* Manuscript

The First Complete Mitochondrial Genome Sequences For Stomatopod Crustaceans: Implications for Phylogeny

Kirsten Swinstrom[1,2], Roy Caldwell[1], H. Matthew Fourcade[2] and Jeffrey L. Boore[1,2]

[1] Department of Integrative Biology, University of California Berkeley, Berkeley, CA

[2] Evolutionary Genomics Department, DOE Joint Genome Institute and Lawrence Berkeley National Lab, Walnut Creek, CA

For correspondence: Jeffrey Boore, DoE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, phone: 925-296-5691, fax: 925-296-5620, JLBoore@LBL.gov

Abstract

We report the first complete mitochondrial genome sequences of stomatopods and compare their features to each other and to those of other crustaceans. Phylogenetic analyses of the concatenated mitochondrial protein-coding sequences were used to explore relationships within the Stomatopoda, within the malacostracan crustaceans, and among crustaceans and insects. Although these analyses support the monophyly of both Malacostraca and, within it, Stomatopoda, it also confirms the view of a paraphyletic Crustacea, with Malacostraca being more closely related to insects than to the branchiopod crustaceans.

Key words: Stomatopod; mitochondrial genome; Crustacea; Arthropod phylogeny; mitochondrial DNA; *Gonodactylus chiragra*; *Lysiosquillina maculata*; *Squilla empusa*

Introduction

Mitochondrial DNA (mtDNA) sequences have been used extensively in phylogenetic analyses to examine relationships among populations or higher taxa. Most of these studies are limited because they use only one or a few genes. More recently however, many complete mitochondrial genomes have been sequenced (Boore, 1999). In particular, a number of phylogenetic analyses using gene order or protein-coding sequences from complete mitochondrial genomes have been conducted to examine relationships within the phylum Arthropoda (e.g. Boore et al., 1998; Garcia-Machado et al., 1999; Wilson et al., 2000; Yamauchi et al., 2002; Nardi et al., 2003). Among the results of these studies is the finding that the Crustacea are a paraphyletic group, with malacostracan crustaceans more closely related to insects than to the branchiopod crustaceans (Garcia-Machado et al., 1999; Wilson et al., 2000; Yamauchi et al., 2002). However, due to the paucity of complete mitochondrial genomes from malacostracans, relationships within this group have not been fully explored using such an expanded data set. In this study we report the first complete mitochondrial genome sequences for malacostracan crustaceans in the subclass Hoplocarida.

Currently, Malacostraca is divided into three subclasses: the Eumalocostraca (the most diverse of the three, including most shrimp, lobsters, and crabs), the Hoplocarida (stomatopods), and the Phyllocarida (leptostracans). The Eumalacostraca and Hoplocarida are traditionally considered sister taxa while the Phyllocarida is thought to be the most basal of the three (Schram, 1986). Thus far, complete mitochondrial genomes have been sequenced only from members of the Eumalacostraca.

Members of the order Stomatopoda are the only living representatives of the Hoplocarida (Schram, 1986) and are abundant, obligate carnivores in shallow, tropical and subtropical seas.

Stomatopods are commonly known as mantis shrimp, a name that refers to their powerful raptorial appendages that are kept folded tightly against their body when not in use. Like their namesake, the preying mantis, mantis shrimp can rapidly extend these powerful appendages like a jack-knife to either spear or smash their prey. At a velocity of over 1,000 centimeters per second in some species, this movement is among the fastest known in the animal kingdom (Caldwell and Dingle, 1976). In addition to their remarkable predatory abilities, stomatopods also have a highly complex visual system (Cronin et al., 1994) as well as exceptionally complex behavior (Caldwell and Dingle, 1975; Reaka and Manning, 1981) and diverse reproductive strategies (Caldwell, 1991). Comparative studies of the evolution of these traits require a phylogeny; however, the few phylogenetic studies of the Stomatopoda conducted to date have yielded somewhat conflicting results with respect to the relationships among the higher order taxa (Ahyong, 1997; Hof, 1998; Ahyong and Harling, 2000). While the two most recent of these analyses have indicated that the Gonodactyloidea are the most basal of the extant stomatopods, a previous analysis indicated that the Squilloidea was the most basal. Thus far, there has been no phylogenetic analysis of stomatopod families using molecular data.

In this study, complete mitochondrial genome sequences were determined for representatives of three of the seven stomatopod superfamilies: *Gonodactylus chiragra* (superfamily Gonodactyloidea), *Lysiosquillina maculata* (superfamily Lysiosquilloidea), and *Squilla empusa* (superfamily Squilloidea). The mitochondrial genome of each species is described and compared with other crustaceans for which complete mtDNA sequences are available. In addition, evolutionary relationships among the stomatopod taxa, and between the stomatopods and other crustaceans, are explored with phylogenetic analyses using the

concatenated protein sequences from all three stomatopods, plus those of other selected crustacean and insect genomes.

Materials and Methods

Specimens

MtDNAs from three stomatopods were sequenced: *Gonodactylus chiragra* (superfamily Gonodactyloidea) collected from Sulawesi, Indonesia; *Lysiosquillina maculata* (superfamily Lysiosquilloidea) collected from Bali, Indonesia; and *Squilla empusa* (superfamily Squilloidea) collected from Port Royal Sound, South Carolina in the United States. All laboratory work took place at the Evolutionary Genomics Laboratory at the Department of Energy's Joint Genome Institute in Walnut Creek, California.

DNA Extraction and Polymerase Chain Reaction (PCR)

Mitochondrial DNA was extracted from the raptorial appendage of each animal using the Mitochondrial DNA Extractor CT Kit (Wako Chemicals USA) following the manufacturer's protocol. Alternatively, genomic DNA was extracted using the DNeasy kit (Qiagen, Inc.) following the manufacturer's protocol. Long PCR was performed using a GeneAmp® PCR System 9700 (Applied Biosystems) thermocycler with either Herculase™ Enhanced DNA Polymerase (Stratagene®) or TaKaRa LA Taq™ (Takara Biomedicals) kits. Primers designed to well conserved regions were used to obtain short, gene-specific sequences from each mtDNA, after which taxon-specific primers were designed to amplify the remaining genome sequence. Concentrations of reagents were according to kit recommendations. Cycling parameters for the Herculase™ kit were: 92°C for 2 min, followed by 37 cycles of 92°C for 30 sec, 50°C for 25 sec, and 68°C for 12 min. The final extension was 15 min at 72°C. Cycling parameters for the TaKaRa™ kit were: 94° C for 2 min, followed by 35 cycles of 94° C for 30 sec, 50° C for 30 sec, and 70° C for 12 min. The final extension was 15 min at 72°C. All PCR products were purified using Ultra-free-MC™ tubes (Millipore).

Shotgun Subcloning

Purified PCR products were sheared at random points to 1.5 Kb fragments with a HydroShear machine (GeneMachines) and blunt-ended with a mixture of Klenow Fragment (New England Biolabs) and T4 DNA polymerase (Roche Molecular Biochemicals). Blunt-ended fragments were gel purified with a QIAquick Gel Extraction Kit (Qiagen) and cloned into pUC18 vectors (Roche Molecular Biochemicals) with the Fast-Link DNA Ligation Kit (Epicentre Technologies). Cloned fragments were electroporated into MAXEfficiency DH10b Electrocompetent Cells (Invitrogen) and plated onto LB-Amp 100 X-gal agar plates.

Rolling Circle Amplification and Cycle Sequencing

Sufficient template for sequencing was generated by amplifying each plasmid clone using TempliPhi™ DNA Sequencing Template Amplification Kit (Amersham Biosciences). TempliPhi™ reactions were performed using a dual 384-Well GeneAmp® PCR System 9700 (Applied Biosystems). Each amplified plasmid clone was sequenced in both the forward and reverse directions. Cycle sequencing reactions were performed using DYEnamic ET Terminator kit (Amersham Biosciences). The ET Terminator cycle sequencing reaction (95° C for 25 sec; 50º C for 10 sec; 60º C for 2 min) was run for 30 cycles, followed by a 4º C hold using a Dual 384-Well GeneAmp® PCR System 9700 thermocycler (Applied Biosystems).

Determination of DNA Sequence

Automated DNA sequencing was performed with MegaBACE 4000 sequencers (Molecular Dynamics). Sequence reads were processed with Cimarron Slim Phredify 2.19.5 as implemented with the MegaBACE 4000 Sequence Analyzer software (Amersham Biosciences). Processed reads were analyzed with phred and assembled with phrap (Phil Green, University of

Washington). Finishing reads were added to phrap assemblies using Sequencher v4.1 (GeneCodes).

Gene content and order was determined by alignment with known genes from the fruit fly, *Drosophila yakuba* (Clary and Wolstenholme, 1985; GenBank accession no. **X03240**), and the giant tiger prawn, *Penaeus monodon* (Wilson et al., 2000; **AF217843**) in the program MacVector version 7.1.1 (Accelrys). The location, secondary structure, and anticodons of the tRNA genes were determined with the help of the program tRNAscan-SE 1.1 (Lowe and Eddy, 1997).

Phylogenetic Analyses

In addition to the three stomatopods sequenced in this study, five other crustaceans, three insects, and one chelicerate were included in the phylogenetic analyses. The crustaceans included three eumalacostracans: *Penaeus monodon* (giant tiger prawn; GenBank accession no. **AF217843**), *Pagurus longicarpus* (hermit crab; **AF150756**), and *Panulirus japonicus* (Japanese spiny lobster; **AB071201**); and two branchiopod crustaceans: *Artemia franciscana* (brine shrimp; **X69067**) and *Daphnia pulex* (water flea; **AF117817**). The insects included were *Locusta migratoria* (migratory locust; **X80245**), *Anopheles gambiae* (African malaria mosquito; **L20934**), and *Drosophila yakuba* (fruit fly; **X03240**). *Limulus polyphemus* (Horseshoe crab; **NC003057**) was chosen to represent the Chelicerata.

The 13 protein-coding genes were conceptually translated using the invertebrate mitochondrial genetic code, then these amino acid sequences were individually aligned using ClustalX (Thompson et al., 1997) using default settings, and then adjusted individually in MacClade (Sinauer). The poorly aligning portions at the N- and C-terminal ends of all sequences were excluded, as were other ambiguously aligning regions within Nad 2, Nad 4 and

Nad 6. The alignments were then concatenated to produce a single multiple alignment of 3,309 amino acid positions.

This alignment was used for both maximum parsimony (MP) and maximum likelihood (ML) analyses. For the MP analysis an exhaustive search was conducted in PAUP 4.0b10 (Swofford, 2001), gaps were considered as missing data. Of the 3,287 total characters, 1,454 were parsimony-informative. The ML analysis was performed with Tree-puzzle 5.0 (Strimmer and von Haeseler, 1996) using the mtREV24 model (Adachi and Hasegawa, 1996) of amino acid sequence evolution. The gamma distribution model of rate heterogeneity was used with an alpha parameter of 0.41 (estimated from the data set by the program). Tree-puzzle used 1,000 quartet-puzzling steps to simultaneously choose from the possible tree topologies and to infer support values for internal branches. Trees from both analyses were rooted with the chelicerate, *Limulus polyphemus*.

Results and Discussion

Mitochondrial Genome Size

For those completely sequenced, mitochondrial genome sizes range in arthropods from 14,535 bases for the European Corn Borer, *Ostrinia nubilalis,* to 19,517 bases for the fruit fly, *Drosophila melanogaster*. For the few published crustacean mitochondrial genomes the range is somewhat narrower, from 14,628 bases for the copepod, *Tigriopus japonicus*, to 15,984 bases for the tiger prawn, *Penaeus monodon*. The stomatopod mitochondrial genomes sequenced here extend the range for crustaceans: *G. chiragra* = 16,279 bases (GenBank accession no. **DQ191682**), *L. maculata* = 16,325 bases (**DQ191683**), and *S. empusa* = 15,828 bases (**DQ191684**) (Table 1).

Most of the size difference between the stomatopod mitochondrial genomes can be attributed to the varying lengths of the A+T-rich control region. In *S. empusa* the control region is 936 bases, while in *G. chiragra* it is 1368 bases, and in *L. maculata* it is 1319 bases. Non-coding regions between genes are also a factor in genome size differences between the three species. *S. empusa* and *G. chiragra* mtDNAs have a total of 89 and 79 intergenic nucleotides, respectively, while in the larger *L. maculata* mtDNA has a total of 178 intergenic nucleotides (Table 1).

Gene Content and Arrangement

All three stomatopod mitochondrial genomes contain the 13 protein-coding, 22 tRNA, and two rRNA genes common to most metazoan animals and all other arthropod mitochondrial genomes. As in other arthropods, there is also an A+T-rich region that contains no known genes. Because the largest non-coding region has been shown to contain signals for regulating

transcription and replication for some mtDNAs, it is likely that this is the functional equivalent (Wolstenholme, 1992).

The order and arrangement of genes in each of the three stomatopod mitochondrial genomes (Figure 1) are identical to each other and to the gene order in *Drosophila* (Clary and Wolstenholme, 1985) and to the crustaceans *Penaeus monodon* (Wilson et al., 2000), *Panulirus japonicus* (Yamauchi et al., 2002), and *Daphnia pulex* (Crease, 1999). The branchiopod crustacean, *Artemia franciscana*, is very similar in gene arrangement with the exception of two tRNA genes (Valverde et al., 1994), while the copepod, *Tigriopus japonicus* (Machida et al., 2002), and the hermit crab, *Pagurus longicarpus* (Hickerson and Cunningham, 2000), have more dramatic differences in gene order.

As is true for many metazoan animals, a number of the genes appear to overlap in all three stomatopod mitochondrial genomes; see Table 1 for details. Overlaps are expected to be tolerated when the genes are on opposite strands except for the necessity of these bases to simultaneously function in each capacity. It is less obvious how overlaps of genes on the same strand would be resolved, since the commonly accepted model has the complete strand generating a polycistronic message, with individual, gene specific RNAs being enzymatically cleaved out. Possible explanations include having multiple transcriptional promoters for the individual genes, translation from a polycistronic RNA with variable initiation sites, or differential transcript cleavage producing only one or the other overlapping RNA.

Base Composition and Codon Usage

The overall A+T content for *S. empusa* mtDNA is 68.4% (A=5587, T=5245, C=2931, G=2064), for *G. chiragra* mtDNA is 67.5% (A=5719, T=5269, C=3264, G=2027), and for *L. maculata* mtDNA is 63.9% (A=5347, T=5076, C=3497, G=2405). In all cases, the non-coding

region is more A+T rich than the rest of the genome, but this effect is less pronounced in *L. maculata* (*S. empusa* = 79.8%, *G. chiragra* = 78.4 %, *L. maculata* = 66.9%). As in other crustacean mtDNAs (i.e. *Daphnia Pulex* [Crease, 1999], *Penaeus monodon* [Wilson et al., 2000], and *Panulirus japonicus* [Yamauchi et al., 2002]), the base composition in the 1st and 3rd codon positions are especially biased towards A and T nucleotides (Table 2).

Translation Initiation and Termination Codons

For all three stomatopod mtDNAs, 12 of the 13 proteins initiate with an ATN codon (Table 1). In each case, the exception is the initiation codon for *cox1*, which appears to be ACG based on alignments with *Drosophila melanogaster*, *D. yakuba*, and *Penaeus monodon*. Two other malacostracan crustaceans, *Penaeus monodon* (Wilson et al., 2000) and *Panulirus japonicus* (Yamauchi et al., 2002) also appear to use ACG as the *cox1* initiation codon. As suggested by Wilson et al. (2000), one possible explanation for this atypical start codon is that RNA editing may convert the ACG codon to an AUG codon in the mRNA, as seen in the tomato (Kadowaki et al., 1995).

Other arthropods are also known to have unconventional initiation codons for the *cox1* gene. Most oddly, in *Drosophila* (Clary and Wolstenholme, 1983) and some crustaceans, such as *D. pulex* (Crease, 1999) and *P. longicarpus* (Hickerson and Cunningham, 2000), a tetranucleotide sequence (ATAA, ATTA, or ATCA) is thought to function as the initiation codon for *cox1* (Clary and Wolstenholme, 1983). However, this does not seem possible in these stomatopods.

All protein-encoding genes have a complete TAA stop codon except for *cox2* and *nad*6 for all three stomatopods, *nad3* for *G. chiragra*, and *cox3* for *S. empusa*. Each of these genes has

an incomplete stop codon of T or TA (Table 1) that is likely converted to a complete TAA termination codon by polyadenylation during RNA processing (Ojala et al., 1980; 1981).

Transfer RNA and Ribosomal RNA Genes

All three stomatopod mitochondrial genomes encode 22 tRNA genes that can fold into cloverleaf secondary structures (Figures 2-4). The anticodons are identical in all three stomatopod species with the exception of tRNA(A), which is CGC for *S. empusa*, and TGC for *G. chiragra* and *L. maculata* (and commonly for other crustaceans). Among the crustacean species for which anticodon usage has been published, there are only two anticodons that otherwise differ: (1) the anticodon for tRNA(S1-nct) (recognizing codon AGN), which is ACT for stomatopods; TCT for *Pagurus longicarpus* (Hickerson and Cunningham, 2000), *Panulirus japonicus* (Yamauchi et al., 2002), and *Tigriopus japonicus* (Machida et al., 2002); and GCT for *Penaeus monodon* (Wilson et al., 2000) and *Daphnia pulex* (Crease, 1999) and (2) the anticodon for tRNA(K), which is TTT for the stomatopods, *P. longicarpus*, *P. japonicus, P. monodon,* and *T. japonicus* and CTT for *D. pulex*. The differences between these anticodons all correspond to the third wobble position.

The small and large subunit ribosomal genes (*rrnS* and *rrnL*, respectively) are in the same relative location and, assuming they extend to the boundaries with flanking genes, are of similar size in all three stomatopods sequenced (Figure 1 and Table 1).

Phylogenetic Analyses

Maximum parsimony (MP) analysis of the amino acid sequences from the concatenated 13 protein-coding genes yields four most parsimonious trees with a length of 7,433 (consistency index = 0.7603, retention index = 0.5135, rescaled consistency index = 0.3904). A strict consensus of these trees is shown in Figure 5. The maximum likelihood (ML) analysis produces

a similar tree topology as the MP analysis, but with greater resolution of the relationships among some groups (Figure 6).

As in previous phylogenetic analyses (Yamauchi et al., 2002), the ML analysis found strong support (81% ML) for the sister taxa relationship between insects and malacostracan crustaceans, but this relationship was not resolved in the parsimony analysis. Both analyses gave strong support for the monophyly of the Malacostraca (100% ML and MP) and for the monophyly of the stomatopods (100% ML and MP). As in the analysis of Yamauchi et al. (2002), the clade including the hermit crab and Japanese spiny lobster had moderate support from ML analysis (77%). However, this clade was not resolved in a strict consensus of the MP trees.

These data support the monophyly of the Malacostraca and a sister taxa relationship between hoplocaridans and eumalacostracans. This confirms that stomatopods, like other malacostracan crustaceans, are more closely related to the insects than to branchiopod crustaceans (Figures 5 and 6).

The few previous phylogenetic analyses of the Crustacea that included stomatopods have also supported a sister taxa relationship between hoplocarids and eumalacostracans (Spears and Abele, 1999; Giribet and Ribera, 2000; Giribet et al., 2001). An exception is the analysis by Schram and Hof (1998), which placed the stomatopod clade in a derived location within the eumalacostracans. However, Schram and Hof (1998) felt this was due to the effects of missing data, and when they removed those characters from the analysis, the stomatopod clade shifted to an unresolved position at the base of the Malacostraca. They concluded that there was no reason to reject the sister group status of Hoplocarida (stomatopods) and the Eumalacostraca.

These data strongly support the monophyly of the Stomatopoda so far sampled. Stomatopods are currently divided into 18 families and seven superfamilies (Ahyong and Harling, 2000), only three of which are sampled here. A preliminary, family-level cladistic analysis of the stomatopods by Ahyong in 1997 concluded that the Squilloidea were the most basal of the extant superfamilies. However, in a more recent analysis, Ahyong and Harling (2000) concluded that the Gonodactyloidea are the most basal of the extant taxa. Hof (1998) reached the same conclusion in his preliminary cladistic analysis. All three of these analyses were based solely on morphological characters.

Of the three superfamilies represented in this study (Gonodactyloidea, Lysiosquilloidea, and Squilloidea) the Squilloidea has moderate support as the most basal in the ML analysis (83%). However, a strict consensus of the MP trees yields no resolution of these groups. Therefore, these data give some support to Ahyong's original conclusion that the Squilloidea are the most basal of the extant stomatopods, and that the Gonodactyloidea and Lysiosquilloidea are sister taxa. However, because few taxa are included in the current study these results must be interpreted with caution.

In addition to more complete sampling within families and a lessening of the impact of missing data, the main difference between Ahyong's (1997) original phylogeny of the Stomatopoda, and the more recent phylogeny published by Ahyong and Harling (2000), was the choice of outgroup. Ahyong and Harling (2000) stated that the placement of the Gonodactyloidea at the base of the extant taxa in their later analysis (in contrast to its more derived location in Ahyong's previous analysis) was the result of outgroup morphology and overall signal in the data set. In his original analysis, Ahyong used members of the Paleostomatopoda and Archaeostomatopodea, both fossil stomatopod lineages, as outgroups.

The more recent analysis used only members from one family within the Archaeostomatopodea, the Tyrannophontidae, as an outgroup. This change in the choice of outgroup was made based on the cladistic analysis of fossil stomatopods by Jenner et al. (1998). Data from the current study suggest that the original outgroup choices may have been more appropriate. However, this study is rather preliminary in that so few taxa are currently represented. As more complete mitochondrial genome sequences become available, including representatives of the Phyllocarida (the third malacostracan subclass yet to be sequenced), and representatives of the other stomatopod superfamilies, more complete analyses can be performed. The complete genomes presented in this study should facilitate the creation of stomatopod specific primers in order to produce these sequences.

Acknowledgements

This work was performed partly under the auspices of the US Department of Energy's Office of Science, Biological and Environmental Research Program, and by the University of California, Lawrence Berkeley National Laboratory under Contract No. DE-AC02-05CH11231.

We wish to thank Dr. Pam Jutte for providing the *Squilla empusa* specimen. This work was performed under the auspices of the US Department of Energy's Office of Science, Biological and Environmental Research Program, and by the University of California, Lawrence Berkeley National Laboratory under contract No. DE-AC02-05CH11231, Lawrence Livermore National Laboratory under Contract No. DE-AC52-07NA27344, and Los Alamos National Laboratory under contract No. DE-AC02-06NA25396.

References

Adachi, J., Hasegawa, M., 1996. Model of amino acid substitution in proteins encoded by mitochondrial DNA. J. Mol. Evol. 42, 459-468.

Ahyong, S.T., 1997. Phylogenetic analysis of the Stomatopoda (Malacostraca). J. Crustacean Biol. 17(4), 695-715.

Ahyong, S.T., Harling, C., 2000. The phylogeny of the Stomatopod Crustacea. Aust. J. Zool. 48, 607-642.

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

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

Caldwell, R.L., Dingle, H., 1975. Ecology and evolution of agonistic behavior in stomatopods. Naturwissenschaften 62, 214-222.

Caldwell, R.L., Dingle, H., 1976. Stomatopods. Scientific American 234, 80-89.

Caldwell, R.L., 1991. Variation in reproductive behavior in Stomatopod Crustacea. In: Bauer, R.T., Martin, J.W. (Eds.), Crustacean Sexual Biology. Columbia Univ. Press, New York, pp. 67-90.

Clary, D.O., Wolstenholme, D.R., 1983. Genes for cytochrome c oxidase subunit I, URF2, and three *tRNA*s in *Drosophila* mitochondrial DNA. Nucleic Acids Res. 11, 6859-6872.

Clary, D.O., Wolstenholme, D.R., 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.

Cronin, T.W., Marshall, N.J., Land, M.F., 1994. The unique visual system of the mantis shrimp. American Scientist 82, 356-365.

Garcia-Machado, E., Pempera, M., Dennebouy, N., Oliva-Suarez, M., Mounolou, J.C., Monnerot, M., 1999. Mitochondrial genes collectively suggest the paraphyly of Crustacea with respect to Insecta. J. Mol. Evol. 49, 142-149.

Giribet, G., Ribera, C., 2000. A review of Arthropod Phylogeny: New data based on ribosomal DNA sequences and direct character optimization. Cladistics 16, 204-231.

Giribet, G., Edgecombe, G.D., Wheeler, W.C., 2001. Arthropod phylogeny based on eight molecular loci and morphology. Nature 413, 157-160.

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

Hof, C.H., 1998. Fossil stomatopods (Crustacea: Malacostraca) and their phylogenetic impact. J. Natural History 32, 1567-1576.

Jenner, R.A., Hof, C.H.J., Schram, F.R., 1998. Palaeo- and Archaeostomatopods (Hoplocarida: Crustacea) from the Bear Gulch Limestone. Mississippian (Namurian), of Central Montana. Contrib. Zool. 67, 155-185.

Kadowaki, K. I., Ozawa, K., Kazama, S., Kubo, N., Akihama, T., 1995. Creation of an initiation codon by RNA editing in the cox1 transcript from tomato mitochondria. Curr. Genet. 28, 415-422.

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

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

Nardi, F., Spinsanti, G., Boore, J.L., Carapelli, A., Romano, D., Frati, F., 2003. Hexapod origins: Monophyletic or Paraphyletic? Science 299, 1887-1889.

Ojala, D., Merkel, C., Gelfand, R., Attardi, G., 1980. The *tRNA* genes punctuate the reading of genetic information in human mitochondrial DNA. Cell 22, 393-403.

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

Reaka, M.L., Manning, R.B., 1981. The behavior of stomatopod crustacea, and its relationship to rates of evolution. J. Crustacean Biol. 1(3), 309-327.

Schram, F.R., 1986. Crustacea. Oxford Univ. Press, Oxford.

Schram, F.R., Hof, C.H.J., 1998. Fossils and the interrelationships of major crustacean groups. In: Edgecombe, G.D. (Ed.), Arthropod Fossils and Phylogeny. Columbia Univ. Press, New York, pp. 233-302.

Spears, T., Abele, L.G., 1999. Phylogenetic relationships of crustaceans with foliaceous limbs: An 18S rDNA study of Branchiopoda, Cephalocarida, and Phyllocarida. J. Crustacean Biol. 19(4), 825-843.

Strimmer, K., von Haeseler, A., 1996. Quartet puzzling: a quartet maximum likelihood method for reconstructing tree topologies. Mol. Biol. Evol. 13, 964-969.

Swofford, D.L., 2001. PAUP*: Phylogenetic Analysis Using Parsimony (*and Other Methods), version 4.0, Sinauer Associates, Sunderland, MA.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.J., 1997. The ClustalX X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 25, 4876-4882.

Valverde, J.R., Batuecas, B., Moratilla, C., Marco, R., Garesse, R., 1994. The complete mitochondrial DNA sequence of the crustacean *Artemia franciscana.* J. Mol. Evol. 39, 400-408.

Wilson , K., Cahill, V., Ballment, E., Benzie, J., 2000. The complete sequence of the mitochondrial genome of the crustacean: *Penaeus monodon*: Are malacostracan crustaceans more closely related to insects than to branchiopods? Mol. Biol. Evol. 17(6), 863-874.

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

Yamauchi, M.M., Miya, M.U., Mutsumi, N., 2002. Complete mitochondrial DNA sequence of the Japanese spiny lobster, *Panulirus japonicus* (Crustacea: Decapoda). Gene 295(1), 89-96.

Figure 1. Linearized representation of the mitochondrial gene arrangement shared by *Squilla empusa*, *Lysiosquillina maculata*, and *Gonodactylus chiragra*. Transfer RNA genes are designated by their single letter amino acid codes, with the two leucine- and two serine-specifying tRNAs differentiated by numeral (L1, L2, S1, and S2 recognizing the codons CUN, UUR, AGN, and UCN, respectively). Genes that are oriented in reverse, i.e., right-to-left, are designated by underlining if protein- or rRNA-encoding and by labeling below the gene map if tRNA genes. All others are oriented left-to-right as drawn.

Figure 2. Putative cloverleaf structures for the mitochondrial tRNAs of *Gonodactylus chiragra*. Lines (-) denote standard Watson-Crick pairs, plus signs (+) denote G-T base pairings.

Figure 3. Putative cloverleaf structures for the mitochondrial tRNAs of *Lysiosquillina maculata*. Lines (-) denote standard Watson-Crick pairs, plus signs (+) denote G-T base pairings.

Figure 4. Putative cloverleaf structures for the mitochondrial tRNAs of *Squilla empusa*. Lines (-) denote standard Watson-Crick pairs, plus signs (+) denote G-T base pairings.

Figure 5. Strict consensus of four most parsimonious trees based on concatenated sequences of the 13 protein coding genes of the mitochondrial genome using maximum parsimony (MP) analysis.

Figure 6. Phylogenetic tree based on maximum likelihood (ML) analysis. Numbers above the branches represent support in percent by puzzle quartets.

Table 1. Features of stomatopod mtDNAs

		Gonodactylus chiragra							*Lysiosquillina maculata*							*Squilla empusa*						
Gene		From	To	L[a]	Start	Stop[b]	AA/ AC[c]	nt[d]	From	To	L[a]	Start	Stop[b]	AA/ AC[c]	nt[d]	From	To	L[a]	Start	Stop[b]	AA/ AC[c]	nt[d]
cox1	H	1	1539	1539	ACG	TAA	512	0	1	1539	1539	ACG	TAA	512	0	1	1539	1539	ACG	TAA	512	0
trnL2	H	1535	1601	67			TAA	-5	1535	1600	66			TAA	-5	1535	1602	68			TAA	-5
cox2	H	1606	2293	688	ATG	T	229	4	1605	2292	688	ATG	T	229	4	1606	2293	688	ATG	T	229	3
trnK	H	2294	2361	68			TTT	0	2293	2360	68			TTT	0	2294	2361	68			TTT	0
trnD	H	2361	2427	67			GTC	-1	2360	2428	69			GTC	-1	2361	2428	68			GTC	-1
atp8	H	2428	2586	159	ATC	TAA	52	0	2429	2587	159	ATT	TAA	52	0	2429	2587	159	ATC	TAA	52	0
atp6	H	2580	3257	678	ATG	TAA	225	-7	2584	3258	675	ATA	TAA	224	-4	2581	3258	678	ATG	TAA	225	-7
cox3	H	3257	4048	792	ATG	TAA	263	-1	3266	4057	792	ATG	TAA	263	7	3258	4045	788	ATG	T/TA*	262	-1
trnG	H	4048	4112	65			TCC	-1	4057	4121	65			TCC	-1	4046	4111	66			TCC	0
nad3	H	4113	4464	353	ATG	T/TA*	117	0	4122	4475	354	ATG	TAA	117	0	4115	4465	351	ATT	TAA	116	3
trnA	H	4465	4529	65			TGC	0	4477	4542	66			TGC	1	4469	4533	65			CGC	3
trnR	H	4537	4601	65			TCG	7	4550	4614	65			TCG	7	4539	4603	65			TCG	5
trnN	H	4606	4674	69			GTT	4	4618	4687	70			GTT	3	4609	4678	70			GTT	5
trnS1	H	4674	4743	70			ACT	-1	4687	4756	70			ACT	-1	4679	4746	68			ACT	0
trnE	H	4743	4810	68			TTC	-1	4756	4824	69			TTC	-1	4749	4814	66			TTC	2

trnF	L	4812	4881	70			GAA	1	4852	4918	67			GAA	27	4815	4882	68			GAA	0
nad5	L	4882	6594	1713	ATG	TAA	570	0	4941	6653	1713	ATA	TAA	570	22	4882	6594	1713	ATA	TAA	570	-1
trnH	L	6613	6678	66			GTG	18	6672	6737	66			GTG	18	6613	6679	67			GTG	18
nad4	L	6678	8018	1341	ATG	TAA	446	-1	6737	8077	1341	ATG	TAA	446	-1	6679	8019	1341	ATG	TAA	446	-1
nad4L	L	8012	8311	300	ATG	TAA	99	-7	8071	8370	300	ATG	TAA	99	-7	8013	8312	300	ATG	TAA	99	-7
trnT	H	8314	8381	48			TGT	2	8373	8439	67			TGT	2	8315	8381	67			TGT	2
trnP	L	8382	8449	68			TGG	0	8440	8505	66			TGG	0	8382	8447	66			TGG	0
nad6	H	8452	8974	524	ATT	T/TA*	174	2	8508	9027	520	ATC	T/TA*	173	2	8457	8976	521	ATT	T/TA*	173	9
cob	H	8975	10111	1137	ATG	TAA	378	0	9028	10164	1137	ATG	TAA	378	0	8977	10113	1137	ATG	TAA	378	0
trnS2	H	10114	10182	69			TGA	2	10164	10232	69			TGA	-1	10112	10181	70			TGA	-2
nad1	L	10212	11153	942	ATA	TAA	313	29	10254	11195	942	ATG	TAA	313	21	10203	11144	942	ATA	TAA	313	21
trnL1	L	11151	11217	67			TAG	-3	11196	11261	66			TAG	0	11142	11207	66			TAG	-3
rrnL	L	11218	12582	1365				0	11262	12608	1347				0	11208	12561	1354				0
trnV	L	12583	12653	71			TAC	0	12609	12679	71			TAC	0	12562	12632	71			TAC	0
rrnS	L	12654	13496	843				0	12680	13535	856				0	12633	13466	834				0
CR	H	13497	14864	1368				0	13536	14854	1319				0	13467	14402	936				0
trnI	H	14865	14933	69			GAT	0	14855	14922	68			GAT	0	14403	14471	69			GAT	0
trnQ	L	14931	14998	68			TTG	-3	14931	14999	69			TTG	8	14469	14538	70			TTG	-3
trnM	H	15006	15074	69			CAT	8	15009	15077	69			CAT	9	14551	14619	69			CAT	12
nad2	H	15075	16079	1005	ATT	TAA	334	0	15078	16088	1011	ATT	TAA	336	0	14620	15621	1002	ATT	TAA	333	0

													-						
trnW	H	16079	16146	67		TCA	-1	16079	16147	69		TCA	10	15620	15688	69		TCA	-2
trnC	L	16146	16209	64		GCA	-1	16149	16213	65		GCA	1	15692	15758	67		GCA	3
trnY	L	16212	16279	68		GTA	2	16260	16325	66		GTA	46	15762	15828	67		GTA	3

[a] Length of the gene in nucleotides (not including the potential terminal A in cases where its being part of the ORF is ambiguous).

[b] Abbreviated stop codons of T or TA are presumably completed by polyadenylation.

[c] Number of inferred amino acids for protein genes and anticodon identities for tRNA genes.

[d] Number of subsequent intervening nucleotides separating this from the next gene, with overlaps indicated by negative numerals.

Table 2. Codon usage for the 13 mitochondrial protein encoding genes for *Gonodactylus chiragra* (Gch), *Lysiosquillina maculata* (Lma), and *Squilla empusa* (Sem).

Amino		Gch		Lma		Sem		Amino		Gch		Lma		Sem	
acid	Codon	N	%	N	%	N	%	acid	Codon	N	%	N	%	N	%
Phe	TTT	242	6.5	167	4.5	223	6.0	Ile	ATT	213	5.7	186	5.0	217	5.8
(gaa)[a]	TTC	67	1.8	139	3.7	95	2.6	(gat)	ATC	81	2.2	89	2.4	65	1.7
Leu	TTA	260	7.0	188	5.1	277	7.5	Met	ATA	153	4.1	157	4.2	184	4.9
(taa)	TTG	82	2.2	103	2.8	95	2.6	(cat)	ATG	77	2.1	78	2.1	58	1.6
Ser	TCT	103	2.8	82	2.2	126	3.4	Thr	ACT	86	2.3	87	2.3	85	2.3
(tga)	TCC	26	0.7	49	1.3	17	0.5	(tgt)	ACC	45	1.2	39	1.0	38	1.0
	TCA	66	1.8	58	1.6	48	1.3		ACA	70	1.9	59	1.6	83	2.2
	TCG	7	0.2	19	0.5	8	0.2		ACG	11	0.3	21	0.6	5	0.1
Tyr	TAT	112	3.0	89	2.4	121	3.3	Asn	AAT	86	2.3	65	1.7	82	2.2
(gta)	TAC	45	1.2	65	1.7	31	0.8	(gtt)	AAC	45	1.2	59	1.6	45	1.2

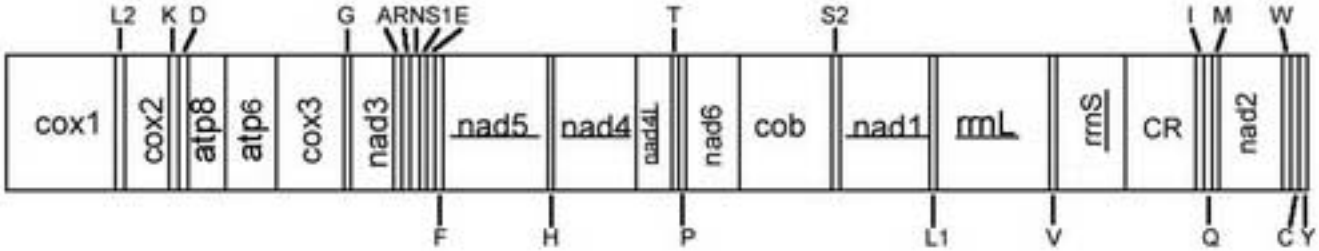
TER	TAA	10	0.3	11	0.3	10	0.3	Lys	AAA	53	1.4	54	1.5	68	1.8
	TAG	0	0.0	0	0.0	0	0.0	(ttt)	AAG	27	0.7	25	0.7	11	0.3
Cys	TGT	35	0.9	24	0.6	35	0.9	Ser	AGT	29	0.8	42	1.1	54	1.5
(gca)	TGC	8	0.2	16	0.4	5	0.1	(gct)	AGC	9	0.2	18	0.5	6	0.2
Trp	TGA	86	2.3	71	1.9	74	2.0		AGA	58	1.6	49	1.3	51	1.4
(tca)	TGG	13	0.3	29	0.8	25	0.7		AGG	33	0.9	20	0.5	28	0.8
Leu	CTT	83	2.2	80	2.1	74	2.0	Val	GTT	94	2.5	94	2.5	98	2.6
(tag)	CTC	21	0.6	31	0.8	13	0.3	(tac)	GTC	27	0.7	34	0.9	28	0.8
	CTA	112	3.0	100	2.7	89	2.4		GTA	100	2.7	94	2.5	94	2.5
	CTG	16	0.4	57	1.5	21	0.6		GTG	23	0.6	50	1.3	24	0.6
Pro	CCT	73	2.0	56	1.5	79	2.1	Ala	GCT	97	2.6	99	2.7	101	2.7
(tgg)	CCC	31	0.8	35	0.9	18	0.5	(tgc)	GCC	50	1.3	48	1.3	49	1.3
	CCA	27	0.7	37	1.0	30	0.8		GCA	65	1.7	65	1.7	69	1.9
	CCG	6	0.2	9	0.2	9	0.2		GCG	17	0.5	27	0.7	13	0.3
His	CAT	51	1.4	49	1.3	46	1.2	Asp	GAT	50	1.3	46	1.2	39	1.0

(gtg)	CAC	36	1.0	35	0.9	37	1.0	(gtc)	GAC	26	0.7	29	0.8	39	1.0
Gln	CAA	51	1.4	51	1.4	62	1.7	Glu	GAA	64	1.7	55	1.5	59	1.6
(ttg)	CAG	19	0.5	22	0.6	9	0.2	(ttc)	GAG	20	0.5	30	0.8	24	0.6
Arg	CGT	15	0.4	12	0.3	19	0.5	Gly	GGT	52	1.4	52	1.4	76	2.0
(tcg)	CGC	5	0.1	5	0.1	1	0.0	(tcc)	GGC	34	0.9	38	1.0	31	0.8
	CGA	27	0.7	28	0.8	33	0.9		GGA	110	3.0	111	3.0	92	2.5
	CGG	15	0.4	17	0.5	9	0.2		GGG	67	1.8	68	1.8	63	1.7

[a] tRNA anticodon identities are shown in parentheses.

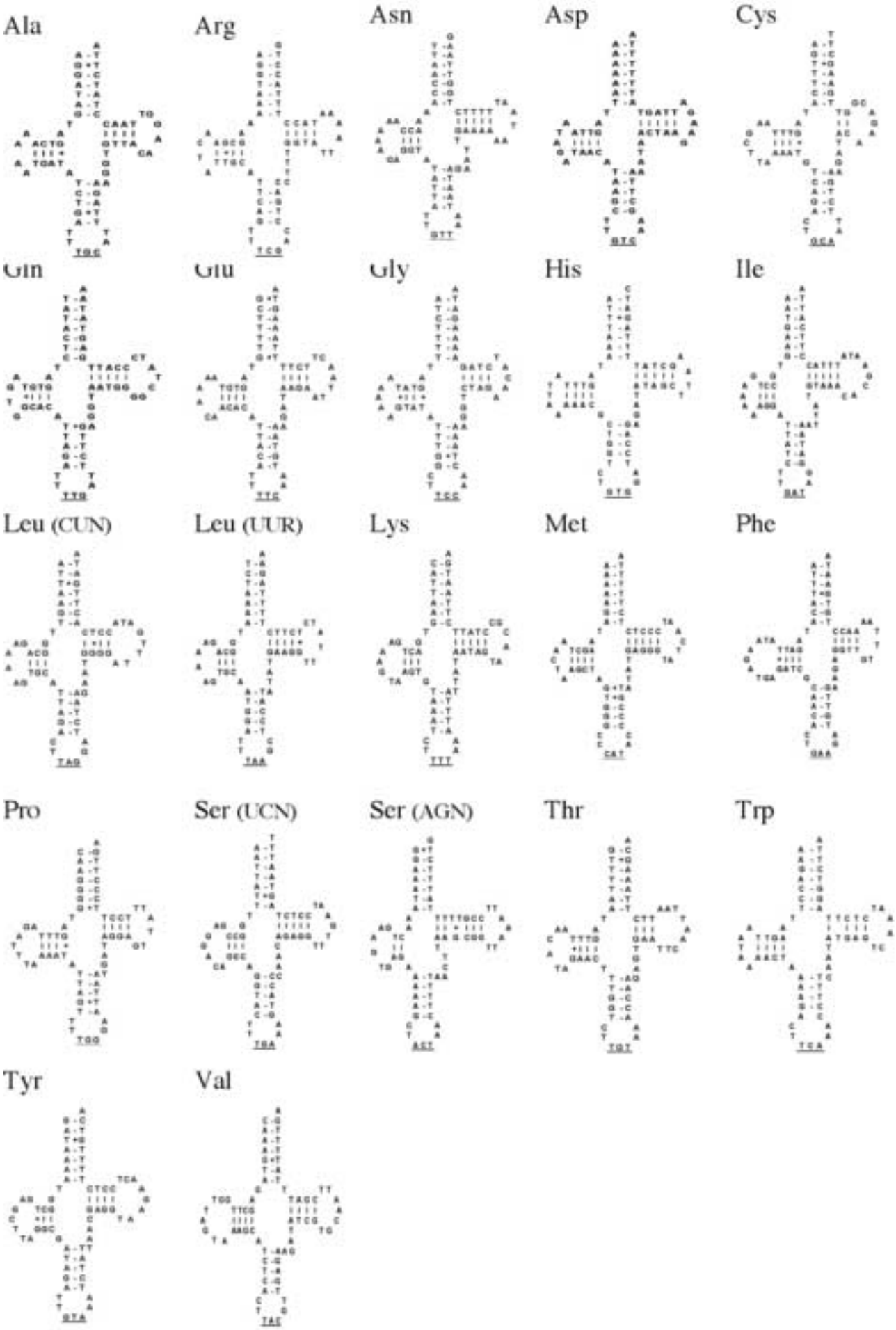
Figure

Click here to download high resolution image



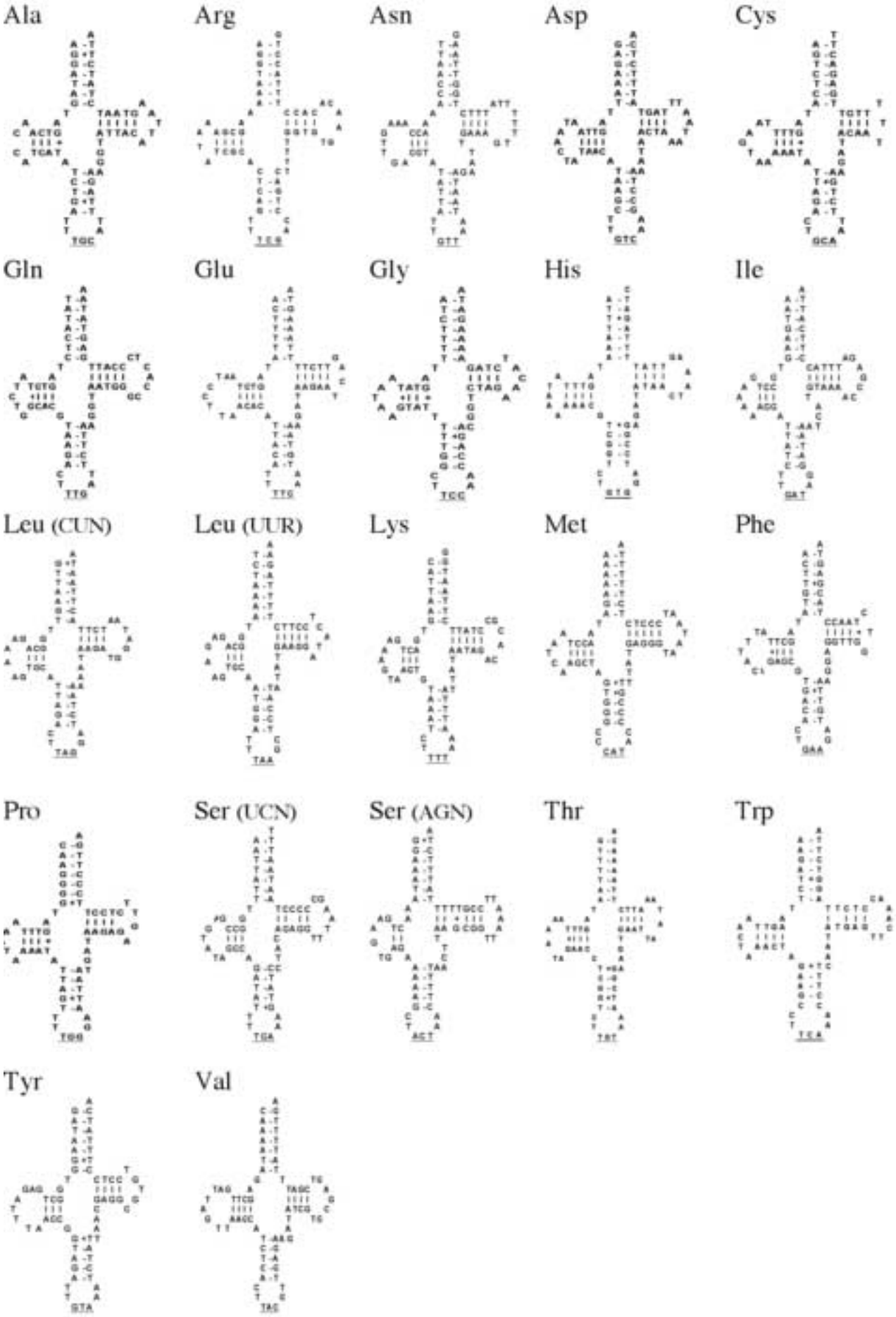
Figure

Click here to download high resolution image



Figure

Click here to download high resolution image



Figure

Click here to download high resolution image

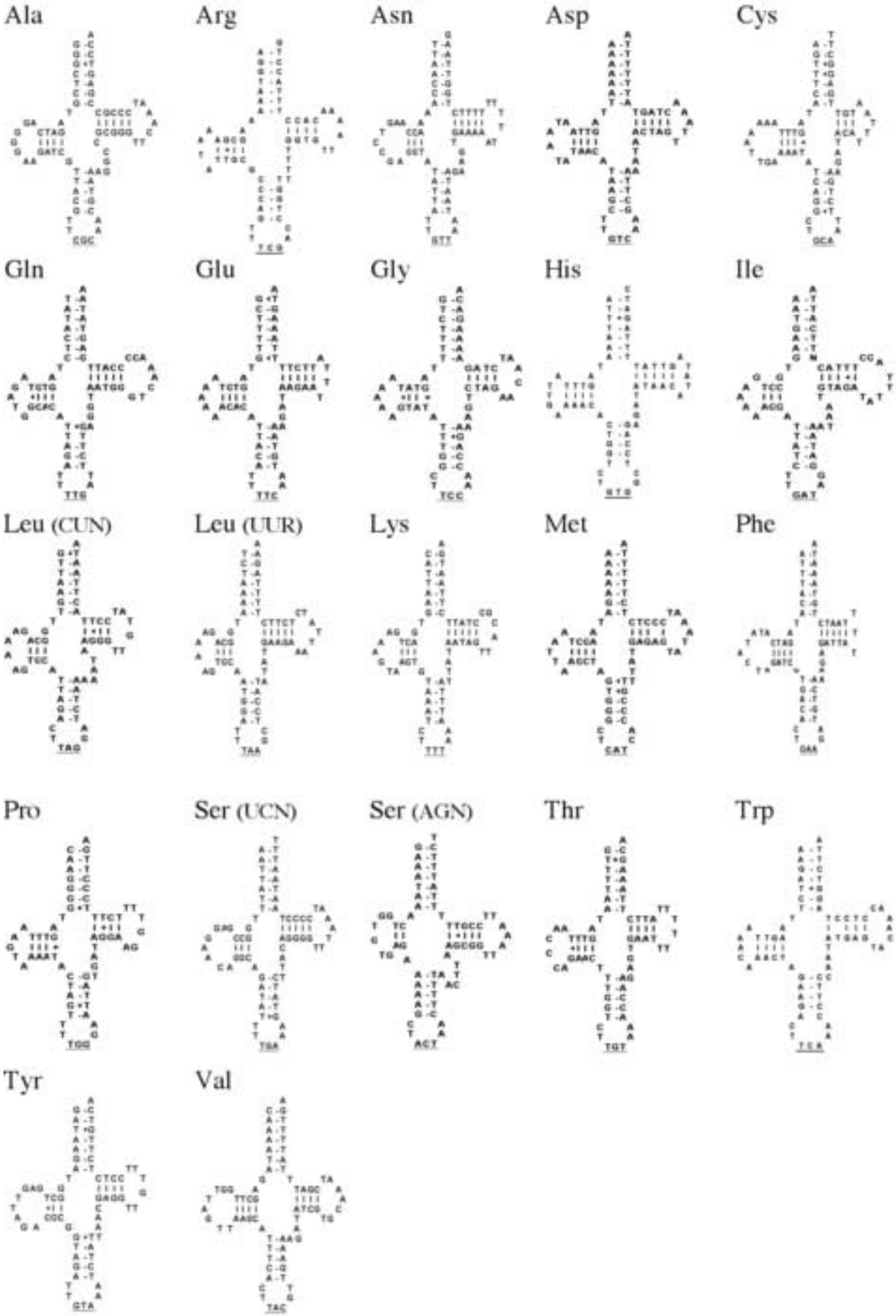


Figure
Click here to download high resolution image

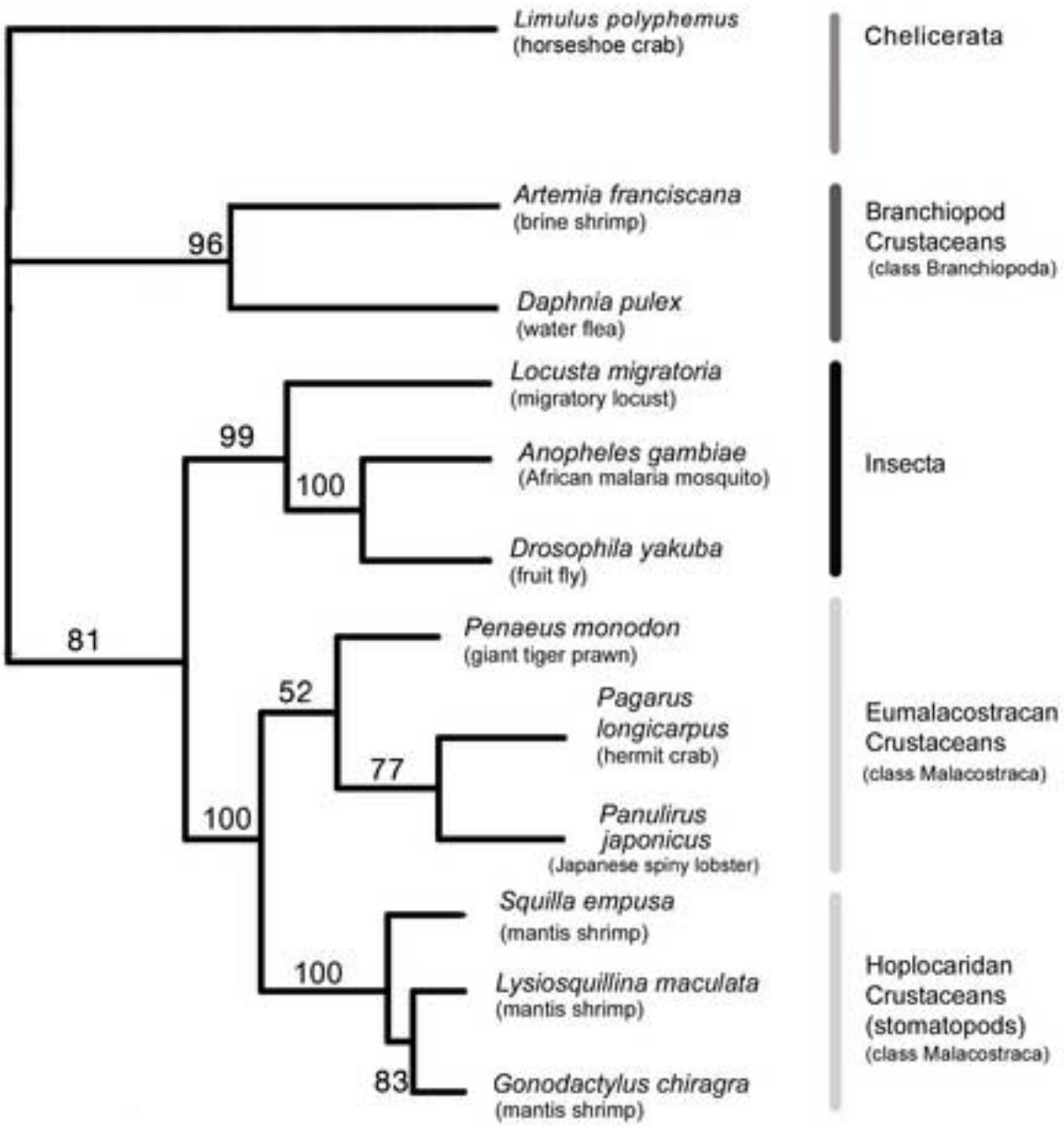
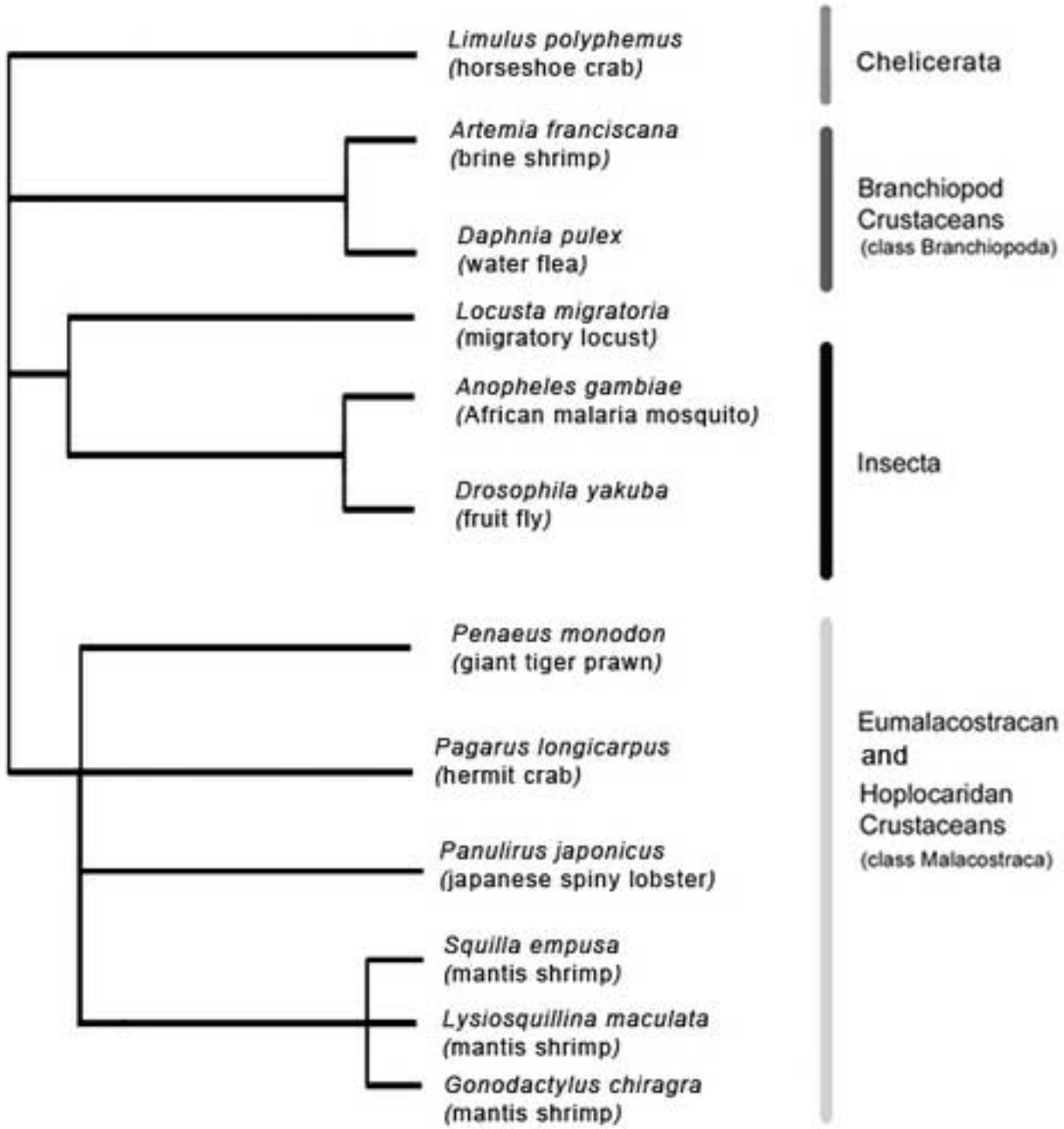



Figure
Click here to download high resolution image