Mitochondrial DNA, 2009; 20(4): 1–12

informa
healthcare

The complete mitochondrial genome of the subterranean crustacean *Metacrangonyx longipes* (Amphipoda): A unique gene order and extremely short control region

MARIA M. BAUZÀ-RIBOT[1], DAMIÀ JAUME[2], CARLOS JUAN[1], & JOAN PONS[2]

[1]*Departament de Biologia, Universitat de les Illes Balears, Palma de Mallorca, Balearic Islands, Spain, and*
[2]*IMEDEA (CSIC-UIB), Instituto Mediterráneo de Estudios Avanzados, Esporles, Balearic Islands, Spain*

(*Received 12 November 2008; revised 3 March 2009; accepted 13 April 2009*)

Abstract
Metazoan mitochondrial genomes usually consist of the same gene set, but some taxonomic groups show a considerable variety in gene order and nucleotide composition. The mitochondrial genomes of 37 crustaceans are currently known. Within the malacostracan superorder Peracarida, only three partial mitogenome sequences and the complete sequence of *Ligia oceanica* (Isopoda) are available. Frequent translocation events have changed the mitochondrial gene order in crustaceans, providing an opportunity to study the patterns and mechanisms of mitogenome rearrangement and to determine their impact on phylogenetic reconstructions. Here we report the first complete nucleotide sequence of an amphipod species, *Metacrangonyx longipes*, belonging to a phylogenetically enigmatic family occurring in continental subterranean waters. The genome has 14,113 bp and contains the usual 13 protein coding genes and two rRNA subunits, but only 21 out of the typical 22 tRNA genes of Metazoa. This is the shortest mitogenome described thus far for a crustacean and also one of the richest in AT (76.03%). The genome compactness results from a very small control region of 76 bp, the occurrence of frequent gene overlap, and the absence of large non-coding fragments. Six of the protein-coding genes have unusual start codons. Comparison of individual protein coding genes with the sequences known for other crustaceans suggests that *nad2*, *nad6*, *nad4L* and *atp8* show the highest divergence rates. *M. longipes* shows a unique crustacean mitogenome gene order, differing even from the condition found in *Parhyale hawaiiensis* (Amphipoda), whose coding sequence has also been completed in the present study.

Keywords: Metacrangonyx longipes*, Amphipoda, mitochondrial genome, control region, gene order*

Introduction

The mitochondrial genome (mitogenome) of metazoans generally comprises a circular double-stranded DNA molecule of 12–20 kb with a highly conserved gene content. It includes 13 protein-coding, two ribosomal and up to 22 transfer RNA genes (Wolstenholme 1992). The Crustacea have more than 52,000 described species, with a range in body plan not matched in any other group of metazoans (Martin and Davis 2001). They include the six recognized classes: Branchiopoda, Cephalocarida, Malacostraca, Maxillopoda, Ostracoda and Remipedia (Martin and Davis 2001). The mitogenome sequences of 37 species of Crustacea have been completed thus far (http://www.ncbi.nlm.nih.gov/genomes), of which 15 correspond to malacostracan decapods (Carapelli et al. 2007; Yang and Yang 2008). Within the malacostracan peracarid order Amphipoda, only a partial mitogenome sequence is currently available in sequence databases: that of *Parhyale hawaiiensis*, although it lacks of about 3 kb including the *rrnS* gene and parts of *rrnL* and *nad2*, and also the control region (Cook et al. 2005). In addition, in the peracarid order Isopoda, one entire (*Ligia oceanica*) and two partial mitogenomes (*Armadillidium vulgare* and *Idotea baltica*) are known (Kilpert and Podsiadlowski 2006;

Correspondence: J. Pons, IMEDEA (CSIC-UIB), Instituto Mediterráneo de Estudios Avanzados, Miquel Marquès 21, Esporles, 07190 Balearic Islands, Spain. Tel: 34 971 611332. Fax: 34 971 611761. E-mail: joan.ponspons@uib.es

ISSN 1940-1736 print/ISSN 1940-1744 online
DOI: 10.1080/19401730902964417

Podsiadlowski and Bartolomaeus 2006; Marcadé et al. 2007). The taxon sampling set for crustacean mitogenomes is quite poor because only 30 out of about 800 known crustacean families are represented (Martin and Davis 2001). Despite this, two important insights into pancrustacean phylogenetics are based on mitogenome data. First, phylogenetic analyses of protein-coding genes (PCGs) including the more intensively sampled mitochondrial genomes of Hexapoda suggest a mutual paraphyly of Crustacea and Hexapoda (Cook et al. 2005; Carapelli et al. 2007). Second, frequent translocation events have apparently changed the mitochondrial gene order in crustaceans compared with the putative ancestral pattern (Kilpert and Podsiadlowski 2006; Yang and Yang 2008). This gene order results from a common inversion of a *trnL2* gene present in a large number of crustaceans and insects, that translocated from a location inferred to be the primitive state as it is found in chelicerates, myriapods, onychophorans, tardigrades, as well as in Pogonophora, Annelida, Echiura, and Mollusca (Boore et al. 1995, 1998). Gene order is not conserved within the superorder Peracarida (for which only information on Isopoda and Amphipoda is currently available), nor is it even conserved within the Isopoda. Despite those differences, the mitogenome of the isopods *L. oceanica*, *I. baltica* and *A. vulgare* share some gene rearrangements (i.e. putative isopod synapomorphies), compared with the arthropod pattern and that of the amphipod *P. hawaiiensis* (Kilpert and Podsiadlowski 2006).

The Metacrangonyctidae (Boutin and Messouli 1988) represent a small family of amphipod crustaceans with two genera: *Metacrangonyx* (Chevreux 1909) (17 species) and *Longipodacrangonyx* (Boutin and Messouli 1988) (monotypic). All members of the family occur only in continental subterranean waters and represent a phylogenetically enigmatic lineage of marine origin showing an extremely disjunct geographic distribution. Two species are found in the Dominican Republic (Hispaniola) (Jaume and Christenson 2001), one is known from Fuerteventura in the Canary Islands (Stock and Rondé-Broekhuizen 1986), 11 from Morocco (Balazuc and Ruffo 1953; Ruffo 1954; Karaman and Pesce 1979; Boutin and Messouli 1988a,b; Messouli et al. 1991; Oulbaz et al. 1998), one from Elba Island, Italy (Stoch 1997), one from the Balearic Islands (Chevreux 1909; Margalef 1952) and two from the Middle East (Ruffo 1982; Karaman 1989). Whereas most species live in interstitial freshwater associated with springs, wells or alluvial sediments, some taxa occur in brackish or athalassohaline waters. Only *Metacangronyx longipes* (Chevreux 1909) from the Balearics and the two Hispaniolan species are ordinary cave dwellers, living in fresh to marine subterranean waters (Jaume and Christenson 2001).

It has been proposed that the Metacrangonyctidae derive from marine littoral ancestors that colonized the continental ground waters during episodes of marine regression (Boutin and Coineau 1990). Although first supposed to be no older than the opening of the northern Atlantic ocean (Boutin 1994), the discovery of *Metacrangonyx* in the Greater Antilles (Jaume and Christenson 2001) suggests a much older origin for the genus: at least before the opening of the northern Atlantic (110 million years before present). Thus its current distribution would be the result of vicariance by plate tectonics and of peripatric speciation associated with episodes of regression in the paleocoastline of Tethys.

We present here the first complete sequence of a mitochondrial genome of an amphipod. We have used the mitogenome of *M. longipes* to compare its gene order with those of other crustaceans, as well as its nucleotide composition and tRNA structure. We especially focus in comparisons on other peracarids such as the amphipod *P. hawaiiensis*, for which we have almost completed the whole mitogenome (except approximately 500 bp of the control region that has not been sequenced because of technical problems), and the isopods *L. oceanica*, *I. baltica* and *A. vulgare*.

Materials and methods

Sampling and DNA extraction

A 3 mm long specimen of *M. longipes* preserved in absolute ethanol was used for DNA extraction by means of the DNeasy Tissue Kit (Qiagen, Hilden, Germany) following the manufacturer's protocol for total genomic DNA purification. The specimen was collected in Cala Varques cave (Mallorca Island, Spain) during fall 2007.

PCR primers and conditions

Gene fragments at opposing ends of the mitochondrial genome were amplified using standard protocols outlined elsewhere (Balke et al. 2005) and universal primers (Table I). Based on the sequence obtained, we designed species-specific long primers (Table I) of about 25–29 bp targeted at the *cox1*/*rrnL* genes to amplify two long fragments of about 4.5 and 10 kb covering the whole circular mitochondrial genome. Long-range PCR amplifications were performed using TaKaRa LA *Taq* polymerase (Takara Bio, Inc., Tokyo, Japan) according to the manufacturer's specifications. The general reaction mixture for each 50 μl was: 5 μl of 10 × LA PCR buffer, 5 μl of 25 mM MgCl, 8 μl of dNTP mixture (2.5 mM each), 2.5 μl forward primer (10 μM), 2.5 μl reverse primer (10 μM), 0.5 μl Takara LA *Taq* (5 U/μl), 24.5 μl distilled H_20 and 2 μl genomic DNA. PCR cycles were as follows: after an initial denaturation step of 94°C for 90 s, 14 cycles were performed at 94°C for 30 s, 57–62°C (depending on primers) for 30 s and 68°C for 5–15 min depending

Table I. Universal and long PCR primers designed to amplify the mitochondrial fragments of *M. longipes*.

Primer name	Forward/reverse	Gene	5′ → 3′ sequence	Reference
16Sa (aka M14)	Forward	*rrnL*	CGCCTGTTTAACAAAAACAT	Xiong and Kocher (1991)
16Sb (aka M74)	Reverse	*rrnL*	CTCCGGTTTGAACTCAGATCA	Xiong and Kocher (1991)
CB3	Forward	*cob*	GAGGAGCAACTGTAATTACTAA	Barraclough et al. (1999)
CB4	Reverse	*cob*	AAAAGAAARTATCATTCAGGTTGAAT	Barraclough et al. (1999)
HCO2198	Reverse	*cox1*	TAAACTTCAGGGTGACCAAAAAATCA	Folmer et al. (1994)
LCO1490	Forward	*cox1*	GGTCAACAAATCATAAAGATATTGG	Folmer et al. (1994)
Metcox1_F2	Forward	*cox1*	TATACGAGTTGGGATAATAGGAATAGACC	Present study
Metcox1_F4	Forward	*cox1*	TAACTGACCGAAACCTAAATACCTCTT	Present study
Met16s_R2	Reverse	*rrnL*	TGAAAAATAGAAAGTATAGCCTGCCC	Present study
Met16S_F1	Forward	*rrnL*	TTCATAAATAGTTCTACCTTTCATTCCAG	Present study
Metcox1_R1	Reverse	*cox1*	CCCTTATAGCAGTTCCTACTATCCTAGC	Present study
Met16S_F2	Forward	*rrnL*	CTATCCAAAATTATTACGCTGTTATCCC	Present study
Metcox1_R2	Reverse	*cox1*	GAAGAAGTCAGTTACCGAACCCTCC	Present study

on the expected fragment size. This was followed by 16 cycles at 94°C for 30 s, 57–62°C for 30 s and 68°C for 5-15 min (increasing by 15 s each cycle) and a final extension for 10 min at 72°C.

Cloning and sequencing

Long mitochondrial fragments were digested independently with *Dra* I, *Rsa* I and *Taq* I restriction enzymes according to the manufacturer's specifications. DNA digestions showed fragments ranging from 150 pb to 1.5 kb when checked on 2% agarose gels. DNA fragments from the three digestions were pooled and purified using the MinElute PCR Purification Kit (Qiagen) and were then cloned into a pJET blunt cloning vector (Fermentas, Glen Burnie, MD, USA) according to the specifications of the manufacturer. One-shot competent *Escherichia coli* cells from Invitrogen (Madison, WI, USA) were used for transformation. Ninety-six recombinant colonies were screened by PCR amplification for inserts of a minimum of 300 bp, and 63 were sequenced in both directions using the pJET vector sequencing primers. Sequences obtained from clones were then used to design specific primers to sequence the long PCR fragments directly by primer walking (list of primers available upon request) to obtain a full contig of the mitogenome. Additional primers were designed to close particular gaps in the sequence. The forward and reverse strands of small PCR amplicons or the long PCR fragments were cycle-sequenced using the ABI BigDye Terminator Cycle Sequencing Kit on an ABI PRISM 3100 Genetic Analyzer (Applied Biosystems, Foster City, CA, USA).

Gene annotation and sequence analysis

Analyses of the quality of chromatograms and contig construction to obtain the whole mitochondrial sequence were performed with the software CodonCode Aligner v2.0 (CodonCode Corp., Denham, MA, USA). Ambiguous nucleotide positions were validated by direct inspection of the chromatograms. Preliminary gene identification was determined by BLAST searching on GenBank databases (http://www.ncbi.nlm.nih.gov) and making multiple alignments to other crustacean nucleotide and amino acid sequences (see Additional File 1 for a list of species and accession numbers). Definitive annotations were performed using the DOGMA webserver (Dual Organellar GenoMe Annotator; http://www.bugmaster.jgi-psf.org/dogma). The 5′ and 3′ ends of protein and ribosomal genes were refined manually by comparison with the complete genes of other crustaceans. Transfer RNA genes were determined with tRNAscan-SE Search Server v1.21 (http://www.lowelab.ucsc.edu/tRNAscan-SE) using a tRNA covariance model (Lowe and Eddy 1997) and by inspection of anti-codon sequences and the predicted secondary structures. Nucleotide frequencies of protein coding and RNA genes were calculated with the DAMBE software package (Xia and Xie 2001), while the effective number of codons was determined according to INCA v1.20 (Supek and Vlahovicek 2004).

Divergence in protein coding genes

Mean nucleotide divergences of individual PCGs were estimated from pairwise comparisons among the complete mitogenomes of crustaceans and were subsequently compared with the values obtained for 35 species representing all major Hexapoda orders for which there are data available. MEGA v4.0.2 (Tamura et al. 2007) was used to calculate corrected distances using the Maximum Composite Likelihood model (Tamura et al. 2004) and among-sites rate variation following a gamma distribution with a shape parameter of 0.4 as estimated in RAxML v7.0.4 (Stamatakis 2006). Gapped positions were not considered in the analysis of each pairwise comparison. Mean divergence values were normalized by dividing the value obtained for each gene by the value of the gene with the highest rate.

Gene rearrangement analyses

We used the program CREx (Bernt et al. 2007) to deduce gene rearrangement scenarios in crustacean mitogenomes based on the detection of strong interval trees on the CREx webserver (http://www.pacosy.informatik.uni-leipzig.de/crex). The strong interval trees reflect genes that appear consecutively in several of the input gene orders; that is, given two gene orders, a set of genes is a common interval if the genes in that set appear consecutively in both gene orders. A certain subset of all common intervals, the "strong common intervals", can be represented as the nodes of a special type of tree. The descendants of a node (strong common interval) are simply the strong common intervals that it includes entirely. If the descendants of a node appear in the same order in both input gene orders, the node is called "linear increasing" (+); if the children of a node appear in exactly the opposite order, it is "linear decreasing" (−); otherwise, the node is called prime (Bernt et al. 2007).

Results and discussion

Genome organization

The mitochondrial sequence of *M. longipes* has an overall length of 14,113 bp (EMBL accession number: AM944817) and shows the usual circular organization found in most metazoans (Figure 1). To our knowledge, this is the smallest mitogenome described so far for a crustacean: close to that of *Tigriopus californicus* (Copepoda, Harpacticoida, Maxillopoda; 14,546 bp) (Machida et al. 2002). Gene annotation reveals the presence of the typical 13 PCGs and the 2 rRNA subunits of metazoan mitochondrial genomes (Table II), but only 21 tRNA genes instead of the typical 22; this is similar to the condition found in the isopod *L. oceanica* (Kilpert and Podsiadlowski 2006). The compactness of the genome is due to the occurrence of frequent gene overlap, since more than 20 genes share borders. These overlapping regions range in size from just 1 bp (several cases) to a maximum of 63 bp (in the gene coding for tRNA-Val, which overlaps with 44 bp of the 5′ end of *rrnL* and with 19 bp of the 3′ end of *rrnS*). Small non-coding sequences or intergenic spacers (range 1–17 bp; see Table II) are also evident in the mitochondrial DNA (mtDNA). A further region of non-coding DNA comprising 76 bp, placed between the *rrnS* and *cob* genes and with an AT content of 84.22% presumably corresponds to the control region and contains the origin of mtDNA replication. The region has a putative secondary structure folding into a hairpin, with a stem of 15 paired nucleotides plus a short loop of four nucleotides (Figure 2). This is similar to other stem-loop structures known to occur in insect mitochondrial control regions (Zhang et al. 1995) and that are presumed to be the origin of replication of mtDNA. The 3′-flanking sequence around the stem region shows

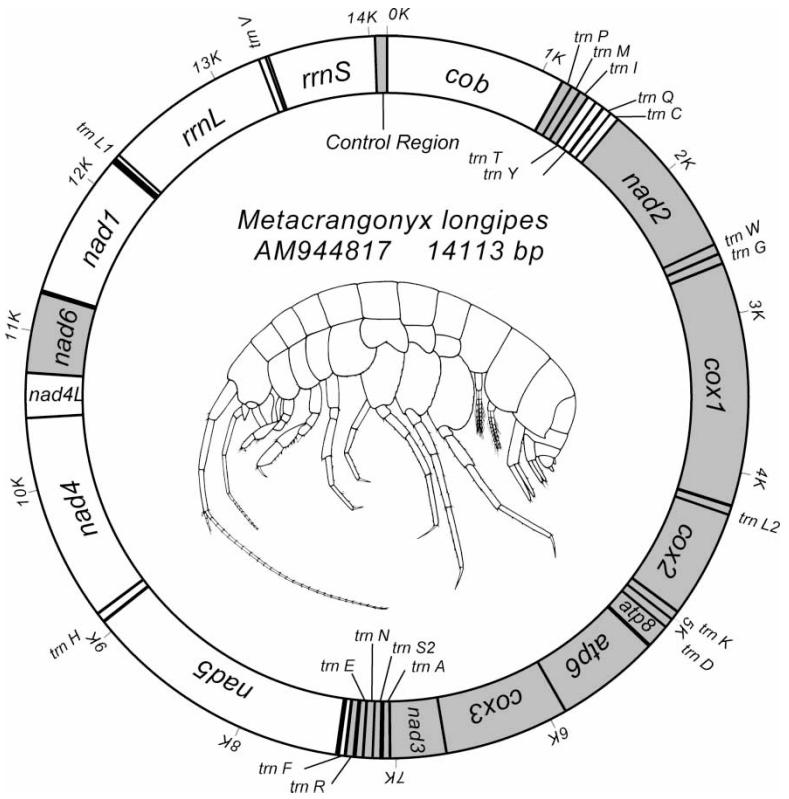


Figure 1. Map of the mitochondrial genome of *M. longipes*. Note: Grey and white segments indicate genes coded on the + strand and − strand, respectively.

Table II. Gene content of the mitochondrial genome of *M. longipes*.

Gene	Position	Strand	Size (bp)	Start codon	Stop codon
cob	1–1137	−	1137	ATG	TAA
Intergenic spacer 1	1138–1140	+	3	n.a.	n.a.
trnP	1141–1204	+	64	n.a.	n.a.
trnM	1205–1265	+	61	n.a.	n.a.
trnI	1264–1326	+	63	n.a.	n.a.
trnT	1325–1383	−	59	n.a.	n.a.
Intergenic spacer 2	1384–1384	+	1	n.a.	n.a.
trnY	1385–1443	−	59	n.a.	n.a.
trnQ	1439–1498	−	60	n.a.	n.a.
trnC	1499–1556	−	58	n.a.	n.a.
nad2	1556–2546	+	991	ATT	T
trnW	2547–2610	+	64	n.a.	n.a.
Intergenic spacer 3	2611–2612	+	2	n.a.	n.a.
trnG	2613–2674	+	62	n.a.	n.a.
cox1	2677–4218	+	1542	ATT	TAA
trnL1	4214–4273	+	60	n.a.	n.a.
cox2	4274–4958	+	685	ATT	T
trnK	4957–5016	+	60	n.a.	n.a.
trnD	5017–5077	+	61	n.a.	n.a.
atp8	5078–5236	+	159	ATC	TAA
atp6	5227–5895	+	669	ATG	TAA
cox3	5895–6683	+	789	ATG	TAA
Intergenic spacer 4	6684–6684	+	1	n.a.	n.a.
nad3	6685–7038	+	354	ATT	TAG
trnA	7037–7097	+	61	n.a.	n.a.
trnS2	7084–7146	+	63	n.a.	n.a.
trnN	7147–7207	+	61	n.a.	n.a.
trnE	7205–7265	+	61	n.a.	n.a.
trnR	7254–7320	+	67	n.a.	n.a.
trnF	7315–7374	−	60	n.a.	n.a.
nad5	7358–9076	−	1719	ATT	TAG
Intergenic spacer 5	9077–9082	+	6	n.a.	n.a.
trnH	9083–9142	−	60	n.a.	n.a.
nad4	9140–10,454	−	1315	ATA	T
nad4L	10,454–10,738	−	285	ATA	TAA
nad6	10,737–11,243	+	507	ATA	TAA
Intergenic spacer 6	11,244–11,260	+	17	n.a.	n.a.
nad1	11,261–12,181	−	921	ATA	TAA
trnL2	12,176–12,235	−	60	n.a.	n.a.
rrnL	12,206–13,342	−	1137	n.a.	n.a.
trnV	13,299–13,361	−	63	n.a.	n.a.
rrnS	13,343–14,037	−	695	n.a.	n.a.
Control region	14,038–14,113	+	76	n.a.	n.a.
All protein coding genes	n.a.	n.a.	11,073	n.a.	n.a.
All tRNAs	n.a.	n.a.	1287	n.a.	n.a.
Complete genome	1–14,113	+	14,113	n.a.	n.a.

the conserved motif GACT present also in the isopod *L. oceanica* and the hoplocarid malacostracan *Squilla mantis* (Kilpert and Podsiadlowski 2006), but the TATA element found in many hexapods at the 5′-flanking region is here replaced by an AATT motif. The low level of non-coding sequences found in the mitogenome of *M. longipes* (< 1%) and the occurrence of frequent gene overlap are indicative of an extremely compact mitogenome.

Protein coding genes: Nucleotide composition and codon usage

The AT content of the protein genes of *M. longipes* is 75.33% (A = 31.25%, C = 11.34%, G = 13.33% and T = 44.08%), while that of the complete mitogenome (+ strand) is 76.03%; this is one of the highest percentages reported in crustaceans and similar to those frequently found in Hexapoda mitochondrial genomes. *Argulus americanus* (Branchiura, Maxillopoda) has the highest AT content found so far in any crustacean at 77.80% (Machida et al. 2002), while the nearly complete amphipod mitochondrial sequence of *P. hawaiiensis* reaches 73.66% (Cook et al. 2005) (authors' own data).

Six of the 13 PCGs of *M. longipes* display unusual start codons for an arthropod mtDNA. The codon ATT is present in five genes (Table II), including *cox1*,

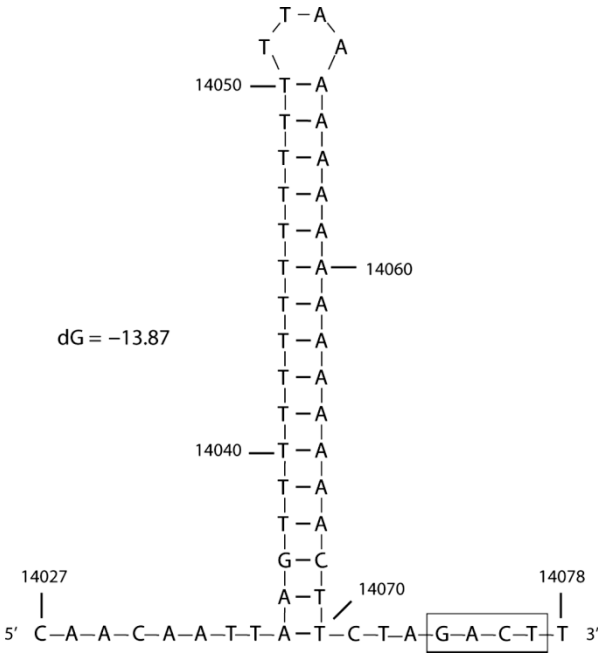


Figure 2. Putative secondary structure of the mitochondrial control region of *M. longipes*. Note: The box indicates the conserved GACT motif present also in the isopod *L. oceanica* and the hoplocarid malacostracan *S. mantis*.

which starts with ACG in other malacostracans (Kilpert and Podsiadlowski 2006 and references therein). In addition, the gene *atp8* starts with the non-canonical codon ATC (Table II). In turn, three of the PCGs show truncated stop codons (Table II). The genes for *nad2*, *nad4* and *cox2* end in a single T. As shown elsewhere, these truncated stop codons are likely to be completed by post-transcriptional polyadenylation, with final transcripts having functional UAA terminal codons (Ojala et al. 1981).

The *M. longipes* mitogenome shows a clear bias in nucleotide frequencies, with similar values in both strands (Table III). Strand bias reflected by GC skew (Perna and Kocher 1995) is slightly negative but close to zero in the genes encoded by the + strand, in contrast to the peracarid isopods studied so far, which show positive values (Kilpert and Podsiadlowski 2006). This has been attributed to the occurrence of inversions of the control region containing the replication origin in *L. oceanica* and *I. baltica*, since most crustaceans show moderate to high negative GC skews in the + strand (Hassanin et al. 2005; Hassanin 2006; Kilpert and Podsiadlowski 2006).

The effective number of codons was calculated for the PCGs. This is a simple measure of codon usage, ranging from 20 when only one codon is used for each amino acid, to 61 (or even 62 in the invertebrate mitochondrial genetic codes since UGA codes there for tryptophan) when all synonymous codons are equally in use (Wright 1990). In the *M. longipes* mitogenome the PCGs show low effective number of codons values (35.38 ± 2.84), so they use about one-half of the possible codons. There is a positive correlation between effective number of codons values and GC content in third codon positions ($r^2 = 0.464$; $P < 0.01$), as described elsewhere (Kilpert and Podsiadlowski 2006). However, the genes *nad2*, *nad3* and *nad6* use a lower number of effective codons than could be expected from their relatively high GC content at third codon positions (18% for *nad2*, and 23% for both *nad3* and *nad6*). Compared with isopods, *M. longipes* displays a considerably lower mean number of effective codons (and hence lower GC content at synonymous sites), and shows values similar to those found in the amphipod *P. hawaiiensis*

Table III. Number of amino acids in the protein coding genes, nucleotide frequencies, AT content, and AT and AG skew for mitochondrial genes of *M. longipes*.

Gene	Strand	Size (amino acids)	%A	%C	%G	%T	%A + T	AT skew	GC skew
cob	−	378	30.17	15.83	10.73	43.27	73.44	−0.18	−0.19
nad2	+	330	31.58	8.38	12.71	47.33	78.91	−0.20	0.21
cox1	+	513	28.34	13.81	16.15	41.7	70.04	−0.19	0.08
cox2	+	228	31.83	12.55	13.72	41.9	73.73	−0.14	0.04
atp8	+	52	42.77	8.8	5.66	42.77	85.54	0.00	−0.22
atp6	+	222	30.64	11.81	12.26	45.29	75.93	−0.19	0.02
cox3	+	262	26.87	13.94	16.22	42.97	69.84	−0.23	0.08
nad3	+	117	29.66	9.32	14.69	46.33	75.99	−0.22	0.22
nad5	−	572	34.03	10.3	12.97	42.7	76.73	−0.11	0.11
nad4	−	438	32.24	8.97	13.76	45.03	77.27	−0.17	0.21
nad4L	−	94	36.84	5.61	12.28	45.27	82.11	−0.10	0.37
nad6	+	168	29.39	11.83	7.89	50.89	80.28	−0.27	−0.20
nad1	−	306	32.14	9.45	14.66	43.76	75.9	−0.15	0.22
rrnL	−	n.a.	39.93	7.92	13.81	38.34	78.27	0.02	0.27
rrnS	−	n.a.	35.83	7.63	12.66	43.88	79.71	−0.10	0.25
Control region	+	n.a.	40.79	7.89	7.89	43.43	84.22	−0.03	0.00
All protein coding genes	n.a.	n.a.	31.25	11.34	13.33	44.08	75.33	−0.17	0.08
All tRNAs	n.a.	n.a.	40.95	9.01	12.98	37.06	78.01	0.05	0.18
Complete genome	+	n.a.	37.36	12.41	11.56	38.67	76.03	−0.02	−0.04

and the more AT-rich mitogenome of *A. americanus* (Branchiura, Maxillopoda) (Machida et al. 2002).

Divergence in protein coding genes

We used the complete dataset of mtDNA sequences of crustacean taxa plus a representation of all major Hexapodan orders for which data are available (35 taxa; see Additional File 1) to assess the relative divergence of individual PCGs. The genes showing lower corrected divergences across Crustacea and Hexapoda were *cox1*, *cox2*, *cox3* and *cob*, while *nad2*, *nad6*, *nad4L* and *atp8* displayed about twice the mean divergence values (Figure 3). There seems to be an association between gene variation and length and, perhaps, strand location, because shorter genes, often present on the − strand (such as *atp8* and *nad4L*), are the most divergent. Nevertheless, *nad2* is placed on the + strand in Hexapoda, and in most crustaceans also shows a high substitution rate. As noted elsewhere (Cameron and Whiting 2007; Salvato et al. 2008), both variability and codon usage analyses of individual PCGs of Isoptera and Lepidoptera reveal that some of the genes most used in molecular systematics, such as *cox1* and *cox2*, have the lowest variability, while the neglected genes *nad2*, *nad3*, *nad4* and *nad6* may prove to be very useful for systematics given their variability and informative nature. Our results show that this could be extended to crustaceans, which show an underlying substitution pattern similar to hexapods at protein coding genes.

Transfer RNA genes

We identified 17 tRNA genes in a general search on the *M. longipes* mitogenome using tRNAscan-SE, and other four (*trnS1*, *trnN*, *trnF* and *trnV*) were inferred from less stringent specific searches in non-coding regions (COVE cut-off score of −20). Despite this, the *trnS2* gene (tRNA-Ser$_{AGN}$) was not found, although it could almost completely overlap with either the *trnG* or *trnW* genes (COVE scores of +0.30 and −3.54, respectively). The trnS2 gene shows unusual characteristics in many arthropods, such as the lack of the DHU arm (Kilpert and Podsiadlowski 2006 and references therein). In addition, in *M. longipes* the tRNA-Thr shows an unusual secondary structure, lacking completely the TΨC arm, whereas the tRNA-Gln lacks the loop normally present at this arm (Figure 4). Nucleotide mismatches were evident in the acceptor stem for tRNA-Gln, tRNA-Arg and tRNA-Ile, and in the anticodon stem for tRNA-Lys (Figure 4). Many cases of mismatches in stems have been described in mitochondrial tRNAs, and are supposed to be modified by RNA editing (Ojala et al. 1981; Xiong and Kocher 1991; Yokobori and Paabo 1995; Kilpert and Podsiadlowski 2006). The tRNA genes are present in both strands although most of them (13 genes) are located in the + strand (Table II and Figure 1).

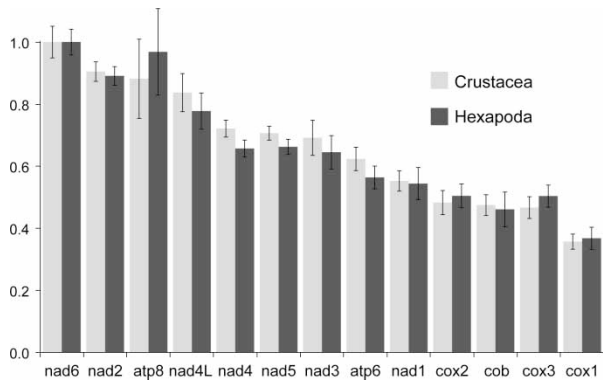


Figure 3. Mean relative corrected divergences of protein coding genes of Crustacea and Hexapoda. Note: DNA divergences of individual genes were estimated from pairwise comparisons among the complete mitogenomes of crustaceans and 35 species representing all major Hexapoda orders.

Ribosomal RNA genes

The *rrnS* and *rrnL* genes are approximately 695 and 1137 bp in length, respectively (Table II), and around 78% AT-rich, thus being considerably shorter than in other crustaceans. This further explains the extreme compactness of the *M. longipes* mitogenome. The *rrnL* gene of *M. longipes* is closest to those of the amphipods *P. hawaiiensis* and *Niphargus rhenorhodanensis* (accession number: EF028415) (75% sequence identity), while the *rrnS* gene does not show any significant similarity to the sequences of other crustaceans. Not enough information on crustacean 12S and 16S rRNAs secondary structure is available to attempt reconstructing their structure based on comparative analyses.

Gene order

M. longipes shows a mitochondrial gene order not found in any other crustacean so far analysed (Figure 5). Although the pancrustacean position of *trnL2* between *cox1* and *cox2* is conserved (Boore et al. 1995), many rearrangements in the *Metacrangonyx* genome compared with the ancestral pattern can be deduced (Boore et al. 1995, 1998). At least three transpositions involving genes *trnR*, *trnG* and *trnC* separately, two shifts of strand (reversals) – one involving the gene *cob* and another the segment including *trnP* and *trnT* – and three complex tandem duplications with subsequent random losses are needed to account for the pattern observed in *M. longipes* compared with the pancrustacean ancestral pattern using heuristic analyses of strong common intervals with CREx (Additional File 2). Alternatively,

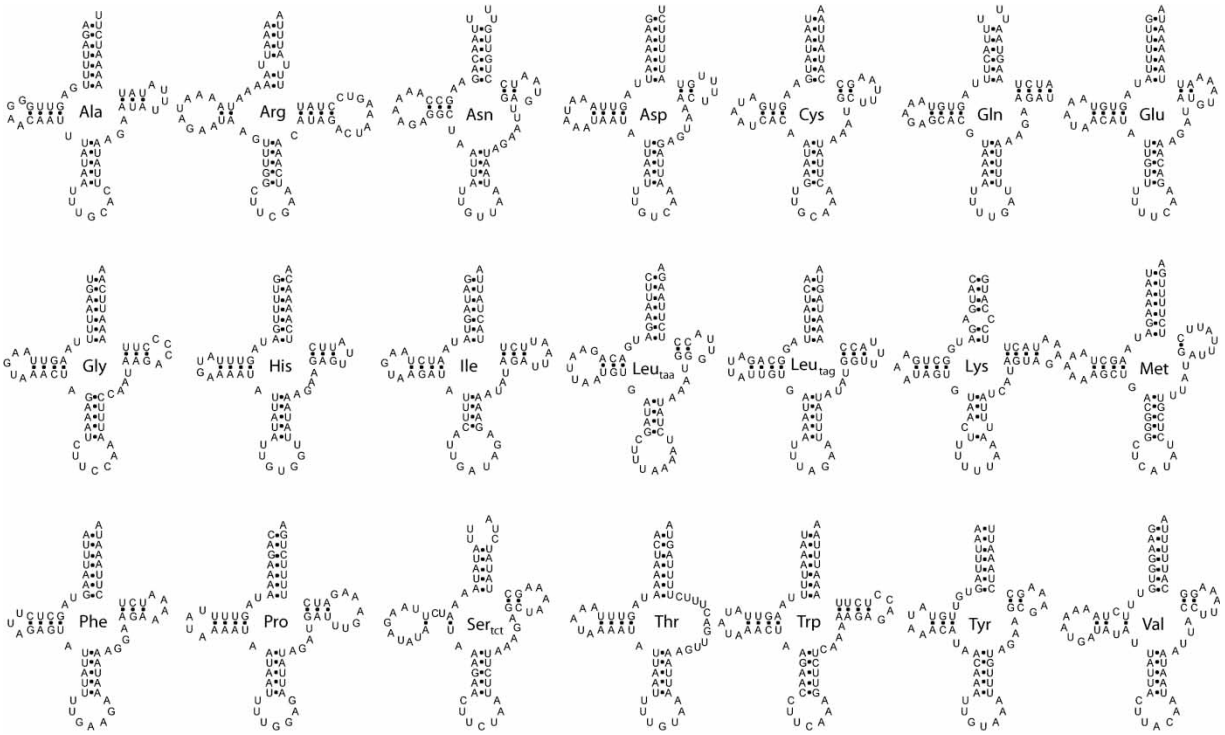


Figure 4. Putative secondary structures of mitochondrial tRNAs in *M. longipes*.

one single reversal of the ancestral pancrustacean segment including *cob nad6 trnP trnT*, followed by a new reversal of the gene *nad6* from the − to the + strand, plus three tandem duplications with subsequent random losses could have produced the *M. longipes* mitogenome gene order. The gene order also differs from the pattern found in the only other amphipod analysed thus far, *P. hawaiiensis* (Cook et al. 2005) (authors' our own data). We have almost completed the sequence for the mitogenome of this species except for a short part of the gene *rrnS* and the control region (EMBL accession numbers: FM957525 and FM957526), annotating the genes for *trnV*, partial *rrnS*, *trnM*, *trnY*, *trnC*, and locating the gene *trnH* between *nad5* and *nad4*, which was absent in the previous annotation (Cook et al. 2005). In addition, based on tRNAscan results, we reannotated the tRNA genes *trnW* and *trnG* previously annotated as *trnC* and *trnW*, respectively (Cook et al. 2005) (accession number: AY639937). In *P. hawaiiensis*, at least 10 of the tRNA genes show positional changes with respect to the pancrustacean pattern. The occurrence of identical transpositions of *trnR* and *trnG* in both *P. hawaiiensis* and *M. longipes* mitogenomes with respect to the ancestral arrangement suggests they could have arisen in the common ancestor of amphipods. The other peracarid mitogenomes known, those of the isopods *L. oceanica* and the uncomplete ones of *I. baltica* and *A. vulgare*, show quite different translocations from the assumed ancestral pancrustacean gene order (Figure 5), with apparently no common shifts derived from the peracarid ancestor being able to explain the observed patterns (Kilpert and Podsiadlowski 2006).

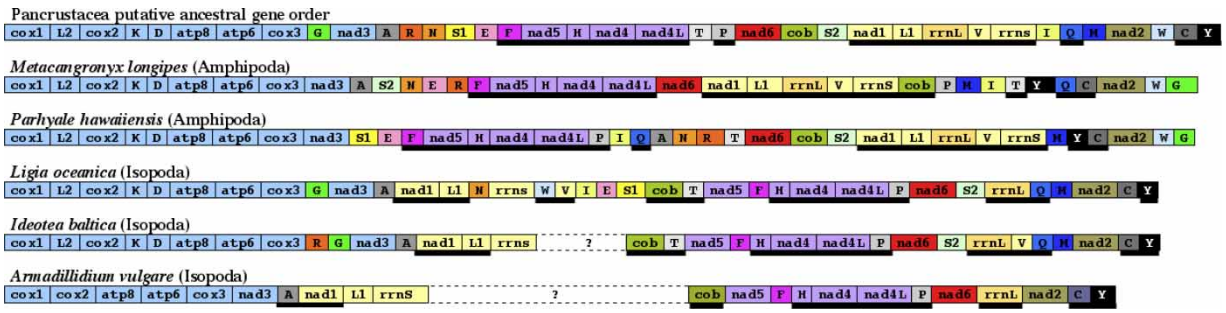


Figure 5. Mitochondrial gene order in Peracarida (Isopoda + Amphipoda) mitogenomes compared with the pancrustacean ancestral pattern. Note: Different colours are used to identify particular conserved and rearranged segments or genes. Genes underlined are present at the − strand.

Conclusions

The sequence of *M. longipes* introduced herein is the first complete mitogenome of a crustacean amphipod and the second for a peracarid thus far obtained, a superorder that is under-represented in the crustacean mitochondrial genome datasets currently available. The mitogenome is very compact, with a short control region, and it appears to be the shortest mitogenome described for a crustacean. Its AT content is high (76.03%), and gene order is not conserved compared with the other four peracarids whose complete or nearly complete mitogenomes are known: the isopods *L. oceanica*, *I. baltica* and *A. vulgare* and the amphipod *P. hawaiiensis*. Common transpositions of *trnR* and *trnG* in both *P. hawaiiensis* and *M. longipes* mitogenomes with respect to the ancestral pancrustacean arrangement suggest that they were present in the common ancestor of these two amphipods. Many differences in gene order are remarkable compared with the condition displayed in isopods. Thus, no inverted strand bias of nucleotide frequencies is found in *M. longipes*, contrary to what is reported for the mitogenomes of *L. oceanica* and *I. baltica* (Kilpert and Podsiadlowski 2006). The data presented herein not only expand the sampling within the crustacean mitochondrial genomes but also will help, when congeneric species from different geographic areas are sequenced, to solve the phylogenetic position and historical biogeography of this enigmatic family found exclusively in subterranean waters.

Acknowledgements

The present work has been financed by the research project CGL2006-01365 of the Spanish Ministerio de Ciencia y Innovación and European Union FEDER funds. MBR and JP are supported by FPI and "Ramon y Cajal" grants from the Ministerio de Ciencia e Innovación of Spain, respectively.

Declaration of interest: The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

References

Balazuc J, Ruffo S. 1953. Due nuove species del genere *Metacrangonyx* Chevreux delle acque interne del Nord Africa Francese. Mem Mus Civ Stor Nat Verona 4:25–33.

Balke M, Ribera I, Beutel RG. 2005. The systematic position of Aspidytidae, the diversification of Dytiscoidea (Coleoptera, Adephaga) and the phylogenetic signal of third codon positions. J Zool Syst Evol Res 43:223–242.

Barraclough TG, Hogan JE, Vogler AP. 1999. Testing whether ecological factors promote cladogenesis in a group of tiger beetles (Coleoptera: Cicindelidae). Proc R Soc Lond B Biol Sci 266:1061–1067.

Bernt M, Merkle D, Ramsch K, Fritzsch G, Perseke M, Bernhard D, Schlegel M, Stadler PF, Middendorf M. 2007. CREx: Inferring genomic rearrangements based on common intervals. Bioinformatics 23:2957–2958.

Boore JL, Collins TM, Stanton D, Daehler LL, Brown WM. 1995. Deducing the pattern of arthropod phylogeny from mitochondrial DNA rearrangements. Nature 376:163–165.

Boore JL, Lavrov D, Brown WM. 1998. Gene translocation links insects and crustaceans. Nature 392:667–668.

Boutin C. 1994. Phylogeny and biogegraphy of metacrangonyctid amphipods in North Africa. Hydrobiologia 287:49–64.

Boutin C, Coineau N. 1990. Regression model. Modèle biphase d'évolution et origine des micro-organismes stygobies interstitiels continentaux. Rev Micropaléontol 33:303–322.

Boutin C, Messouli M. 1988a. *Metacrangonyx gineti* n. sp. d'une source du Haut-Atlas marocain et la famille des Metacrangonyctidae n. fam. (Crustacés Amphipodes stygobies). Vie Milieu 38:67–84.

Boutin C, Messouli M. 1988b. *Longipodacrangonyx maroccanus* n. gen. n. sp., nouveau représentant du groupe *Metacrangonyx* dans les eaux souterraines du Maroc. Crustaceana 13:256–271.

Cameron SL, Whiting MF. 2007. Mitochondrial genomic comparisons of the subterranean termites from the genus Reticulitermes (Insecta: Isoptera: Rhinotermitidae). Genome 50:188–202.

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

Chevreux E. 1909. Amphipodes (1ère série). Arch Zool Exp Gén 5:27–42.

Cook CE, Yue Q, Akam M. 2005. Mitochondrial genomes suggest that hexapods and crustaceans are mutually paraphyletic. Proc R Soc Lond B Biol Sci 272:1295–1304.

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

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.

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

Jaume D, Christenson K. 2001. Amphi-Atlantic distribution of the subterranean amphipod family Metacrangonyctidae (Gammaridea). Contrib Zool 70:99–125.

Karaman GS. 1989. *Metacrangonyx ortali* n sp, a new subterranean member of the family Crangonyctidae, from the Dead Sea region. Stud Mar 20:33–49.

Karaman GS, Pesce GL. 1979. Researches in Africa by the Zoological Institute of l'Aquila, Italy V: On three subterranean amphipods from North Africa (Amphipoda: Gammaridea). Bull Zool Mus Univ Amsterdam 7:197–207.

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

Lowe TM, Eddy SR. 1997. tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. Nucl Acids Res 25:955–964.

Machida RJ, Miya MU, Nishida M, Nishida S. 2002. Complete mitochondrial DNA sequence of *Tigriopus japonicus* (Crustacea: Copepoda). Mar Biotechnol 4:406–417.

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

Margalef R. 1952. Materiales para la hidrobiología de la isla de Menorca. Publ Inst Biol Apl 11:5–112.

Martin JW, Davis GE. 2001. An updated classification of the recent Crustacea. Los Angeles, CA: Natural History Museum of Los Angeles County. p 123.

Messouli M, Boutin C, Coineau N. 1991. Phylogénie et biogéographie évolutive d'un groupe de Metacrangonyctidae, Crustacés Amphipodes stygobies du Maroc. I. Les espèces du groupe *panousei*. Mém Biospéol 25:247–262.

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

Oulbaz Z, Coineau N, Boutin C. 1998. *Metacrangonyx knidirii* n sp et *Metacrangonyx remyi* Balazuc et Ruffo, 1953 (Crustacés Amphipodes Metacrangonyctidae) des eaux souterraines du Maroc. Mém Biospéol 25:35–43.

Perna NT, Kocher TD. 1995. Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes. J Mol Evol 41:353–358.

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

Ruffo S. 1954. *Metacrangonyx longicaudus* n sp (Amphipoda Gammaridae) delle acque sotterranee del Sahara Marocchino. Mem Mus Civ Stor Nat Verona 4:127–130.

Ruffo S. 1982. Una nouvelle espèce de *Metacrangonyx* Chevreux (Amphipoda: Gammaridae) du désert du Sinaï. Isr J Zool 31: 151–156.

Salvato P, Simonato M, Battisti A, Negrisolo E. 2008. The complete mitochondrial genome of the bag-shelter moth *Ochrogaster lunifer* (Lepidoptera, Notodontidae). BMC Genom 9:331.

Stamatakis A. 2006. RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22:2688–2690.

Stoch F. 1997. *Metacrangonyx ilvanus* n sp, the first Italian representative of the family Metacrangonyctidae (Crustacea: Amphipoda). Ann Limnol 33:255–262.

Stock JH, Rondé-Broekhuizen B. 1986. A new species of *Pygocrangonyx*, an amphipod genus with African affinities, from Fuerteventura. Bijdr Dierk 56:247–266.

Supek F, Vlahovicek K. 2004. INCA: Synonymous codon usage analysis and clustering by means of self-organizing map. Bioinformatics 14:2329–2330.

Tamura K, Nei M, Kumar S. 2004. Prospects for inferring very large phylogenies by using the neighbor-joining method. Proc Natl Acad Sci USA 101:11030–11035.

Tamura K, Dudley J, Nei M, Kumar S. 2007. MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0. Mol Biol Evol 24:1596–1599.

Wolstenholme DR. 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.

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

Xiong B, Kocher TD. 1991. Comparison of mitochondrial DNA sequences of seven morphospecies of black flies (Diptera: Simuliidae). Genome 34:306–311.

Yang JS, Yang WJ. 2008. The complete mitochondrial genome sequence of the hydrothermal vent galatheid crab *Shinkaia crosnieri* (Crustacea: Decapoda: Anomura): A novel arrangement and incomplete tRNA suite. BMC Genom 9:257.

Yokobori S, Paabo S. 1995. tRNA editing in metazoans. Nature 377:490.

Zhang DX, Szymura JM, Hewitt GM. 1995. Evolution and structural conservation of the control region of insect mitochondrial DNA. J Mol Evol 40:382–391.

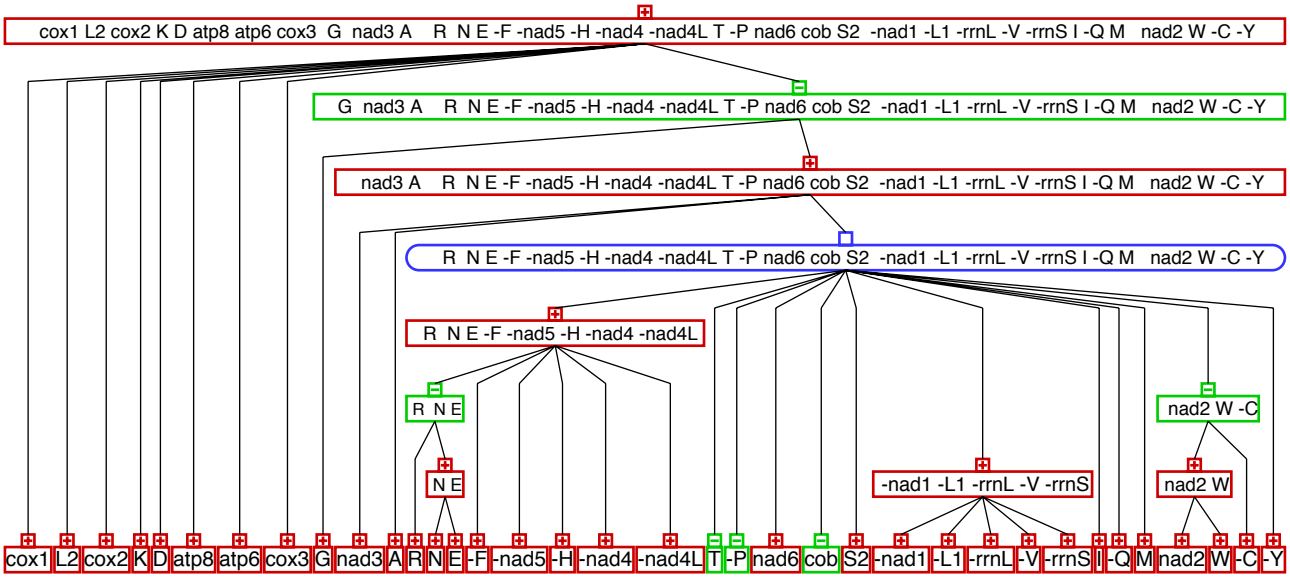
Additional Files

Additional File 1. Taxon names and EMBL accession numbers of the crustacean and hexapod mitogenomes used for gene annotation and gene divergence analyses.

Species	Accession no.	Taxonomy
CRUSTACEA		
Argulus americanus	NC_005935	Maxillopoda Branchiura
Armadillidium vulgare	EF643519	Malacostraca Peracarida Isopoda
Armillifer armillatus	NC_005934	Pentastomida
Artemia franciscana	NC_001620	Branchiopoda Anostraca
Callinectes sapidus	NC_006281	Malacostraca Eucarida Decapoda Pleocyemata Brachyura
Cherax destructor	NC_011243	Malacostraca Eucarida Decapoda Pleocyemata
Daphnia pulex	NC_000844	Branchiopoda Anomopoda
Eriocheir sinensis	NC_006992	MalacostracaEucarida Decapoda Pleocyemata Brachyura
Euphausia superba	AB084378	Malacostraca Eucarida Euphausiacea
Fenneropenaeus chinensis	NC_009679	Malacostraca Eucarida Decapoda Dendrobranchiata
Geothelphusa dehaani	NC_007379	Malacostraca Eucarida Decapoda Pleocyemata Brachyura
Gonodactylus chiragra	NC_007442	Malacostraca Hoplocarida Stomatopoda
Halocaridina rubra	NC_008413	Malacostraca Eucarida Decapoda Pleocyemata Caridea
Harpiosquilla harpax	NC_006916	Malacostraca Hoplocarida Stomatopoda
Hutchinsoniella macracantha	NC_005937	Cephalocarida
Idotea baltica	DQ442915	Malacostraca Peracarida Isopoda
Lepeophtheirus salmonis	NC_007215	Maxillopoda Copepoda
Ligia oceanica	NC_008412	Malacostraca Peracarida Isopoda Oniscidea
Litopenaeus vannamei	NC_009626	Malacostraca Eucarida Decapoda Dendrobranchiata
Lysiosquillina maculata	NC_007443	Malacostraca Hoplocarida Stomatopoda
Macrobrachium rosenbergii	NC_006880	Malacostraca Eucarida Decapoda Pleocyemata Caridea
Marsupenaeus japonicus	NC_007010	Malacostraca Eucarida Decapoda Dendrobranchiata
Megabalanus volcano	NC_006293	Maxillopoda Thecostraca Cirripedia Thoracica
Pagurus longicarpus	NC_003058	Malacostraca Eucarida Decapoda Anomura
Panulirus japonicus	NC_004251	Malacostraca Eucarida Decapoda Pleocyemata
Parhyale hawaiiensis	AY639937	Malacostraca Peracarida Amphipoda
Penaeus monodon	NC_002184	Malacostraca Eucarida Decapoda Dendrobranchiata
Pollicipes mitella	NC_008742	Maxillopoda Thecostraca Cirripedia Thoracica
Pollicipes polymerus	NC_005936	Maxillopoda Thecostraca Cirripedia Thoracica
Portunus trituberculatus	NC_005037	Malacostraca Eucarida Decapoda Brachyura
Pseudocarcinus gigas	NC_006891	Malacostraca Eucarida Decapoda Pleocyemata Brachyura
Pseudosquilla ciliata	AY947836	Malacostraca Hoplocarida Stomatopoda
Speleonectes tulumensis	NC_005938	Remipedia
Squilla empusa	NC_007444	Malacostraca Hoplocarida Stomatopoda
Squilla mantis	NC_006081	Malacostraca Hoplocarida Stomatopoda
Tetraclita japonica	NC_008974	Maxillopoda Thecostraca Cirripedia Thoracica
Tigriopus californicus	NC_008831	Maxillopoda Copepoda
Tigriopus japonicus	NC_003979	Maxillopoda Copepoda Harpacticoida
Triops cancriformis	NC_004465	Branchiopoda Notostraca
Triops longicaudatus	NC_006079	Branchiopoda Notostraca
Vargula hilgendorfii	NC_005306	Ostracoda Myodocopa
Metacrangonyx longipes	AM944817	Malacostraca Peracarida Amphipoda
HEXAPODA		
Aleurodicus dugesii	NC_005939	Insecta Hemiptera Hemimetabola
Anopheles gambiae	NC_002084	Insecta Diptera Holometabola
Antheraea pernyi	NC_004622	Insecta Lepidoptera Holometabola
Apis mellifera ligustica	NC_001566	Insecta Hymenoptera Holometabola
Bombyx mori	NC_002355	Insecta Lepidoptera Holometabola
Ceratitis capitata	NC_000857	Insecta Diptera Holometabola
Chrysomya putoria	NC_002697	Insecta Diptera Holometabola
Crioceris duodecimpunctata	NC_003372	Insecta Coleoptera Holometabola
Drosophila melanogaster	NC_001709	Insecta Diptera Holometabola
Gomphiocephalus hodgsoni	NC_005438	Collembola
Gryllotalpa orientalis	NC_006678	Insecta Orthoptera Hemimetabola
Haematobia irritans irritans	NC_007102	Insecta Diptera Holometabola
Heterodoxus macropus	NC_002651	Insecta Phthiraptera Hemimetabola
Homalodisca coagulata	NC_006899	Insecta Hemiptera Hemimetabola
Japyx solifugus	NC_007214	Diplura
lepidopsocid RS-2001	NC_004816	Insecta Psocoptera Hemimetabola
Locusta migratoria	NC_001712	Insecta Orthoptera Hemimetabola
Melipona bicolor	NC_004529	Insecta Hymenoptera Holometabola
Nesomachilis australica	NC_006895	Insecta Archaeognatha Ametabola

Additional File 1 – *continued*

Species	Accession no.	Taxonomy
Onychiurus orientalis	NC_006074	Collembola
Orthetrum triangulare melania	AB126005	Insecta Odonata Hemimetabola
Ostrinia nubilalis	NC_003367	Insecta Lepidoptera Holometabola
Pachypsylla venusta	NC_006157	Insecta Hemiptera Hemimetabola
Periplaneta fuliginosa	NC_006076	Insecta Dictyoptera Hemimetabola
Philaenus spumarius	NC_005944	Insecta Hemiptera Hemimetabola
Podura aquatica	NC_006075	Collembola
Pteronarcys princeps	NC_006133	Insecta Plecoptera Hemimetabola
Pyrocoelia rufa	NC_003970	Insecta Coleoptera Holometabola
Schizaphis graminum	NC_006158	Insecta Hemiptera Hemimetabola
Thermobia domestica	NC_006080	Insecta Thysanura Ametabola
Thrips imaginis	NC_004371	Insecta Thysanoptera Hemimetabola
Triatoma dimidiata	NC_002609	Insecta Hemiptera Hemimetabola
Tribolium castaneum	NC_003081	Insecta Coleoptera Holometabola
Tricholepidion gertschi	NC_005437	Insecta Thysanura Ametabola
Xenos vesparum	DQ364229	Strepsiptera Holometabola



Additional File 2. Rearrangement steps deduced using detection of strong interval trees to account for the gene order of *M. longipes* mitogenome compared with the ancestral pancrustacean order.

This paper was first published online on iFirst on *5 June 2009*