J Mol Evol (1999) 49:142–149

JOURNAL OF MOLECULAR EVOLUTION



# Mitochondrial Genes Collectively Suggest the Paraphyly of Crustacea with Respect to Insecta

**Erik García-Machado,[1] Malgorzata Pempera,[2,*] Nicole Dennebouy,[2] Mario Oliva-Suarez,[1] Jean-Claude Mounolou,[2] Monique Monnerot[2]**

[1] Centro de Investigaciones Marinas, Universidad de La Habana, Calle 1ra No. 2808, esquina 30, Miramar, C. Habana, Cuba
[2] Centre de Génétique Moléculaire, CNRS, 91198 Gif-sur-Yvette Cedex, France

Received: 25 August 1998 / Accepted: 8 March 1999

**Abstract.** Complete sequences of seven protein coding genes from *Penaeus notialis* mitochondrial DNA were compared in base composition and codon usage with homologous genes from *Artemia franciscana* and four insects. The crustacean genes are significantly less A + T-rich than their counterpart in insects and the pattern of codon usage (ratio of G + C-rich versus A + T-rich codon) is less biased. A phylogenetic analysis using amino acid sequences of the seven corresponding polypeptides supports a sister-taxon status for mollusks–annelid and arthropods. Furthermore, a distance matrix-based tree and two most-parsimonious trees both suggest that crustaceans are paraphyletic with respect to insects. This is also supported by the inclusion of *Panulirus argus* COII (complete) and COI and COIII (partial) sequence data. From analysis of single and combined genes to infer phylogenies, it is observed that obtained from single genes are not well supported in most topologies cases and notably differ from that of the tree based on all seven genes.

**Key words:** mtDNA — Protein genes — *Panulirus argus* — *Penaeus notialis* — Protostomes — Crustacea — Phylogeny

## Introduction

Crustacea, Chelicerata, Insecta, and Myriapoda are the major taxonomic components of the animal phylum Arthropoda. The long evolutionary history and overwhelming morphological diversity of arthropods have prompted a continuous debate about their phylogenetic relationships for over a century (Snodgrass 1938; Weygoldt 1986; Barnes 1987; Willmer 1990). Historically, many and varying evolutionary scenarios have been suggested but none has been formally established (Briggs and Fortey 1989; Hessler and Newman 1975; Manton 1977; Snodgrass 1938). The question is now being reinvestigated with molecular phylogenies. However, results from various studies are not congruent (Wägele and Stanjek 1995; Turbeville et al. 1991), and in some cases (Ballard et al. 1992; Friedrich and Tautz 1995; Regier and Shultz 1997) there is conflict with morphological considerations (Eernisse et al. 1992; Willmer 1990). This lack of consistency between morphological and molecular data is probably due to the complexity of biological diversity and/or some flaws of classical approaches. This issue should be clarified. A search for appropriate markers was necessary to test all current hypotheses.

For approximately the last 15 years mitochondrial DNA (mtDNA) has been used by population and evolutionary biologists as a valuable source of data. At present the mtDNA from almost 70 animal species is totally sequenced; less than one-third of them correspond to invertebrate taxa. Complete mtDNA sequences are known for six species of insects—*Drosophila yakuba*

* *Present address:* Polish Academy of Science, Marine Biology Center, ul. SW Wojciech S, 81-347, Gdynia, Poland
*Correspondence to:* Pr. Monique Monnerot; *e-mail:* monnerot@cgm.cnrs-gif.fr

(Clary and Wolstenholme 1985), *Drosophila melanogaster* [composite sequence (Lewis et al. 1996)], *Anopheles gambiae* (Beard et al. 1993), *Anopheles quadrimaculatus* (Mitchell et al. 1993), *Apis mellifera* (Crozier and Crozier 1993), and *Locusta migratoria* (Flook et al. 1995)—and only one species of Crustacea, *Artemia franciscana* (Valverde et al. 1994). Preliminary sequence results on a restricted domain of *Penaeus notialis* mtDNA (partial sequences of protein coding genes and the small subunit of the ribosomal RNA) and comparisons with homologous regions in genomes of other animal taxa led us to suspect a paraphyletic status of crustaceans with respect to insects (García-Machado et al. 1996).

New data from *Penaeus notialis* and *Panulirus argus* mtDNA provided an opportunity to reinvestigate the evolutionary organization of arthropods. We focused on three major questions: (a) Are arthropods (insects and crustaceans in this study) monophyletic with respect to other protostomes (mollusks and annelid) when several mitochondrial protein coding genes are used for phylogenetic inference? (b) Are these markers able to identify the Eutrochozoa assemblage (Annelida + Mollusca)? and (c) Do more sequence information support paraphyly of crustaceans?

## Materials and Methods

### *Species and Sequences*

The sequence of a segment of the mtDNA of the malocostracan crustacean *Penaeus notialis* [partially analyzed in a previous work (García-Machado et al. 1996)] was achieved (4553 additional bp). The sequence thus available (9276 bp) represents more than the half of the mitochondrial genome of this species. Nearly all of the mtDNA of the spiny lobster *Panulirus argus* was cloned, and part of it sequenced (3570 bp). Strategies followed for sequence determination are presented by García-Machado et al. (1996) and García-Machado (1997). The sequences reported above are available at EMBL/GenBank under accession numbers X84350 and AJ133049–54.

The phylogenetic analysis involved the sequences of seven genes (COI, COII, COIII, ATP6, ATP8, ND2, and ND3) from the following species (the accession numbers of the sequences retrieved from EMBL or GenBank are given in parentheses): the insects *Drosophila yakuba* (X03240; Clary and Wolstenholme 1985), *Anopheles gambiae* (L20934; Beard et al. 1993), *Anopheles quadrimaculatus* (L04272; Mitchell et al. 1993), and *Locusta migratoria* (X80245; Flook et al. 1995); the crustacean (branchiopod) *Artemia franciscana* (X69067; Valverde et al. 1994); the mollusk (polyplacophoran) *Katharina tunicata* (U09810; Boore and Brown 1994); the gastropods *Albinaria coerulea* (X83390; Hatzoglou et al. 1995) and *Cepaea nemoralis* (U23045; Terrett et al. 1996); the oligochaete annelid *Lumbricus terrestris* (U24570; Boore and Brown 1995); and three deuterostome taxa as a composite outgroup—the echinoderm *Strongylocentrotus purpuratus* (X12631; Jacobs et al. 1988), the cyclostome *Petromyzon marinus* (U11880; Lee and Kocher 1995), and the mammal *Mus musculus* (J01420; Bibb et al. 1982).

### *Alignment Strategy*

The inferred amino acid sequences [Translate Program from the Wisconsin Package, version 9.1; Genetics Computer Group (GCG), Madison, WI] of the seven genes were aligned pairwise using the CLUSTAL V multiple sequence aligner (Higgins et al. 1992). The resulting alignments were refined by eye using the Aligner Sequence Editor (Eernisse 1995). Considering the long divergence time of the groups analyzed and the gene size variations for some of the taxa, we took special care in maximizing similarities between sequences of the most closely related taxa (monophyletic assemblages), e.g., *Cepaea nemoralis*–*Albinaria coerulea*, gastropods–*Katharina tunicata*, and so on. The alignments of sequences for the seven genes in their entirety were then concatenated to generate a full data set. ATP6 and ATP8, of which a few amino acids overlap, were treated as independent genes.

### *Data Analysis*

Distances between taxa were estimated using gamma probabilities for amino acid substitutions implemented in the MEGA package (Kumar et al. 1993). A gamma parameter equal to 1 was used in all cases and calculations were made with the pairwise deletion option in MEGA. Phylogenies were inferred using the neighbor-joining method (Saitou and Nei 1987) and maximum-parsimony analyses were carried out using the PAUP package, version 3.1.1 (Swofford 1993). In all cases the most-parsimonious trees were constructed with the heuristic search option; the sequences were added at random and branch swapping was carried out using a tree bisection–reconnection procedure. The characters were considered as equally weighted and the alignment gaps were treated as missing data. When multiple (two or three) most-parsimonious trees were obtained we constructed a strict consensus tree. The confidence of nodes in the trees was evaluated by bootstrap (Felsenstein 1985). To estimate differences between neighbor-joining or maximum-parsimony trees inferred from single genes or various gene combinations and those obtained from the full data set, topological distances ($d_T$) were calculated as by Russo et al. (1996).

## Results and Discussion

Sequencing of a 8051-bp *Bgl*II cloned fragment (previously estimated as 7.9 kb by restriction analysis) of the mitochondrial genome of *Penaeus notialis* was completed (see García-Machado et al. 1996). mtDNA from *Panulirus argus* was almost totally cloned, and part of it (3570 bp) sequenced. For subsequent phylogenetic analysis the whole sequence of seven mitochondrial protein coding genes (COI, COII, COIII, ATP6, ATP8, ND2, and ND3) from *Penaeus notialis* and the sequences of the COI and COIII genes (partial) and of the COII gene (complete) from *Panulirus argus* were thus available.

### *Composition and Pattern of Codon Usage in Crustaceans and Insects*

The nucleotide composition was calculated for the seven protein genes from *Penaeus notialis* mtDNA and the homologous genes from *Artemia franciscana* and the four insects, *Drosophila yakuba*, *Anopheles gambiae*, *Anopheles quadrimaculatus*, and *Locusta migratoria* (Table 1). The A + T percentages for *Penaeus notialis* and *Artemia franciscana* protein genes are almost-identical (64%) but far lower than in insects (73.5%). *Panulirus argus* genes (partial COI and COIII and com-

**Table 1.** Percentages of A + T in the studied mitochondrial protein coding genes and ratios of G + C-rich to A + T-rich codons

| Species | Composition, A + T (%) | Ratio of G + C-rich/A + T-rich codons[a] |
|---|---|---|
| *Penaeus notialis* | 64.3 | 0.66 |
| *Panulirus argus* | 58.1 | 0.77 |
| *Artemia franciscana* | 64.1 | 0.53 |
| *Drosophila yakuba* | 74.7 | 0.43 |
| *Anopheles gambiae* | 74.1 | 0.42 |
| *Anopheles quadrimaculatus* | 73.8 | 0.42 |
| *Locusta migratoria* | 71.3 | 0.38 |

[a] G + C-rich codons—Pro, Ala, Arg, and Gly; A + T-rich codons—Phe, Ile, Met, Tyr, Lys, and Asn (Crozier and Crozier 1993).

plete COII) also appear to be less A + T-rich (58.1%) than their *Penaeus notialis* counterparts. It should be noted that the values calculated from seven genes of *Penaeus notialis* represent a good estimator of the composition of the whole set of protein genes of this species (see also García-Machado et al. 1996).

Several studies have demonstrated that the base composition of mtDNA is highly correlated with the use of codons and the evolutionary rates of mitochondrial genes (Crozier and Crozier 1993; Crozier et al. 1989; Jermiin and Crozier, 1994). Insect mitochondrial protein genes exhibit a preference for using A + T-rich codons (Table 1), in contrast to *Artemia franciscana,* where G + C-rich codons are relatively more frequent (Flook et al. 1995; Valverde et al. 1994). This situation was examined in the *Penaeus notialis* and *Panulirus argus* protein genes by calculating the ratio of G + C-rich codons (Pro, Ala, Arg, Gly) to A + T-rich ones (Phe, Ile, Met, Tyr, Lys, Asn) as suggested by Crozier and Crozier (1993) (Table 1). The values obtained for the three crustaceans range from 0.53 in *Artemia franciscana* to 0.77 in *Panulirus argus* and they are remarkably higher than those obtained for insects (from 0.38 in *Locusta migratoria* to 0.43 in *Drosophila yakuba*).

We also determined the relative synonymous codon usage (RSCU) according to Sharp et al. (1986) to compare the overall usage in the six arthropods (*Panulirus argus* genes not included). Considering an equal usage of synonymous codons (a premise of this analysis), we determined the ratio of the variances of the respective uses between pairs of species and the values obtained were compared to tabulate *F* values for significance under the hypothesis of equal variances ($H_o$: $S1^2 = S2^2$). Both insects and crustaceans appear to be homogeneous groups in this respect. However, as expected from simple eye inspection, in all cases differences between insects and crustaceans are highly significant ($p < 0.001$), which is well correlated with the general use of some A + T-rich codons in insect codon families.

Base composition differences are currently considered to be an effect of mutational directional pressures (Sueoka 1962) that obscure the phylogenetic inferences and yield erratic branching orders when phylogenies are estimated (Pesole et al. 1995). Considering the compositional differences of the nucleotide sequences among the mtDNA of the taxa studied in the present work, as well as their ancient relationships, we decided to use inferred amino acid sequences for a phylogenetic study. It was expected that amino acid sequences should be more conservative due to functional constraints and preserved historical phylogenetic information.

### *Phylogenetic Analysis*

*Building Trees.* Alignments of the amino acid sequences of the seven genes from 13 of the taxa studied yielded a total of 1862 positions. The tree inferred from this full data set, using the neighbor-joining method, is presented in Fig. 1a. Maximum-parsimony analysis produces two most-parsimonious trees (length = 5851, CI = 0.775, RI = 0.521). Both tree-building procedures generate almost-identical topologies which strongly support the three major nodes: one node (■) grouping insects and crustaceans (mandibulate arthropods), a second node (●) including annelid and mollusks, and a third node (○) which relates the ingroup (protostomes) with the deuterostomes. The bootstrap percentage values for these three nodes differ in relation to the method used but are significantly always high. The differences between the results obtained by the two methods are restricted to the branching order within the clade composed of the annelid and mollusks.

The monophyly of Arthropoda (crustaceans and insects) and that of the Annelida–Mollusca assemblage (Eutrochozoa) are well supported (■ and ●, respectively in Fig. 1a). These results are in full agreement with previous cladistic studies based on morphology (Brusca and Brusca 1990; Eernisse et al. 1992; Weygoldt 1986) and molecular or molecular/morphological data (Ballard et al. 1992; Kim et al. 1996; Turbeville et al. 1991; Wägele and Stanjek 1995; Wheeler et al. 1993).

The monophyly of Mollusca could be questioned: the neighbor-joining tree proposes a *Katharina–Lumbricus* association when *Katharina* is placed as basal in one most-parsimonious tree and closely related to the gastropods in the other. In all cases bootstrap values are not significant and the strict consensus tree of the two most-parsimonious trees results in an unresolved polytomy. Characters weighted with the rescaled consistency index produce a single most-parsimonious tree which supports the monophyly of Mollusca.

The paraphyly of crustaceans with respect to insects is well supported. Within crustaceans, the malacostracan decapod *Penaeus notialis* is closely allied to the insects, while the branchiopod *Artemia franciscana* appears to be an early-diverging taxon with respect to the assemblage insects–crustaceans. In a preliminary survey we had investigated the phylogenetic relationships among *Penaeus*

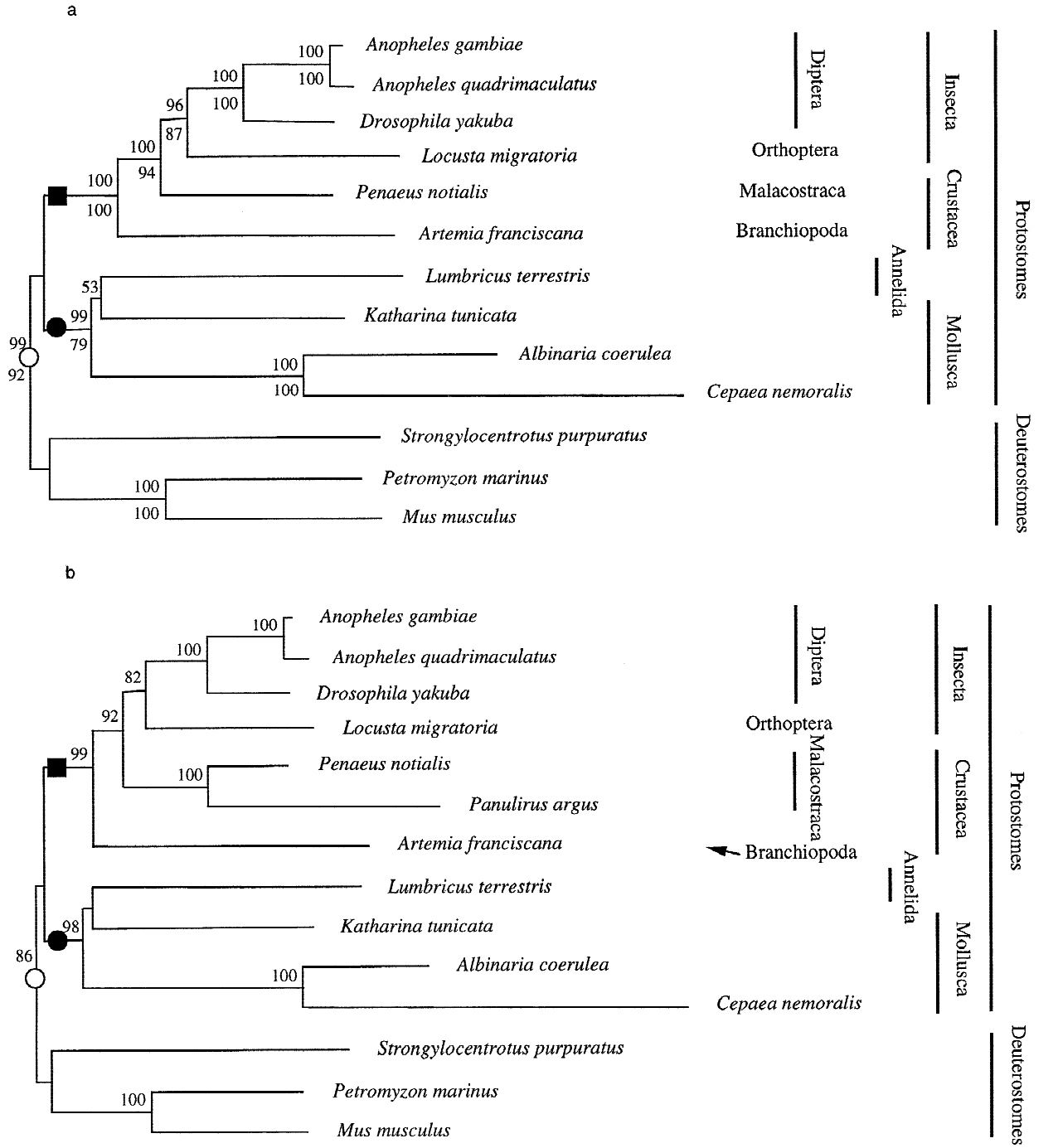


**Fig. 1.** Phylogenetic trees inferred from the sequences analyzed. **a** Neighbor-joining tree obtained with the seven mitochondrial protein coding genes, using gamma distances with a gamma parameter equal to 1. A strict consensus tree obtained from two most-parsimonious trees is almost-identical to the tree shown. The differences are restricted to the branching pattern on the Annelida–Mollusca clade. Bootstrap support of the different nodes is depicted *above* and *below internal branches* for distance (10000 replicates) and parsimony (100 replicates), respectively. **b** Neighbor-joining tree obtained including *Panulirus argus* using the partial amino acid sequences of COI and COIII and the total sequence of COII. Gamma distances were calculated with a gamma parameter equal to 1. Bootstrap support of the different nodes is given *on internal branches* (1000 replicates). (■) Node for insects and crustaceans; (●) node for annelids and mollusks; (○) branch joining deuterostomes and protostomes.

*notialis*, a few other arthropods, and a mollusk by using partial sequences of the COI, COII, and COIII genes and the small subunit of the mitochondrial rRNA (García-Machado et al. 1993, 1996). The results suggested a paraphyly of crustaceans with respect to insects, however, in any case, nodes were well supported by bootstrap. The availability of more sequences (1737 amino acids from seven genes all together) and the consideration of a new species in the analysis clearly allow us to confirm these observations.

**Table 2.** Stability of some particular nodes on trees based on different gene groupings and phylogenetic approaches

| Particular node | Gene grouping[a] | | | | | | | | All seven genes |
|---|---|---|---|---|---|---|---|---|---|
| | COI–II–III | COI–II–III–ATP6 | COI–II–III–ATP8 | COI–II–III–ND2 | COI–II–III–ND3 | COI–II–III–ATP6–ND2 | COI–II–III–ATP6–8–ND2 | COI–II–III–ATP6–ND2–3 | ATP6–ND2 | |
| (Diptera, *Locusta*)[b] | 74/– | 88/+ | 68/– | 96/+ | 79/– | 97/+ | 96/+ | 98/+ | 99/+ | 95/+ |
| (*Penaeus*, insects)[b] | 93/– | 100/+ | 95/– | 97/+ | 100/– | 100/+ | 100/+ | 100/+ | 94/+ | 100/+ |
| (*Artemia*, (*Penaeus* + insects))[b] | 100/– | 100/+ | 100/– | 100/+ | 100/– | 100/+ | 100/+ | 100/+ | 99/+ | 100/+ |
| (Annelids, mollusks)[b] | 98/+ | 100/+ | 100/+ | 84/– | –/– | 100/+ | 100/+ | 99/+ | 99/+ | 100/+ |
| (*Katharina*, gastropods)[b] | –/+ | –/– | –/+ | 88/– | –/– | 64/+ | 52/– | 50/– | 83/+ | –/– |
| (Protostomes, deuterostomes)[b] | 89/+ | 97/+ | 82/+ | 100/+ | –/– | 100/+ | 100/+ | 99/+ | 100/+ | 100/+ |
| Total number of amino acid positions | 1034 | 1281 | 1105 | 1407 | 1171 | 1654 | 1725 | 1791 | 620 | 1862 |
| $d_T$[c] | 0/6 | 0/0 | 0/6 | 2/4 | 8/9 | 2/2 | 2/3 | 2/1 | 2/1 | n.a.[d]/1 |

[a] COI–II–III, ATP6–8, and ND2–3 stand for cytochrome *c* oxidase subunits I, II, and III, ATPase subunits 6 and 8, and NADH dehydrogenase subunits 2 and 3 genes, respectively.

[b] Numbers on the left-hand side of the shill correspond to bootstrap values for the corresponding node on neighbor joining-based trees (as percentages from 500 replicates). A – on the left-hand side of the shill corresponds to the absence of the corresponding node on neighbor joining-based trees. A – or + on the right-hand side of the shill stands, respectively, for the presence or absence of the corresponding node on maximum parsimony-based trees.

[c] Topological distances are calculated by comparing the tree obtained with each gene grouping to that obtained from the seven genes taken altogether (full data set). Values given on the left- or right-hand side of the shill correspond to neighbor joining- or maximum parsimony-based trees, respectively.

[d] Not applicable.

*Method Liability.* The effect of the inclusion of another malacostracan was tested using sequences of *Panulirus argus* mtDNA (partial COI and COIII and total COII). A neighbor-joining examination produces a tree with an arrangement identical to that of the tree derived from the previous data set, in which *Panulirus* and *Penaeus* branch together with 100% bootstrap confidence (Fig. 1b). The tree (length = 2411, CI = 0.725, and RI = 0.515) obtained by maximum parsimony also reveals a dipteran–crustacean clade at the exclusion of the orthopteran (data not shown). However, the bootstrap analysis offers support only for the nodes of the dipterans and malacostracans, respectively. It appears that the number of sites or genes used in this analysis is not appropriate to perform accurate phylogeny reconstructions by the two methods. In spite of this, the very major arrangements are conserved, and the coincidence of the distance tree with the full data set trees suggests that close relationships between malacostracan crustaceans and insects may not be an artifact.

*Paraphyly of Crustacea with Respect to Insecta.* One of the major results of this work is the confirmation of the paraphyly of Crustacea with respect to Insecta suggested in previous analyses (García-Machado et al. 1996). Indeed, the relative position of crustaceans and insects has long been a matter of discussion (Hessler and Newman 1975; Manton 1977; Weygoldt 1986). However, recent evidence based on a conserved arrangement of major mitochondrial genes and the presence of unique shared gene boundaries (Boore et al. 1995, 1998; García-Machado et al. 1996; García-Machado, 1997), shared patterns of axogenesis (Thomas et al. 1984; Whitington et al. 1993), appendage development and similar expression of the genes involved in this process (Panganiban et al. 1995), or ommatidia structure (Paulus 1979) all suggests that these two taxa are more closely related to each other than to other arthropod taxa. This may have been overlooked previously, in most of the studies in which the molecular phylogeny of arthropods was investigated, because of the choice of crustacean representatives. Moreover, insect origin is not clear at the present time, and a potential Remipedia-like ancestor is still considered by some authors (see details given by Brusca and Brusca 1990). A recent study of 17 arthropod species using the amino acid sequences of the nuclear gene encoding the elongation factor 1-α provided support for a clade Hexapoda/Branchiopoda as well as evidence for polyphyly of Crustacea (Regier and Shultz 1997). These results, together with those from the present work, are in good agreement with the previous suggestion of paraphyly of crustaceans with respect to Insecta by Friedrich and Tautz (1995).

Relationships inside the crustacean groups are not very well understood either, which may be not surprising in a group that had probably arisen by the early Cambrian (Briggs and Fortey 1989; Hessler 1982; Schramp

**Table 3.** Stability of some particular nodes on trees based on different genes and phylogeny approaches

| | Gene[a] | | | | | | |
|---|---|---|---|---|---|---|---|
| Particular node | COI | COII | COIII | ATP6 | ATP8 | ND2 | ND3 |
| (Diptera, *Locusta*)[b] | 57/− | 59/− | 62/− | 71/− | −/− | 99/+ | 33/− |
| (*Penaeus*, insects)[b] | −/− | 97/− | 48/− | 99/− | −/− | 48/+ | 80/− |
| (*Artemia* (*Penaeus* + insects))[b] | −/− | 37/− | 91/− | 95/+ | −/− | −/+ | 71/− |
| (Annelids, mollusks)[b] | 94/+ | 56/− | −/− | 100/+ | −/− | −/− | −/− |
| (*Katharina*, gastropods)[b] | −/− | −/− | −/− | −/− | −/− | −/− | −/− |
| (Protostomes, deuterostomes)[b] | 76/+ | 50/+ | 39/− | 92/+ | −/− | 99/+ | −/− |
| Total number of amino acid positions | 530 | 232 | 272 | 247 | 71 | 373 | 137 |
| $d_T$ | 2/4 | 2/4 | 2/8 | 0/3 | 10/12 | 8/4 | 8/9 |

[a] COI, COII, and COIII, ATP6 and ATP8, and ND2 and ND3 stand for cytochrome oxidase subunits I, II, and III, ATPase subunits 6 and 8, and NADH dehydrogenase subunits 2 and 3 genes, respectively.

[b] Numbers on the left-hand side of the shill correspond to bootstrap values for the corresponding node on neighbor joining-based trees (as percentages from 500 replicates. A − on the left-hand side of the shill corresponds to the absence of the corresponding node on neighbor joining-based trees. A − or + on the right-hand side of the shill stands, respectively, for the presence or absence of the corresponding node on maximum parsimony-based trees.

1982, 1986). Molecular studies using sequences of the 18S rRNA and the mitochondrial l-rRNA (16S) gene have provided clear information only on taxa which diverged recently, i.e., more recently than the subphylum taken as a whole such as Decapoda suborders (Cunningham et al. 1992; Kim and Abele 1989). The position and status of Branchiopoda, for instance, are still a matter of debate (Brusca and Brusca 1990). Fryer (1992) suggests that this group, in particular, the Anostraca, of which *Artemia* is a member, represents the most primitive group of the extant crustaceans considering various morphological and ontogenetic characteristics. In contrast, malacostracans appear in most of studies as the most derived group of crustaceans [see Bowman and Abele (1982) for details on the members and classification of Crustacea]. Sequence comparison of the COI and COII genes including the maxillopodan cirriped *Lepas anatifera* (unpublished sequence established by one of us) also placed *Artemia* as an early-diverging taxon and the cirripede in an intermediate position with respect to malacostracans and branchiopods (data not shown).

### *Phylogenies from Single or Combined Gene Sequences*

In a further analysis of individual genes and different gene combinations, we considered the topology of the neighbor-joining tree obtained with the full data set as a reference pattern, without considering whether or not the taxon relationships observed are correct. Table 2 presents the results and the three main nodes: one grouping insects and crustaceans (mandibulate arthropods), one grouping annelid and mollusks, and a third joining Deuterostomes and Protostomes (■, ●, and ○, respectively, in Fig. 1). They are supported in all cases except in the combinations COI–II–III, COI–II–III–ATP8, COI–II–III–ND2, and COI–II–III–ND3 as indicated by the topological distances and bootstrap values.

All differences between trees are related to the placement of *Katharina tunicata* and to the eventual monophyly or paraphyly of Mollusca as well as to the relative arrangement of *Artemia*, *Penaeus*, and insects. However, support is weak in all topologies alternative to the clustering evidenced above: *Artemia*, (*Penaeus* + insects).

Otherwise, the paraphyletic condition of Crustacea is again revealed with the strongest confidence in all combinations using the distance approach and in most cases using maximum parsimony.

When genes are analyzed one by one (Table 3), only ATP6 and COII generate a topology identical to that of the full data set when using distance approach. Small genes (ATP8 and ND3) are particularly prone to generate different patterns (high $d_T$ values) and ND2 also performs poorly, with $d_T = 8$ and $d_T = 4$ for distance and maximum parsimony, respectively. The three cytochrome oxidase subunit sequences (COI, COII, and COIII) give similar trees, and curiously, the one obtained with the COI sequence differs from the full data set tree in proposing monophyly of Crustacea, however, this arrangement is not strongly supported. As seen in other studies these results indicate that some single-gene comparisons are not appropriated for phylogenetic inferences when distant relationships are investigated (Cao et al. 1994; Liu and Beckenbach 1992; Russo et al. 1996).

### *Conclusions*

Notwithstanding the evidence we have presented here, the results must be considered cautiously. The number of taxa analyzed in this study represents no more than a tiny portion of the enormous diversity of the arthropods, annelids, and mollusks: this may be a major factor of misunderstanding in phylogenetic inferences. In the case of the crustaceans studied we have paid some attention to the effect of the base composition bias, which may obscure phylogenetic information, and have shown that *Penaeus* and *Artemia* do not differ in base composition

and/or in their relative use of codons. Again, this does not exclude the possibility of differences in evolutionary rates, which, in fact, are suggested by inspection of the distance data matrix, when, for example, *Lumbricus* and *Katharina* appear to be much more similar to each other than gastropods are among themselves.

Molecular phylogenies of metazoans have been questioned in the last few years because of frequent conflicts with morphological considerations. Most of the work presented to date is based on one kind of information, mainly the 18S rRNA gene sequence, which has proved to have limitations when some distant relationships are investigated (Aboudheif et al. 1998). The revelation of new sequences opens up new possibilities. The congruence of the results obtained with various combinations of mitochondrial genes suggests that some, but not all, genes contain information useful for inferring phylogenies. It is expected that those genes will vary from study to study, however, the use of amino acid-deduced sequences from genes such as COI, COII, COIII, ATP6, and ND2 appears to be a good additional approach to assess protostomian phylogeny from mitochondrial protein coding genes.

*Acknowledgments.* The authors thank the two anonymous referees for their helpful comments on the manuscript.

## References

Abouheif E, Zardoya R, Meyer A (1998) Limitations of Metazoan 18S rRNA sequence data: Implications for reconstructing a phylogeny of the animal kingdom and inferring the reality of the Cambrian explosion. J Mol Evol 47:394–405

Ballard JW, Olsen GJ, Faith DP, Odgers WA, Rowell DM, Atkinson PW (1992) Evidence from 12S ribosomal RNA sequences that onychophorans are modified arthropods. Science 253:1345–1348

Barnes RD (1987) Invertebrate zoology. 5th ed. W.B. Saunders, Philadelphia

Beard CB, Hamm DM, Collins FH (1993) The mitochondrial genome of the mosquito *Anopheles gambiae:* DNA sequence, genome organization, and comparisons with mitochondrial sequences of other insects. Insect Mol Biol 2(2):103–124

Bibb MJ, Van Etten RA, Wright CT, Walberg MW, Clayton DA (1981) Sequence and gene organization of mouse mitochondrial DNA. Cell 26:167–180

Boore JL, Brown WM (1994) Complete DNA sequence of the mitochondrial genome of the black chiton, *Katharina tunicata.* Genetics 138:423–443

Boore JL, Brown WM (1995) Complete sequence of the mitochondrial DNA of the annelid worm, *Lumbricus terrestris.* Genetics 141:305–319

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 DV, Brown WM (1998) Gene translocation links insects and crustaceans. Nature 392:667–668

Bowman TE, Abele LG (1982) Classification of the recent Crustacea. In: Abele LG (ed) The biology of the Crustacea, Vol 1. Systematics, the fossil record, and biogeography. Academic Press, New York, pp 1–27

Briggs DEG, Fortey RA (1989) The early radiation and relationships of the major arthropods group. Science 246:1670–1673

Brusca RC, Brusca GL (1990) Invertebrates. Sinauer, Sunderland, MA

Cao Y, Adachi J, Janke A, Pääbo, Hasegawa M (1994) Phylogenetic relationships among Eutherian orders estimated from inferred sequences of mitochondrial proteins: Instability of tree based on a single gene. J Mol Evol 39:519–527

Clary DO, Wolstenholme DR (1985) The mitochondrial DNA molecule of *Drosophila yakuba:* Nucleotide sequence, gene organization and genetic code. J Mol Evol 22:252–271

Crozier RH, Crozier YC (1993) The mitochondrial genome of the honeybee *Apis mellifera:* Complete sequence and genome organization. Genetics 133:97–117

Crozier RH, Crozier YC, MacKinlay AG (1989) The COI and COII region of honeybee mitochondrial DNA: Evidence for variation in insect mitochondrial evolutionary rates. Mol Biol Evol 6(4):399–411

Cunningham CW, Blackstone NW, Buss LW (1992) Evolution of king crabs from hermit crab ancestors. Nature 355:539–542

Eernisse DJ (1995) DNA translator and aligner: HyperCard utilities and phylogenetic analysis of molecules. CABIOS 8:177–184

Eernisse DJ, Albert JS, Anderson FE (1992) Annelida and arthropoda are not sister taxa: A phylogenetic analysis of spiralian metazoan morphology. Syst Biol 41(3):305–330

Felsenstein J (1985) Confidence limits on phylogenies: An approach using bootstrap. Evolution 39:783–791

Flook PK, Rowell CHF, Gellinssen G (1995) The sequence organization and evolution of the *Locusta migratoria* mitochondrial genome. J Mol Evol 41:928–941

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

Fryer G (1992) The origin of the Crustacea. Acta Zool 73(5):273–286

García-Machado E (1997) Génome mitochondrial de la crevette *Penaeus notialis.* Réflexion sur la phylogénie des Arthropodes. PhD thesis, Université Paris-Sud, Orsay

García-Machado E, Dennebouy N, Oliva-Suárez M, Mounolou J-C, Monnerot M (1993) Mitochondrial 16S rRNA gene of two species shrimps Sequence variability and secundary structure. Crustaceana 65(3):279–286

García-Machado E, Dennebouy N, Oliva-Suárez M, Mounolou J-C, Monnerot M (1996) Partial sequence of the shrimp *Penaeus notialis* mitochondrial genome. CR Acad Sci Paris 319:473–486

Hatzoglou E, Rodakis GC, Lecanidou R (1995) Complete sequence and gene organization of the mitochondrial genome of the land snail *Albinaria coerulea.* Genetics 140:1353–1366

Hessler RR (1982) Evolution within the Crustacea. In: Abele LG (ed) The biology of the Crustacea, Vol 1. Systematics, the fossil record, and biogeography. Academic Press, New York, pp 150–185

Hessler RR, Newman WA (1975) A trilobitomorph origin for the Crustacea. Fossils Strata 4:437–459

Higgins DG, Bleasby AJ, Fuchs R (1992) CLUSTALV: Improved software for multiple sequence alignment. CABIOS 8:189–191

Jacobs HT, Elliott DJ, Math VB, Farquharson A (1988) Nucleotide sequence and gene organization of sea urchin mitochondrial DNA. J Mol Biol 202:185–217

Jermiin LS, Crozier RH (1994) The cytochrome b region in the mitochondrial DNA of the ant *Tetraponera rufoniger:* Sequence control. J Mol Evol 38:282–294

Kim CB, Moon SY, Gelder SR, Kim W (1996) Phylogenetic relationships of annelids, molluscs, and arthropods evidenced from molecules and morphology. J Mol Evol 43:207–215

Kim W, Abele LG (1989) Molecular phylogeny of selected decapod crustaceans based on 18S rRNA nucleotide sequences. J Crustacean Biol 10(1):1–3

Kumar S, Tamura K, Nei M (1993) MEGA version 1 01. Pennsylvania State University, University Park

Lee WJ, Kocher TD (1995) Complete sequence of a sea lamprey (*Pet-*

*romyzon marinus*) mitochondrial genome: Early establishment of the vertebrate genome organization. Genetics 139:873–887

Lewis DL, Farr CL, Kaguni LS (1995) *Drosophila melanogaster* mitochondrial DNA: Completion of the nucleotide sequence and evolutionary comparisons. Insect Mol Biol 4:263–278

Liu H, Beckenbach AT (1992) Evolution of the mitochondrial cytochrome oxidase II gene among 10 orders of insects. Mol Phyl Evol 1(1):41–52

Manton SM (1977) The Arthropoda: Habits, functional morphology, and evolution. Clarendon Press, Oxford

Mitchell SE, Cockburn AF, Seawright JA (1993) The mitochondrial genome of *Anopheles quadrimaculatus* species A: Complete nucleotide sequence and gene organization. Genome 36:1058–1073

Panganiban G, Sebring A, Nagy L, Carroll S (1995) The development of crustacean limbs and the evolution of arthropods. Science 270: 1363–1366

Paulus HF (1979) Eye structure, the monophyly of Arthropoda. In: Gupta AP (ed) Arthropod phylogeny. Van Nostrand Reinhold, New York, pp 299–377

Pesole G, Dellisanti G, Preparata G, Saccone C (1995) The importance of base composition in the correct assessment of genetic distances. J Mol Evol 41:1124–1127

Regier CJ, Schultz JW (1997) Molecular phylogeny of the major arthropod groups indicates polyphyly of crustaceans, new hypothesis for the origin of Hexapods. Mol Biol Evol 14(9):902–913

Russo CAM, Takesaki N, Nei M (1996) Efficiencies of different genes, different tree-building methods in recovering a known vertebrate phylogeny. Mol Biol Evol 13(3):525–536

Saitou N, Nei M (1987) The neighbor-joining method: A new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425

Schramp FR (1982) The fossil record, the evolution of Crustacea In: Abele LG (ed) The biology of the Crustacea, Vol 1. Systematics, the fossil record, and biogeography. Academic Press, New York, pp 93–147

Schramp FR (1986) Crustacea. Oxford University Press, London

Sharp PM, Tuohy TMF, Mosurski, KR (1986) Codon usage in yeast: Cluster analysis differentiates highly and lowly expressed genes. Nucleic Acids Res 14:5125–5143

Snodgrass RE (1938) Evolution of Annelida, Onychophora and Arthropoda. Smithson Misc Coll 97:1–159

Sueoka N (1962) On the genetic basis of variation and heterogeneity of DNA base composition. Proc Natl Acad Sci USA 48:582–592

Swofford DL (1993) PAUP 3 1. Illinois Natural History Survey, Champaign

Terrett JA, Milles S, Thomas RH (1996) Complete DNA sequence of the mitochondrial genome of *Cepaea nemoralis* (Gastropoda: Pulmonata). J Mol Evol 42:160–168

Thomas JB, Bastiani MJ, Bate M, Goodman CS (1984) From grasshopper to *Drosophila:* A common plan for neuronal development. Nature 310:203–207

Turbeville JM, Pfeifer DM, Field KG, Raff RA (1991) The phylogenetic status of arthropods, as inferred from rRNA sequences. Mol Biol Evol 8(5):669–686

Valverde JR, Batuecas B, Moratilla C, Marco R, Garesse R (1994) The complete mitochondrial DNA sequence of the crustacean *Artemia fransciscana.* J Mol Evol 4:400–408

Wägele JW, Stanjek G (1995) Arthropod phylogeny inferred from partial 12SrRNA revisited: Monophyly of the Tracheata depends on sequence alignment. J Zool Syst Evol Res 33:75–80

Weygoldt P (1986) Arthropod interrelationships—The phylogenetic-systematic approach. Z Zool Syst Evol 24:19–35

Wheeler WC, Cartwright P, Hayashi C (1993) Arthropod phylogeny: A combined approach. Cladistic 9:1–39

Whitington PM, Leach D, Sandeman R (1993) Evolutionary change in neuronal development within the arthropods: Axogenesis in the embryos of two crustaceans. Development 118:449–461

Willmer P (1990) Invertebrate relationships patterns in animal evolution. University Press, Cambridge