Volume 16 Number 14 1988 Nucleic Acids Research

Genome organization of *Artemia* mitochondrial DNA

Beatriz Batuecas, Rafael Garesse, Manuel Calleja, José Ramón Valverde and Roberto Marco

Departamento de Bioquímica de la UAM and Instituto de Investigaciones Biomédicas de CSIC, Facultad de Medicina, Universidad Autónoma de Madrid, C/Arzobispo Morcillo 4, Madrid 28029, Spain

Received March 5, 1988 Accession nos: see footnote

ABSTRACT

To extend to the crustacean class the information concerning the genomic organization of the mitochondrial DNA (mtDNA) a double strategy has been used: Southern blot analysis with cloned *Drosophila* mtDNA probes and sequence comparison to the *Drosophila* mtDNA of the sequenced termini of different subclones along the *Artemia* mitochondrial genome, probably the smallest mtDNA studied at this level to date. These approaches have allowed us to localize the 16S rRNA gene, two tRNA genes and eleven protein genes. The genome organization is surprisingly similar to the *Drosophila* mtDNA, with the 16S rRNA and the protein genes located in the same positions and orientations as their *Drosophila* counterparts. The only changes detected are at the level of tRNA genes, although the position and orientation of some of these are also conserved. These results contrast with the important rearrangements detected among other invertebrates mtDNAs and suggest that the genome organization of the mitochondrial DNA may be more conserved in the arthropods than in other invertebrate phyla.

INTRODUCTION

In contrast to the multiple forms and much larger size of mitochondrial DNA (mtDNA) from fungi, plants and protists, animal mtDNA consists of a circular double stranded molecule ranging in size from 15 to 23 Kb (1), with no introns and an exceptionally compact organization with few and very short intergenic spacing between genes. Most of the animal mtDNAs sequenced so far (2-7) code for two ribosomal RNAs, 22 tRNAs and a group of 13 polypeptide. These proteins are all components of the electron transfer chain of the inner mitochondrial membrane: three subunits of the cytochrome oxidase (CO I, II and III), two subunits of the ATPase (6 and 8), seven subunits of the NADH oxidoreductase (ND1, 2, 3, 4, 4L, 5 and 6) and the cytochrome b (cyt b). The organization of the vertebrate mitochondrial genome is highly conserved, with the gene order and orientation strictly maintained from *Xenopus* to man (2-5). In the invertebrate phyla, the situation is different. Although the gene content is conserved, except in the case of the nematodes *Caenorabditis elegans* and *Ascaris suum* which lack the ATPase 8 gene (8,9), the genome organization is different in practically every species examined up to date (6-11), with the genes changing in position and orientation. All these variations should reflect the evolution of animal mitochondria, and can be explained as a successive series of reorganization events that have taken place during animal evolution (12). This "reshuffling" phenomena of the genes is especially remarkable in the case of the tRNA genes

(12), and is in agreement with the high rate of evolution and extreme adaptability of this molecule (9,13). The genetic code has also suffered important variations. In addition to the changes observed in vertebrate mtDNA (2-5), namely, TGA specifying tryptophan, AUA specifiying methionine and ATN used as initiation codon, in most invertebrates mtDNA AGA(G) specifies serine (6,10,11) instead of arginine (standard code) or termination (vertebrate mtDNA code). In starfish, another deviation has been recently reported with AAA specifying asparagine instead of lysine (11).

In order to examine the degree of change in the organization of the mitochondrial genome throughout the animal kingdom, it was particularly important to obtain more molecular information from other animals, specially invertebrates. In particular, in the class Crustacea, very few studies about the characteristics of mtDNA have been reported to date (14). We have recently initiated the characterization of the mtDNA from the brine shrimp *Artemia*, cloning the entire molecule in the vector pBR322 and constructing its physical map. In this paper we present the localization of the 16S rRNA, eleven of the protein genes and two tRNA genes. With these data it is possible to construct a genetic map showing the organization of the major part of the *Artemia* mtDNA molecule. A preliminary and partial report of this work has been published (15).

MATERIAL AND METHODS.

Mitochondrial DNA subcloning. Brine shrimp cysts were originally obtained from Metaframe Corporation, San Francisco Bay. *Artemia* mtDNA was prepared by the procedure of Schmitt et al (14), from cysts developed during 12-14 hr at 30ºC.The cysts were homogenized under disruptive conditions (16) to isolate mitochondria in high yields. The homogenates were centrifuged at 2,000 g x 10 min to sediment nuclei and the remaining yolk granules. The supernatant was centrifuged at 10,500 g x 30 min to prepare a crude mitochondrial fraction. This fraction was further purified through a discontinuous step gradient to recover the free mitochondria at the appropriate density. After a short DNAse I treatment, to digest contaminating external DNA, the mitochondria were washed and lysed. The supercoiled mtDNA was finally purified in an ethidium bromide CsCl gradient. Recombinant plasmids pMA2 and pMA3 contain the BamH I, 9.6 and 5.7 Kb, fragments of *Artemia* mt DNA, respectively, cloned in the BamH I site of pBR322. They cover the entire mitochondrial genome and were used for subcloning the smaller EcoR I, Hind III, Hind III-BamH I, EcoR I-BamH I and Ava I-Hind III fragments.

Digestion with restriction enzymes was carried out as indicated by the manufacturers. DNA was electrophoresed on 0.7 to 1.2% agarose gels in TBE buffer (Tris 100 mM, Boric Acid 100 mM, EDTA 2 mM, pH 8). To purify specific fragments we have used electroelution with DEAE paper (17). The purified fragments were ligated to the appropriate vector restricted with the same enzyme in the presence of T_4 DNA ligase (Amersham). EcoR I and Hind III fragments were subcloned in mp10 or mp11 M13 RF using standard conditions (18) and an E. coli JM 101 host. Several recombinants were randomly selected from each ligation in order to obtain both orientations of the insert with respect to the M13 universal primer. The BamH I-Hind III, Eco R I-BamH I or Ava I-Hind III purified fragments were directly cloned in both orientations in mp10 and mp11 M13 RF vectors. Single stranded DNA was prepared from M13 recombinant plasmids according to Messing (18). Supercoiled plasmids or M13 RFs were prepared using the alkaline procedure as described (19).

Heterologous *Drosophila* mtDNA probes. We have used four different specific cloned fragments of *Drosophila* mtDNA previously described: B, M.3, 62F9 (7) and M 2/8 (20) The clone B contains a 5386 bp fragment including the DNA for cytb, ND4, ND4L and most of ND5

and ND1. The clone M.3 contains a 1.7 Kb fragment encoding the 3' ends of ND5 and CO III genes and the complete ND3 gene. The clone 62F9 contains a 1.6 Kb fragment encoding the 3' end of the 16S rRNA gene and the 5' end of the ND1 gene. The clone M 2/8 contains a 4869 bp fragment encoding the CO I, CO II, ATPase 8 and ATPase 6 genes, most of the CO III gene and the 3' end of the ND2 gene. In addition, the four clones contain several tRNA genes.

Blotting and Hybridization. The DNA fragments generated in the digestion of pMA2 and pMA3 clones with restriction enzymes were separated in agarose gels and blotted onto nitrocellulose filters (21). Nick translated DNA probes (specific activities > 10^7 cpm/μg) were hybridized in 4 x SSC, 0,1% SDS, 1 x Denhart solution, 50% formamide and 100 mg/ml denatured salmon sperm DNA at 42ºC for 48 hours.

Sequencing of recombinant clones. We have obtained the sequences of both ends of the different recombinant clones using the dideoxy termination method of Sanger (22) with the 17-mer universal primer and [α-^{32}P] dATP (> 400 Ci/mmol) or [α-^{35}S thio] dATP (>400 Ci/mmol) obtained from Amersham. Electrophoresis and autoradiography development was performed using standard conditions (23-24). Genes within the different sequences were identified by direct comparison of both nucleotide and amino acid predicted *Artemia* sequences to the complete *Drosophila melanogaster* sequences (7,20). The comparison were carried out in a VAX/VMS computer using programs developed by Staden (25-26).

RESULTS

Southern Blot Analysis

As an initial step to characterize the *Artemia* mtDNA we have cloned the entire molecule derived from San Francisco brine shrimp in two BamH I recombinants plasmids (pMA2 and

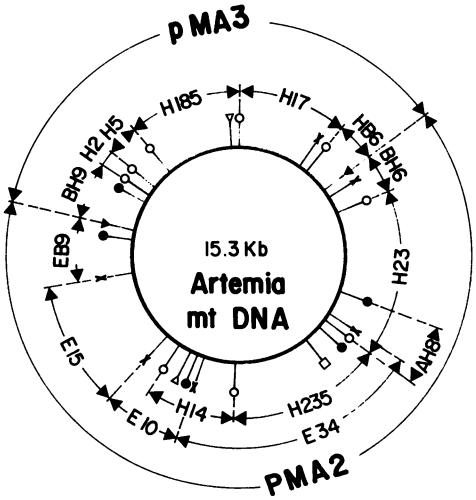


Fig. 1 - *Artemia* mt DNA Restriction Map and Subcloning strategy.

pMA2 and pMA3 correspond to two BamH I fragments cloned in pBR322 as described in Materials and Methods. These BamH I clones have been used for subcloning different smaller fragments. The limits of the subclones obtained are indicated, the name corresponding to the initial(s) of the restriction enzyme(s) used and a number indicative of its size (see Table 1). (⚲) Hind III: (⚸) EcoR I; (⚲) Ava I; (⚲) Pvu II; (⚲) Pst I; (⚲) BamH .

TABLE 1- Restriction fragments of *Artemia* mtDNA

Restriction enzyme	Sal I	Sma I	Pvu II	Pst I	BamH I	Ava I	EcoR I	Hind III
	- - -	- - -	15.3	8.9	9.6	6.0	5.8	3.8
				6.4	5.6	3.3	3.4	2.35
						3.25	2.7	2.3
						0.9*	1.5	1.85
						0.8	1.0	1.7
						0.2	0.9	1.4
								1.2
								0.5
								0.2
Total (kb)			15.3	15.3	15.2	15.35	15.3	15.3

* Double bands, as deduced by the increased Ethidium fluorescence.

pMA3). The physical map of the molecule using different restriction enzymes is schematically shown in Figure 1. It has no restriction site for Sal I and Sma I, a single site for Pvu II, two sites for Pst I and BamH I and several sites for Ava I, Hind III and EcoR I. The restriction sites are scattered around the molecule yielding multiple fragments ranging in size from 0.2 to 9.6 Kb (Table 1). Adding the fragment sizes generated in the digestion with each enzyme, the *Artemia* mitochondrial genome appear to be 15.3 $\pm$ 0.1 Kb, one of the smaller mtDNAs described to date (1).

Our initial approach to localize the genes in the *Artemia* mtDNA was to hybridize with specific probes from *Drosophila* mtDNA the pMA2 and pMA3 clones restricted with several enzymes (see Material and Methods) In Figure 2, the results obtained using the clones 62F9 and M2/8 are shown. The 62F9 clone allows the localization of the 16S rRNA gene in the 1.7 Kb Hind III and 0.6 Kb Hind III-BamH I fragments of the *Artemia* mtDNA. It hybridizes with the fragments 0.75 Kb EcoR I-BamH I, 0.6 Kb Hind III-BamH I and 1.7 Kb Hind III of the pMA3 clone. The 3.9 Kb Ava I fragment is not visible because this region of the gel was excised to eliminate the signal produced by the 4 Kb pBR322 fragment (see Legend of Figure 2). The 62F9 *Drosophila* clone contains the 3' end of the 16S rRNA gene, a region highly conserved in different species (3,7) and 350 bp of the poorly conserved ND1 gene (7).

The results obtained with the clone M 2/8 are also shown in Figure 2. This cloned fragment contains a series of genes well conserved among the mtDNAs of different species including invertebrates (3,6). Positive signals were detected in the 0.9 Kb BamH I-EcoR I, 1.5 Kb and 1 Kb EcoR I fragments of the pMA2 clone indicating that in *Artemia*, the genes CO I, CO II,

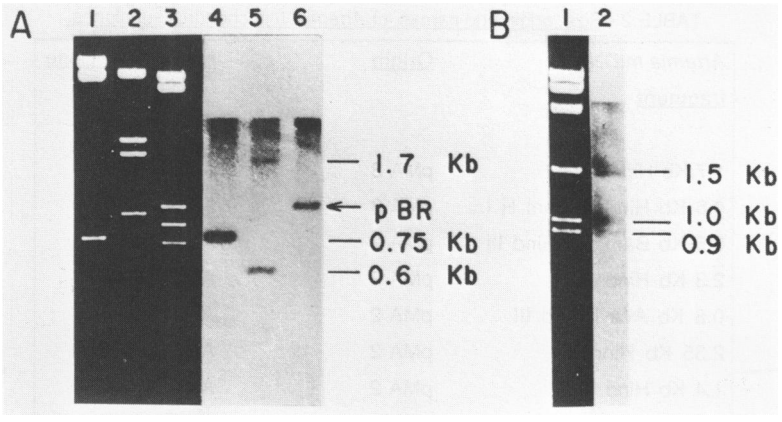


Fig. 2 - Southern Blot Analysis of Restricted pMA2 and pMA3 clones.

A- Southern blot containing pMA3 digested with BamH I-EcoR I (1), BamH I-Hind III (2), or BamH I- Ava I (3) was hybridized with the probe 62F9 (Materials and Methods). Lanes 1, 2 and 3 show the ethidium bromide stained agarose gel and lanes 4, 5 and 6 the autoradiography of the corresponding Southern blot. In addition to the pBR322 band (lane 6) that crosshybridize with the probe, a positive signal is detected from 0.75 Kb BamH I-EcoR I , 1.7 Kb Hind III and 0.6 Kb BamH I-Hind III fragments.

B- Southern blot containing pMA2 digested with BamH I and EcoR I was hybridized with the probe M 2/8 (Materials and Methods). Lane 1 shows the ethidium bromide stained agarose gel and lane 2 the autoradiography of the corresponding Southern blot. The fragments of 0.9 Kb, 1 Kb and 1.5 Kb show positive signals.

In both panels (A and B), the upper region of the gel was excised before transfer in order to eliminate the 4 Kb pBR322 fragment generated in the plasmid digestions.

ATPase 8 and ATPase 6 may be also contiguous and localized in the 2.8 Kb BamH I-Hind III region of the mitochondrial genome. On the other hand, with the less conserved regions contained in the B and M.3 *Drosophila* clones (7) we only obtained faint hybridization signals with different fragments located in *Artemia* mtDNA in between the previously identified genes.In order to overcome these difficulties, we decided to use a different approach, namely, subcloning fragments of the *Artemia* mtDNA covering the entire genome and sequencing the termini of the subclones to compare them with the *D. melanogaster* sequence (7, 20).

Subcloning of *Artemia* mtDNA

In the subcloning procedure, the pMA2 and pMA3 DNA was digested with the appropriated enzyme(s), and after isolation, the fragment of interest was ligated into M13 mp10 and/or M13 mp11 as described in Materials and Methods. We have subcloned 15 different fragments (Table 2) covering completely the *Artemia* mtDNA. We designate them AMSC (Artemia Mitochondrial SubClones), followed by the first letter(s) of the restriction site(s) used for cloning and a number indicative of its size (i.e., AMSC-H17 stands for 1.7 Kb Hind III fragment). The position of the clones along the molecule are schematically shown in Figure 1. They include

TABLE 2 - Size, origin and names of *Artemia* mitochondrial subclones.

Artemia mtDNA fragment	Origin	Name of the clone
1.7 Kb Hind III	pMA 3	AMSC-H17
0.6 Kb Hind III-Bam H I	pMA 3	AMSC-HB6
0.6 Kb Bam H I-Hind III	pMA 2	AMSC-BH6
2.3 Kb Hind III	pMA 2	AMSC-H23
0.8 Kb Ava I-Hind III	pMA 2	AMSC-AH 8
2.35 Kb Hind III	pMA 2	AMSC-H235
1.4 Kb Hind III	pMA 2	AMSC-H14
3.4 Kb EcoR I	pMA 2	AMSC-E34
1.0 Kb EcoR I	pMA 2	AMSC-E10
1.5 Kb EcoR I	pMA 2	AMSC-E15
0.9 Kb EcoR I-Bam H1	pMA 2	AMSC -EB9
0.9 Kb Bam H1-Hind III	pMA 3	AMSC-BH9
0.2 Kb Hind III	pMA 3	AMSC-H2
0.5 Kb Hind III	pMA 3	AMSC-H5
1.85 Kb Hind III	pMA 3	AMSC-H185

seven Hind III fragments, three EcoR I fragments, three BamH I-Hind III fragments, one Ava I-Hind III and one EcoR I-BamH I fragments (see Table 2).

Partial Sequencing of AMSC clones

We have obtained the nucleotide sequences from both ends of each clone using the dideoxy termination method as described in Materials and Methods. From these, the amino acid sequences in both orientations were deduced. Finally, the nucleotide and amino acid sequences were compared with the *D. melanogaster* mtDNA sequences (7,20). The results are shown in Figure 3, where the sequences are presented consecutively starting from the 1.7 Kb Hind III fragment. They range in size from 93 to 369 nucleotides. In some cases, a clear homology was detected with one or two genes encoded in the *Drosophila* mitochondrial genome (Table 3). At nucleotide (16S rRNA) or amino acid level (protein genes) this homology was high enough to identify by simple inspection of the data the genes in some of the subclones termini: AMSC-H17 (2) (69% homologous to 16S rRNA), AMSC-BH6 (H) (76% to ND1), AMSC-H23 (1) (50% to cytb), AMSC-H23 (2) (44,5% to ND4L), AMSC-H235 (1) (59.1% to ND4), AMSC-H14 (2) (74.8% to CO III), AMSC-E34 (2) (65,8% to ND3), AMSC-E10 (2) (68. 2% to ATPase 6) AMSC-EB9 (B) (85% to CO I). Three additional sequences, located in AMSC-H14 (1), AMSC-H5(1) and AMSC-AH8 (A) showed a low level of homology to ND5, ND2 and ND6 genes

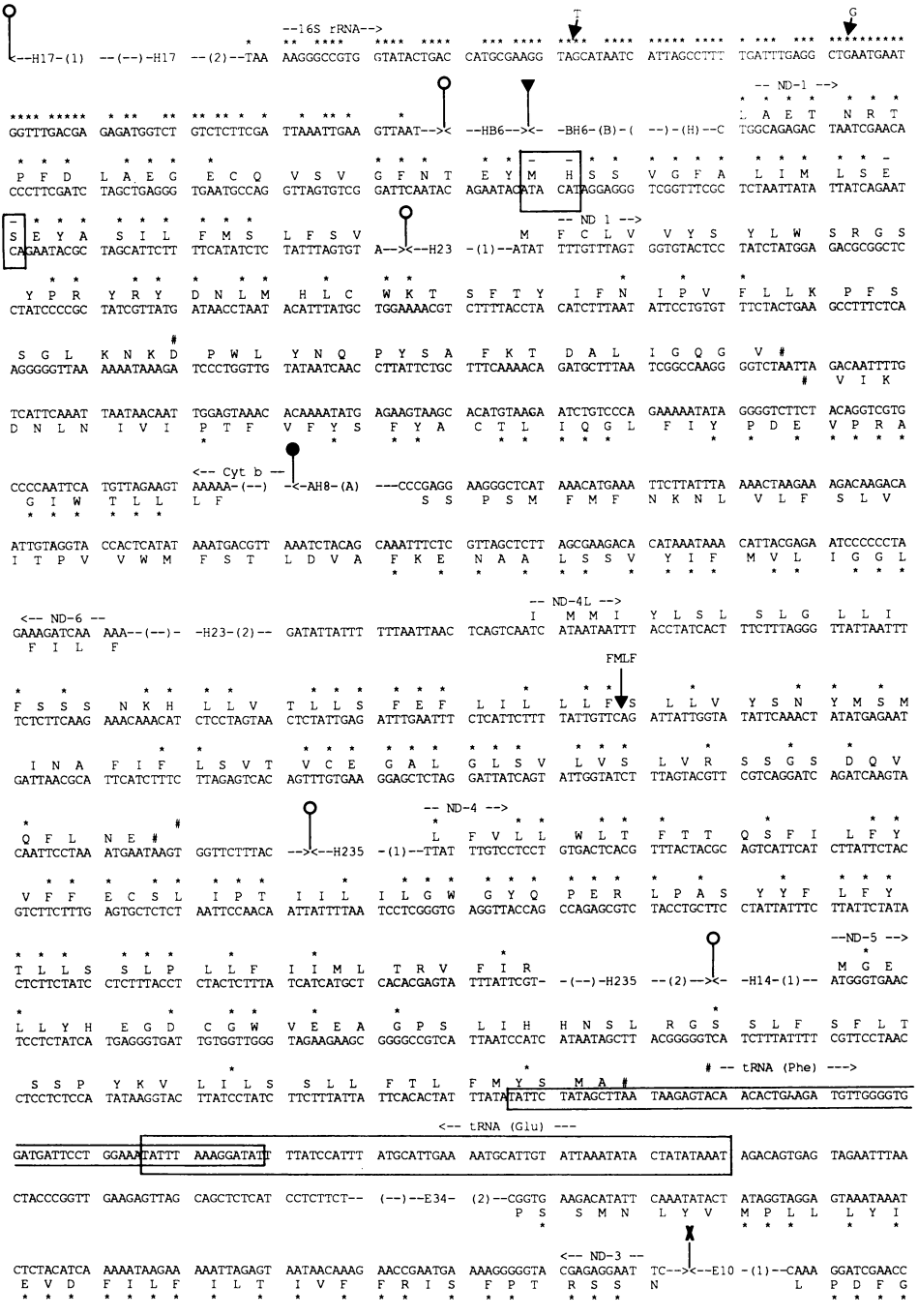
TABLE 3- Gene identification in *Artemia* subclones

Artemia# subclone	Length of the sequence (nucleotides)	Homologous *Drosophila* mt gene	Position in the *Drosophila* sequence (reference)	Percentage* of homology (amino acids)
AMSC-H17 (1)	120	- - - -	- - - - - -	- - - -
AMSC-H17 (2)	119	16S RNA	7941-8060 (7)	68**
AMSC-HB6 (H)	126	- - - - -	- - - - - -	- - - -
AMSC-HB6 (B)	- - -	- - - - -	- - - - - -	- - - -
AMSC-BH6 (B)	174	- - - - -	- - - - - -	- - - -
AMSC-BH6 (H)	162	ND-1	6662-6812 (7)	76
AMSC-H23 (1)	369	ND-1	6452-6617 (7)	31.5
		Cyt b	6227-6366 (7)	50
AMSC-H23 (2)	300	ND-4L	4280-4567 (7)	44.5
AMSC-AH8 (A)	170	ND-6	4865-5032 (7)	35.7
AMSC-H235 (1)	213	ND-4	3781-3992 (7)	59.1
AMSC-H235 (2)	211	- - - -	- - - - - -	- - - -
AMSC-H14 (1)	348	ND-5	1136-1318 (7)	18.6
AMSC-E34 (1)	- - -	- - - -	- - - - - -	- - - -
AMSC-E34 (2)	117	ND-3	485-598 (7)	65.8
AMSC-E10 (1)	93	ND-3	389-484 (7)	53.1
AMSC-H14 (2)	345	CO III	4717-4869 (20)	
			+ 6-201 (7)	74.8
AMSC-E10 (2)	132	ATPase 6	4259-4328 (20)	68.2
AMSC-E15 (1)	168	ATPase 6	4091-4258 (20)	58.6
AMSC-E15 (2)	231	- - - - -	- - - - - - - -	- - - -
AMSC-EB9 (E)	- - -	- - - - -	- - - - - - - -	- - - -
AMSC-EB9 (B)	175	COI	1743-1883 (20)	85
AMSC-BH9 (B)	- - -	- - - -	- - - - - - -	- - - -
AMSC-BH9 (H)	205	- - - -	- - - - - - -	- - - -
AMSC-H5 (2)	249	ND-2	20-268 (20)	30.1
AMSC-H5 (1)	161	ND-2	368-529 (20)	13
AMSC-H2 (1)	100	- - -	- - - - - - -	- - -
AMSC-H2 (2)	110	- - -	- - - - - - -	- - -
AMSC-H185 (1)	158	- - -	- - - - - - -	- - -
AMSC-H185 (2)	132	- - -	- - - - - - -	- - -

* excluding insertions/deletions
** homology at nucleotide level
# in parenthesis is indicated the sequenced terminus: 1 or 2 for the two ends of the fragment: H, Hind III site; A, Ava I site; B, BamH I site; E, EcoR I site.

respectively (Table 3). This is consistent with the poor conservation of the NADH dehydrogenase subunits among related species (3,7, 27).

There are additional reasons indicating that the detected homologies are correct, namely, the sequence of the end 2 of the AMSC-H14 clone contains the 3' end of a gene (Figure 3) encoding a presumptive protein 18.6% homologous to the carboxy terminal region of the ND5 *Drosophila* polypeptide. It contains a stretch of nine consecutive amino acids (D C G W V E E A G), well conserved (55%) when compared with a similar array of amino acids present in the *D.*


--16S rRNA-->
T
G
<--H17-(1) --(--)-H17 --(2)--TAA AAGGGCCGTG GTATACTGAC CATGCGAAGG TAGCATAATC ATTAGCCTTT TGATTTGAGG CTGAATGAAT
-- ND-1 -->
GGTTTGACGA GAGATGGTCT GTCTCTTCGA TTAAATTGAA GTTAAT--><- --HB6--><- -BH6-(B)-(--)-(H)--C TGGCAGAGAC TAATCGAACA
CCCTTCGATC TAGCTGAGGG TGAATGCCAG GTTAGTGTCG GATTCAATAC AGAATACATA CATAGGAGGG TCGGTTTCGC TCTAATTATA TTATCAGAAT
-- ND 1 -->
CAGAATACGC TAGCATTCTT TTCATATCTC TATTTAGTGT A--><--H23 -(1)--ATAT TTTGTTTAGT GGTGTACTCC TATCTATGGA GACGCGGCTC
CTATCCCCGC TATCGTTATG ATAACCTAAT ACATTTATGC TGGAAAACGT CTTTTACCTA CATCTTTAAT ATTCCTGTGT TTCTACTGAA GCCTTTCTCA
AGGGGGTTAA AAAATAAAGA TCCCTGGTTG TATAATCAAC CTTATTCTGC TTTCAAAACA GATGCTTTAA TCGGCCAAGG GGTCTAATTA GACAATTTTG
TCATTCAAAT TAATAACAAT TGGAGTAAAC ACAAAATATG AGAAGTAAGC ACATGTAAGA ATCTGTCCCA GAAAAATATA GGGGTCTTCT ACAGGTCGTG
CCCCAATTCA TGTTAGAAGT <-- Cyt b -- AAAAA-(--) -<-AH8-(A) ---CCCGAGG AAGGGCTCAT AAACATGAAA TTCTTATTTA AAACTAAGAA AGACAAGACA
ATTGTAGGTA CCACTCATAT AAATGACGTT AAATCTACAG CAAATTTCTC GTTAGCTCTT AGCGAAGACA CATAAATAAA CATTACGAGA ATCCCCCCTA
-- ND-4L -->
<-- ND-6 -- GAAAGATCAA AAA--(--)- -H23-(2)-- GATATTATTT TTTAATTAAC TCAGTCAATC ATAATAATTT ACCTATCACT TTCTTTAGGG TTATTAATTT
FMLF
TCTCTTCAAG AAACAAACAT CTCCTAGTAA CTCTATTGAG ATTTGAATTT CTCATTCTTT TATTGTTCAG ATTATTGGTA TATTCAAACT ATATGAGAAT
GATTAACGCA TTCATCTTTC TTAGAGTCAC AGTTTGTGAA GGAGCTCTAG GATTATCAGT ATTGGTATCT TTAGTACGTT CGTCAGGATC AGATCAAGTA
-- ND-4 -->
CAATTCCTAA ATGAATAAGT GGTTCTTTAC --><--H235 -(1)--TTAT TTGTCCTCCT GTGACTCACG TTTACTACGC AGTCATTCAT CTTATTCTAC
GTCTTCTTTG AGTGCTCTCT AATTCCAACA ATTATTTTAA TCCTCGGGTG AGGTTACCAG CCAGAGCGTC TACCTGCTTC CTATTATTTC TTATTCTATA
--ND-5 -->
CTCTTCTATC CTCTTTACCT CTACTCTTTA TCATCATGCT CACACGAGTA TTTATTCGT- -(--)-H235 --(2)--><- -H14-(1)-- ATGGGTGAAC
-- tRNA (Phe) --->
TCCTCTCCA TATAAGGTAC TTATCCTATC TTCTTTATTA TTCACACTAT TTATATATTC TATAGCTTAA TAAGAGTACA ACACTGAAGA TGTTGGGGTG
<-- tRNA (Glu) ---
GATGATTCCT GGAAATATTT AAAGGATATT TTATCCATTT ATGCATTGAA AATGCATTGT ATTAAATATA CTATATAAAT AGACAGTGAG TAGAATTTAA
CTACCCGGTT GAAGAGTTAG CAGCTCTCAT CCTCTTCT-- (--)--E34- (2)--CGGTG AAGACATATT CAAATATACT ATAGGTAGGA GTAAATAAAT
<-- ND-3 --
CTCTACATCA AAAATAAGAA AAATTAGAGT AATAACAAAG AACCGAATGA AAAGGGGGTA CGAGAGGAAT TC--><--E10 -(1)--CAAA GGATCGAACC

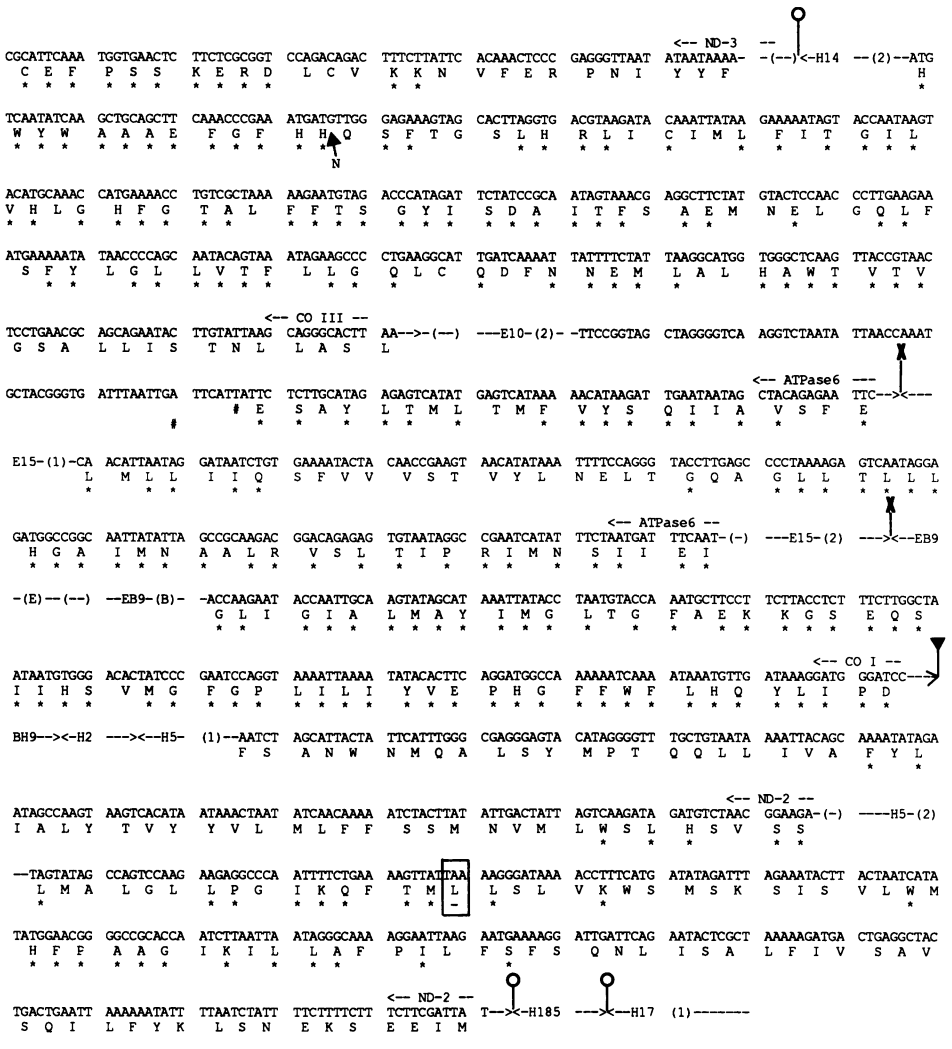


Fig. 3- *Artemia* mtDNA sequences of the subclones termini.

The sequences obtained from both ends of various *Artemia* mt DNA subclones are presented, starting from one end of the AMSC-H17 clone. They are shown in the same order as their position in the *Artemia* mitochondrial genome. The termini of the clones are indicated in brackets: E for EcoR I, H for Hind III, B for BamH I, A for Ava I ends and 1 and 2 for the two different ends of fragments bound by EcoR I or Hind III sites. The name of the genes are shown in arrows indicating the direction of the reading frame. The tRNA genes are shown in boxes The predicted sequences of the protein genes using one letter amino acid code are shown above the nucleotide letters when they correspond to the sense sequence or below when the sense sequence is on the opposite strand. The astericks represent identical amino acids in the same position when compared to the equivalent *Drosophila* gene sequence, # representing stop signals. The nucleotides inserted in *Artemia* mtDNA in protein, tRNA or 16S rRNA genes are underlined and boxed. The deletions in *Artemia* of nucleotides or amino acids are indicated with an arrow pointing to the position where the nucleotides are present in the *D. melanogaster* sequence. The limits of the clones are indicated with the same symbols used in Fig 1.

melanogaster sequence (7). This conserved region include two glycines and one tryptophan, the best conserved amino acids among the *D. melanogaster* and *D. yakuba* mtDNAs encoded proteins (7). Furthermore, overlapping 15 nucleotides with the end of the presumptive *Artemia* ND5 gene, the tRNA(phe) gene can be identified (Figure 3). Also overlapping 14 nucleotides with this tRNA gene, but encoded on the opposite strand the tRNA(glu) gene can be identified (Figure 2). Both tRNA genes are located in the *Drosophila* mtDNA contiguously to the ND-5 gene and in the same orientation as detected in the *Artemia* mt DNA (6,7). Considering the high degree of conservation of the gene order among *Artemia* and *Drosophila* detected in other regions of the molecule, it makes sense to conclude that the 3' region of the ND-5 gene is located at the end 2 of the AMSC-H14 clone (Table 3).

In the clone AMSC-H5, both ends contain sequences that can be translated in phase in a series of consecutive amino acids showing a low level of homology (13% and 30.1%) to two regions of the ND2 *Drosophila* protein (Figure 3). These two regions are separated by a number of nucleotides that fit quite well with the distance between the two ends of the AMSC-H5 clone. The central sequence of the end 2 of this clone (Figure 3) contains a stretch of 22 consecutive amino acids (S F S F L I P F A L L I K I G A A P F H M W) 68% homologous to the central region of the *Drosophila* ND2 protein. These data allow the localization of the ND2 protein in the 0,5 Kb Hind III fragment. Unfortunately, neither the clone AMSC-H185 nor the AMSC-H2 shows any homology to the ND2 or any other*Drosophila* gene, and consequently we have not been able to deduce the orientation of the ND2 gene in the *Artemia* mitochondrial genome. As mentioned before, this is not surprising considering the low level of homology detected between the *D. melanogaster* and *D. yakuba* ND2 genes (6).

Finally, the Ava I end of the clone AMSC-AH8 contains a sequence 36 % homologous at the amino acid level to the *Drosophila* ND6 protein. Although this percentage is not very high, the first 25 amino acids of the sequence are extremely well conserved (76%) when compared to a similar region of the ND6 *Drosophila* protein. Considering the relative poor conservation of the ND6 gene among related species (7,27), this result localize the ND6 gene in the 0.8 Kb Ava I - Hind III *Artemia* mitochondrial fragment.

We have not been able to detect homologies in the sequenced ends of some clones (Table 3). This is especially noteworthy in the case of end 2 of the AMSC-E15, a clone that lie contiguous to the 3' end of the CO I gene (Fig 3). In a similar position of the *Drosophila* mitochondrial genome is located the 5' end of the CO II gene, well conserved among related vertebrate and most invertebrate species (6,27). This result could be interpreted in one of two ways, either this region is poorly conserved between *Artemia* and *Drosophila* or in the *Artemia* genome there is a spacer region contiguous to the CO I gene.

Therefore, the hybridization and sequencing results allow us to localize 11 protein genes in the *Artemia* mitochondrial genome, namely, ND1, Cyt b, ND6, ND4L, ND4, ND5, ND3, CO I, CO III, ATPase 6 and ND2. In all the cases except one (ND2), the orientation of the genes can be

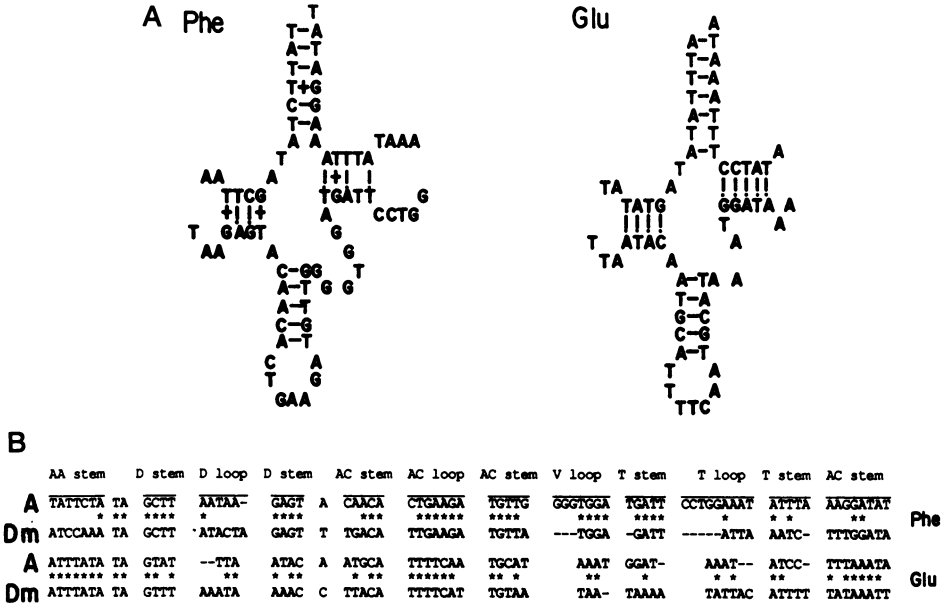


Fig. 4 - Structure and homology with *Drosophila* of the identified *Artemia* tRNAs genes.

A. The sequences of tRNA(phe) and tRNA(glu) are presented in their presumptive secondary cloverleaf structure.

B. The sequences of *Drosophila* (Dm) and *Artemia* (A) tRNAs genes are aligned with asterisks indicating positions of identity. Spaces have been introduced to maximize homology and the sequences are presented in blocks of the various stems and loops, as indicated.

directly deduced. There are still two protein genes that remain to be localized: CO II and ATPase 8. If we consider the size of the equivalent genes in other animal mtDNAs there are two possible locations for the CO II and ATPase 8 genes in the *Artemia* mitochondrial genome: one, between the ND-2 and 16S rRNA genes, and the other between the CO I and ATPase 6 genes. The latter option is more probable considering the high degree of conservation between *Artemia* and *Drosophila* in the organization of the other mtDNA genes. Furthermore, the gene order CO I - CO II - ATPase 8 - ATPase 6 - CO III is in principle conserved in vertebrate and most invertebrate mtDNAs. The results obtained in the Southern experiments also support this possibility. Thus, although it will be unequivocally confirmed when the sequence is completed, the more probable location for the CO II and ATPase 8 genes is in the middle of the 1.5 Kb EcoR I fragment.

Location of two transfer tRNA genes.

We have carried out a systematic search of tRNA genes in the *Artemia* mtDNA sequences reported in this paper using the computer program TRNA developed by Staden (28). Only the 3' upstream region of the ND5 gene contains sequences with a potential cloverleaf secondary structure (Figure 4). They have the anticodons GAA and TTC respectively, corresponding to the

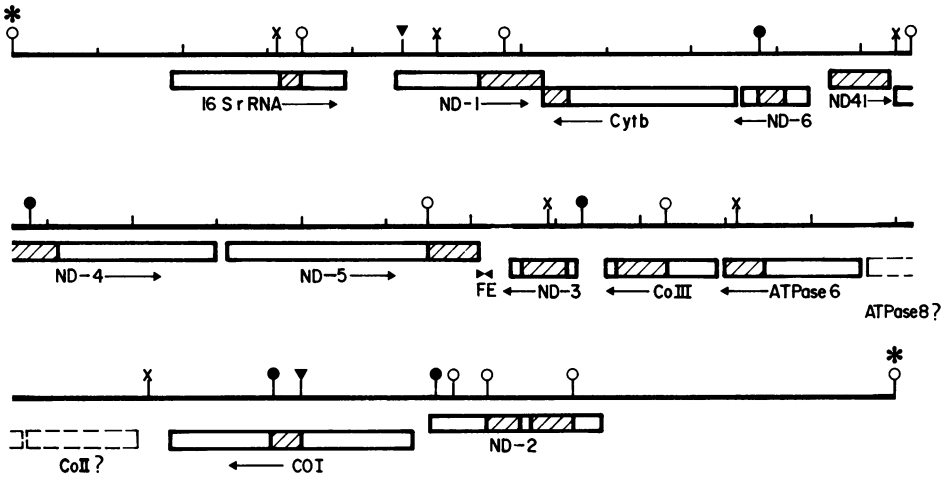


Fig. 5 - Present status of the knowledge about the genome organization of *Artemia* mtDNA.
The genome organization of *Artemia* mtDNA is schematically represented starting arbitrarily from one extreme of the 1.7 Kb Hind III fragment (•). The 16S rRNA and protein genes are boxed, with the sequenced region hatched. The presumptive limits of the genes are calculated considering a similar size to that of their *Drosophila* counterparts. When known, the orientation of the genes is shown (--->). The two identified tRNA genes correspond to phenylalanine (F) and glutamic acid (E) and are encoded on opposite strands. The tentative location of the ATPase 8 and CO II genes is also shown in broken line boxes.
( ), Hind III; ( ), EcoR I; ( ), BamH I; ( ) Ava I.

genes of tRNA(phe) and tRNA(glu). They strongly resemble their vertebrate and insect counterparts, lacking a number of features well conserved in the cytoplasmic tRNAs (2,5,29). The homology level when compared to the equivalent *D. melanogaster* genes is considerable, 67 % for the tRNA(phe) gene and 69 % for the tRNA(glu) gene. They showed several insertions/deletions, mainly in loops. The tRNA(phe) gene is 9 nucleotides larger in *Artemia* that in *Drosophila* (Fig 4). This gene contains in *Artemia* two exceptionally long V and T loops with 3 and 5 extra nucleotides respectively, due to a GGG insertion in the V loop and a CCTGG insertion in the T loop when compared to its *Drosophila* counterpart. As expected, the more conserved regions are the AC loop and D and AC stems, while the V, T and D loops are the more variable ones (Fig 4).

DISCUSSION

In this paper we present the characterization of *Artemia* mtDNA at the gene level.The data make it possible to establish the genomic organization of the molecule (Figure 5), indicating that it is remarkably similar to that present in *Drosophila* mtDNA. These results are the first data about the overall mtDNA organization in the whole Crustacean class.

The only genes that have changed their positions in relation to *Drosophila* are the tRNA genes. For example, the tRNA(ser ^{UCN}) gene is located between the cytb and ND-1 genes in *Drosophila*, probably acting as punctuation signal in the maturation of the transcripts as suggested for vertebrate and invertebrate mtDNAs (3, 7, 30). However in *Artemia*, the cytb and ND-1 genes are adjacent, without any spacing region between them. This situation is also found in other mtDNAs, where several alternative ways to specify translation have been proposed (3). However, two other tRNA genes identified in this work are located in the same position as their homologous *Drosophila* counterparts, the tRNA(phe) and tRNA(glu) genes, which are encoded 3' upstream of the ND-5 gene (Fig 5).

The level of homology between the *Artemia* and *Drosophila* protein gene regions at the amino acid level is variable, ranging between 13 % and 75% (Table 3). Since the sequence available is only partial, it does not allow definitive conclusions about the degree of conservation among the protein genes of the two genomes. However, there is a tendency to observe a poorer conservation in the NADH - dehydrogenase genes, as detected among similarly related species (7,27). In practically all the genes, several insertions/deletions have been detected when compared to their *Drosophila* counterparts. They are located at the 3' end of several genes (cytb, ND4L, ND5 and ATPase 6) and in the middle sequences of the ND1, ND4L, CO III and ND2 genes (Figure 3). In all cases, the insertions/deletions are shorter than 12 nucleotides.

Recently, the organization of a fragment of *Locusta migratoria* mtDNA has been reported (31). It contains the 3' end of the tRNA(leu ^{CUN}) gene, the ND1 gene, the tRNA (ser ^{UCN}) gene and the 3' end of the cyt b gene (31). They are arranged in the same order and orientation as reported for *Drosophila* (6,7), suggesting a high conservation in insects. The results presented in this paper extend the conservation to a different branch of the arthropod phylum. It has to be remembered that the arthropods are the more evolved phylum of the protostomian branch of evolution (32). Furthermore, the vertebrates and more particularly the mammals are also the more evolved class of the deuterostomian branch of evolution and these animals present a more or less frozen mitochondrial genome organization and genetic code. These considerations lead to the intriguing suggestion that the conservation detected in the arthropod phylum may indicate that the flexible mtDNA molecule is finally "fixed" maintaining gene organization and genetic code in the more recently and highly evolved organisms. This "fixed" organization would be different for the two branches of animal evolution.

As previously discussed, *Drosophila* mtDNA presents several changes in relation to the universal and vertebrate mitochondrial genetic code (6,20), with TGA specifying tryptophan instead of termination, AUA specifying methionine instead of isoleucine and AGA(G) specifying serine instead of termination (vertebrate mitochondrial code) or arginine (standard code). In the *Artemia* mitochondrial genes, TGA and AGA(G) triplets appear several times as internal codons. Nine out of eleven times, TGA corresponds to a tryptophan residue in the corresponding position of the *Drosophila* sequence. AGA appear 13 times, and six of them corresponds to a

serine in the equivalent *Drosophila* positions. Finally, AGG only appears once, and also correspond to a serine in *Drosophila* . These findings support the idea that *Artemia* and *Drosophila* mitochondrial DNAs use the same genetic code, in agreement with our suggestion of the high level of conservation of mtDNA in the arthropod phylum.

Finally, the size of *Artemia* mtDNA, as deduced by restriction enzyme analysis is 15.3 ± 0,1 Kb (Table 1), in the range of the smallest animal mitochondrial genomes described to date. This situation imposes additional constraints on the genome organization, and should have the consequence that intergenic regions would be shorter than in the majority of the animal species. According to this idea in *Paracentrotus lividus*, a species with a 15.5 Kb mtDNA genome, the intergenic region containing the origin of replication and the transcription signals, has a minimum size (400-500 bp) compared to other animal mt DNAs (9). The results presented here also suggest an extremely compact organization in *Artemia* mtDNA This is observed for example in the 3' end of the ND5 gene, that contain two tRNA genes overlapping several nucleotides with each other and with the ND5 gene (Fig 3). In order to complete the detailed genomic organization of the *Artemia* mtDNA additional sequencing work is now in progress.

Acknowledgments.

The work was possible through the economic support of grants from the Comisión Asesora de Investigación Científica y Técnica and Fondo de Investigaciones Sanitarias (Spain). The critical reading of the manuscript by Drs.L. Sastre and D. Mathog is gratefully acknowledged.

Accession nos: E34(2), X07664; E1(1), X07665; H14(2), X07666; E1(2), X07667; E15(1), X07668; EB9(B), X07669; H5(1), X07670; H5(2), X07671; H17(2), X07657; BH6(H), X07658; H23(1), X07659; AH8(A), X07660; H23(2), X07661; H235(1), X07662; H14(1), X07663.

REFERENCES

1. Wallace, D.C (1982). *Microbiol Rev.* **46**, 208-240.
2. Anderson, S., Bankier, A.T., Barrell, B.G., de Bruijn, M.H.L., Coulson, A.R., Drouin, J., Eperon, I.C., Nierlich, D.P., Roe, B.A., Sanger, F., Schreier, P.H. Smith, A.J.H., Staden, R. and Young, I.G. (1981). *Nature* **290**, 257-465
3. Anderson, A., de Bruijn, M.H.L., Coulson, A.R., Eperon, I.C., Sanger, F. and Young, I.G. (1982). *J. Mol. Biol.* **156**, 683-717.
4. Bibb, M.J., Van Etten, R.A., Wright, C.T., Walberg, M.W. and Clayton D.A. (1981) *Proc. Nat. Acad. Sci. USA* **78**, 3.789-3.793.
5. Roe, A.B., Ma, D.P., Wilson, R.K. and Wong, J.F.H. (1985). *J. Biol. Chem.* **260**, 979-9774.
6. Wolstenholme, D.R. and Clary D.O. (1985). *Genetics* **105**, 725-744
7. Garesse, R. (1988). *Genetics* (in the press).
8. Jacobs, H.T. and Lonsdale, D.H. (1987). *Trends in Genet.* **3**, 337-341.
9. Wolstenholme, D.R., MacFarlane, J.L.,Okimoto, R.,Clary, D.O. and Wahleitnher, D.A. (1987) *Proc.Nat.Acad. Sci. USA* **84**,1324-1328

10. Cantatore, P., Roberti, M., Morisco, P., Renaldi, G., Gadaleta, M.N. and Saccone, C. (1987). *Gene* **53**, 41-54.
11. Himeno, H., Masaki, H., Kawai, T., Ohta, T.,Kumagai, I., Miura, K.I. and Watanade, K. (1987) *Gene* **56**,219-230
12. Cantatore, P., Gadaleta, M.N., Roberti, M., Saccone, C. and Wilson, A.C. (1987). *Nature* **329**, 853-855.
13. Barrell, B.G., Bankier, A.T. and Drouin, J. (1979) *Nature* **282**,189-194
14. Schmitt, H., Beckman, J.S., and Littauer, U.Z. (1974). *Eur. J. Biochem.* **47**, 227-234.
15. Batuecas, B., Marco, R., Calleja, M. and Garesse, R. (1987) in *Artemia Research and its applications*. Decleir, W., Moens, L., Slegers, H., Jaspers, E. and Sorgeloos, P. (eds) vol. 2, Physiology. Biochemistry, Molecular Biology. Universa Press, Wetteren, Belgium, 355-367.
16. Vallejo, C.G., Perona, R., Garesse, R. and Marco, R. (1981). *Cell Differ.* **10**, 343-356.
17. Lizardi, P.M., Binder, R. and Short, S.A. (1984). *Gene Anal Techn.* **1**, 33-39.
18. Messing, J. (1982) in Wu, R., Grossman, L. and Moldave, K. (eds). *Methods in Enzymology* **101**, 20-77. Academia Press, New York.
19. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982). *"Molecular Cloning a Laboratory Manual"*. Cold Spring Harbor Laboratory. Cold Spring Harbor. New York.
20. De Bruijn, M.H.L. (1983). *Nature*, **304**, 234-241.
21. Southern, E.M. (1975). *J. Mol. Biol.* **98**, 503-517.
22. Sanger, F., Nicklen, S., and Coulson, A.R. (1977). *Proc Nat. Acad. Sci. USA* **74**, 5436-5467.
23. Sanger, F. and Coulson, A.R. (1978). *FEBS Letters* **87**, 197-210.
24. Biggins, M.D., Gibson, T.J. and Hong, G.F. (1983). *Proc. Nat. Acad. Sci. USA.* **50**, 3963-3965.
25. Staden, R. (1982). *Nucleic Acids Res.* **10**, 2.951-2.961.
26. Staden, R. (1984). *Nucleic Acids Res.* **12**, 521-538.
27. Brown, W.N., Prager, A., Wang, A. and Wilson, A.C. (1982). *J. Mol. Evol.* **18**, 225-239.
28. Staden, R. (1980). *Nucleic Acids Res* **8**, 817-825.
29. Clary, D.O. and Wolstenholme, D.R. (1985). *Nucleic Acids Res* **12**, 2.367-2.379.
30. Ojala, D., Montoya, J. and Attardi, G. (1981). *Nature* **290**, 470-474.
31. Mc Cracken, A., Uhlenbusch, I. and Gellisen, G. (1987). *Curr. Genet. 11*, 625-630.
32. Gardiner, M.S., (1972). *"The Biology of invertebrates"*. McGraw-Hill. New York.