Short communication

Major rearrangements characterize the mitochondrial genome of the isopod *Idotea baltica* (Crustacea: Peracarida)

Lars Podsiadlowski *, Thomas Bartolomaeus

Institut für Biologie, Freie Universität Berlin, Germany

Received 13 December 2005; revised 11 April 2006; accepted 11 April 2006
Available online 25 April 2006

1. Introduction

Mitochondrial genomes represent an important data source for phylogenetic analyses. Thirty-seven genes encoding for 13 protein subunits, 2 rRNAs and 22 tRNAs are usually present in a bilaterian mitochondrial genome (Boore, 1999; Wolstenholme, 1992). Additionally, a non-coding control region can be identified in most cases, probably bearing the transcription initiation sites. Different rates of evolutionary change between different parts of the mitochondrial genome make them useful for a variety of phylogenetic questions ranging from studies at the population level (mitochondrial control region, cytochrome c oxidase subunit I and cytochrome b, e.g., Liebers et al., 2004) up to the interrelationships of animal phyla (concatenated amino acid alignments of protein-coding genes, e.g., Helfenbein et al., 2004; Lavrov et al., 2005). Besides nucleotide or amino acid sequences other types of characters may be useful in phylogenetic studies, like for instance gene order (e.g., Boore et al., 1995; Roehrdanz et al., 2002), RNA secondary structure (e.g., Macey et al., 2000) or changes of mitochondrial genetic code (e.g., Castresana et al., 1998).

Mitochondrial gene order is not stable among Crustacea. Major gene rearrangements involving several protein-coding genes have been reported in copepods (Machida et al., 2002, 2004; Tjensvoll et al., 2005), an ostracod (Ogoh and Ohmiya, 2004), a branchiuran (Lavrov et al., 2004), and among Malacostraca in the freshwater crayfish *Cherax destructor* (Miller et al., 2004) and the crab *Eriocheir sinensis* (Sun et al., 2005). Taxon sampling regarding complete mitochondrial genomes among Malacostraca remains rather poor. Besides several decapod and stomatopod species only one euphausiacean and one amphipod species have been sequenced so far. Cladistic analyses of morphological data have led to alternative hypotheses about interrelationships of living eumalacostracan taxa. The hypotheses of Schram (1986) and Richter and Scholtz (2001) differ in the position of Decapoda (sister group to Euphausiacea or to a clade combining the latter with Peracarida and Syncarida). Wills (1998) places Syncarida and then Peracarida as sister groups to the remainder Eumalacostraca, while other authors favor a basal split between Stomatopoda and all other eumalacostracan taxa (Richter and Scholtz, 2001) or Peracarida and all other Eumalacostraca (Wheeler, 1998). In some studies, peracarids are not monophyletic, as Mysidacea (Watling, 1999) do not cluster with the other peracarid taxa. The latter hypothesis is also confirmed by a recent molecular analysis based on 18S rRNA sequences (Spears et al., 2005).

In our study, we report DNA sequence for the major part of the mitochondrial genome of the isopod *Idotea baltica*, including all protein-coding and rRNA-coding genes and 17 tRNAs. This is the first report on mitochondrial gene order in isopods and it is highly derived in comparison to other crustaceans. We discuss gene rearrangements in Peracarida and try to resolve phylogenetic relationships of Malacostraca using sequence data from mitochondrial protein-coding genes.

2. Materials and methods

2.1. Animals, DNA extraction

Specimens of the isopod *Idotea baltica* (F. Idoteidae) were collected in Roscoff (France) and preserved in 99% ethanol. Total genomic DNA was extracted from legs and the head of the animal, using the DNeasy Tissue Kit (Qiagen, Hilden, Germany) following the manufacturers’ protocol.
2.2. PCR and sequencing

All PCR amplifications were performed with HotMasterTaq polymerase (Eppendorf, Germany). PCR primer pairs for amplification of several mitochondrial protein-coding genes were designed using nucleotide alignments of crustacean and insect species (Podsiadlowski and Bartolomaeus, 2005). In a first step, these primers were used to get sequence information for some of the mitochondrial genes. Amplification was successful for cox1, cox3, and 12S (for primer sequences and annealing temperatures see Table 1). In addition, PCR primer pairs designed to amplify larger fragments were tested according to Roehrdanz et al. (2002). PCR conditions were as follows: an initial denaturation step (1 min, 94 °C) was followed by 40 cycles of 1 min denaturation at 94 °C, 1 min annealing at a temperature specific for each primer pair (Table 1), and 1 min (cox1, cox3, 12S) or 7 min (Roehrdanz primers) extension time at 68 °C; in the end a final extension step (68 °C, 5 min) was performed. Using the primers suggested by Roehrdanz et al. (2002) amplification was successful with two primer pairs, CB2H-N4(87) and N4-16S2, each yielding a fragment of about 2 kb.

In the last step, PCR primers were designed for amplification of the missing fragments located between the already known ones. Several combinations of primers were tested to do so. PCR conditions for these secondary steps were as follows: an initial denaturation step (94 °C, 2 min) is followed by 40 cycles of denaturation (94 °C, 30 s), annealing (1 min, 52–55 °C) and extension (5 min, 68 °C, 90 s) and in the end a final extension step (68 °C, 5 min). Combinations that yielded PCR products were: 16Sr-cox1r, cox1f-cox3r, and cox3f-12Sr (for primer sequence information see Table 1).

PCR products were purified using the PCR purification kit (Qiagen) or if necessary the Gel extraction kit (Qiagen) following manufacturers’ protocols. All sequencing was done on a CEQ 8000 capillary sequencer (Beckman–Coulter). DCTS Quick start kits (Beckman–Coulter) were used following manufacturers’ protocols: 4 μl DCTS solution was mixed with 1 μl sequencing primer (10 nM), 1 μl purified PCR product and 4 μl molecular grade water. Sequencing PCR was performed on an Eppendorf Mastercycler gradient with 30 cycles of denaturation (20 s, 96 °C), annealing (20 s, primer specific temperature), and extension (2 min 60 °C). Sequence editing and assemblage was done using CEQ software and Bioedit (Hall, 1999). To determine gene identity BLAST searches on NCBI Blast Entrez databases was used. Starting points and ends of ribosomal RNAs were assumed to extend to the boundaries of flanking genes. Start codons and stop codons of protein-coding genes were presumed to be the nearest start or stop codon to the beginning of the sequence alignment of homologous genes with other malacostracan species. tRNAs were identified using tRNAscan-SE 1.21 (Lowe and Eddy, 1997), some were only found after eye inspection of the suspected regions. Transfer-RNA identities were specified by their anticodon sequence. The nucleotide sequence of the partial mitochondrial genome of I. baltica (14,247 bp) was deposited in GenBank, Accession No. DQ442915.

2.3. Phylogenetic analysis

Concatenated amino acid alignments were used for phylogenetic analyses. Mitochondrial genome data from the following species (with accession numbers) were used during phylogenetic analyses: Artemia franciscana (Anostraca; NC001620), Daphnia pulex (Cladocera; NC000844), Triops cancrigormis (Notostraca; NC004465), Vargula hilgendorf (Ostracoda; NC005306), Hutchinsoniella macracantha (Cephalocarida; NC005937), Megabalanus volcano (Cirripedia; NC006293), Tetraclita japonica (Cirripedia; NC008974), Pollicipes polymerus (Cirripedia; NC005936), Pseudosquilla ciliata (Stomatopoda; AY947836), Squilla
mantis (Stomatopoda; NC006081), Lysiosquillina maculata (Stomatopoda; NC007443), Harpiosquilla harpax (Stomatopoda; NC006916), Macrobrachium rosenbergii (Caridea; NC006880), Penaeus monodon (Penaeidea; NC002184), Marsupenaeus japonicus (Penaeidea; NC007010), Cherax destructor (Astacidea; NC011243), Pagurus longicarpus (Anomura, NC003058), Panulirus japonicus (Palaemonidea, NC004251), Portunus trituberculatus (Brachyura; NC007379), Callinectes sapidus (Brachyura; NC006281), Geothelphusa dehaani (Brachyura; NC007379), Eriocheir sinensis (Brachyura; NC003979), Pseudocarcinus gigas (Brachyura; NC007379), Parhyale hawaiiensis (Peracarida: Amphipoda; AY639937). We used only sequences from protein-coding genes which are located on the same strand in all of the species, leaving eight genes for the final analysis (atp6, cox1-3, nad1-4). Alignments were done using CLUSTAL X (version 1.81; Jeanmougin et al., 1998) with default settings (gap opening: 10; gap extension: 0.2; protein weight matrix: Gonnet 250). Ambiguously aligned proportions were discarded aided by the Geneblocks software (Version 0.91b; Castresana, 2000) with the following settings: 27 taxa; minimum number of sequences for a conserved position: 14; minimum number of sequences for a flank position: 22; maximum number of contiguous non-conserved positions: 8; minimum length of a block: 10; allowed gap positions: “with half.” The final alignments consisted of 1966 amino acids or 5898 nucleotides, respectively, for 27 taxa. Two taxa (Speleonectes tulumensis, Remipedia, NC005938 and Tigriopus japonicus, Copepoda, NC003979) originally included in the analysis failed to pass the afii9851 test of homogeneity in amino acid composition compared to the complete alignment (implemented in tree-puzzle), and were therefore excluded from the final analysis.

Phylogenetic inference was estimated by the following methods: (1) distance analysis using logdet distances with DAMBE (version 4.2.13; Xia and Xie, 2001), with 1000 bootstrap replicates; (2) parsimony using PAUP* (version 4.0b10; Swofford, 2001), with 1000 bootstrap replicates; (3) maximum likelihood, performed with PHYML (Guindon and Gascuel, 2003; http://atgc.lirmm.fr/phyml) for the amino acid alignment (with 100 bootstrap replicates under mtREV model, one of the few amino acid models derived from mitochondrial protein sequences) and PAUP* for the nucleotide alignment (model GTR+I+J; chosen according to the AIC with modeltest ver. 3.7; Posada and Crandall, 1998). (4) Bayesian inference using MrBayes (ver. 3.1.2; Ronquist and Huelsenbeck, 2003) was used for estimation of clade support for the nucleotide alignment (one million generations were run under the model GTR+I+J; 100 out of 1000 trees were discarded as burn-in).

3. Results and discussion

3.1. Organization of the mitochondrial genome of I. baltica

A large segment, comprising about 90% of the mitochondrial genome has been reconstructed from sequencing results. Several attempts to amplify the missing region between 12S rRNA and cyt b failed. There are some reports suggesting linearization of the mitochondrial DNA in another isopod, Armadillidium vulgare (Raimond et al., 1999; Rigaud et al., 1999). This would provide a convincing explanation for the failure of closing this sequence gap. However, PCR amplification of the mitochondrial control region is known to be difficult, and there is yet no other...
evidence for a linear organization of the mitochondrial genome in *I. baltica*.

Relative positions of the genes are shown in Fig. 1. All 13 protein coding and 2 rRNA genes that are usually present in metazoan mitochondrial genomes were identified in the sequenced fragment, while we identified only 17 out of 22 tRNA genes that are usually present. The remaining 5 tRNA genes as well as the control region are probably located in the missing part between 12S rRNA and cytb.

At least six gene translocations have to be assumed to change the putative ancestral state of Pancrustacea into the derived gene order of the *I. baltica* mitochondrial genome (Fig. 1). Four translocations affected single genes (*tRNA-T, tRNA-R, cytb, 12S*), while the other two affected blocks of two adjacent genes (*nad1/tRNA-L; nad5/tRNA-F*). Three translocations also led to an inversion of the gene or gene block (*cytb, 12S, nad1/tRNA-L*). The ancestral state of Pancrustacea is still present in several insects, some decapods, stomatopods and the phyllopods *Daphnia pulex* and *Triops cancriformis*. Compared to that gene order the position of the control region must also have changed in *I. baltica* because there is no noncoding region present between 16S rRNA and *nad2*. It is likely that its location is in the part between 12S and cytb where sequence data are missing. While lots of cases have been reported where tRNA translocations are involved, such a diversity of rearrangements (protein coding genes, control region) is reported only from few other crustaceans: the ostracod *Vargula hilgendorfi* (Ogoh and Ohmiya, 2004), the copepod *Tigriopus japonicus* (Machida et al., 2002), the branchiuran *Argulus americanus* (Lavrov et al., 2004) and the decapods *Cherax destructor* (Miller et al., 2004) and *Eriocheir sinensis* (Sun et al., 2005).

Fig. 2. Maximum likelihood tree depicting crustacean relationships. Concatenated amino acid sequences of eight protein-coding genes (*atp6, cox1-3, nad1-4*) were used to infer phylogeny. Numbers above branches are bootstrap percentages from maximum likelihood (upper), minimum evolution logdet distance (middle), and parsimony analyses (lower). Malacostracan taxa indicated in bold. See text for full species names and accession numbers.
The only other peracarid crustacean with almost complete mitochondrial genome sequence published is the amphipod *Parhyale hawaiiensis* (Cook et al., 2005; the sequence lacks the control region and five tRNA genes). In *P. hawaiiensis* several gene translocations have to be presumed, with only tRNA genes involved (tRNA-C, tRNA-Y, tRNA-Q, g, tRNA-R, tRNA-N, tRNA-H, tRNA-T; Fig. 1). Only tRNA-T was translocated in both peracarids, *P. hawaiiensis* and *I. baltica*, but to different final positions (between nad4/nad6 in *P. hawaiiensis* and between cyt/b/nad5 in *I. baltica*), suggesting independent translocation events in the two species. We conclude that no gene translocation is shared by *I. baltica* and *P. hawaiiensis*, in other words, that none of these events took place before the split between isopods and amphipods. Shorter fragments from other amphipods show no differences with the supposedly plesiomorphic pancrustacean gene order (Fig. 1; several *Cyamus* species, according to Kaliszewska et al., 2005; *Gammarus minus*, according to GenBank AF228046, Waits et al., unpublished manuscript). Therefore, the last common ancestor of isopods and amphipods, and the ancestor of all peracarids must have shared the ancestral pancrustacean gene order. More data from other peracarid taxa, especially from the more basal taxon Mysida, would be a good test to confirm this hypothesis. Furthermore, gene translocations may be helpful as phylogenetic markers to resolve amphipod or isopod ingroup relationships or to determine their closest relatives.

3.2. Phylogenetic analysis

The phylogenetic analysis of malacostracan mitochondrial genomes (concatenated amino acid sequences from eight protein-coding genes, Fig. 2; first and second codon positions from eight protein-coding genes, Fig. 3) did not support a peracarid clade consisting of *I. baltica* and *P. hawaiiensis* and did not even place them among the
remainder of Malacostraca. The latter group (Malacostraca without peracarids), as well as Stomatopoda, Brachyura, and Cirripedia find good support in the analyses of both datasets. Reptantia is well supported in the nucleotide-based analyses, but only weakly supported in analyses of amino acid data. Decapoda is well supported only in the nucleotide-based analysis, while in the analysis using amino acid data it is paraphyletic with respect to Stomatopoda. The position of the peracarid taxa varies among the analyses. Using amino acid sequences I. baltica clusters with Phyllopoda in the maximum likelihood tree, but without good bootstrap support, while P. hawaiiensis is placed together with an ostracod and an anostracan, likewise not well supported by bootstrap analysis. Using the nucleotide dataset, I. baltica, P. hawaiiensis and an ostracod form a clade with the remainder of Malacostraca in the maximum likelihood tree, but their relationships are not well supported by bootstrapping or through Bayesian posterior probabilities.

In a recent study of malacostracan phylogeny using nuclear 18S and 28S rRNA monophyly of Peracarida was not supported (Babbitt and Patel, 2005), but the peracarid taxa form the most basal branches of Malacostraca. As there were ostracods missing in that analysis these findings are similar to our results obtained with the nucleotide dataset. Similar results come from the most accurate analysis concerning Peracarida (Spear et al., 2005). In that study the authors show that Mysida do not belong to the remainder of peracarids, but rather forms the sister group to all other Eumalacostraca (excl. Peracarida). In our study, Peracarida (excl. Mysida) is the sister group to all other Eumalacostraca. A similar position of peracarids among Malacostraca was also obtained in an analysis by Wills (1998) based on a morphological dataset including many fossil taxa.

Our findings that peracarids do not cluster with the other Eumalacostraca is therefore not surprising, but mitochondrial genome data of I. baltica and P. hawaiiensis additionally seem biased through reversed nucleotide frequencies. Recently, Hassanin et al. (2005) and Hassanin (2006) demonstrated that strand reversal of mitochondrial genes leads to dramatic shifts in nucleotide frequencies. Depending on the orientation of the mitochondrial control region, one strand shows a bias in favor of (C/A), the other of (G/T). In order to check if there are suspicious differences between the taxa and genes under study we determined nucleotide frequencies in third codon positions from protein-coding genes encoded on the (+)-strand (cox1) and (−)-strand (nad4), respectively (Table 2). While most taxa show a clear bias towards (C/A) in the (+)-strand encoded gene and for (G/T) in the (−)-strand, four taxa show a reversed bias: the two peracarid species, the cephalocarid Parhyale hawaiiensis and the ostracod I. baltica may additionally a...


