Genetic markers in blue crabs (*Callinectes sapidus*) II: Complete Mitochondrial Genome Sequence and Characterization of Genetic Variation

Allen R. Place[1], Xiaojun Feng[1], Colin R. Steven[1], H. Matthew Fourcade[2], and Jeffrey L. Boore[2]

[1]Center of Marine Biotechnology, University of Maryland Biotechnology Institute, 701 East Pratt Street, Baltimore, Maryland 21202

[2]DOE Joint Genome and Lawrence Berkeley National Lab, 2800 Mitchell Drive, Walnut Creek, CA 94598

Corresponding author: Allen Place, Tel: (410) 234-8828 Fax: (410) 234-8896 E-mail address: place@umbi.umd.edu

Keywords: mitochondrial genome, blue crab, nucleotide variation, long PCR, shotgun sequencing, gene rearrangement

Abbreviations: *atp6*, *atp8*, ATPase subunits 6 and 8; *cox1*, *cox2*, *cox3*, cytochrome oxidase subunits 1-3; CR, control region; *cob*, cytochrome b; mtDNA, mitochondrial DNA; *nad1-6*, *4L*, NADH dehydrogenase subunits, 1-6 and 4L; PCR, polymerase chain reaction; RCA, rolling circle amplification; *rrnL, rrnS*, large and small subunit ribosomal RNA; *trnX,* tRNA genes, where X stands for the one letter amino acid code, with tRNAs for L and S differentiated by anticodon in parentheses.

Abstract

Given the commercial and ecological importance of the dwindling Chesapeake Bay blue crab (*Callinectes sapidus*) fishery there is a surprising scarcity of information concerning the molecular ecology of this species. The few studies published to date are based on allozyme data and indicate a single, panmictic population along the Atlantic coast. To address this short-coming we have initiated the development of genetic markers from both the nuclear and mitochondrial genomes of the blue crab. Here we report the entire nucleotide sequence for the blue crab mitochondrial genome, which is a 16,263 bp in length, circular, and A+T-rich (69.1%). We have identified all of the normal complement of 37 genes (for 13 proteins, 22 tRNAs, and two rRNAs) plus a large (1,434 bp), putatively hypervariable region that is 78.2 % A+T. Gene arrangement is similar to those of other arthropods (e.g. *Artemia*) but dramatically different from that of the hermit crab (*Pagurus longicarpus*), which has a unique gene order among arthropods. As in the mtDNA of the swimming crab, *Portunus trituberculatus, trnH* is located between *trnE* and *trnF*, rather than at its primitive position upstream of *nad5*. Genetic variation is matrilineally inherited based on parent/offspring screening for nucleotide variation in the putative control region.

1. Introduction

Genetic analysis of population structure in the blue crab (*Callinectes sapidus*) has received surprisingly little attention, considering the commercial and ecological importance of the species. Early studies of protein polymorphisms (allozymes) suggested similar gene frequencies in Chesapeake Bay and Chincoteague Bay populations, but differences between these populations and with those from South Carolina (reviewed in Burton and Feldman, 1982). Heterozygote deficiencies in the Chesapeake Bay and Chincoteague Bay populations reported by Cole and Morgan (1978) may reflect the mixing of genetically differentiated subpopulations, but may otherwise have been a technical artifact resulting from inconsistencies in scoring of gels.

There are at least 11 described species of the blue crab genus Callinectes that exist in temperate and tropical waters of the Atlantic and Pacific oceans. They are members of the crustacean group Brachyura, one of the most specious (Warner, 1977), comprising almost 6,000 species placed in 47 families (Bowman and Abele, 1982). Brachyurans are found worldwide, largely in marine habitats, from tropical mountains to deep-sea hydrothermal vents.

Because of matrilineal inheritance, high evolutionary rates of change, and apparent lack of intermolecular genetic recombination, mtDNAs have been the mainstay for studying population structure, phylogeography, and phylogenetic relationships of animals at various taxonomic levels (Avise, 2000). In animals, these genomes are generally small (15-20 kb), circular DNAs containing the same 37 genes plus a single large non-coding region which contains controlling elements for replication and transcription. Animal mtDNAs have extensive intraspecific polymorphism (especially in the non-coding control region) and generally evolve faster than nuclear genes, although to date no studies have investigated blue crab mitochondrial polymorphisms. Most mtDNA variants involve nucleotide substitutions or small length changes,

whereas gene order can remain stable for long periods of time in some lineages. We have determined the complete sequence of the *Callinectes sapidus* mitochondrial genome, which we describe here, in part to identify variable regions for distinguishing the mothers of hatchery derived juveniles from those in the wild.

2. Materials and Methods

2.1 Blue crab mitochondrial genome isolation and amplification

A mtDNA-enriched preparation was isolated from the sponge (i.e. egg mass) of a single female sampled from Davis Bay, Mississippi. Ten grams of eggs were taken from the sponge and homogenized in 100 mL of ice-cold buffer (100 mM Tris-Cl, pH 7.4, 250 mM sucrose, 10 mM EDTA). Cellular debris was pelleted and discarded following two ten-minute centrifugations at 1,500 x g. The mitochondria were pelleted from this supernatant with a ten-minute centrifugation at 10,000 x g, resuspended in 500 µL Buffer I (50 mM glucose, 25 mM Tris-Cl pH8.0, 10 mM EDTA), then lysed with 1 mL of Buffer II (0.2N NaOH, 1% SDS) for five minutes on ice, which causes DNA denaturation. Following lysis, 750 µL of Buffer III (3 M KOAc, 11.5 M Acetic Acid) was added to neutralize the pH of the sample and allow DNA reannealing. Since the mtDNA consists of two toplogically constrained strands, it reanneals very rapidly, leaving long strings of nuclear DNA to become entwined in membranous debris, which was pelleted with a five-minute centrifugation at 12,000 x g. The supernatant, containing mtDNA, was ethanol precipitated, treated with RNAse, extracted with phenol-chloroform, then ethanol precipitated a second time.

To obtain blue crab specific mtDNA sequences we first amplified a portion of the mtDNA *rrnL* using primers 16SARL [CGC CTG TTT ATC AAA AAC AT] and 16SB [CCG GTT GAA CTC AGA TCA]. We obtained a 572 base pair fragment that was cloned into the

pCR4 vector using a TOPO-TA Cloning Kit (Invitrogen, Carlsbad, CA) and sequenced in both directions using vector specific primers on an ABI 377 DNA Sequencer (Applied Biosystems, Foster City, CA).

After BLAST analysis confirmed the sequence's identity, we designed specific primers facing "out" for long range PCR to amplify the whole mtDNA: HPK 16Saa [ATG CTA CCT TTG CAC GGT CAA GAT ACC GCG GC] and HPK 16Sbb [CTT ATC AAA GGA AAA GTT TGC GAC CTC GAT GTT G]. The blue crab mtDNA was amplified in a one-step, long range PCR (Cheng, Higuchi and Stoneking, 1994) using the High Fidelity PCR Kit (Catalog # 1732641, Roche Diagnostics, Indianapolis, IN). PCR conditions were as detailed in the product literature, and briefly as follows: ~400 ng template DNA, 0.3 μM each of forward and reverse primers, 200 μM dNTPs, 1X Buffer 3 (from Kit), 2.2 mM $MgCl_2$, 2.6 U Expand High Fidelity Enzyme mix. The reactions were cycled at 92°C for 2 minutes, 10 cycles of 92°C for 10 seconds; 64°C for 30 seconds and 68°C for 13 minutes, followed by 20 cycles of 92°C for 10 seconds, 64°C for 30 seconds, and 68°C for 13 minutes + 20 seconds/cycle, then a final extension step at 68°C for 7 minutes.

2.2 Cloning and sequencing

Approximately 3 μg (Figure 1) of PCR product was sheared randomly into fragments of about 1.5 kb by forcing it repeatedly through a narrow aperture using a Hydroshear device (Gene Machines, San Carlos, CA). Following enzymatic end repair and gel purification, these fragments were ligated into pUC18 and transformed into *E. coli* to create plasmid libraries, all using standard techniques (Sambrook, et al. 1989). Q-Pix automated colony pickers (Genetix USA, Boston, MA) were used to select and transfer individual colonies into 384-well plates, forming the glycerol stock and providing the input for rolling circle amplification (RCA)

reactions. After overnight incubation, an aliquot was processed robotically through RCA amplification of plasmids, sequencing reactions using ET terminators (Amersham Bioscience, Piscataway, NJ), reaction clean up using SPRI, and electrophoretic separation on a Megabace 4000 automated DNA sequencer (Amersham Biosciences, Piscataway, NJ) to produce sequencing reads from each end of each plasmid. Sequences were fed automatically from sequencing instruments into a UNIX-based folder system, where they were processed using Phred, trimmed for quality, and assembled using Phrap. Quality scores were assigned automatically, and the electropherograms and assembly were viewed and verified using Sequencher (GeneCodes). A total of 700 sequencing reads averaging 736 bp in length (of quality Q20 or better) were assembled to yield a 15,924 bp contig. The long-range primer sequences were trimmed from the ends and this contig was merged with the intervening sequence of *rrnL* obtained earlier.

2.3 Sequence analysis and annotation

DNA sequences were analyzed using the computer software MacVector ver. 7.1.1 (Accelrys Inc.). Locations of the genes for the 13 proteins and two rRNAs were determined by comparisons to nucleotide or amino acid sequences of previously-determined complete mtDNA sequences from other crustaceans. The 22 tRNA genes were identified by their proposed cloverleaf secondary structures and anticodon sequences and corroborated when possible using tRNAscan-SE (Lowe and Eddy, 1997). Sequence data are available from GenBank under accession number **AY363392**.

2.4 Phylogenetic analysis

In addition to the sequences from *C. sapidus*, we included in our phylogenetic analysis mtDNA sequences from seven other crustaceans, brine shrimp (*Artemia franciscana*, GenBank

accession number X69067), water flea (*Daphnia pulex*, AF117817), copepod (*Tigriopus japonicus*, AB060648), giant tiger prawn (*Penaeus monodon*, AF217843), spiny lobster (*Panulirus japonicus,* AB071201), hermit crab (*Pagurus longicarpus*, AF150756), blue swimming crab (*Portunus trituberculatus*, AB093006), a chelicerate, the horseshoe crab (*Limulus polyphemus*, AF216203), and three insects, African malaria mosquito (*Anopheles gambiae*, L20934), fruit fly (*Drosophila yakuba*, X03240), and migratory locust (*Locusta migratoria*, X80245). The horseshoe crab and two branchiopod crustaceans (*A. franciscana* and *D. pulex*) were selected as collective outgroups with reference to the recent analyses of arthropod phylogenies (Wilson et al, 2000). Amino acid sequences from the individual protein-coding genes were aligned using ClustalW with default gap penalties. Ambiguous regions of the alignment, mainly the N- and C-termini of genes, were excluded from the alignments.

The concatenated amino acid sequences from the 13 protein-coding regions were subjected to maximum parsimony (MP), and neighbor-joining (NJ) analyses. All phylogenetically uninformative sites were ignored and gaps were considered as missing data. Branch-and-bound MP analyses were conducted with 1000 bootstrap replicate searches using PAUP4B10a (Swofford, 2001). A minimum evolution tree was estimated using the NJ method with 2000 bootstrap replicates using MEGA 2.1 (Kumar et al. 2001).

2.5 Nucleotide variation among individuals

In order to assess the nucleotide variation in and inheritance of the mitochondrial genome in the blue crab, a male and female pair were raised and mated entirely in captivity at the Aquaculture Research Center-II (ARC-II) at the Center of Marine Biotechnology. DNA was isolated from the father and mother as well as from 30 of the offspring from the first brood. The father was captured from the Rhode River, Maryland and the mother was hatched and raised in

captivity to maturity in ARC-II. Her mother had been captured from the York River, Virginia. The offspring were sampled as juveniles.

DNA samples were prepared from walking leg muscle using the FastDNA Kit (QBIOgene, Carlsbad, CA). The following primers were used to PCR amplify a 455 bp fragment (positions 14,859-15,314) of the putative control region of the blue crab mitochondrial genome: Forward AATCTCTTCAACTACGCC; Reverse TATGTGATGCGTCCTTGG. After TA-cloning, the nucleotide variation in a 273 bp sequence (15,015-15,288) was determined as above for *rrnL* and compared among parents and offspring as well as from unrelated individuals of the Chesapeake Bay and Mississippi.

3. Results and Discussion

3.1 *Genome Organization*

The mitochondrial genome of *C. sapidus* is 16,263 bp in length and contains the genes for the same 13 proteins, 22 tRNAs, and two rRNAs (Figure 2; Table 1) as found in other metazoans. Some of the genes overlap, as reported for other crustacean mtDNAs (Table 1). In addition, there is a 1,435 bp non-coding region, rich in A+T (>78%), between the *rrnS* and *trnI* genes, which corresponds to the common position of control region sequences for other arthropods. No notable reduction or extension of gene length compared to other decapods was observed (Table 2).

The overall A+T content of the *C. sapidus* mtDNA is 69.1%, similar to that of other decapods (Table 2). This value is not as high as some other arthropods, particularly insects (75.3 – 84.9%; see Crease, 1999). This pattern of base composition holds for the protein coding,

rRNA, and tRNA genes (Table 2) when considered separately . The A+T content of the control region is higher in *C. sapidus* than the other decapods except for the penaid shrimp (Table 2).

The gene order of *C. sapidus* (Figure 2) is nearly identical to that observed in most other arthropods, including having the translocation of *trnL(taa)*, from the arrangement *rrnL, trnL(taa), trnL(tag), nad1* found in *L. polyphemus* to the arrangement *cox1, -trnL(taa), cox2*. (This included its movement to the opposite strand, indicated by the minus symbol.) This translocation occurred at the base of an insect-crustacean clade and has been shown to be a strong indicator of the close relationship of these groups to the exclusion of the myriapods, chelicerates, tardigrades and onychoporans (Boore, et al. 1995; Boore and Brown 1998).

Generally, few rearrangements have been observed in arthropod mtDNAs and these have generally been translocations of only tRNA genes. The crustacean *A. fransiscana* has a translocation of *trnI-trnQ*, apparently derived for this lineage only, since another branchiopod, *D. pulex* has these genes in the same arrangement as *L. polyphemus* (Figure 2) and *D. yakuba*.

Both of the brachyuran species (*C. sapidus* and *P. trituberculatus*) share a derived translocation of *trnH* from between *nad4* and *nad5* to between *trnE* and *trnF* (the latter two tRNAs being located downstream of *nad5*; Figure 2). Although several mechanisms have been proposed for mitochondrial gene rearrangements (Levinson and Gutman, 1987; Moritz and Brown, 1987; Macey et al., 1998), one of the most widely-accepted mechanisms is tandem duplication of gene regions as a result of slipped-pair mispairing, followed by deletion of duplicated genes. Yamauchi et al (2003) has previously proposed this mechanism to account for the translocation of *trnH* for *P. trituberculatus* (Figure 3). The two deletion events appear to be incomplete in both *P. trituberculatus* and *C. sapidus* because two short, non-coding regions remain between *trnE* and *trnF* and between *nad5* and *nad4* (19 and 21 bp, respectively, for *P.*

trituberculatus and 19 and 16 bp, respectively, for *C. sapidus*), perhaps as vestiges of the presumed gene duplications.

3.2 *Protein coding genes*

Among the 13 protein-coding genes of *C. sapidus* mtDNA, there are two reading-frame overlaps on the same strand for the same pairs as observed in other crustaceans (*atp6* and *atp8* share four nucleotides; *nad4* and *nad4L* share seven nucleotides) (Table 1). Two protein encoding genes (*atp6* and *nad3*) start with variations previously seen for other animal mt genes (ATA and ATT, respectively), whereas the other eleven start with ATG (Table 1). Six *C. sapidus* genes end with TAA or TAG (*atp6*, *8*, *nad1*, *3*, *4*, *4L*), and the remaining genes have incomplete stop codons, either TA (*cox1*, *nad5*, *6*) or T (*nad2*, *cox1*, *2*, *cob*), presumably completed by polyadenylation (Ojala et al., 1981) as is common for animal mt genes.

In protein-coding genes, base composition is biased to nucleotides of A+T in both the 1st and 3rd codon positions (Table 3). Furthermore, remarkable differences in the C+G usage between strands is observed. These traits were also observed in *L. polyphemus* (Lavrov et al., 2000), *P. japonicus* (Yamauchi et al., 2002), and *P. trituberculatus* (Yamauchi et al., 2003).

3.3 *Transfer RNA genes*

C. sapidus mtDNA encodes 22 tRNA genes ranging in size from 63 to 73 nucleotides which can fold into clover-leaf secondary structures (Table 1). The *trnS(tct)* gene in *C. sapidus* lacks the DHU arm but this is commonly observed in metazoan mtDNAs (Wolstenholme, 1992).

The AAN codon family is two-fold degenerate. AA(C/U) specifies the amino acid N, carried by a tRNA with the anticodon GUU (G pairing with the C or U in the mRNA wobble position), whereas AA(A/G) specifies K, carried by a tRNA with the anticodon UUU (U pairing with the A or G in the mRNA wobble position. This is the case for *C. sapidus* and for all other

sampled crustaceans (Valverde et al., 1994; Crease, 1999; Hickerson and Cunningham, 2000; Wilson et al. 2000; Machida et al., 2002; Yamauchi et al., 2002, 2003). However, for all studied non-crustacean arthropod mtDNAs, *trnK* appears to use a CTT anticodon instead. How the anticodon C would pair with A in addition to G of the AA(A/G) codons is not clear, although it is possible that post-transcriptional tRNA modification occurs. A more full reconstruction of the molecular evolution of this system awaits further taxon sampling of mtDNA sequences and experimentation.

The only other *C. sapidus* anticodon nucleotides that differ from those of other arthropods is for *trnS(tct)*. Here, as for the other crustaceans, the anticodon is UCU instead of GCU as found for other arthropod mt tRNAs. UCU is the anticodon expected to best decode the four-fold degenerate AGN codon family for serine; it remains unclear how this is accomplished for arthropods with GCU anticodons for tRNA(S).

3.4 *Ribosomal RNA genes*

The *rrnS* and *rrnL* genes of *C. sapidus* are arranged as typically found for arthropod mtDNAs (Table 1, Figure 2). If we assume that the rRNA genes occupy all of the available spaces between adjacent genes, the gene lengths are 785 and 1323 bp for *rrnS* and *rrnL*, respectively (Table 1). The partial *rrnL* sequences previously deposited in GenBank (CSA298190 and CSU75269) for *C. sapidus* have very little sequence similarity to that presented here for the *rrnL* from the Mississippi specimen.

An alignment to *rrnS* of *Drosophila virilis* (X05914, 784 bp) and *D. pulex* (753 bp) shows 356 nucleotide identities, most being in the 3' part of the molecule, with 33 gaps introduced. Preliminary assessment of the potential secondary structure of *C. sapidus* rRNAs indicates that it could be reasonably superimposed (i.e., they have a similar location of helices)

on those proposed for *Drosophila virilis* (Cannone et al. 2002) and *Daphnia pulex* (Crease, 1999).

3.5 Phylogenetic Relationships

Maximum parsimony (MP) analysis of the amino acid sequences from the concatenated 13 protein-coding genes yields a single most parsimonious tree (Figure 4A), with a length of 11,319 steps (consistency index = 0.779; retention index =0.462; rescaled consistency index =0.359). Of the 3,813 sites, 948 are constant and 1,935 are parsimony informative. NJ analysis produced the same tree with similar bootstrap values (Figure 4B). All internal branches except for those between the spiny lobster and hermit crab (56 –65%) are supported by moderate to high (75 – 100%) bootstrap values. The close phylogenetic relationship between the portunid crabs is clearly evident in both analyses.

Among the brachyuran species, it appears that *atp8, nad2,* and *nad6* are diverging fastest while *cob*, *cox1*, and *atp8* are retaining a higher degree of identity (Table 4).

3.6 Allelic Variation

Since the major motivation for this work was to obtain sufficient sequence information to design primers to screen for nucleotide variation in the mitochondrial genome, we focused initially on screening the putative control region (positions 14,694-16,128), since this region is highly variable in many other organisms. A comparison of the sequence obtained from the original Mississippi female used to determine the entire mitochondrial genome sequence with that of randomly selected females from Mississippi and Virginia found a unique 24 bp insertion (5' CATTTATGTGCATATATATATATA-3'). We don't believe this is a PCR or sequencing artifact since more than ten clones covered this region in the sequence assembly. Moreover, other frequent insertion/deletions (indels) were observed between individuals in the control region. A

comparison of the sequence in the region from 15,015 to 15,288 among five offspring, the mother, and father, found all variation (including indels) to be maternally inherited (Figure 5).

4 Conclusions

The mitochondrial genome of *C. sapidus* is similar to other metazoans having 37 genes (13 proteins, 22 tRNAs, and 2 rRNAs) including a large (1,434 bp; 78.2 % AT) putatively hypervariable region. Gene order is identical to *P trituberculatus*, a portunid crab, with the *trnH* gene being translocated to a location between the *trnE* and *trnF* genes. The putative control region is highly variable (including indels) and individual variation is maternally inherited. Given these findings we can now begin to characterize the genetic structure of the blue crab females in the Chesapeake Bay and along the Eastern seaboard.

Acknowledgements

This is contribution number 04-597 from the University of Maryland Biotechnology Institute, Center of Marine Biotechnology. This work was funded by a grant from NOAA Award #NA17FU2841 Blue Crab Advanced Research Consortium Project to A.R.P. and by contract No. DE-AC03-76SF00098 between the U.S. Department of Energy, Office of Biological and Environmental Research, and the University of California, Lawrence Berkeley National Laboratory.

References

Avise, J. C., 2000. Phylogeography, Harvard Univ. Press, London, p 447.

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

Boore, J. L., and W. M. Brown. 1995. Complete sequence of the mitochondrial DNA of the annelid worm *Lumbricus terrestris.* Genetics 141:305-319.

Bowman, T. E., and Abele, L. G., 1982. Classification of recent Crustacea. In: Abele, L. G., (Ed.), The Biology of Crustacea, Vol. 1. Systematics, the Fossil Record, and Biogeography, Academic Press, New York, pp. 1-92.

Burton, R.S., and Feldman, M.W., 1982. Population genetics of coastal and estuarine invertebrates: does larval behavior influence population structure? Pp. 537-551 *in* V. S. Kennedy, ed. Estuarine Comparisons. Academic Press, New York.

Cannone, J. J., S. Subramanian, M. N. Schnare, J. R. Collett, L. M. D'Souza, Y. Du, B. Feng et al. 2002. The Comparative RNA Web (CRW) Site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs: Correction. BMC Bioinformatics 3:15.

Clary, D.O., and Wolstenholme, D.R., 1985. The mitochondrial DNA molecule of *Drosophila yakuba*: nucleotide sequence, gene organization and genetic codes. J. Mol. Evol. 2, 52-271.

Cole, M.A., and Morgan II, R.P., 1978. Genetic variation in two populations of the blue crab, *Callinectes sapidus*. Estuaries 1, 202-205.

Crease, T. J. 1999. The complete sequence of the mitochondrial genome of *Daphnia pulex* (Cladocera: Crustacea). Gene 233:89-99.

Cheng, S., Higuchi, R., and Stoneking, M., 1994. Complete mitochondrial genome amplification. Nat. Genet. 7(3), 350-351.

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

Lavrov, D. V., J. L. Boore, and W. M. Brown. 2000. The complete mitochondrial DNA sequence of the horseshoe crab *Limulus polyphemus*. Mol Biol Evol 17:813-824.

Levinson, G., and Gutman, G. A., 1987. Slipped-strand mispairing: a major mechanism for DNA sequence evolution. Mol. Biol. Evol., 4, 203-221.

Lowe, T.M. & Eddy, S.R. (1997) "tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence", Nucl. Acids Res., 25, 955-964.

Macey, J. R., J. A. Schulte, 2nd, A. Larson, and T. J. Papenfuss. 1998. Tandem duplication via light-strand synthesis may provide a precursor for mitochondrial genomic rearrangement. Mol Biol Evol 15:71-75.

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

McMillen-Jackson, A.L., Bert, T.M., and Steele, P., 1994. Population genetics of the blue crab *Callinectes sapidus*: modest population structuring in a background of high gene flow. Mar. Biol. 118, 53-65.

Moritz, C., and Brown, W. M., 1987. Tandem duplications in animal mitochondrial DNAs: variation in incidence and gene content among lizards. Proc. Natl. Acad. Scie, USA 84, 7183-7187.

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

Sambrook, J., Fritsch, E., and Maniatis, T., 1989. Molecular cloining: a laboratory manual. 2nd Edition. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Swofford, D. L., 2001. PAUP*: Phylogenetic Analysis Using Parsimony (*and Other Methods), version 4B10a Sinauer Associates, Sunderland, MA.

Sudhir Kumar, Koichiro Tamura, Ingrid B. Jakobsen, and Masatoshi Nei (2001) MEGA2: Molecular Evolutionary Genetics Analysis software, Bioinformatics Vol. 17, 12:1244-1245.

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

Warner, G. F., 1977. The Biology of Crabs, Elek Science, London, p 202

Wilson, K., V. Cahill, E. Ballment, and J. Benzie. 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 17:863-874.

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

Wuyts, J., Van de Peer, Y., Winkelmans, T., De Wachter R. (2002) The European database on small subunit ribosomal RNA. Nucleic Acids Res. **30,** 183-185.

Wuyts J., De Rijk P., Van de Peer Y., Winkelmans T., De Wachter R. (2001) The European Large Subunit Ribosomal RNA database. Nucleic Acids Res. 29(1):175-177

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

Yamauchi, M. M., M. U. Miya, and M. Nishida. 2003. Complete mitochondrial DNA sequence of the swimming crab, *Portunus trituberculatus* (Crustacea: Decapoda: Brachyura). Gene 311:129-135.

Table 1
Location of genes in the mitochondrial genome of the blue crab *C. sapidus*

Features	Position number		Size (bp)[a]	Protein initiation and termination		tRNA anticodon	Intergenic nucleotides[b]
	From	To		Start	Stop		
trnM	1	69	69			CAT	0
nad2	70	1075	1006	ATG	T-		0
trnW	1076	1142	67			TCA	-1
trnC	1142	1209	68(L)			GCA	0
trnY	1210	1275	66(L)			GTA	0
cox1	1276	2809	1534	ATG	T-		0
trnL(taa)	2810	2874	65			TAA	13
cox2	2888	3572	685	ATG	T-		0
trnK	3573	3640	68			TTT	0
trnD	3641	3705	65			GTC	0
atp8	3706	3867	162	ATG	TAG		-4
atp6	3864	4538	675	ATA	TAA		-1
cox3	4538	5328	791	ATG	TA-		0
trnG	5329	5391	63			TCC	0
nad3	5392	5745	354	ATT	TAA		3
trnA	5749	5812	64			TGC	3
trnR	5816	5878	63			TCG	1
trnN	5880	5947	68			GTT	2
trnS(tct)	5950	6016	67			TCT	0
trnE	6017	6084	68			TTC	19
trnH	6104	6166	63(L)			GTG	1
trnF	6168	6232	65(L)			GAA	0
nad5	6233	7959	1727(L)	ATG	TA-		16
nad4	7976	9310	1335(L)	ATG	TAA		-7
nad4l	9304	9606	303(L)	ATG	TAA		2
trnT	9609	9673	65			TGT	0
trnP	9674	9739	66(L)			TGG	2
nad6	9742	10247	506	ATG	TA-		0
cob	10248	11382	1135	ATG	T-		0
trnS(tga)	11383	11449	67			TGA	34
nad1	11484	12421	938(L)	ATG	TAA		22
trnL(tag)	12444	12512	69(L)			TAG	0
rrnL	12513	13835	1323(L)				0
trnV	13836	13908	73(L)			TAC	0
rrnS	13909	14693	785(L)				0
CR	14694	16128	1435				0
trnI	16129	16196	68			GAT	-3
trnQ	16194	16262	69(L)			TTG	

[a] (L) denotes that the gene is encoded on the L-strand.

[b] Number of nucleotides separating from the downstream gene. Negative numbers indicate overlapping nucleotides between adjacent genes.

Table 2
Characteristics of decapod mitochondrial DNAs

Species	L-strand		13 protein-coding genes		*rrnL*		*rrnS*		22 tRNA genes		Putative Control Region	
	Length (bp)	A + T (%)	Number of amino acids	A + T (%)	Length (bp)	A + T (%)	Length (bp)	A + T (%)	Length (bp)	A + T (%)	Length (bp)	A + T (%)
Callinectes sapidus	16,263	69.1	3712	67.0	1323	71.8	785	70.3	1463	71.6	1435	78.2
Portunus trituberculatus	16,026	70.6	3715	68.8	1332	73.8	840	70.1	1468	72.0	1104	76.3
Penaeus monodon	15,984	70.6	3716	69.3	1365	74.9	852	71.6	1494	68.0	991	81.5
Panulirus japonicus	15,717	64.5	3715	62.6	1355	69.2	855	67.1	1484	68.9	786	70.6
Pagurus longicarpus	------[a]	------[a]	3698	69.6	1303	77.1	789	77.2	1458	74.1	-----[a]	-----[a]

[a]15,630 bp was sequenced except approximately 300 of 850 bp putative control region (Hickerson and Cunningham, 2000)

Table 3
Base composition (%) of the 13 protein-coding genes for the mitochondrial genome of the blue crab *C. sapidus*.

	A	C	G	T
All genes				
1st	29.9	16.1	22.6	31.4
2nd	18.3	20.9	15.7	45.1
3rd	35.6	14.6	8.8	40.9
Total	27.9	17.2	15.7	39.1
Genes encoded on H-strand[a]				
1st	27.8	21.7	13.0	37.4
2nd	27.7	21.8	12.7	37.7
3rd	27.9	21.8	12.6	37.6
Total	27.8	21.8	12.8	37.6
Genes encoded on L-strand[b]				
1st	28.1	9.9	20.2	41.8
2nd	28.3	9.9	20.1	41.7
3rd	28.2	9.9	20.2	41.7
Total	28.2	9.9	20.2	41.7

Table 4
Percent amino acid identity (similarity) with *Callinectes sapidus*

Protein	*L. polyphemus*	*A. franciscana*	*P. longicarpus*	*P. trituberculatus*
ATP6	60 (14)	52 (19)	76 (12)	93 (3)
ATP8	33 (18)	25 (20)	37 (24)	72 (12)
COX1	80 (11)	80 (12)	90 (5)	98 (0)
COX2	70 (13)	57 (21)	83 (7)	92 (3)
COX3	68 (10)	62 (30)	83 (7)	95 (3)
COB	71 (12)	69 (14)	76 (12)	97 (1)
NAD1	57 (18)	51 (18)	74 (12)	89 (5)
NAD2	38 (25)	28 (20)	45 (19)	73 (14)
NAD3	54 (20)	45 (16)	72 (12)	89 (7)
NAD4	51 (16)	34 (17)	68 (17)	87 (7)
NAD4L	45 (15)	30 (17)	55 (19)	88 (4)
NAD5	46 (17)	36 (18)	61 (15)	88 (6)
NAD6	34 (20)	30 (22)	48 (16)	74 (10)

Figure Legends

Figure 1

Long-Range mtDNA PCR Product. Agarose gel showing the migration pattern of blue crab mtDNA amplicon resulting from HPK16Saa and HPK16Sbb primers in a one-step amplification of nearly the entire blue crab mtDNA. The left lane contains DNA standards (Lamda DNA/HindIII Fragments, Invitrogen/Life Technologies, Carlsbad, CA). The right lane contains the one-step mtDNA PCR product which is approximately 16 kb in length.

Figure 2

Linearized representation of the circular mitochondrial gene arrangements for the known Crustacea compared with that of an outgroup, the chelicerate *Limulus polyphemus*. All genes are transcribed from left to right except for those underlined to indicate opposite orientation. tRNA genes are designated by the corresponding single-letter amino acid code, except for the denotations L1, L2, S1, and S2, which indicate the genes *trnL(tag)*, *trnL(taa)*, *trnS(tct)*, and *trnS(tga)*, respectively. Asterisks indicate genes that are located differently from their positions in *L. polyphemus* mtDNA.

Figure 3

Proposed mechanism of the *trnH* translocation in *Callinectes sapidus* under a model of tandem duplication of gene regions and subsequent gene deletions (Yamauchi et al., 2003).

Figure 4

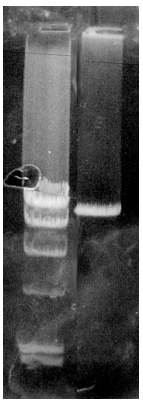
A single most-parsimonious (MP) tree (A) for the Crustacea taxa inferred from concatenated amino acid sequences of the 13 protein-coding genes (tree length = 11,319; consistency index = 0.779, retention index = .462; rescaled consistency index = 0.359) The neighbor-joining (NJ) tree (B) is identical in topology to that of the MP tree. Numbers beside internal branches indicate bootstrap values for each analysis obtained for 1000 (MP), and 2000 (NJ) replicates.

Figure 5

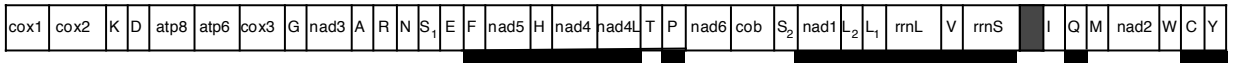
NJ-tree topology showing the genetic relationships among five offspring, mother, dad, and the Mississippi mitochondrial genome based on partial nucleotide sequences of the control region (positions 15,015-15,288).

23kb
9.4kb

4.3kb

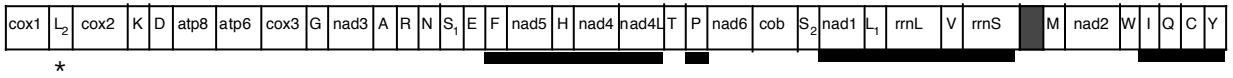


Limulus polyphemus (Chelicerata)



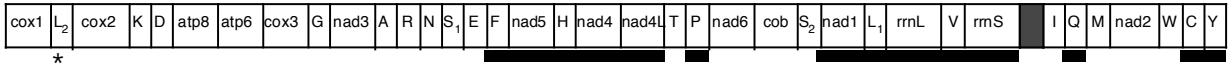
14,985 bp

Artemia franciscana (Crustacea, Branchiopoda)



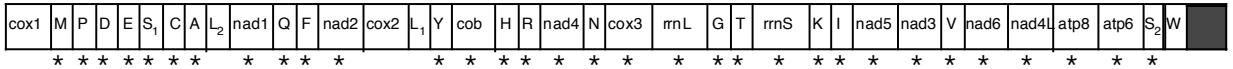
15,822 bp

Daphnia pulex (Crustacea, Branchiopoda)



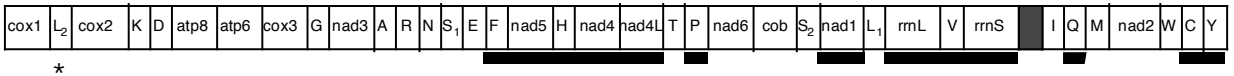
15,333 bp

Tigriopus japonicus (Crustacea, Maxillopoda)



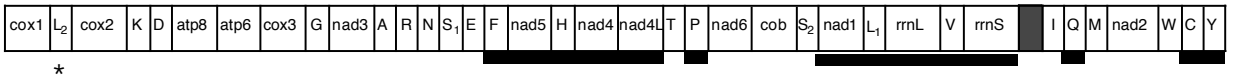
14,627 bp

Panulirus japonicus (Crustacea, Malacostraca)



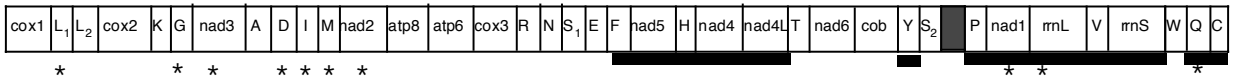
15,716 bp

Penaeus monodon (Crustacea, Malacostraca)



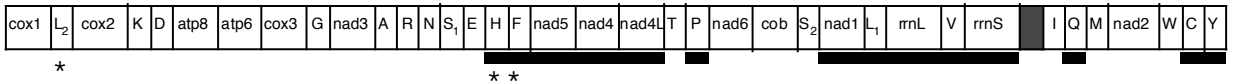
15,984 bp

Pagurus longicarpus (Crustacea,Malacostraca)



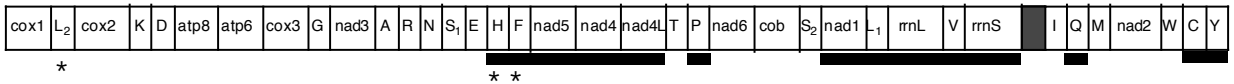
15,630 bp

Portunus trituberculatus (Crustacea, Malacostraca)



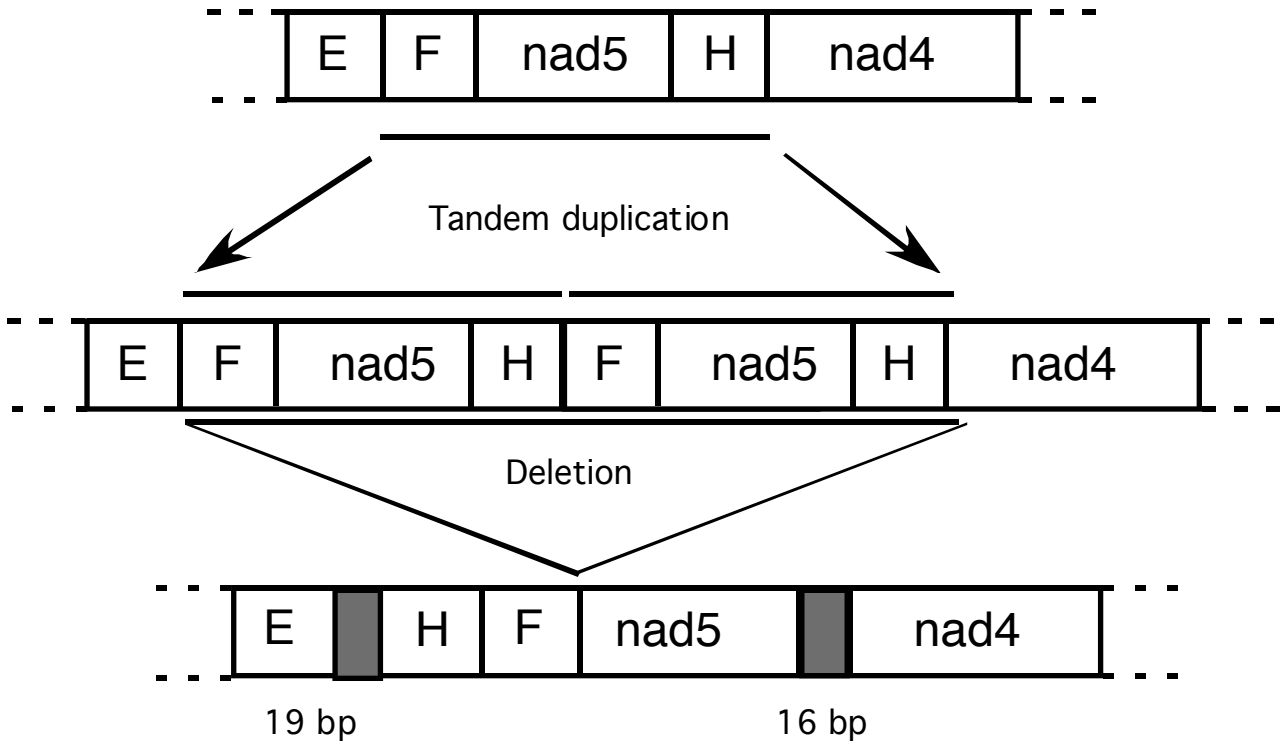
16,026 bp

Callinectes sapidus (Crustacea, Malacostraca)



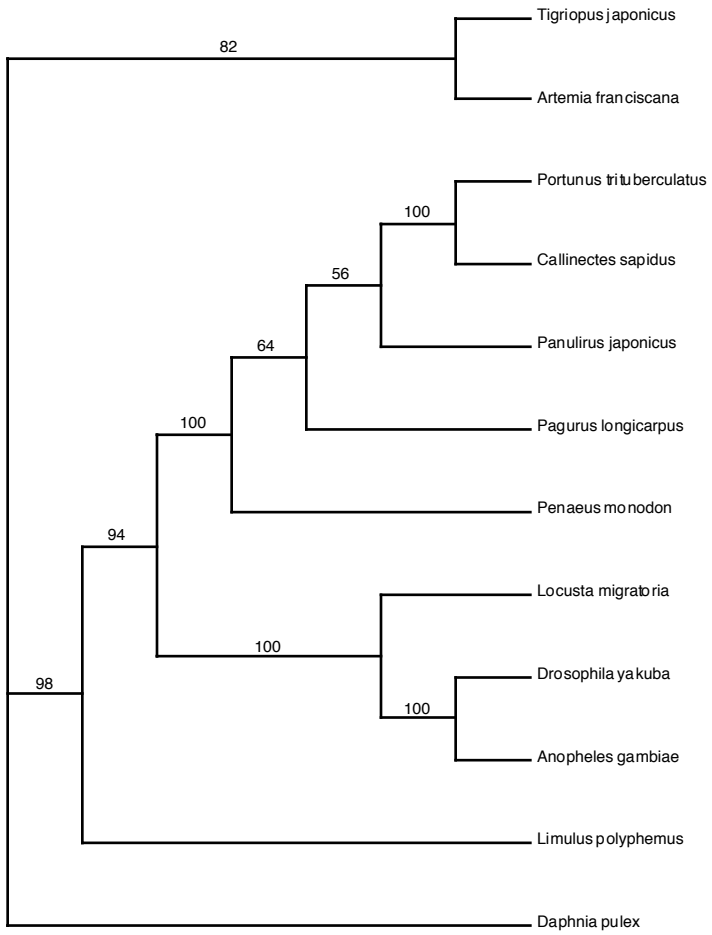
16,263 bp

1

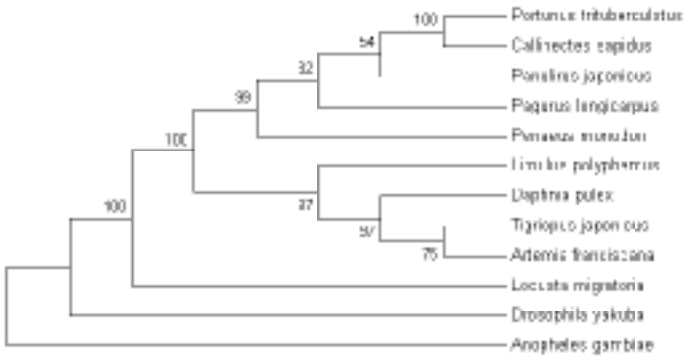

E
F
nad5
H
nad4
Tandem duplication
E
F
nad5
H
F
nad5
H
nad4
Deletion
E
H
F
nad5
nad4
19 bp
16 bp

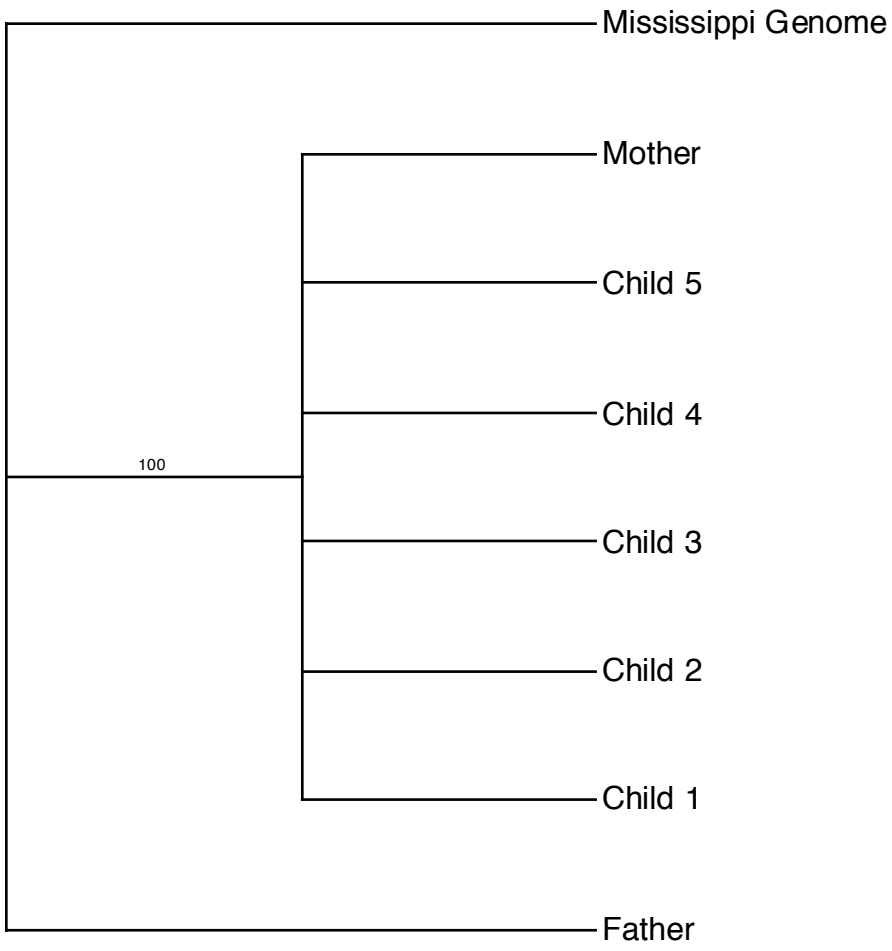
2

A


Tigriopus japonicus
Artemia franciscana
Portunus trituberculatus
Callinectes sapidus
Panulirus japonicus
Pagurus longicarpus
Penaeus monodon
Locusta migratoria
Drosophila yakuba
Anopheles gambiae
Limulus polyphemus
Daphnia pulex
82
100
56
64
100
94
100
98
100

B


Portunus trituberculatus
Callinectes sapidus
Panulirus japonicus
Pagurus longicarpus
Penaeus monodon
Limulus polyphemus
Daphnia pulex
Tigriopus japonicus
Artemia franciscana
Locusta migratoria
Drosophila yakuba
Anopheles gambiae
100
54
32
99
100
100
77
57
75


Mississippi Genome
Mother
Child 5
Child 4
100
Child 3
Child 2
Child 1
Father