

Dramatic Mitochondrial Gene Rearrangements in the Hermit Crab *Pagurus longicarpus* (Crustacea, Anomura)

M. J. Hickerson and C. W. Cunningham

Department of Zoology, Duke University

The entire mitochondrial gene order of the crustacean *Pagurus longicarpus* was determined by sequencing all but approximately 300 bp of the mitochondrial genome. We report the first major gene rearrangements found in the clade including Crustacea and Insecta. At least eight mitochondrial gene rearrangements have dramatically altered the gene order of the hermit crab *P. longicarpus* relative to the putatively ancestral crustacean gene order. These include two rearrangements of protein-coding genes, the first reported for any nonchelicerate arthropod. Codon usage and amino acid sequences do not deviate substantially from those reported for other crustaceans. Investigating the phylogenetic distribution of these eight rearrangements will add additional characters to help resolve decapod phylogeny.

Introduction

The metazoan mitochondrial genome is a circular, double-stranded DNA molecule that is highly variable in DNA sequence but conservative in gene content (Wolstenholme 1992). Mitochondrial gene order has been found to vary across the Metazoa, generating interest in using mitochondrial DNA (mtDNA) gene order for phylogenetic inference (Boore, Lavrov, and Brown 1998). Although the apparent rarity of gene rearrangements provides few characters for phylogenetic reconstruction (Curole and Kocher 1999), useful characters will accumulate as more mitochondrial genomes are sequenced. Of the nine arthropods whose complete mitochondrial gene order has been determined, gene rearrangements have been restricted to tRNA genes, with the exception of protein-coding gene movements in two lineages of ticks. The two complete crustacean mitochondrial genomes sequenced to date have revealed no major deviation from the gene order found in *Drosophila yakuba*, aside from two tRNA gene movements in *Artemia franciscana*.

The phylogenetic position of crustaceans within the arthropods is not agreed upon, and the monophyletic status of the Crustacea has even been called into question (Ballard et al. 1992; Friedrich and Tautz 1995; Regier and Shultz 1997). Here, we describe the mitochondrial gene order of the hermit crab *Pagurus longicarpus* and report the first major gene rearrangements in any nonchelicerate arthropod. Comparisons with partial crustacean genomes suggest that the rearrangements in the hermit crab *P. longicarpus* appear to have arisen within the Decapoda. Because they are restricted to the Decapoda, these rearrangements will not help resolve the issue of crustacean monophyly. On the other hand, exploring their phylogenetic distribution may provide additional characters to help discern relationships among decapod groups.

Key words: crustacean mitochondrial DNA, gene order, gene rearrangements, hermit crab, *Pagurus longicarpus*.

Address for correspondence and reprints: Cliff Cunningham, Department of Zoology, Duke University, Durham, North Carolina 27708-0325. E-mail: cliff@duke.edu.

Mol. Biol. Evol. 17(4):639–644. 2000

© 2000 by the Society for Molecular Biology and Evolution. ISSN: 0737-4038

Materials and Methods

DNA was extracted from a live *P. longicarpus* individual obtained from Beaufort, N.C., using the Genome kit (BIO 101) according to the manufacturer's instructions. With the exception of approximately 300 bp of an 850-bp AT-rich region, 15,630 bp of the mtDNA molecule was sequenced directly from a combination of overlapping 0.5–2.0-kb PCR products and a long PCR 7.5-kb product cloned into TOPO XL vectors (Invitrogen). The overlapping 0.5–2.0-kb fragments were PCR-amplified in 50 mM Tris HCl (pH 9.0), 20 mM ammonium sulfate, 0.005% BSA, and 2.5 mM MgCl₂, with 200 μM of each dNTP, 0.5 mM of each primer, and 1.25 U Amplitaq Gold polymerase (Perkin Elmer) in a total volume of 50 μl.

The 7.5-kb fragment was initially amplified using the Boehringer Mannheim Expand Long Template PCR system with conditions specified for system 3. PCR products were inserted into the pCR[®]XL plasmid vector using the Invitrogen TOPO XL cloning kit. Plasmid DNA was purified from each clone using the QIAGEN plasmid miniprep kit. The 7.5-kb fragment was sequenced with M13 and T7 reverse primers and then subsequently with newly designed internal primers. As a check against PCR error, three different clones were sequenced, and a two-out-of-three rule was used to choose between base ambiguities. The 0.5–2.0-kb fragments were cycle-sequenced with the PCR amplification primers and internal primers when necessary. All cycle-sequencing was performed with ABI PRISM BigDye terminator chemistry and analyzed on ABI 373 or 377 automated sequencers (ABI/Perkin-Elmer).

Mitochondrial protein and RNA gene sequences were initially identified using BLAST searches. Confirmation of gene sequences was then done by alignment with *Drosophila yakuba* mitochondrial sequences using CLUSTAL W, version 1.4. The amino acid sequences of the protein-coding genes were inferred from the *Drosophila* genetic code. A preliminary screen for tRNA genes was done using tRNAscan-SE 1.1 (Lowe and Eddy 1997). The default search mode was used (cove cutoff score = 15), specifying mitochondrial/chloroplast DNA as the source and using the invertebrate mitochondrial genetic code for tRNA structure prediction.

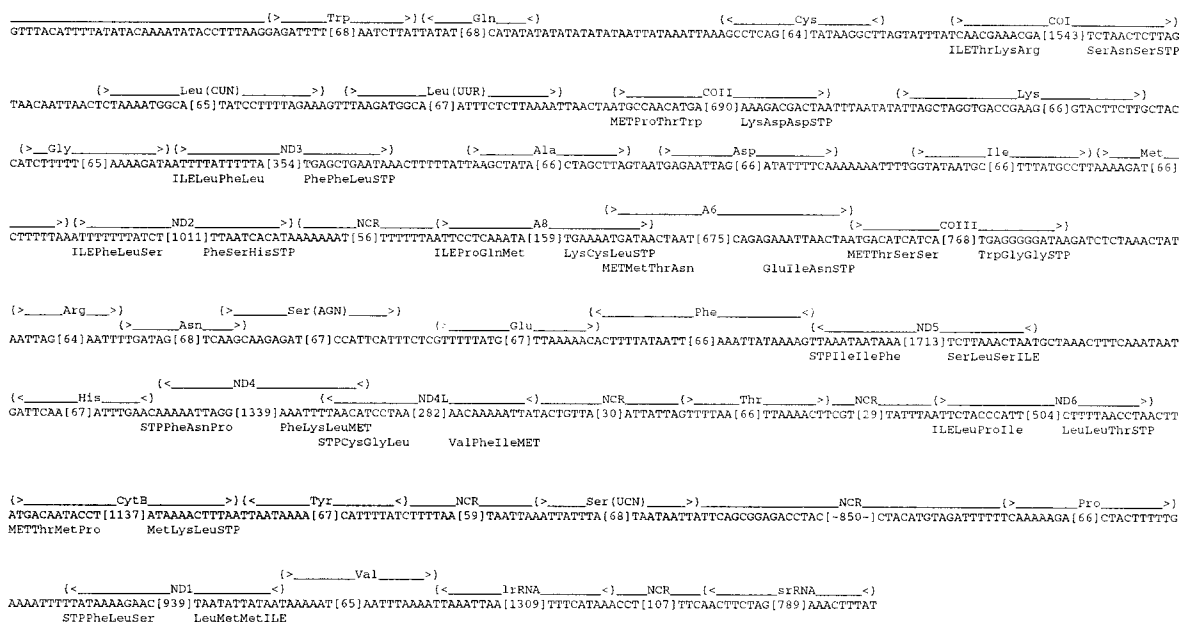


FIG. 1.—A schematic map of the of the mtDNA sequence of *Pagurus longicarpus*. Numbers within brackets are the numbers of nucleotides of each gene or noncoding region. With the exception of approximately 300 bp of the noncoding region, the entire circular mtDNA was sequenced and is represented here as a linear molecule. The direction in which each gene is transcribed is indicated by the direction of darts within the braces that delineate the extent of each gene. The inferred amino acids at the amino- and carboxy-termini of protein genes are given below the corresponding DNA sequence. Noncoding regions are labeled NCR if they are longer than 28 bp.

tRNA secondary structures were detected in this manner for 20 out of the 22 tRNA's, the two remaining tRNA's were detected by inspecting noncoding sequences for tRNA-like secondary structures by eye.

Contiguous sequences were obtained for the entire mitochondrial genome with the exception of approximately 300 bp within the AT-rich noncoding region. This sequence is deposited in GenBank under accession number AF150756.

Results and Discussion

The Mitochondrial Genome of the Hermit Crab *P. longicarpus*

Aside from approximately 300 bp within the AT-rich region that proved difficult to sequence, we obtained the entire mitochondrial genome of *P. longicarpus*. Our sequences include both rRNA genes, all 22 tRNA genes, and all 13 protein-coding genes commonly found in the mitochondria of metazoan animals (Wolstenholme 1992). All orientations and gene sequence junctions are shown in figure 1. Each of the 22 putative tRNA sequences can be folded into cloverleaf structures similar to those found in other metazoan animals (fig. 2). The weakest inferred structures are tRNA^{Ser(AGN)} and tRNA^{Val}, with six base mispairings each (fig. 2). Although technical difficulties (most likely secondary structure) prevented sequencing of the entire AT-rich region (estimated to be approximately 850 bp in size), we obtained the sequence of its flanking regions (38 bp 5' and 500 bp 3'). Before discussing the gene rearrangements, we will discuss characteristics of the mitochondrial genome.

Base Composition and Gene Content

For the gene sequences encoding the 13 protein genes, the nucleotide composition is as follows: A = 3,239 (29.1%), T = 4,509 (40.5%), G = 1,716 (15.4%), and C = 1,669(15.0%). The 30.40% GC content falls within the range found in arthropod mtDNA genomes so far. The lower end of this range is represented by the insect *Apis mellifera* (14.7%; Crozier and Crozier 1993) and the chelicerates *Ixodes hexagonus* (27.4%) and *Rhipicephalus sanguineus* (22.0%; Black and Roehrdanz 1998), while the higher end is represented by the crustaceans *Artemia franciscana* (35.5%; Valverde et al. 1994) and *Daphnia pulex* (37.7%; Crease 1999).

Translation Initiation and Termination Signals

All termini of the 13 protein-coding genes are shown in figure 1 and summarized in table 1. As with the *Drosophila* genetic code, ATT, ATA, and ATG initiate translation in 12 out of 13 of the protein-coding genes (table 1). For the initiation of COI in *P. longicarpus*, we propose a novel 4-bp initiation codon (ATCA) based on the locations of COI initiation in *D. yakuba* and *D. pulex* (Clary and Wolstenholme 1985; Crease 1999) and the fact that there are no other initiation codons near the beginning of COI in *P. longicarpus*. A 4-bp initiation codon (ATTA) has also been proposed for the COI of *Daphnia pulex* (Van Raay and Crease 1994). For 11 of the 13 protein-coding genes, the 3' sequence regions revealed a termination codon of TAA, while COI used TAG. For ND4, we could find no complete termination codon. Following Crease (1999), we inferred an incomplete termination codon beginning with the base T. This association

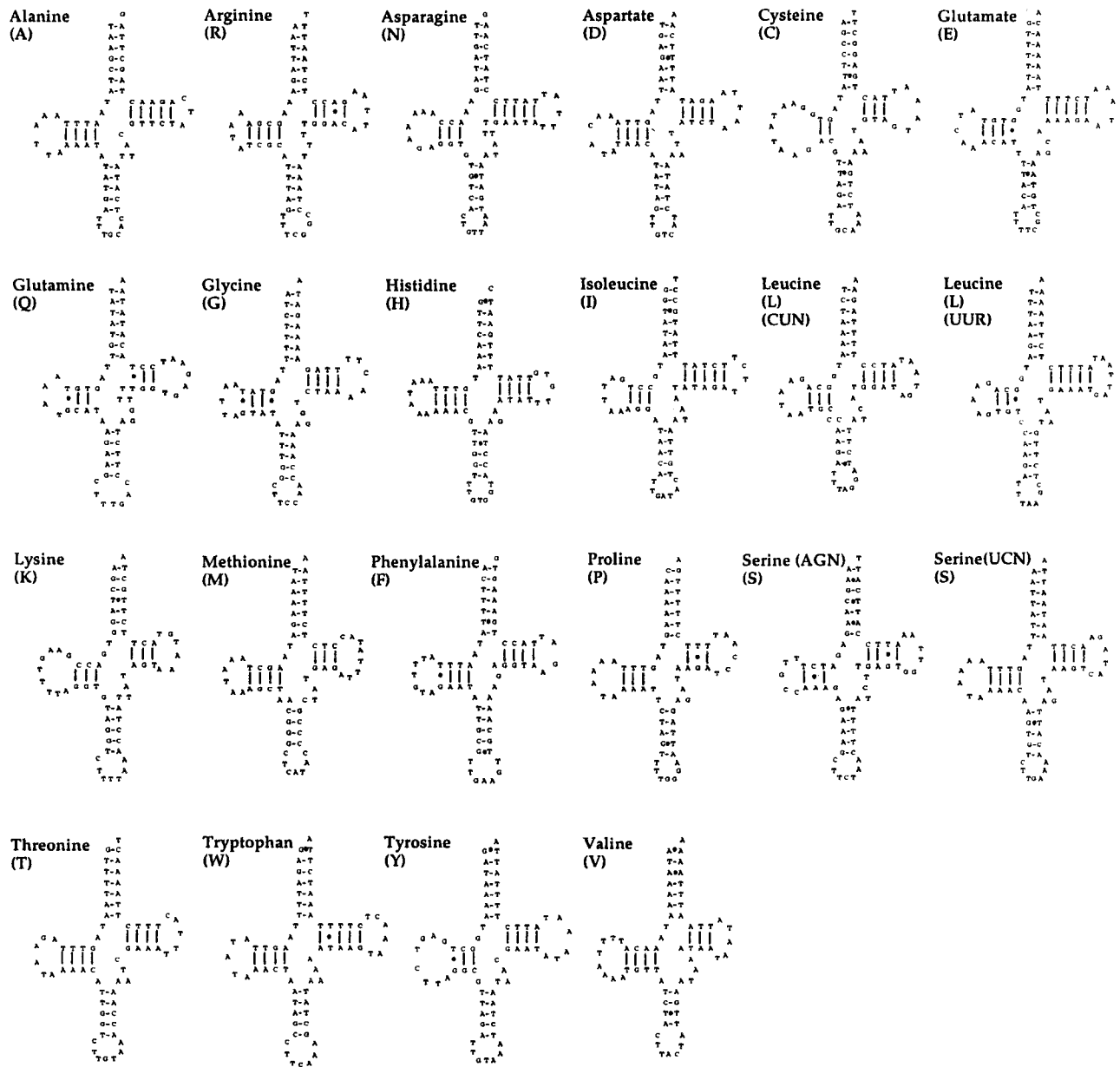


FIG. 2.—Putative cloverleaf structures for the mitochondrial tRNA's from *Pagurus longicarpus*. Lines denote standard Watson-Crick pairs; dots denote suboptimal base pairings.

between incomplete termination codons and tRNA genes is common in metazoan mtDNA (Wolstenholme 1992). One possible explanation is that complete termination codons are formed by polyadenylation of transcripts, as has been observed for some human mtDNA genes (Ojala, Montoya, and Attardi 1981).

As is common in many mtDNA genomes, a number of the genes overlap (Wolstenholme 1992). Of the eight overlapping genes reported here, two are encoded on the opposite strand (fig. 1). The A8-A6 overlap falls into the pattern consistent with both proteins being translated from a single bicistronic message by initiation at a 5'-terminal start site for A8 and at an internal start site for A6 (Wolstenholme 1992). The terminal end of the ND4L transcript overlaps with the initiation end of

ND4 by 7 bp, although they are on different reading frames. The other five overlaps were found to be no longer than 3 bp. All 13 protein-coding genes for *P. longicarpus* were close in size to those for *Drosophila* and *Daphnia*, although the gene sizes of *Artemia* differed markedly from those of these three taxa in five cases. Each of nine protein-coding genes in *Pagurus* was closer in amino acid sequence to *Drosophila* and *Daphnia* than to *Artemia* (table 1), consistent with the large amount of divergence found in branchiopod crustaceans (e.g., Crease 1999).

tRNAs and Codon Usage

Twenty tRNAs have anticodons identical to those of *D. yakuba* and *D. pulex*, while two differ: tRNA^{Lys}

Table 1
Comparisons of the Mitochondrial Protein-Coding Genes of the Hermit Crab (*Pagurus longicarpus*) with Those of the Brine Shrimp (*Artemia franciscana*) and the Fruit Fly *Drosophila yakuba*

Protein	NO. OF AMINO ACIDS PER PROTEIN				% AMINO ACID IDENTITY WITH <i>P. LONGICARPUS</i>			PREDICTED INITIATION AND TERMINATION CODONS (<i>P. LONGICARPUS</i>)	
	<i>P. longicarpus</i>	<i>A. franciscana</i>	<i>D. yakuba</i>	<i>Daphnia pulex</i>	<i>A. franciscana</i>	<i>D. yakuba</i>	<i>D. pulex</i>		
A6	224	219	224	224	70.3	82.2	75.9	ATG	TAA
A8	52	52	53	53	19.2	36.5	Not significant	ATT	TAA
COI	513	512	512	512	81.5	81.3	81.3	(A)TCA	TAG
COII	229	228	227	226	80.5	88.5	85.3	ATG	TAA
COIII	263	257	262	262	70.2	76.6	74.2	ATG	TAA
Cytb	378	381	378	377	78.9	83.2	79.0	ATG	TAA
ND1	310	298	324	311	63.6	71.1	70.0	ATT	TAA
ND2	336	298	341	329	52.3	63.6	59.8	ATT	TAA
ND3	117	111	117	117	70.2	69.4	60.0	ATT	TAA
ND4	446	386	446	440	61.8	72.6	58.9	ATG	T--
ND4L	93	85	96	91	69.2	77.8	70.1	ATA	TAA
ND5	570	541	572	569	53.4	67.9	59.3	ATT	TAA
ND6	167	219	174	17	Not significant	55.1	67.6	ATA	TAA

NOTE.—Gene lengths are as inferred in the text and depicted in figure 1 or obtained from GenBank. Actual gene length could be slightly different due to ambiguity in determining start and stop codons. Percentage of identity is the number of identical inferred amino acids in a pairwise alignment divided by the mean length of the two compared sequences using the BLAST search engine.

(TTT instead of CTT) and tRNA^{Ser(AGN)} (TCT instead of GCT). This usage of TTT for tRNA^{Lys} is consistent with most metazoan mtDNA systems. As with tRNA^{Lys}, the difference for Serine^{Ser(AGN)} corresponds to the third base in the wobble position. The presence of these two anticodons in *P. longicarpus* could be due to codon usage. According to the codon usage inferred for the 13 *P. longicarpus* amino acid sequences, tRNA^{Lys} is coded by AAA (rather than AAG) in 85.0% of the codons, while tRNA^{Ser} is coded by AGA 52.0% of the time. This codon usage makes the anticodon of tRNA^{Lys} more efficient than the anticodon found in *D. yakuba* (table 2).

Fourteen anticodons have T in the wobble position, seven have G, and tRNA^{Met} has C. All nine tRNAs with anticodons that recognize fourfold-degenerate codons have T in the wobble position. Of the 11 anticodons that recognize twofold-degenerate codons, six have G and five have T in the first position. Comparisons of codon usage (table 2) indicate that for 6 of the 11 twofold-degenerate codons, the tRNAs match the less used co-

don. All six of these cases involve G/T mispairings, which are the most allowable type of base mispairings (Wolstenholme 1992).

Phylogenetic Utility of Mitochondrial Gene Rearrangements in the Decapoda

To demonstrate a mitochondrial gene rearrangement, we must know the original gene order. Inferring the primitive gene order in crustaceans is relatively straightforward, since several insects (e.g., *D. yakuba*) have a gene order identical to that of one of the two previously reported crustacean genomes (*D. pulex*; Crease 1999). The other previously reported crustacean gene order (the brine shrimp *A. franciscana*; Valverde et al. 1994) shows two tRNA rearrangements that are unique relative to all other arthropods. This suggests that the *A. franciscana* tRNA rearrangements are likely autapomorphies with respect to the taxa sampled to date.

Table 2
Number of Occurrences of the 3,711 Codons in the 13 Protein-Encoding Genes of *Pagurus longicarpus* mtDNA

Codon (aa)	<i>N</i>	Codon (aa)	<i>N</i>	Codon (aa)	<i>N</i>	Codon (aa)	<i>N</i>
TTT (Phe)	244	TCT (Ser)	141	TAT (Tyr)	127	TGT (Cys)	45
TTC (Phe)	87	TCC (Ser)	24	TAC (Tyr)	24	TGC (Cys)	11
TTA (Leu)	293	TCA (Ser)	76	TAA (Stop)	11	TGA (Trp)	89
TTG (Leu)	47	TCG (Ser)	9	TAG (Stop)	1	TGG (Trp)	10
CTT (Leu)	97	CCT (Pro)	59	CAT (His)	50	CGT (Arg)	23
CTC (Leu)	16	CCC (Pro)	20	CAC (His)	30	CGC (Arg)	4
CTA (Leu)	94	CCA (Pro)	61	CAA (Gln)	53	CGA (Arg)	29
CTG (Leu)	17	CCG (Pro)	4	CAG (Gln)	7	CGG (Arg)	5
ATT (Ile)	268	ACT (Thr)	77	AAT (Asn)	116	AGT (Ser)	42
ATC (Ile)	35	ACC (Thr)	20	AAC (Asn)	34	AGC (Ser)	6
ATA (Met)	167	ACA (Thr)	76	AAA (Lys)	85	AGA (Ser)	78
ATG (Met)	49	ACG (Thr)	3	AAG (Lys)	15	AGG (Ser)	24
GTT (Val)	99	GCT (Ala)	94	GAT (Asp)	51	GGT (Gly)	97
GTC (Val)	6	GCC (Ala)	25	GAC (Asp)	17	GGC (Gly)	10
GTA (Val)	103	GCA (Ala)	73	GAA (Glu)	51	GGA (Gly)	95
GTG (Val)	19	GCG (Ala)	8	GAG (Glu)	30	GGG (Gly)	26

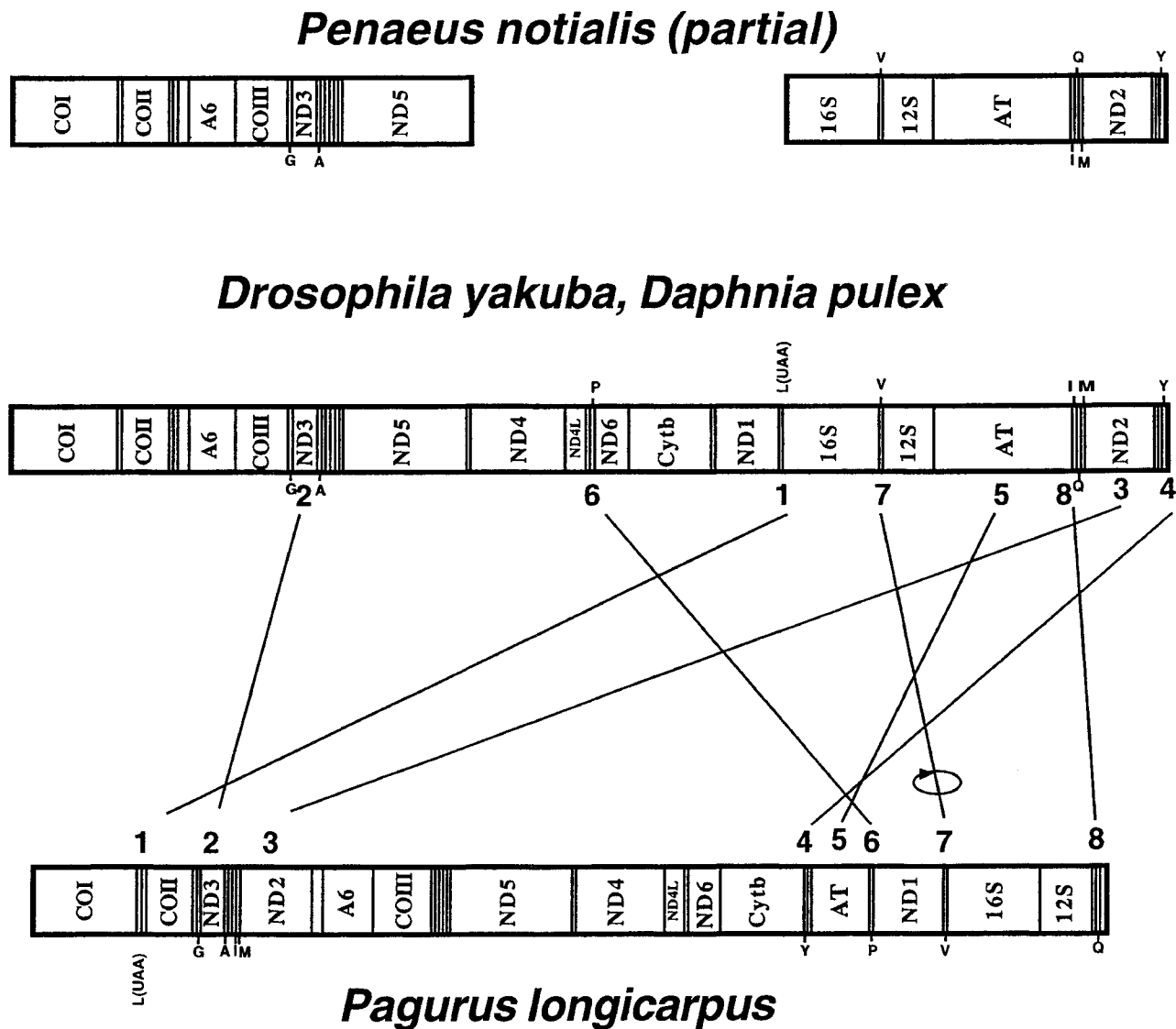


FIG. 3.—Gene arrangements in several arthropods. For the sections shown, the gene order of *Penaeus notialis* is identical to the *Drosophila yakuba/Daphnia pulex* gene order. Rearrangements between the putatively primitive *D. yakuba/D. pulex* gene order and the hermit crab *Pagurus longicarpus* are shown for eight contiguous blocks of genes: (1) tRNA^{Leu(UAA)} (L); (2) tRNA^{Gly} (G) ND3 tRNA^{Ala} (A); (3) tRNA^{Ile} (I) tRNA^{Met} (M) ND2; (4) tRNA^{Tyr} (Y); (5) the AT-rich region; (6) tRNA^{Pro} (P); (7) tRNA^{Val} (V); and (8) tRNA^{Gln} (Q). Only tRNAs that took part in rearrangements are shown.

In contrast, we found a total of eight rearrangements in the decapod hermit crab *P. longicarpus* relative to the putatively primitive *D. pulex/D. yakuba* gene order (fig. 3). Two of these include the first reported rearrangements of arthropod protein-coding genes outside of the Chelicerata (Black and Roehrdanz 1998; Campbell and Barker 1998, 1999). These two rearrangements involve NADH2 and NADH3 (regions 2 and 3 in fig. 3). Only one rearrangement in *P. longicarpus* involved an inversion of the direction of transcription (tRNA^{Val}, fig. 3).

All eight of these rearrangements appear to have arisen within the Decapoda. Not only are they absent in the crustaceans *D. pulex* and *A. franciscana*, but the partial sequence of the decapod shrimp *Penaeus notialis* (Garcia-Machado et al. 1996) is sufficient to show that it possesses none of the eight rearrangements (fig. 3).

The presence of so many rearrangements increases the number of characters available to resolve decapod phylogeny. As with any character data, identically derived gene orders can arise through homoplasy. Homoplasy in mitochondrial gene order has been found in a number of taxa (Flook, Rowell, and Gellissen 1995; Mindell, Sorenson, and Dimcheff 1998; Downton and Austin 1999). The strongest inferences will be made when mitochondrial gene rearrangements are congruent with other sources of phylogenetic data (Cuore and Kocher 1999).

With eight rearrangements, it may be possible to demonstrate the intermediate gene orders that led to the arrangement found in *P. longicarpus*. Using this information in conjunction with other molecular data will help to evaluate the probabilities of homology and convergence in gene order. An investigation of the phylo-

genetic distribution of these rearrangements, together with a multigene sequencing effort, is currently in progress.

Acknowledgments

We must thank K. Tieu, who first discovered that there is something unusual about the hermit crab mitochondrial genome. We also thank C. Morrison for help in the cloning procedure and D. Snyder for running the ABI 373 and 377 automated sequencers (ABI/Perkin-Elmer). This work was supported by NSF grant DEB-9615461 to C.W.C.

LITERATURE CITED

- BALLARD, J. W. O., G. J. OLSEN, D. P. FAITH, W. A. ODGERS, D. M. ROWELL, and P. W. ATKINSON. 1992. Evidence from 12S ribosomal-RNA sequences that onychophorans are modified arthropods. *Science* **258**:1345–1348.
- BLACK, W. C. I., 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.
- BOORE, J. L., D. V. LAVROV, and W. M. BROWN. 1998. Gene translocation links insects and crustaceans. *Nature* **392**:667–668.
- CAMPBELL, N. J. H., and S. C. BARKER. 1998. An unprecedented major rearrangement in an arthropod mitochondrial genome. *Mol. Biol. Evol.* **15**:1786–1787.
- . 1999. The novel mitochondrial gene arrangement of the cattle tick, *Boophilus microplus*: fivefold tandem repetition of a coding region. *Mol. Biol. Evol.* **16**:732–740.
- CLARY, D. O., and D. R. WOLSTENHOLME. 1985. The mitochondrial DNA molecule of *Drosophila yakuba*: nucleotide sequence, gene organization and genetic code. *J. Mol. Evol.* **22**:252–271.
- CREASE, T. J. 1999. The complete sequence of the mitochondrial genome of *Daphnia pulex* (Cladocera: Crustacea). *Gene* **233**:89–99.
- 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.
- CURIOLE, A. P., and T. D. KOCHER. 1999. Mitogenomics: digging deeper with complete mitochondrial genomes. *Trends Ecol. Evol.* **14**:394–398.
- DOWTIN, M., and A. D. AUSTIN. 1999. Evolutionary dynamics of a mitochondrial rearrangement “hot spot” in the Hymenoptera. *Mol. Biol. Evol.* **16**:298–309.
- FLOOK, P., H. ROWELL, and G. GELLISSEN. 1995. Homoplastic rearrangements of insect mitochondrial transfer-RNA genes. *Naturwissenschaften* **82**:336–337.
- FRIEDRICH, M., and D. TAUTZ. 1995. Ribosomal DNA phylogeny of the major extant arthropod classes and the evolution of myriapods. *Nature* **376**:165–167.
- GARCIA-MACHADO, E., N. DENNEBOUY, M. O. SUAREZ, J. C. MOUNOLOU, and M. MONNEROT. 1996. Partial sequence of the shrimp *Penaeus notialis* mitochondrial genome. *C. R. Acad. Sci. D Sci. Nat.* **319**:473–486.
- LOWE, T., 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.
- MINDELL, D. P., M. D. SORENSON, and D. E. DIMCHEFF. 1998. Multiple independent origins of mitochondrial gene order in birds. *Proc. Natl. Acad. Sci. USA* **95**:10693–10697.
- OJALA, D., J. MONTOYA, and G. ATTARDI. 1981. tRNA punctuation model of RNA processing in human mitochondria. *Nature* **290**:470–474.
- REGIER, J. C., and J. W. SHULTZ. 1997. Molecular phylogeny of arthropods and the significance of the Cambrian “explosion” for molecular systematics. *Am. Zool.* **38**:918–928.
- 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.
- VAN RAAY, T. J., and T. J. CREASE. 1994. Partial mitochondrial DNA sequence of the crustacean *Daphnia pulex*. *Curr. Genet.* **25**:66–72.
- WOLSTENHOLME, D. R. 1992. *Animal mitochondrial DNA: structure and evolution*. Academic Press, New York.

ROSS CROZIER, reviewing editor

Accepted December 21, 1999