF	+	2542-2601	60				3
trnH	-	2605-2666	62				-3
nad4	-	2664-4005	1342	447	ATG	T	-7
nad4l	-	3999-4286	288	95	ATT	TAA	15
trnP	-	4302-4366	65				1
nad6	+	4368-4868	501	166	ATT	TAG	-2
trnS-UCN	+	4867-4927	61				*
rrnL	-	4928-6142	1215				*
trnV	-	6143-6203	61				2
trnQ	-	6206-6268	63				-2
trnM	+	6267-6330	64				0
nad2	+	6331-7329	999	332	ATT	TAG	-9
trnC	-	7321-7369	49				1
trnY	-	7369-7430	62				0
cox1	+	7431-8969	1539	512	ACG	TAA	-5
trnL-UUR	+	8965-9027	63				0
cox2	+	9028-9711	684	227	GTG	TAA	0
trnK	+	9712-9772	61				-2
trnD	+	9771-9834	64				0
atp8	+	9835-9993	159	52	TTG	TAA	-7
atp6	+	9987-10661	675	224	ATG	TAG	2
cox3	+	10664-11449	786	261	ATA	TAG	7
trnR	+	11457-11526	70				-13
trnG	+	11514-11575	62				0
nad3	+	11576-11929	354	117	ATA	TAA	-2
trnA	+	11928-11992	65				21
nad1	-	12014-12943	930	309	ATT	TAG	0
trnL-CUN	-	12994-13005	62				-5
trnN	+	13001-13064	64				*
rrnS [a]	+	13065-13809 [a]	745 [a]				

* Gene borders determined by borders of adjacent genes

[a] Incomplete gene sequence

d) Gene content of the mitochondrial genome of *Armadillidium vulgare*.

Gene	Strand	GenBank position no.	Size (nts)	Size (aa)	Start codon	Stop codon	Intergenic nucleotides
cob [a]	-	1-366 [a]	366 [a]	122 [a]	ATA	? [a]	-26
trnT	-	341-402	62				27
nad5	+	430-2109	1680	559	ATA	TAA	-2
trnF	+	2108-2164	57				-26
trnH	-	2139-2201	63				-16
nad4	-	2186-3541	1356	451	ATG	TAA	5
nad4L	-	3547-3825	279	92	ATA	TAA	-23
trnP	-	3803-3877	75				1
nad6	+	3879-4373	495		ATT	TAA	-13
trnS-UCN	+	4361-4425	65				*
rrnL	-	4426-5468	1043				*
trnQ	-	5469-5523	55				1
trnM	+	5525-5587	63				-6
nad2	+	5582-6583	1002	333	ATA	TAA	-14
trnC	-	6570-6622	53				-23
trnY	-	6600-6668	69				-3
cox1	+	6666-8199	1534	511	ATG	T	55
cox2	+	8255-8932	678	225	ATA	TAA	43
trnD	+	8976-9047	72				-22
atp8	+	9026-9181	156	51	GTT	TAA	-7
atp6	+	9175-9843	669	222	ATG	TAA	2
cox3	+	9846-10637	792	263	ATG	TAA	-5
trnG	+	10633-10697	65				-13
nad3	+	10685-11038	354		ATA	TAA	-2
trnV	+	11037-11101	65				-26
nad1	-	11076-12005	930	309	ATA	TAA	*
rrnS	+	12006-12831	826				*
trnW	+	12832-12890	59				

* Gene borders determined by borders of adjacent genes

[a] Incomplete gene sequence

e) Gene content of the mitochondrial genome of *Asellus aquaticus*.

Gene	Strand	GenBank position no.	Size (nts)	Size (aa)	Start codon	Stop codon	Intergenic nucleotides
cob [a]	-	1-837 [a]	837 [a]	279 [a]	ATT	? [a]	0
trnT	-	838-896	59				15
nad5	+	912-2606	1695	564	ATA	TAG	0
trnF	+	2607-2675	69				11
trnH	-	2665-2724	60				0
nad4	-	2725-4051	1327	442	ACA	T	7
nad4L	-	4045-4347	303	100	ATA	TAA	0
trnP	-	4348-4407	60				10
nad6	+	4418-4897	480	159	ATA	TAA	1
trnS-UCN	+	4899-4958	60				*
rrnL	-	4959-6126	1168				*
trnQ	-	6127-6188	62				-9
trnM	+	6180-6241	62				0
nad2	+	6242-7219	978	325	ATT	TAA	-15
trnC	-	7205-7253	49				-1
trnY	-	7253-7316	64				0
cox1	+	7317-8855	1539	512	ACC	TAA	-5
trnL-CUN	+	8851-8911	61				0
trnL-UUR	+	8912-8971	60				0
cox2	+	8972-9649	678	225	ATC	TAA	-2
trnK	+	9648-9710	63				-1
trnD	+	9710-9774	65				0
atp8	+	9775-9933	159	52	CTG	TAA	-7
atp6	+	9927-10595	669	222	ATG	TAA	-1
cox3	+	10595-11380	786	261	ATG	TAA	-1
trnG	+	11380-11439	60				0
nad3	+	11440-11793	354	117	ATT	TAA	-2
trnA	+	11792-11852	61				2
trnR	+	11855-11915	61				-1
trnV	+	11915-11974	60				-4
nad1	-	11971-12885	915	304	ATT	TAA	16
trnN	+	12902-12964	63				*
rrnS [a]	+	12965-13639 [a]	675 [a]				

* Gene borders determined by borders of adjacent genes

[a] Incomplete gene sequence

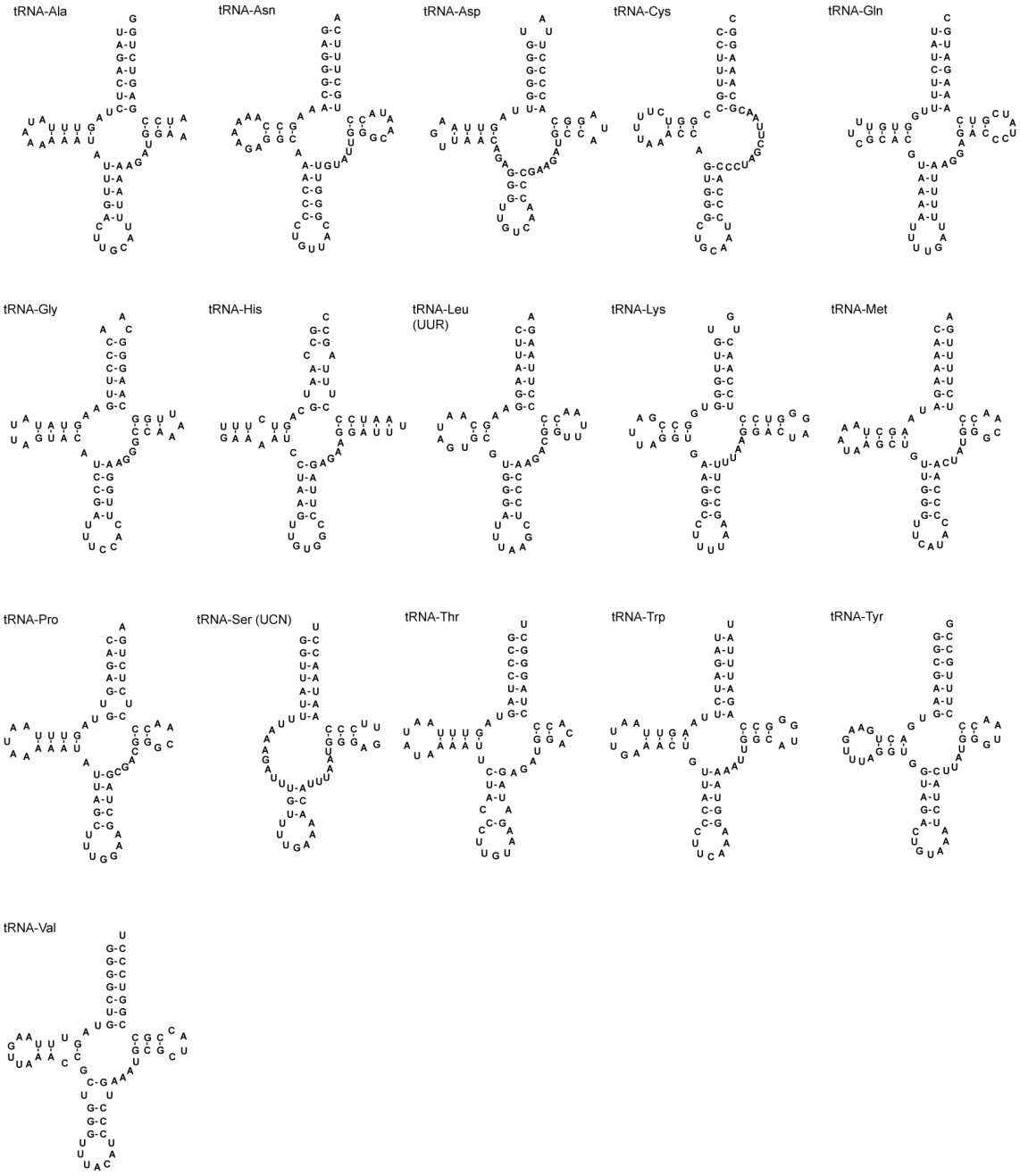
f) Gene content of the mitochondrial genome of *Janira maculosa*.

Gene	Strand	GenBank position no.	Size (nts)	Size (aa)	Start codon	Stop codon	Intergenic nucleotides
cob [a]	-	1-331 [a]	331	110	ATA	? [a]	-18
trnT	-	314-385	72				-13
trnH	-	373-443	71				-5
nad4	-	439-1740	1302	433	ATG	TAA	-7
nad4L	-	1734-2030	297	98	ATT	TAA	-3
trnP	-	2028-2086	59				3
nad6	+	2090-2581	492	163	ATA	TAA	-1
trnS-UCN	+	2581-2650	70				*
rrnL	-	2651-3752	1102				*
trnV	-	3753-3812	60				-27
trnQ	-	3787-3862	76				0
trnM	+	3863-3930	68				-10
nad2	+	3921-4880	960	319	ATA	TAA	-15
trnC	-	4866-4916	51				-12
trnY	-	4905-4973	69				-4
cox1	+	4970-6503	1534	511	ACG	T	94
trnL-UUR	+	6598-6657	60				215
trnL-CUN	+	6873-6932	60				-1
cox2	+	6932-7603	672	223	ATT	TAA	61
trnD	+	7665-7728	64				4
cox3	+	7733-8527	795	264	ATG	TAA	37
trnG	+	8565-8638	74				-18
nad3	+	8621-8968	348	115	ATT	TAA	5
trnA	+	8974-9040	67				*
rrnS [a]	+	9041-9871	831 [a]				

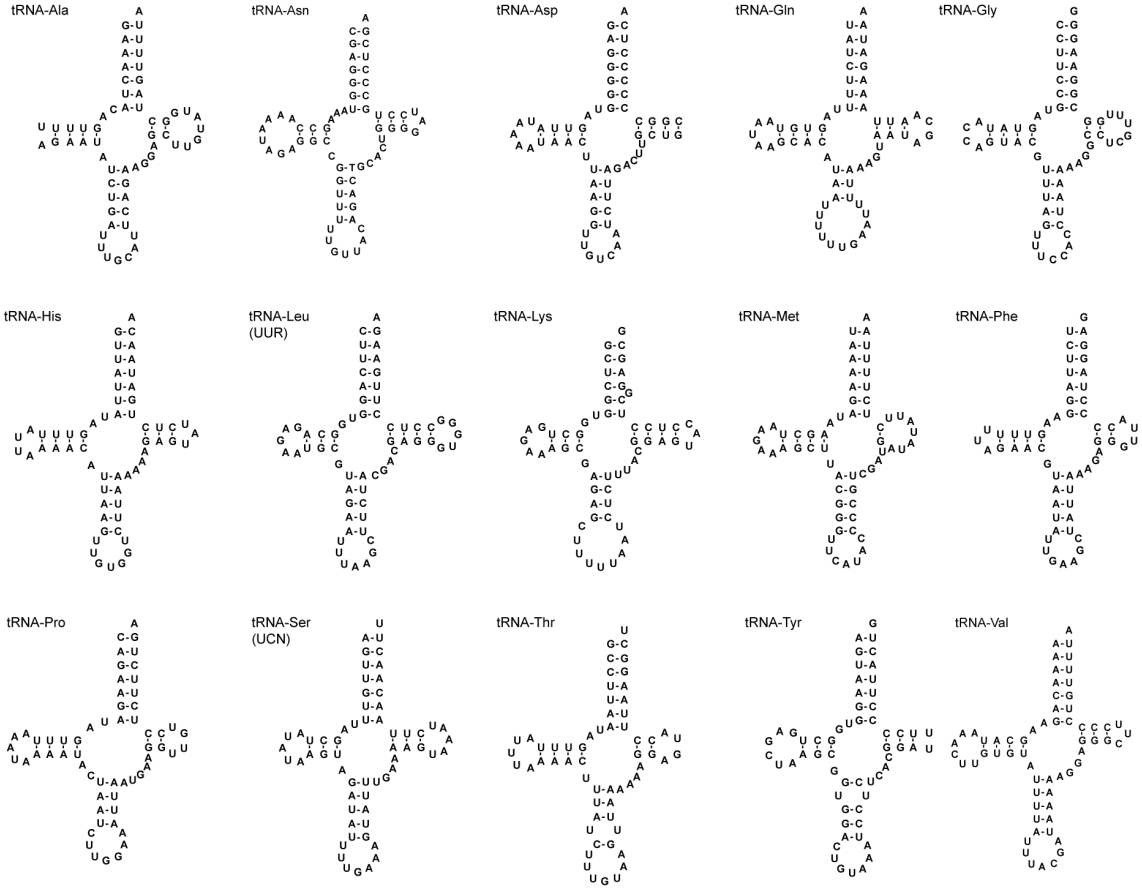
* Gene borders determined by borders of adjacent genes

[a] Incomplete gene sequence

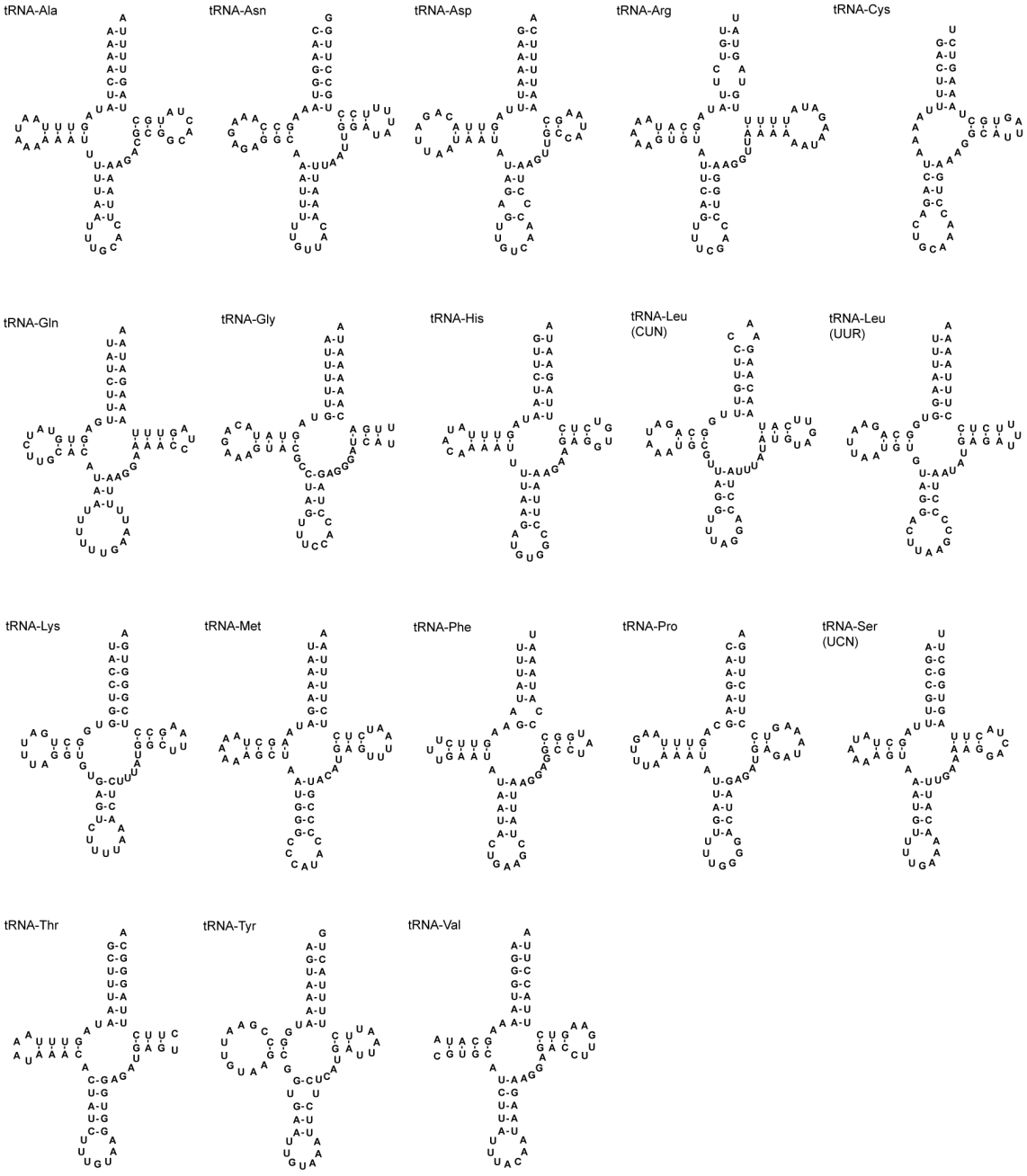
Supplementary file 4.5



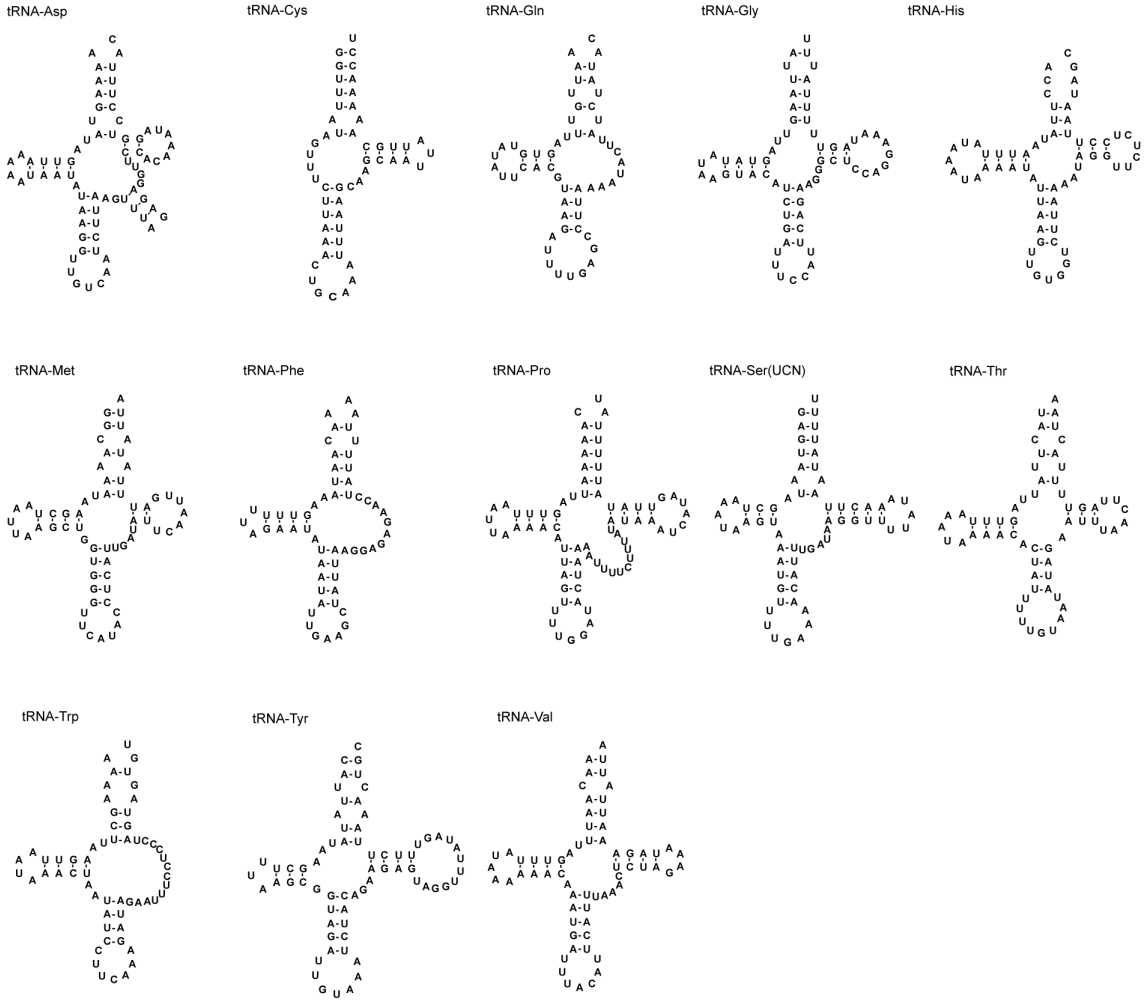
a) tRNA structure plots of *Eurydice pulchra*.



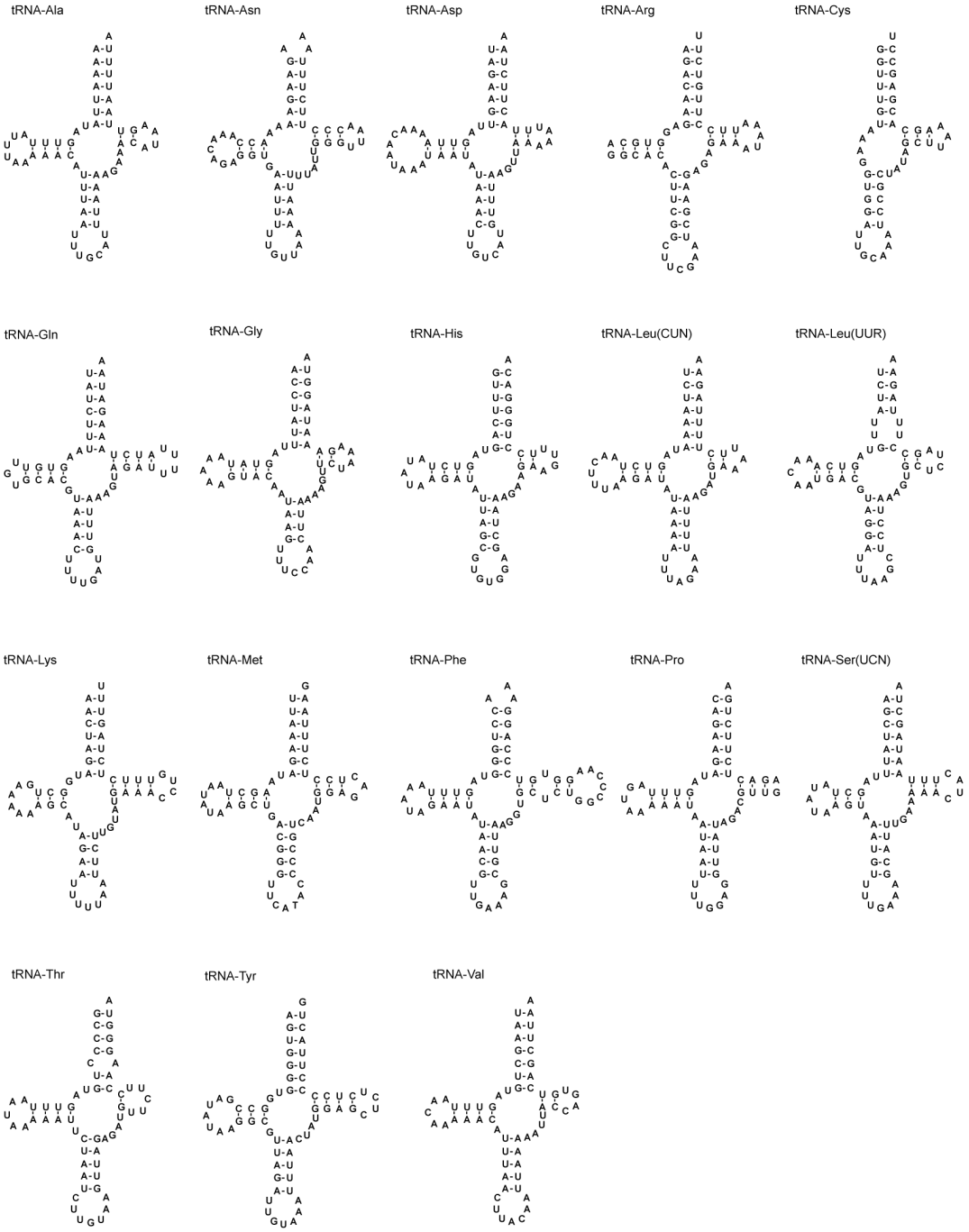
b) tRNA structure plots of *Sphaeroma serratum*.



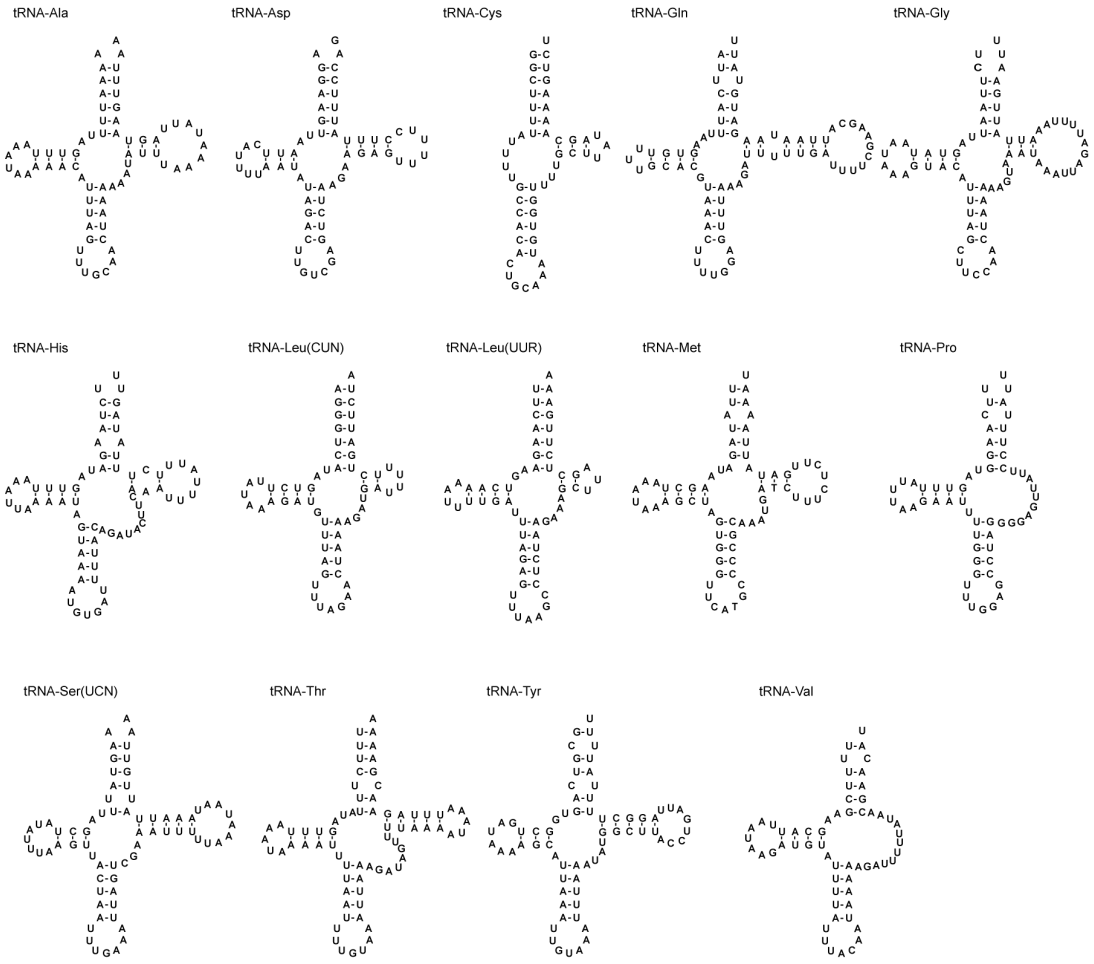
c) tRNA structure plots of *Glyptonotus* cf. *antarcticus*.



d) tRNA structure plots of *Armadillidium vulgare.*



e) tRNA structure plots of *Asellus aquaticus*.



f) tRNA structure plots of *Janira maculosa***.**

Supplementary file 4.6

a) Nucleotide frequencies, AT content, AT- and GC-skew for mitochondrial genes of *Eurydice pulchra*.

Gene(strand)	Nucleotide frequency				%AT	AT skew	GC skew
	A	C	G	T			
atp6(+)	0.2083	0.1786	0.2783	0.3348	54.3	-0.23	0.22
atp8(+)	0.2157	0.1699	0.2484	0.3660	58.2	-0.26	0.19
cob(-)*	0.2409	0.2145	0.1980	0.3465	58.7	-0.18	-0.04
cox1(+)	0.2099	0.1813	0.2632	0.3457	55.6	-0.24	0.18
cox2(+)	0.2214	0.1891	0.2654	0.3240	54.5	-0.19	0.17
cox3(+)	0.2193	0.1850	0.2573	0.3384	55.8	-0.21	0.16
nad1(-)	0.2222	0.2244	0.2211	0.3322	55.4	-0.20	-0.01
nad2(+)	0.2212	0.1583	0.3008	0.3197	54.1	-0.18	0.31
nad3(+)	0.2110	0.1416	0.3295	0.3179	52.9	-0.20	0.40
nad4(-)	0.2247	0.2446	0.1803	0.3503	57.5	-0.22	-0.15
nad4L(-)	0.1905	0.2687	0.1803	0.3605	55.1	-0.31	-0.20
nad5(+)	0.2556	0.1281	0.3018	0.3146	57.0	-0.10	0.40
nad6(+)	0.2201	0.1300	0.3187	0.3312	55.1	-0.20	0.42
prot. cod. total*	0.2201	0.1857	0.2572	0.3371	55.7	-0.21	0.16
rrnL(-)	0.3036	0.2307	0.1756	0.2901	59.4	0.02	-0.14
rrnS(+)*	0.2607	0.2249	0.2708	0.2436	50.4	0.03	0.09
rRNA total*	0.2822	0.2278	0.2232	0.2669	54.9	0.03	-0.02
total*	0.2649	0.1769	0.2645	0.2937	55.9	-0.05	0.20

* Incomplete sequence

b) Nucleotide frequencies, AT content, AT- and GC-skew for mitochondrial genes of *Sphaeroma serratum*.

Gene(strand)	Nucleotide frequency				%AT	AT skew	GC skew
	A	C	G	T			
atp6(+)	0.1949	0.1949	0.3244	0.2857	48.1	-0.19	0.25
atp8(+)	0.2179	0.1987	0.3718	0.2115	42.9	0.01	0.30
cob(-)*	0.2497	0.2575	0.1612	0.3316	58.1	-0.14	-0.23
cox1(+)	0.2010	0.2115	0.2746	0.3129	51.4	-0.22	0.13
cox2(+)	0.2188	0.2129	0.2746	0.2937	51.3	-0.15	0.13
cox3(+)	0.1703	0.2452	0.2834	0.3011	47.1	-0.28	0.07
nad1(-)	0.1775	0.2759	0.2524	0.2941	47.2	-0.25	-0.04
nad2(+)	0.2067	0.1429	0.2958	0.3546	56.1	-0.26	0.35
nad3(+)	0.1433	0.2063	0.3926	0.2579	40.1	-0.29	0.31
nad4(-)	0.2422	0.2809	0.1490	0.3279	57.0	-0.15	-0.31
nad4L(-)	0.2582	0.2484	0.1536	0.3399	59.8	-0.14	-0.24
nad5(+)	0.2425	0.1082	0.2940	0.3553	59.8	-0.19	0.46
nad6(+)	0.2284	0.1214	0.2778	0.3724	60.1	-0.24	0.39
prot. cod. total*	0.2116	0.2081	0.2696	0.3107	52.2	-0.19	0.12
rrnL(-)	0.3105	0.2133	0.1614	0.3147	62.5	-0.01	-0.14
rrnS(+)*	0.2413	0.2427	0.3236	0.1925	43.4	0.11	0.14
rRNA total*	0.2759	0.2280	0.2425	0.2536	53.0	0.05	0.00
total*	0.2533	0.1782	0.2778	0.2907	54.4	-0.07	0.22

* Incomplete sequence

c) Nucleotide frequencies, AT content, AT- and GC-skew for mitochondrial genes of *Glyptonotus* cf. *antarcticus.*

Gene(strand)	Nucleotide frequency				%AT	AT skew	GC skew
	A	C	G	T			
atp6(+)	0.2815	0.1748	0.1556	0.3881	67.0	-0.16	-0.06
atp8(+)	0.3774	0.1321	0.1447	0.3459	72.3	0.04	0.05
cob(-)*	0.2342	0.1961	0.1842	0.3855	62.0	-0.24	-0.03
cox1(+)	0.2411	0.1845	0.2092	0.3652	60.6	-0.20	0.06
cox2(+)	0.2865	0.1988	0.1594	0.3553	64.2	-0.11	-0.11
cox3(+)	0.2125	0.2020	0.2074	0.3779	59.0	-0.28	0.01
nad1(-)	0.2570	0.1462	0.1892	0.4075	66.5	-0.23	0.13
nad2(+)	0.2743	0.1552	0.1692	0.4014	67.6	-0.19	0.04
nad3(+)	0.2684	0.1554	0.2006	0.3757	64.4	-0.17	0.13
nad4(-)	0.2498	0.1782	0.1611	0.4109	66.1	-0.24	-0.05
nad4L(-)	0.2639	0.1563	0.1632	0.4167	68.1	-0.22	0.02
nad5(+)	0.2697	0.1383	0.2199	0.3721	64.2	-0.16	0.23
nad6(+)	0.2894	0.1178	0.1856	0.4072	69.7	-0.17	0.22
prot. cod. total*	0.2697	0.1643	0.1807	0.3853	65.5	-0.18	0.05
rrnL(-)	0.3770	0.1366	0.1366	0.3498	72.7	0.04	0.00
rrnS(+)*	0.3356	0.1919	0.1866	0.2859	62.2	0.08	-0.01
rRNA total*	0.3563	0.1643	0.1616	0.3179	67.4	0.06	-0.01
total*	0.3160	0.1665	0.1797	0.3378	65.4	-0.03	0.04

* Incomplete sequence

d) Nucleotide frequencies, AT content, AT- and GC-skew for mitochondrial genes of *Armadillidium vulgare.*

Gene(strand)	Nucleotide frequency				%AT	AT skew	GC skew
	A	C	G	T			
atp6(+)	0.3139	0.1196	0.1734	0.3931	70.7	-0.11	0.18
atp8(+)	0.3269	0.0897	0.1218	0.4615	78.8	-0.17	0.15
cob(-)*	0.3060	0.1694	0.1366	0.3880	69.4	-0.12	-0.11
cox1(+)	0.2653	0.1525	0.1897	0.3924	65.8	-0.19	0.11
cox2(+)	0.3083	0.1386	0.1711	0.3820	69.0	-0.11	0.10
cox3(+)	0.2525	0.1654	0.1793	0.4028	65.5	-0.23	0.04
nad1(-)	0.4140	0.1495	0.1398	0.2968	71.1	0.16	-0.03
nad2(+)	0.3094	0.0898	0.1776	0.4232	73.3	-0.16	0.33
nad3(+)	0.2712	0.0960	0.2090	0.4237	69.5	-0.22	0.37
nad4(-)	0.3119	0.1593	0.1091	0.4196	73.2	-0.15	-0.19
nad4L(-)	0.2867	0.1219	0.1039	0.4875	77.4	-0.26	-0.08
nad5(+)	0.3149	0.0798	0.1851	0.4202	73.5	-0.14	0.40
nad6(+)	0.3030	0.0606	0.1939	0.4424	74.5	-0.19	0.52
prot. cod. total*	0.3065	0.1225	0.1608	0.4102	71.7	-0.14	0.14
rrnL(-)	0.3883	0.1112	0.1179	0.3826	77.1	0.01	0.03
rrnS(+)*	0.3535	0.1380	0.1901	0.3184	67.2	0.05	0.16
rRNA total*	0.3709	0.1246	0.1540	0.3505	72.1	0.03	0.09
total*	0.3389	0.1182	0.1683	0.3747	71.4	-0.05	0.17

* Incomplete sequence

e) Nucleotide frequencies, AT content, AT- and GC-skew for mitochondrial genes of *Asellus aquaticus.*

Gene(strand)	Nucleotide frequency				%AT	AT skew	GC skew
	A	C	G	T			
atp6(+)	0.2855	0.1928	0.1390	0.3827	66.8	-0.15	-0.16
atp8(+)	0.2642	0.3019	0.1321	0.3019	56.6	-0.07	-0.39
cob(-)*	0.2186	0.2079	0.2366	0.3369	55.6	-0.21	0.06
cox1(+)	0.2521	0.2385	0.1761	0.3333	58.5	-0.14	-0.15
cox2(+)	0.2832	0.2330	0.1593	0.3245	60.8	-0.07	-0.19
cox3(+)	0.2506	0.2188	0.1768	0.3537	60.4	-0.17	-0.11
nad1(-)	0.2350	0.1257	0.2120	0.4273	66.2	-0.29	0.26
nad2(+)	0.2894	0.1963	0.1728	0.3415	63.1	-0.08	-0.06
nad3(+)	0.2712	0.2090	0.1582	0.3616	63.3	-0.14	-0.14
nad4(-)	0.2344	0.1665	0.2238	0.3753	61.0	-0.23	0.15
nad4L(-)	0.2475	0.1551	0.2145	0.3828	63.0	-0.21	0.16
nad5(+)	0.2248	0.2307	0.2212	0.3233	54.8	-0.18	-0.02
nad6(+)	0.2729	0.2208	0.1479	0.3583	63.1	-0.14	-0.20
prot. cod. total*	0.2561	0.2075	0.1823	0.3541	61.0	-0.16	-0.06
rrnL(-)	0.3322	0.1476	0.1931	0.3270	65.9	0.01	0.13
rrnS(+)*	0.3917	0.1721	0.1335	0.3027	69.4	0.13	-0.13
rRNA total*	0.3620	0.1599	0.1633	0.3149	67.7	0.07	0.00
total*	0.3105	0.2132	0.1669	0.3094	62.0	0.00	-0.12

* Incomplete sequence

f) Nucleotide frequencies, AT content, AT- and GC-skew for mitochondrial genes of *Janira maculosa.*

Gene(strand)	Nucleotide frequency				%AT	AT skew	GC skew
	A	C	G	T			
atp6(+)	-	-	-	-	-	-	-
atp8(+)	-	-	-	-	-	-	-
cob(-)*	0.2719	0.1390	0.1964	0.3927	66.5	-0.18	0.17
cox1(+)	0.2771	0.1617	0.1695	0.3918	66.9	-0.17	0.02
cox2(+)	0.3155	0.1637	0.1518	0.3690	68.5	-0.08	-0.04
cox3(+)	0.2704	0.1535	0.1698	0.4063	67.7	-0.20	0.05
nad1(-)	-	-	-	-	-	-	-
nad2(+)	0.3000	0.1365	0.1229	0.4406	74.1	-0.19	-0.05
nad3(+)	0.2816	0.1178	0.1638	0.4368	71.8	-0.22	0.16
nad4(-)	0.2939	0.1338	0.1516	0.4207	71.5	-0.18	0.06
nad4L(-)	0.3333	0.0909	0.1582	0.4175	75.1	-0.11	0.27
nad5(+)	-	-	-	-	-	-	-
nad6(+)	0.3089	0.1159	0.1199	0.4553	76.4	-0.19	0.02
prot. cod. total*	0.2947	0.1348	0.1560	0.4145	70.9	-0.17	0.07
rrnL(-)	0.3869	0.1090	0.1299	0.3742	76.1	0.02	0.09
rrnS(+)*	0.3498	0.1846	0.1882	0.2774	62.7	0.12	0.01
rRNA total*	0.3684	0.1468	0.1591	0.3258	69.4	0.07	0.05
total*	0.3419	0.1467	0.1411	0.3703	71.2	-0.04	-0.02

* Incomplete sequence

- gene is not available

g) Nucleotide frequencies, AT content, AT- and GC-skew for mitochondrial genes of *Idotea balthica.*

Gene(strand)	Nucleotide frequency				%AT	AT skew	GC skew
	A	C	G	T			
atp6(+)	0.2365	0.1916	0.1961	0.3757	61.2	-0.23	0.01
atp8(+)	0.2692	0.1731	0.2115	0.3462	61.5	-0.13	0.10
cob(-)*	0.2463	0.2562	0.1665	0.3310	57.7	-0.15	-0.21
cox1(+)	0.2280	0.1816	0.2319	0.3586	58.7	-0.22	0.12
cox2(+)	0.2328	0.1728	0.2430	0.3514	58.4	-0.20	0.17
cox3(+)	0.2168	0.2092	0.2219	0.3520	56.9	-0.24	0.03
nad1(-)	0.2296	0.2370	0.1979	0.3354	56.5	-0.19	-0.09
nad2(+)	0.2413	0.1253	0.2322	0.4012	64.3	-0.25	0.30
nad3(+)	0.2308	0.1481	0.2877	0.3333	56.4	-0.18	0.32
nad4(-)	0.2455	0.2493	0.1507	0.3545	60.0	-0.18	-0.25
nad4L(-)	0.2823	0.2211	0.1463	0.3503	63.3	-0.11	-0.20
nad5(+)	0.2429	0.1437	0.2375	0.3758	61.9	-0.21	0.25
nad6(+)	0.2742	0.1149	0.1875	0.4234	69.8	-0.21	0.24
prot. cod. total*	0.2443	0.1865	0.2085	0.3607	60.5	-0.19	0.06
rrnL(-)	0.3347	0.1900	0.1546	0.3207	65.5	0.02	-0.10
rrnS(+)*	0.2885	0.1893	0.2337	0.2885	57.7	0.00	0.10
rRNA total*	0.3116	0.1897	0.1942	0.3046	61.6	0.01	0.00
total*	0.2819	0.1633	0.2267	0.3281	61.0	-0.08	0.16

* Incomplete sequence

Supplementary file 4.7

tRNA availability and secondary structure modification.

tRNA	gene	*Eophreatoicus* sp. 14	*Janira maculosa*	*Asellus aquaticus*	*Ligia oceanica*	*Armadillidium vulgare*	*Glyptonotus* cf. *antarcticus*	*Idotea balthica*	*Sphaeroma serratum*	*Eurydice pulchra*
tRNA-Ala	*trnA*	●	●	●	●	?	●	●	●	●
tRNA-Arg	*trnR*	●	?	●	?	?	●	●	?	?
tRNA-Asn	*trnN*	●	?	●	●	?	●	?	●	●
tRNA-Asp	*trnD*	●	●	●	●	**V**	●	**R**	●	●
tRNA-Cys	*trnC*	**L**	**L**	**L**	**L**	**L**	**L**	**L**	?	**R**
tRNA-Gln	*trnQ*	●	●	●	●	**R**	●	●	●	●
tRNA-Glu	*trnE*	●	?	?	●	?	?	?	?	?
tRNA-Gly	*trnG*	●	●	●	●	●	●	●	●	●
tRNA-His	*trnH*	●	●	●	●	**A**	●	●	●	●
tRNA-Ile	*trnI*	**L**	?	?	**R**	?	?	?	?	?
tRNA-Leu(CUN)	*trnL1*	●	●	●	●	?	●	●	?	?
tRNA-Leu(UUR)	*trnL2*	●	●	●	●	?	●	●	●	●
tRNA-Lys	*trnK*	●	?	●	●	?	●	●	●	●
tRNA-Met	*trnM*	●	●	●	●	●	●	●	●	●
tRNA-Phe	*trnF*	●	?	●	●	**R**	●	●	●	?
tRNA-Pro	*trnP*	●	**R**	●	●	●	●	●	●	●
tRNA-Ser(AGY)	*trnS1*	**L**	?	?	**L**	?	?	?	?	?
tRNA-Ser(UCN)	*trnS2*	●	●	●	●	●	●	**R**	●	**L**
tRNA-Thr	*trnT*	●	●	●	●	●	●	●	●	●
tRNA-Trp	*trnW*	●	?	?	●	**R**	?	?	?	●
tRNA-Tyr	*trnY*	●	●	●	●	●	●	●	●	●
tRNA-Val	*trnV*	●	**R**	●	●	●	●	●	●	●

●: usual clover-leaf; A: poor acceptor arm pairings; L: left arm (DHU-arm) missing; R: right arm (TΨC-arm) missing; V: variable loop elongation; ?: missing data

Supplementary file 5.1

Amplification and sequencing primers used to obtain the complete mt genome of *Caprella mutica*.

Primer	Orientation	Sequence (5'-3')	Reference
Isop-16sf	Reverse	AARAAWGATTGCGACCTCGATGTTGAATTG	Kilpert & Podsiadlowski 2009
Isop-16sr	Forward	TATGCTACCTTAGCACAGTYAGRATACTGCGGC	Kilpert & Podsiadlowski 2009
Pera-Co1r	Reverse	AADGCTATATCAGGAGCCCCAATTATTAAAGG	Kilpert & Podsiadlowski 2009
Crust-12sf	Reverse	CAGCAKYCGCGGTTAKAC	Podsiadlowski & Bartolomaeus 2005
Crust-12sr	Forward	ACACCTACTWTGTTACGACTTATCTC	Podsiadlowski & Bartolomaeus 2005
Crust-16sf	Reverse	TGACYGTGCDAAGGTAGC	Podsiadlowski & Bartolomaeus 2005
Crust-16sr	Forward	CCGGTCTGAACTCAYATC	Podsiadlowski & Bartolomaeus 2005
Crust-16sf2	Reverse	GCGACCTCGATGTTGGATTAA	Kilpert & Podsiadlowski 2009
Crust-16sr2	Forward	CCGGTCTGAACTCAYGTA	Kilpert & Podsiadlowski 2009
Crust-cox1f	Forward	ACTAATCACAARGAYATTGG	Podsiadlowski & Bartolomaeus 2005
Crust-cox1r	Reverse	TAGTCTGAGTANCGTCGWGG	Podsiadlowski & Bartolomaeus 2005
Crust-cox3f	Forward	ATAATTCAATGATGACGAGA	Podsiadlowski & Bartolomaeus 2005
Crust-cox3r	Reverse	CCAATAATWACATGWAGACC	Podsiadlowski & Bartolomaeus 2005
Crust-CytB-f1	Forward	CGAGATGTAAAYTAYGGSTGAC	Kilpert & Podsiadlowski 2009
Crust-CytB-f2	Forward	CGAGATGTAAAYTAYGGWTGAC	Kilpert & Podsiadlowski 2009
Crust-CytB-r	Reverse	CTACGGGAGTGCACCRATYC	Kilpert & Podsiadlowski 2009
Crust-N4f	Reverse	TTGAGGTTAYCAGCCYG	Podsiadlowski & Bartolomaeus 2005
Crust-N4r	Forward	ATATGAGCYACAGAAGARTAAGC	Podsiadlowski & Bartolomaeus 2005
Crust-N5f	Forward	AGAATTCTACTAGGDTGRGATGG	Podsiadlowski & Bartolomaeus 2005
Crust-N5r	Reverse	AAAGAGCCTTAAATAAAGCATG	Podsiadlowski & Bartolomaeus 2005
16S2	Reverse	GCGACCTCGATGTTGGATTAA	Roehrdanz et al. 2002
N4	Forward	GGAGCTTCAACATGAGCTTT	Roehrdanz et al. 2002
N4(87)	Reverse	TCAGCTAATATAGCAGCTCC	Roehrdanz et al. 2002
L39-Met (S1-S3)	Forward	AAGCTHVTGGGCTCATACCCC	Yamauchi et al. 2004
L329-ND2 (S5)	Forward	GGWGCHGCHCCNTTWCATTTTTG	Yamauchi et al. 2004
H718-ND2 (S1)	Reverse	AABCCHGDGAAMGGDGGHAVHCCHCC	Yamauchi et al. 2004
H1368-CO1 (S2)	Reverse	TATAGAGTTCCAATRTCYTTGTGATT	Yamauchi et al. 2004
L1564-CO1 (S10)	Forward	ATGGTWATACCGATTWTRATTGG	Yamauchi et al. 2004
H1602-CO1 (S3)	Reverse	GGGAADGCTATGTCWGGGGC	Yamauchi et al. 2004
L2020S-CO1 (S11)	Forward	GACCCTGCGGGWGGRGGRGATCC	Yamauchi et al. 2004
H2619-CO1 (S10,S11)	Reverse	GGTATWCCWGCKAGWCCTAAGAAATGTTG	Yamauchi et al. 2004
H3514-CO2 (S15)	Reverse	CCACAAATTTCKGAACATTGWCCATAAAA	Yamauchi et al. 2004
L4268-A6 (S16)	Forward	GGDTGGDTTAAWMAHACWCAACA	Yamauchi et al. 2004
H4375-A6 (S15)	Reverse	GCDATCATGTTDGCDGMWAGTCG	Yamauchi et al. 2004
H4806-CO3 (S16)	Reverse	GADGHAATGAAHAGGATTATDCCTCA	Yamauchi et al. 2004
L5170-CO3 (S20)	Forward	TTGTDGCNACHGGMTTTCATGG	Yamauchi et al. 2004
L5928-Asn (S25)	Forward	GAAGCTAATWGGTAGCRTWTCATTGTTAATG	Yamauchi et al. 2004
H6110-Phe (S20)	Reverse	AAGAGTATAGCACTGAAGATG	Yamauchi et al. 2004
H6721-ND5 (S25)	Reverse	TTTTWTTCWAAGGATTTAATT	Yamauchi et al. 2004
L13337-12S (S48)	Reverse	YCTACTWTGYTACGACTTATCTC	Yamauchi et al. 2004
H13845-12S (S48)	Forward	GTGCCAGCAGCTGCGGTTA	Yamauchi et al. 2004
Cm-464	Reverse	ACGATAAAATAAGTACAGACGACATAGC	This study
Cm-839	Reverse	AAATAATAAAGTAAACAACTCAAG	This study
Cm-943	Forward	AATAATGAATATACTGGCAATCGC	This study
Cm-1108	Forward	ATCCTTGATAGATTCCTCG	This study
Cm-1682	Reverse	AAAGTTAAAGAGGGAGGTAGTAGCC	This study
Cm-1878	Forward	ATACTTAGACCGAATGCCTTTG	This study
Cm-2309	Reverse	ATTTTGATTCCTGTGGG	This study
Cm-2390	Forward	TATTTACTTTAGGTGGTCTCACTGG	This study
Cm-2861	Forward	GCTACGACGACATTCCTTTAC	This study
Cm-3943	Reverse	GAGGGGTCAAAAATTCTAAATAAG	This study
Cm-4175	Forward	TACTTTTACAGCCACATCGCAC	This study
Cm-4966	Reverse	GATTTCAAATGGGTTAAAAGAC	This study
Cm-5051	Forward	GAAACAAAAACAGCCCTAATCG	This study
Cm-5981	Forward	TCGGATTGACAAATGCTC	This study

Primer	Orientation	Sequence (5'-3')	Reference
Cm-6441	Forward	TGCCCACTATTTTTCCGAC	This study
Cm-7613	Reverse	GGATTAACCACCCTTTGTCTG	This study
Cm-8002	Reverse	CAGATTGTTGGAAGACCCG	This study
Cm-8151	Forward	ACCCTTCTGTCACTCCTCACC	This study
Cm-10080	Reverse	AGTACAGCCTGCCCAGTGC	This study
Cm-10342	Forward	CTAAATCGAAAGACTCTCACCG	This study
Cm-10769	Reverse	GAGGTGAGGATATACAGAAGGAG	This study
Cm-10871	Forward	TAACAAACAAACTCCTCTGAATAG	This study
Cm-11832	Forward	ACTTGTTTACAATCCTTAGACTCCC	This study
Cm-11863	Reverse	AGAATAAGGGAGTCTAAGGATTGTAAAC	This study
Cm-12019	Forward	CTTTGTTAGAGGCGTCCCG	This study
Cm-14034	Forward	ATAAAAACAGGTAAAACACTACACTCG	This study
Cm-14669	Forward	AATCAGAAAGAAGGTTCCCC	This study
Cm-15233	Reverse	ACGACCCATCTACTTACTCTTGTTG	This study
Cm-15378	Reverse	TAACAAAAGAGCTGGGACC	This study

Supplementary file 5.2

Method of tRNA detection used for mt genome of *Caprella mutica*. All 22 tRNA genes were identified by software or by manual search of non-coding parts of the mt genome. Standard search with tRNAscan-SE performed with 'Invertebrate Mito' genetic code and 'Mito/Chloroplast' model. When no tRNAs were detected, the less strict 'Nematode Mito' model was used instead. ARWEN was used with 'Invertebrate mitochondrial genetic code' option. All results were manually checked afterwards.

tRNA	tRNAscan-SE 1.21 (software)	ARWEN 1.2.2.c (software)	manual
tRNA-Ala	detected	detected	accepted
tRNA-Asn	detected	detected	accepted
tRNA-Asp	detected	detected	accepted
tRNA-Arg	detected	detected	accepted
tRNA-Cys	detected	detected	accepted
tRNA-Gln	detected	detected	accepted
tRNA-Glu	detected	detected	accepted
tRNA-Gly	detected	detected	accepted
tRNA-His	detected	detected	accepted
tRNA-Ile	detected	detected	accepted
tRNA-Leu(CUN)	detected	detected	accepted
tRNA-Leu(UUR)	detected	detected	accepted
tRNA-Lys	detected	detected	accepted
tRNA-Met	detected	detected	accepted
tRNA-Phe	detected	detected	accepted
tRNA-Pro	detected*	detected	accepted
tRNA-Ser(AGY)	not detected	not detected	detected
tRNA-Ser(UCN)	not detected	not detected	detected
tRNA-Thr	detected	detected	accepted
tRNA-Trp	detected	detected	accepted
tRNA-Tyr	detected	detected	accepted
tRNA-Val	detected	detected	accepted

* Detection with 'Nematode Mito' setting only

Supplementary file 5.3

Comparison of the two control regions of the *Caprella mutica* mt genome. Complementary sections within the control region are colored (red and blue).