

Mitochondrial DNA in the Sea Urchin *Arbacia lixula*: Evolutionary Inferences from Nucleotide Sequence Analysis¹

Carla De Giorgi, Cecilia Lanave, Maria Donata Musci,
and Cecilia Saccone

Dipartimento di Biochimica e Biologia Molecolare, Centro di Studio sui Mitochondri e Metabolismo Energetico (CNR), University of Bari

From the stirodont *Arbacia lixula* we determined the sequence of 5,127 nucleotides of mitochondrial DNA (mtDNA) encompassing 18 tRNAs, two complete coding genes, parts of three other coding genes, and part of the 12S ribosomal RNA (rRNA). The sequence confirms that the organization of mtDNA is conserved within echinoids. Furthermore, it underlines the following peculiar features of sea urchin mtDNA: the clustering of tRNAs, the short noncoding regulatory sequence, and the separation by the ND1 and ND2 genes of the two rRNA genes. Comparison with the orthologous sequences from the camarodont species *Paracentrotus lividus* and *Strongylocentrotus purpuratus* revealed that (1) echinoids have an extra piece on the amino terminus of the ND5 gene that is probably the remnant of an old leucine tRNA gene; (2) third-position codon nucleotide usage has diverged between *A. lixula* and the camarodont species to a significant extent, implying different directional mutational pressures; and (3) the stirodont-camarodont divergence occurred twice as long ago as did the *P. lividus*-*S. purpuratus* divergence.

Introduction

Sea urchins have been the subject of many molecular studies because of the knowledge of their embryological development (Giudice 1986). Their taxonomy, however, is very difficult to establish because of an insufficient fossil record which does not allow tracing of a reliable phylogenetic history.

The complete sequence of mitochondrial DNA (mtDNA) has already been determined in two camarodont sea urchins [*Paracentrotus lividus* (Cantatore et al. 1989) and *Strongylocentrotus purpuratus* (Jacobs et al. 1988)] and is under investigation in another camarodont species, *S. franciscanus* (W. Brown, personal communication). The other echinoderms whose mtDNA has been isolated and extensively sequenced are the sea stars (Asteroidea) *Asterina pectinifera* (Himeno et al. 1987) and *Pisaster ochraceus* (Jacobs et al. 1989; Smith et al. 1989), but their mtDNA sequences are too divergent from those of sea urchin to allow inferences about their evolutionary relationships. Accordingly, it was decided to study the genetic organization and the nucleotide sequence of the mtDNA in another sea urchin, *Arbacia lixula*, which belongs to the order Stirodonta, whose divergence from Camarodonta is still uncertain (Smith 1984).

In the present paper we report the nucleotide sequence, of two mtDNA fragments, totaling 5,127 nucleotides (nt), from *A. lixula*. These fragments comprise 18 tRNA

1. Key words: *Arbacia lixula*, sea urchin, evolution, mitochondrial DNA sequences.

Address for correspondence and reprints: Cecilia Saccone, 4 Trav. via ReDavid 200, 70126 Bari, Italy.

Mol. Biol. Evol. 8(3):515-529, 1991.

© 1991 by The University of Chicago. All rights reserved.

0737-4038/91/0803-0009\$02.00

a

→ COIII

A G N R T E A V Q A L F L T V L G I Y F T I L Q A W E Y Y D S P 100
GCAGGAAATCGGACAGAAGCAGTCCAAGCTTTATTCTTGACTGTTGCCTAGGGATTTATTTACTATACTTCAAGCATGAGAATATATGACTCTCCTT

F T I A D S V Y G S T F F V A T G F H G L H V I N S T T F L L V C L 200
TTACCATAGCAGATAGTGTCTACGGCTCTACTTTCTTCGTAGCGACAGGATTTTCATGGCCTTCACGTTATTAATAGAACAACATTTCTCTTAGTCTGCCT

F R L I N F H F S A H H H F G F E A A A W Y W D F V D V V V A F S 300
TTTTGACTTATTAATTCATTCTCAGCTCATCACCACCTTTGGATTTGAGGCAGCTGTTTGATATTGGGACTTCGTGGACGTAGTGGTGGCTTTTTTCT

L Y M H I W W G S * tRNA SER (UCN) ← 400
CTATATATGCATATATGATGAGGCTCCTAAGAGAGAAAGAGGAATTGAACCTCCATCAAACGGTTTCAAACCGCACGCATTACTATCTGCCACTTCTCCTA

→ ND3 500
M T S M T F L I I I T I I I A S I L G L A A H I L P S R N
ATTATATATAAGAAATGACCTCAATGACTTTCTGATAATTATTACAATTATAATAGCATCTATTCTTGGTTTGGCGGCACATATTTGGCCCTCCCGAAA

A D A E K S S P Y E C G F D P L N S A R L P F S F R F F L V A I L 600
AGCAGATGCCGAGAAGACTCCCCTTACGAATGTGGATTTGACCCCTTAAACTCTGCTCGCCTACCCTTTCTTTTCGATTCTTTCTGGTGGCAATATTA

F L L F D L E I A L L F P L P G A V L I T N P S N L I I A A T I F 700
TTTTTATTGTTTGATCTGGAAATAGCACTTCTATTTCCCTCCAGGAGCTGTTTTAATAACCAACCCAAGAAATCTTATTATAGCTGCTACTATATTTA

M F I L T F G L V F E W I N G G L E W A E * → ND4 800
TGTTTATCCTAACCTTTGGTTTAGTGTGGTGAATAAAAGGCGGATTAGAGTGAGCAGAATAAATCCATATAAAATGATTACTCTTATATTATTATAAG

I G M T L T S F V V S K N N L W P V T I L Q S A F L S I A A I A L 900
TATTGGTATGACTTTAAACAAGTTTTGTGTTTCCAAGAAAACCTTTGGCCTGTTACTATTCTACAAAGTGCCTTTTTGTCTATCGTGCATCGCACTT

I N N H W I S K W H N F I N S V D S M Q L P L L V L S C W L T P L 1000
ATAAATAAACACTGAATATCTAAGTGACATAACTTCATTAACCTCAGTGGACTCTATGCAACTACCCCTATTAGTATTAAGTTGTTGACTAACTCCCTTGG

A L I A S K G H L N N L P I T N Q R T F I V L V I I I T T S L I I T 1100
CTTTAATAGCTAGTAAGGGCCACCTAAAAAATTTACCAATAACAAAACAACGAACCTTTTATTGTATTAGTTATCATAATAACAACCTTCTCTTATAATAAC

TTTCAGCTCTCTTGAATTAATCTATTTTACATTGCGCTTGCAGAC1ACTCTTGTCTCTACTTTAATATTAATTACTCGTTGAGGTGCGCTGATGGAGCGT 1200

F Q A G L Y F I F Y T L F G S L P L L I S L I A L Y F S S N S L S
TTTCAAGCTGGATTATATTTTATATTTTATACTTATTGGCTCTCTTCTCTTCTAATTAGCCCTATTGCTCTTTATTTTCCAGTAAATCATTATCAA 1300

I P N V E L V W L T T N S S T S L T V W W L L S I L A F L V K M P I
TTCCAAACGTAGAGTTAGTATGGTTAACAACTAAAAGTAGAACTTCTCTTACTGTTTGATGACTTTTATCCATACTAGCCCTTCTAGTAAAGATGCCCAT 1400

Y G F H L W L P K A H V E A P V A G S M I L A A I L L K L G G Y G
ATATGGGTTTCATCTATGGTTGCCAAGGCCCATGTAGAGGCCCTGTAGCAGGCTCTATGATACTGGCAGCAATTTTGTTAAAGTTAGTGGATACGGA 1500

L M R L I S L F S T T S L N F S S L P L V V F C C W G A L V T S I
TTAATGCGCTTGATTAGTTTGTCTCACTACATCACTAAATTTTCTCTTTGCCGTTGGTAGTATTTTGTGTGGAGGCCCTTAGTAACCAAGTATA 1600

I C I R Q T D L K A L I A Y S S V G H M S I V A A G V F S Q T I W G
TATGTATCCGCCAAACAGATCTAAAGGCCCTTAATTGCTTATTTCATCAGTAGGCCACATGAGCATAGTTGCTGCTGGAGTCTTCTCCCAAACATCTGAGG 1700

I N G A L M L M I A H G L V S S A L F A L A N T M Y E R S G T R T
AATAAAAGGTCTTTAATGCTAATGATTGCCCATGGACTAGTCTCCTCAGCGTTATTTGCCCTTGCGAAAACATGTACGAGCGAAGAGGAACCCGAACC 1800

L V I T R G M K L I L P L S T F W W L I M C A A N L G F P Y S N L
TTAGTAATAACTCGAGGAATGAAGCTTATACTTCCACTGTCTACCTTCTGGTGGTTAATAATGTGTGCTGCAATCTCGGCTTCCCCTACTCTAAATTA 1900

I G E I L Y I S W Y G W S I W C S Y F S N Y N C V W G V Y S L M I F
TAGGAGAAATATTATATATCTCTTGATACGGATGATCTATTTGATGTTCTATTTTAGGAATTACAACTGTGTTTGGGGCGTATACTCTCTAATGATTTT 2000

Q V S Q Q G P S S H F L L N V P T S F S R E H L L F L L H L L P L
TCAGGTTTCTCAACAGGCCCTCTCTCATTTTCTGCTAAAAGTCCCTACCTCATTTTCTCGAGAGCACCTACTATTCTTATTACATCTACTACCTTT 2100

L L I I P T P N L V L I S → tRNA **W L K *** **HIS**
CTACTTATAATACCAACACCTAAGTTAGTTTAAATATCATGACTAAAGTAGTTTAGCAAATATCACGCTGTGGCACTGAAGACGCTAGTTAAACTCTAG 2200

517

FIG. 1.—Nucleotide sequence of sense strand of *Arbacia lixula* mtDNA fragments. Arrows indicate the direction of transcription of each gene. In the case of protein-coding genes, the predicted amino acid sequences are indicated by the single-letter code, and the stop codons are indicated by an asterisk. a, Sequence of 2,847-nt *EcoRI-HpaII* subfragment of largest *BamHI* fragment, containing carboxy-terminal end of COIII gene, Ser(UCN)-tRNA, ND3 and ND4 genes, His-tRNA and Ser(AGN)-tRNA genes, and amino-terminal end of ND5 gene. b, Sequence of 2,280 nt of small *BamHI* fragment containing 664 nt of 3' end of 12S rRNA gene, cluster of 15 tRNA genes, 136-nt-long noncoding region, and amino-terminal end of ND1 gene.

CCTTAAGTCCGAAGTCTATGTGTTTGTCTTAGTCCTGCTAAGTCTAAAGACTGCGGTTCAACTCCGTAGATGCTTCGATGGTTATAAGCCCTTCTACATT 2300
 L V S I T L S I I C L I V S I L Y T S K S F V A Q R N F L T S G N
 ACTTGTTC AATAACTCTCAGTATTATTTGTTTGATAGTAAGAATCCTTTACACTTCTAAGTCATTGTAGCTCAGCGAAATTTTAACTAGCGGTAAA 2400
 I A F S G A S L N I T S D G S A V Y S W T N G P F S I N I L K F L
 ATAGCCTTTTCAGGAGCAAGCTTAAATATAACTTCCGATGGAAGAGCTGTTTACTCTTGGACAAAAGGACCCTTTCTATAAAATATTCTTAAGTTTCTAG 2500
 A F L S L I N L F L F V G L E F Q E T N V T F S I W L S N T A A N V
 CCTTTCTTTCGTTAATAAAATTGTTTTGTTTGTGCGATTAGAATTTCAAGAAACAAACGTAACCTTCTCTATTTGACTAAGAAATACAGCTGCTAAAGT 2600
 S L S I L F D H Y F I V F L T V A L V V T W S I M N F S L L Y G E
 CTCATTAAGCATTATTTGACCACTATTTTATAGTTTTTTGACCGTTGCATTGGTTGTTACCTGATCAATAATGAATTTCTCATTATTATATGGAGAA 2700
 D P N K N V F L L L T I F L L N M L I L T C S N S L F L L F L G W
 GATCCTAAAAGAAAGTATTCCTCTTATTAACAATCTTCCTATTA AAAATGTTAATCCTCACTTGCTCTAAAAGACTATTCTTACTATTTTTAGGATGAG 2800
 E G V G F L S F L L I K M M N H
 AAGGAGTTGGTTTTTTATCTTTCCTTCTTATAAAGATGATGAACCAC 2847

FIG. 1 (Continued)

b

→ 12S rRNA

CCGATCCAGTAAACGAAACAAAGAAGGAGGTAAAATACGTGCCAGCCACCGCGTTAAACGTATGCCCTAAAGTAAAAGTAACTAGACCGGTGTAAGGGTGG 100

TTAGAATGTTACAGCCTAGCTATAGTTTTTATGATAGTGGTAGAAACTTACCTAAAATAAAATAGCTAGCTATACTTTGACCCACGAAAAGTAACTAGGCATA 200

AACCAGATTAGATACCCTGTTATACCTAGAAGTAAACAACCTATAAACACCAGAGAACTACTTAACTTAAAGTTAAACTCAAGGACTTGCCTTTCCAACC 300

TCCTGAGGAGCTTGCCTCGGATCGATAATCCACGCTATACCTTACCGATTCTTGCCTTCACAGCTTATACATCTGCGAAATTTCTTCTAAGAAAGTTA 400

ACACAAAGAGTCAATCTCTTCACATCAGATCAAGTGCAGCTTATGAATCGGGGACAGGTGAGCTACAATGTTAATAAAAACCAGTGAATAAGGGTTGAA 500

AATAGTCCTTAAGAAATTGGATTCAGCAGTAAGCCCCAGAAGAGAATGGGGCTGAAGAAAGCTCTGAATGCGTACACATCGCCCCTCACTCTCGCCTAGC 600

CAAAGGTAAGGGGAGAAAAGTCGTAACATAGTAGGTATACCGGAAGGTGCACCTGGAAAATGTTCTCTGTAGTTGAACCACAACAAGAGCTTTTCACGCT 700

CTAGGGTTTGAGTGAGACTCTCAACAGGGATAGCTTCGAAAGTCAATAGCACAGACGCTCTGGTCTTGTAAACATGAGTGAGGGCTAAAATCCCTCTCGAA 800

GCTAAACAGAACTCCCATCATAACCTATAACCCCTCTCTACGCTCGGGGCTCTGGGTCTGGTATTACCTCTTTTCTTATTGGGGGGGGGGGGGGGGGG 900

GGGGGGGATTTCCTAACTAACTATATATAATAAAGGGCCAAAGGATAGTTTAAATAAAAATCATAGCTTTGGGAGTTATAAATATAAGTTAGACCTTA 1000

TTCCTTGAATACTGCCTTTAAGGAGACGGGGATCGAACCCGTAATAAGAAATCAAAGTTCTTCGTTTTTCTTATTAACTACTGCTTATGGGTTGTAG 1100

→ tRNA GLU

→ tRNA THR

→ tRNA PRO

tRNA GLN ←

← tRNA ASN

CCTAAATGAGAAGGCGCTTGGCCGTTAACCAAGAGACAATAGGATAAAAACCTATCCTCCCAGACTAAGATAGCAAAGTGGTAAATGCAAAAAGATTAGG 1200
 tRNA LEU (CUN)
 tRNA ALA
 ATCTTATACCGTAGGTTCAATTCTTCTTAGTTGTTAGTTCCTAAGAATTTAACTTAGGGCTTCTGCTTGCAAAGCAGATATTTTCTTAACTAGAACC 1300
 tRNA TRP
 tRNA CYS
 ACAAGGGCTTAAGTTAAGCCAAACTGATAGCCTTCAAAGCTTTTAATAAGAATGAAAATTTCTTAGCCCTTTGGGCTTTATAGTGTAAAGCAACATCGTGG 1400

 ATTGCAAATCCTCAGATACAATTTAAATATTGTTAAAGCCTCAAGATAGCCTGACTTGCACAGACGTTGACTCCGTGTAAAAGAGGTGCTTGGACTAACT 1500
 tRNA VAL
 tRNA f-MET
 AGCTATATCTTGAGTATATATATAAAGTAGGGTAAGCTAAGTGACAAGCTTTGGGCTCATACCCCAAAAATGGAAGGATAGAAACCTCCCCCTACTT 1600
 tRNA ASP
 tRNA TYR
 TCTAGAAGTCGCTGGAGTTTAAACCAGCAAATTTGGCCTGACAACCCAAAGTTATTTTTTAACTAGACCTCCTCTTAAATAAGCTTGAACAAGATGGCT 1700
 tRNA GLY
 GAGACATAAAGCGGTGGATTGTA AACCCATAAAATATAGGTGAAACTCCTTTTCTTGTATTACTCTATGAGTACATCAGTATATTTGACTTCCAATCAGA 1800
 tRNA LEU (UUR)
 TGGTCTTGATTA AAAACTTCAAGATAGAGTATGCTGCTAGAATAGCAAATGGTAAATGCAGAAGGCCAAGACCTTCCCTATCAAGGGTTCAAATCCTTTTTT 1900

 ND1
 I Y A Y I F A F F E L I T F L V P V L L A V A F L T L V E R K V
 TAGTTATATATGCTTACATATTTGCTTTTTTGAAGCTTATTACATTTCTAGTCCCCGATATTATTAGCAGTAGCTTTTTTAAACCCCTAGTAGAACGGAAGGT 2000

 L G Y M Q F R R G P N V V G L T D F C N L F A D G L K L F I K E T
 CCTAGGCTATATGCAATTTCCGAAGGGCCCTAAAGTAGTTGGGCTTACGGACTTTTGCAACCTGTTTCGCTGACGGCCTTAAAGCTCTTTATTAAGGAAACA 2100

 L K P S S A S P Y L F F A S P V L F L T L A L L L W N F M P V T S
 CTTAAGCCTTCTTCGGCTTCCCCCTATTATTTTTGCTTCCCCTGTTTATTTTTAACTCTAGCCCTTGTCTCTGAAAATTTATGCCGGTAACAGCC 2200

 P A L D L Q L S L L L V L G L S S L S V Y A I L G S G
 CTGCCCTGGACCTTCAAGTATCTTTACTATTGGTTCTAGGTTTGCCAGATTGTCGGTCTACGCTATATTAGGATCGGGG 2280

Fig. 1 (Continued)

Table 1
Nucleotide Usage in Third Position of Fourfold-degenerate Codons

SPECIES	% IN THIRD-CODON POSITION			
	A	C	G	T
<i>Strongylocentrotus purpuratus</i>	37	25	11	26
<i>Paracentrotus lividus</i>	44	23	10	23
<i>Arbacia lixula</i>	34	20	6	40

NOTE.—Data are percentages (rounded to the nearest whole number) of occurrence of each fourfold codon across the protein-coding genes: COIII, ND1, ND3, ND4, and ND5.

genes, two complete genes, portions of four others, and a noncoding region (see fig. 1 for greater detail).

Material and Methods

Arbacia lixula mtDNA was obtained from the eggs of pooled individuals by using the technique described by Cantatore et al. (1987). The plasmid used for cloning the two *Bam*HI restriction fragments was pAT153 (Maniatis et al. 1982), and we sequenced the entire smaller *Bam*HI fragment and part of the larger *Bam*HI fragment after subcloning them in pUC8. Sequences were determined according to the dideoxynucleotide chain-termination method of Sanger, modified for double-stranded templates (Sanger et al. 1980). Unidirectional deletions of predictable size were produced by digesting the recombinant DNA with a combination of exonuclease III (Stratagene) and mung bean nuclease (Stratagene) (Cantatore et al. 1989). Linearized recombinant DNA was treated at 30°C for various periods of times with 4 U of exonuclease III/μg followed by 3 U of mung bean nuclease/μg for 30 min. A series of nested deletions was generated. The deletion products were religated, and the sequence of the first 200–300 bases of each insertion was determined. Both strands were sequenced. All the positions were detected on several overlapping clones, each nucleotide being sequenced an average of 2.8 times. Sequence analysis was carried out on a VAX/VMS 5.2 operating system by using the software package GLORIA, recently developed in our laboratory (M. Altimonelli and C. Lanave, unpublished data).

Results

Sequence Features

We demonstrated that the overall gene organization of *Arbacia lixula* is identical to that reported for the sea urchins *Paracentrotus lividus* (Cantatore et al. 1989) and *Strongylocentrotus purpuratus* (Jacobs et al. 1988), by using restriction mapping and Southern hybridization (C. DeGiorgi and C. Saccone, unpublished data). Furthermore, the fact that the clustering of 15 tRNAs between the 12S rRNA and ND1 genes is present in *A. lixula* confirms that no major rearrangements have occurred since the divergence of the orders Stirodonta and Camarodonta.

Figure 1a shows the sequence of the 2,847-nt *Eco*RI-*Hpa*II subfragment of the largest *Bam*HI fragment of *A. lixula* mtDNA. It contains tRNA and protein-coding genes in the following order: 109 amino acids at the carboxy-terminal end of the COIII gene, the Ser(UCN)-tRNA gene in the opposite orientation, the complete sequences of the ND3 and ND4 genes, the His-tRNA and Ser(AGN)-tRNA genes, and 190 amino acids of the amino-terminal end of the ND5 gene.

Table 2
T/T' between Species

Species and Fragment Compared to <i>Arbacia lixula</i>	T/T' ^a
<i>Stronglyocentrosus purpuratus</i> :	
ND1	2.41 ± 1.11
COIII	1.41 ± 0.42
ND5	2.40 ± 0.50
Mean	2.0 ± 0.50
<i>Paracentrotus lividus</i> :	
ND1	1.35 ± 0.67
COIII	2.92 ± 0.58
ND5	3.40 ± 0.67
Mean	2.9 ± 0.60

^a Values are computed from the second-codon positions of the protein-coding genes contained in the sequenced fragment of *A. lixula* (ND3 excluded) compared with the orthologous genes from *S. purpuratus* and *P. lividus*. The computation is from Lanave et al. (1984). T/T' is relative to *S. purpuratus* vs. *P. lividus*, which is set equal to 1. The weighted mean computed from the two means here reported is 2.5 ± 0.54.

A reiterated sequence, TATATATAA, is present in the same position [Ser(UCN)-tRNA, ND3 junction] in *A. lixula* as in the other sea urchins. In figure 1b the sequence of the 2,280-nt small *Bam*HI fragment is shown. This sequence contains 662 nt of the 3' end of the 12S rRNA gene, a cluster of 15 tRNA genes, a 136-nt noncoding region, and 375 nt at the 5' end of the ND1 gene. The order of the tRNA genes is the same as that described for *P. lividus* and *S. purpuratus*, in which the noncoding sequence (located between the Thr-tRNA and the Pro-tRNA genes) is 132 and 121 nt long, respectively.

The 136-nt noncoding region is characterized by the presence of a run of 26 consecutive G's. It has already been reported that there is a tendency for heterogeneity to arise in the length of homopolymeric tracts, both in vivo and when propagated in bacterial cells (Hauswirth et al. 1984). Thus, the precise number of G residues is difficult to quantify. However, in two independently isolated clones we found exactly the same number.

The G-rich region is followed by an A+T-rich sequence in which the conserved sequence motif TATATATAA is present in the same location as in *P. lividus* and *S. purpuratus*. The conserved consensus sequence TATATATAA is also found between the Val-tRNA and Met-tRNA genes, again as in *P. lividus* and *S. purpuratus*.

The presence of the conserved sequence motif TATATATAA in the same places in *A. lixula* as in the other sea urchins suggests that it represents a recognition signal for enzymes involved in transcription or processing (Cantatore et al. 1989).

mRNA-based Quantitative Estimate of Sequence Divergence

For a quantitative estimate of sequence divergence in the three echinoids, we applied our stationary Markov model to the protein-coding genes (Lanave et al. 1984). This method analyzes the evolutionary dynamics of the first and second codon positions of protein-coding genes, one at a time. If the change occurring at the third-codon positions is silent, this position is also analyzed. Our method requires that the sequence

sites under comparison be “stationary,” i.e., that their base frequencies (q_i ; $i = 1, 2, 3, 4 = A, C, G, T$) coincide within statistical fluctuations. If such a condition is fulfilled, we can calculate the relative times-of-divergence ratio (T/T') of a pair of sequences (by using at least three sequences). The absolute time of divergence for the pairs of sequences can be estimated by using as a calibration point a suitable time of divergence from the paleontological record (Saccone et al. 1990).

Comparison of *A. lixula* protein genes with those of either *P. lividus* or *S. purpuratus* reveals that only the second-codon position is stationary for all genes. The lack of stationarity of the third-codon position can be inferred from the data in table 1, where the relative usage of the third base in the fourfold codons of protein genes is reported. In particular, while the relative usage of A and C is roughly similar in all three species, the usage of G drops to 6% in *A. lixula* while the usage of T dramatically increases to 40%.

Table 2 shows the T/T' values of the three species pairs, calculated on the basis of the second-codon positions of the protein-coding genes. We analyzed each individual gene, as well as the “supergene” obtained by linking together all the protein genes except ND3, the divergence of which has already reached saturation in the two camarodont species. Table 2 shows that the average divergence of *A. lixula* from either *P. lividus* or *S. purpuratus* is at least twice that between *P. lividus* and *S. purpuratus*.

tRNA Genes and the 136-nt Noncoding Region

The conservation of tRNA genes (table 3) seems to be great in echinoderms, particularly for the Ser(UCN)-tRNA gene. The differences between *P. lividus* and *S. purpuratus* need to be checked, as in several instances (e.g., Pro- and His-tRNA genes) *P. lividus* differs more from *S. purpuratus* than either one of them does from *A. lixula* (the sequences of *P. lividus* and *A. lixula* have been carefully checked). The strong conservation of tRNA gene sequences suggests that they evolve under strong functional constraints in echinoderms.

The 136 nt is the only significant block of unassigned sequence. It has already been identified, in both *P. lividus* and *S. purpuratus*, as the main noncoding region.

Sequence alteration in the main noncoding region of the three echinoderms under study is reported in table 4. We observe few insertions/deletions and base substitutions in the three species; moreover, transversions outnumber transitions. The conservation of this region in the three species indicates that it is highly constrained and thus should play a fundamental role in the replication and expression of the genome.

ND5 Gene

The ND5 gene is 90 nt longer in sea urchins than the corresponding gene in vertebrates (Cantatore et al. 1989). The 72 nt at the 5' end of this extension may once have been a tRNA gene for leucine (CUN) codon (Cantatore et al. 1987; but see Jacobs et al. 1989). The nucleotide sequence corresponding to the extension domain can be folded into a secondary cloverleaf-like structure that is reminiscent of and may be derived from a tRNA gene that was present in the common ancestor (fig. 2).

Analysis of the base substitutions and amino acid replacements in the 90-nt extension and in the 480-nt internal region of the three sea urchin species demonstrates that the extension evolved faster than the internal region (table 5). A χ^2 test shows that the differences in the number of amino acid replacements observed in the two parts of the gene are statistically significant. A similar result was obtained by using Grantham's (1974) method, by which an amino acid-difference formula gives a relative

Table 3
Percentages of Similarity

SPECIES AND tRNA	% SIMILARITY FOR		
	<i>Strongylocentrotus purpuratus</i>	<i>Arbacia lixula</i>	Starfish
<i>Paracentrotus lividus</i> :			
Phe	93	...	78 ^a
Leu-UUR	91	92	...
Ile	90	...	75 ^a
Met	93	90	...
Trp	90	88	76 ^a
Lys	91
Gly	85	83	...
Arg	85	...	68 ^a
Leu-CUN	93	82	69 ^a
Thr	83	73	...
Asn	94	92	77 ^a
Cys	88	78	...
Glu	94	86	...
Pro	75	80	72 ^a
Tyr	87	82	...
His	78	79	73 ^b
Ser-AGN	83	85	71 ^b
Val	90	80	...
Ala	88	93	71 ^a
Asp	96	93	...
Gln	83	84	60 ^a
Ser-UCN	94	90	71 ^b
<i>S. purpuratus</i> :			
Phe	78 ^a
Leu-UUR		88	...
Ile	75 ^a
Met		89	...
Trp		88	84 ^a
Lys
Gly		83	...
Arg	68 ^a
Leu-CUN		82	72 ^a
Thr		83	...
Asn		89	75 ^a
Cys		73	...
Glu		86	...
Pro		76	78 ^a
Tyr		77	...
His		82	71 ^b
Ser-AGN		80	64 ^b
Val		74	...
Ala		85	79 ^a
Asp		91	...
Gln		81	56 ^a
Ser-UCN		88	72 ^b
<i>A. lixula</i> :			
Phe
Leu-UUR
Ile
Met

Table 3 (Continued)

SPECIES AND tRNA	% SIMILARITY FOR		
	<i>Strongylocentrotus purpuratus</i>	<i>Arbacia lixula</i>	Starfish
<i>A. lixula</i> (Continued)			
Trp			78 ^a
Lys
Gly
Arg
Leu-CUN			69 ^a
Thr
Asn			76 ^a
Cys
Glu
Pro			72 ^a
Tyr
His			72 ^b
Ser-AGN			69 ^b
Val
Ala			70 ^a
Asp
Gln			65 ^a
Ser-UCN			71 ^b

^a Data are for *Pisaster ochraceus* (Smith et al. 1989).

^b Data are for *Asterina pectinifera* (Himeno et al. 1987).

value to the physical and chemical difference between pairs of amino acids. On the other hand, when the comparison is carried out between the two species of the same order, the 5' extension and the rest of the gene show equal rates of evolution. Similarly, in the comparison of five closely related sea urchin species there are no differences between the 72-nt extension and the internal region (Thomas et al. 1989). This seems to suggest that the amino acid differences between orders are saturated in the internal region because of selection for gene function, whereas such constraints are not operating in the extension.

Discussion

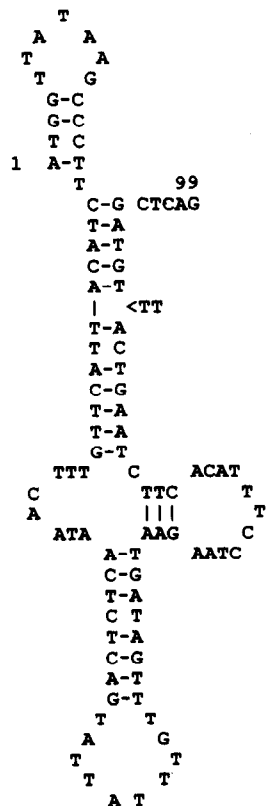
Gene Boundaries

Comparison of the intergenic regions of mitochondrial genes of *Arbacia lixula* with those of the two camarodont species allowed a definition of gene borders. When there was a choice between two possible start codons, making intervening lengths

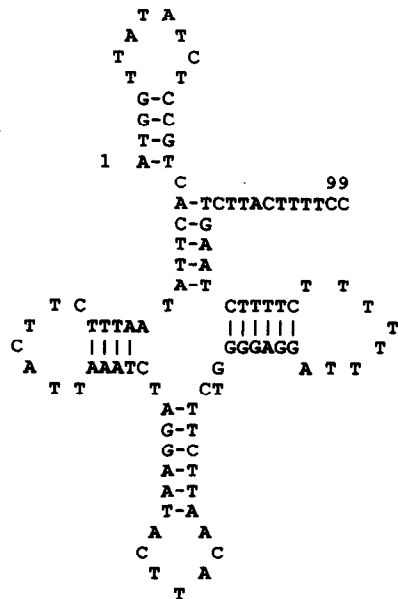
Table 4
Sequence Alterations in Main Noncoding Regions

Comparison (no. of sites analyzed)	% Similarity	No. of Transitions	No. of Transversions	No. of Insertions/Deletions
<i>Arbacia lixula</i> vs. <i>P. lividus</i> (127)	76	10	20	14
<i>A. lixula</i> vs. <i>Strongylocentrotus purpuratus</i> (114)	72	13	19	28
<i>Paracentrotus lividus</i> vs. <i>S. purpuratus</i> (113)	77	7	19	28

A. lixula



S. Purpuratus



P. lividus

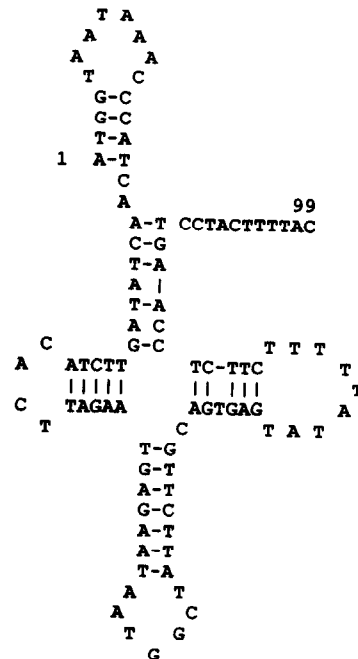


FIG. 2.—Proposed secondary structure of 5' end domain of ND5 gene. The sequence from nucleotide 1 to nucleotide 99 can be folded into a cloverleaf-like structure in the three species *Arbacia lixula*, *Strongylocentrotus purpuratus*, and *Paracentrotus lividus*.

Table 5
Analysis of Base Substitutions and Amino Acid Replacements in NAS Gene

	% BASE SUBSTITUTION										
	<i>Arbacia lixula</i> vs. <i>Strongylo-</i> <i>centrotus</i> <i>purpuratus</i>				<i>A. lixula</i> vs. <i>Paracentrotus</i> <i>lividus</i>				% AMINO ACID REPLACEMENT ^a		
	Codon Position		Codon Position		<i>A. lixula</i> vs. <i>S.</i> <i>purpuratus</i>	<i>A. lixula</i> vs. <i>P.</i> <i>lividus</i>	<i>S.</i> <i>purpuratus</i> vs. <i>P.</i> <i>lividus</i>				
	First	Second	First	Second							
Extension ^b	56.7	40.0	46.7	43.3	70.0 (47.6)	73.3 (52.0)	36.7 (20.7)				
Internal region ^c	39.4	29.4	41.9	36.2	46.9 (30.0)	52.5 (38.1)	36.2 (22.0)				

NOTE.—Statistical evaluation of the amino acid replacements was as follows: *A. lixula* vs. *S. purpuratus*, $\chi^2 = 4.519$, $P(\chi^2) = 0.03$; *A. lixula* vs. *P. lividus*, $\chi^2 = 3.641$, $P(\chi^2) = 0.06$; and *S. purpuratus* vs. *P. lividus*, $\chi^2 = 0.003$, $P(\chi^2) = 0.85$. Statistical evaluation of the amino acid replacements measured according to the method of Grantham (1974) was as follows: *A. lixula* vs. *S. purpuratus*, $\chi^2 = 3.11$, $P(\chi^2) = 0.08$; *A. lixula* vs. *P. lividus*, $\chi^2 = 3.82$, $P(\chi^2) = 0.05$; and *S. purpuratus* vs. *P. lividus*, $\chi^2 = 0.06$, $P(\chi^2) = 0.80$.

^a Values in parentheses are the average values of the amino acid differences, according to Grantham (1974).

^b From 1 to 90 nt of ND5 (30 amino acids).

^c From 91 to 570 nt of ND5 (160 amino acids).

equal in three cases seemed appropriate, especially since the result was consistent with Gadaleta et al.'s (1989) rule that ATH and GTG are only used as vertebrate mitochondrial initiator codons if there are no nucleotides between the two neighboring genes. This led to the specification of ATA as initiator for ND1, and of ATG as initiator for ND3 and ND4.

The secondary structure of each tRNA defines the anticodon unambiguously; however, the precise positions of the termini in tRNA genes are often debatable. When more than one secondary structure could be drawn, we chose that in which the aminoacyl stem was 7 nt long, as in the study by Cantatore et al. (1989), even though the aminoacyl stem could have been 2 nt longer in Thr-tRNA, Asn-tRNA, Ala-tRNA, Trp-tRNA, and Val-tRNA and 1 nt longer in Met-tRNA.

In *Paracentrotus lividus* the ND4 and His-tRNA genes overlap for a length of 10 nt (Cantatore et al. 1989). The same could apply to *A. lixula*. However, when the His-tRNA gene is made 1 nt longer at each end (as suggested by the findings in *Strongylocentrotus purpuratus*), the overlap between the ND4 and His-tRNA genes in both *A. lixula* and *P. lividus* becomes 11 nt. Similarly, the Ser(AGN)-tRNA and ND5 genes become adjacent in all three species if the tRNA gene has an aminoacyl stem 8 nt long. Obviously, direct analysis of the RNA products is necessary if definite conclusions are to be drawn.

tRNA Genes

In invertebrate genomes the organization of tRNA genes shows a much higher variability compared with that of rRNA or protein genes. We have suggested that during the evolution of sea urchins a tRNA lost its function and became part of a protein-coding gene (Cantatore et al. 1987). According to Jacobs et al. (1989), a clustered organization of tRNA genes may have been present in the ancestral mitochondrial genome. In vertebrate mtDNA, tRNAs may serve as signals for the processing of the mitochondrial multicistronic transcripts (De Giorgi and Saccone 1989).

Our results reported in table 3 indicate that tRNA genes are highly conserved in the echinoderm lineage. The degree of conservation of the echinoid tRNA genes compared with those of the asteroid *Asterina pectinifera* is particularly relevant, considering the 400–500-Mya time of divergence of the two classes. These observations indicate that, whatever their role, tRNA genes must evolve under evolutionary pressures different from those operating on protein and rRNA genes. The evolution depends both on the particular position of tRNAs in the genome and on the modalities of gene expression in a given taxonomic group (also see Thomas et al. 1989).

Directional Mutation Pressure

That the existence of directional mutation pressure can have a great influence of phylogenetic inferences from sequence data was first suggested by Freese (1962) and Sueoka (1962). According to this theory, the effect of a mutation on a genome has a directionality toward higher or lower G+C content. Nonrandom G+C content at the first two positions within a codon in a protein-coding gene may reflect the constraints on amino acid composition that are required for the biological function of the molecule, while the directional change in G+C is more marked in a selectively neutral part of the genome. As we report in table 1, the *A. lixula* protein-coding genes show a marked drift of G→T in the codon usage, which is responsible for the lack of stationarity (C. Saccone and R. Holmquist, unpublished data). This is the reason why our quantitative method for estimating sequence divergence could be applied only to the second-codon positions of the mRNA coding genes.

The data in table 2 indicate that the average divergence of *A. lixula* from either *P. lividus* or *S. purpuratus* is at least twice that which exists between *P. lividus* and *S. purpuratus*. Smith (1988) has recently estimated phylogenetic relationships for several camarodont species, from which a tentative date of ~50 Mya can be derived as the time of divergence between *P. lividus* and *S. purpuratus*. We estimate that the Stirodonta-Camarodonta time of separation is 125 ± 27 Mya, a value 38% lower than the 200 Mya reported elsewhere (Smith 1984). Either value might be right.

Sequence Availability

The DNA sequences have been deposited in EMBL under accession numbers X53727 and X53726 for the sequences shown in figures 1a and 1b, respectively.

Acknowledgments

This work has been funded by a grant from Ministero della Pubblica Istruzione.

LITERATURE CITED

- CANTATORE, P., M. N. GADALETA, M. ROBERTI, C. SACCONI, and A. C. WILSON. 1987. Duplication and remoulding of tRNA genes during the evolutionary rearrangement of mitochondrial genomes. *Nature* **329**:853–855.
- CANTATORE, P., M. ROBERTI, G. RAINALDI, M. N. GADALETA, and C. SACCONI. 1989. The complete nucleotide sequence, the gene organization and the genetic code of the mitochondrial genome of *Paracentrotus lividus*. *J. Biol. Chem.* **264**:10965–10975.
- DE GIORGI, C. 1988. Mitochondrial DNA polymorphism in the eggs of the sea urchin *Arbacia lixula*. *Cell Biol. Int. Rep.* **12**:407–412.
- DE GIORGI, C., and C. SACCONI. 1989. Mitochondrial genome in animal cells: structure, organization and evolution. *Cell Biophys.* **14**:67–78.
- FREESE, E. 1962. On the evolution of the base composition of DNA. *J. Theor. Biol.* **3**:82–101.

- GADALETA, G., G. PEPE, G. D. CANDIA, C. QUAGLIARIELLO, E. SBISÀ, and C. SACCONI. 1989. Complete nucleotide sequence of the *Rattus norvegicus* mitochondrial genome: cryptic signals revealed by comparative analysis between vertebrates. *J. Mol. Evol.* **28**:497–516.
- GIUDICE, G. 1986. The sea urchin embryo: a developmental biological system. Springer, Berlin.
- GRANTHAM, R. 1974. Amino acid difference formula to help explain protein evolution. *Science* **185**:862–864.
- HAUSWIRTH, W. W., M. J. VANDE WALLE, P. J. LAIPIS, and P. D. OLIVO. 1984. Heterogeneous mitochondrial DNA D-loop sequences in bovine tissue. *Cell* **37**:1001–1007.
- HIMENO, H., H. MASAKI, T. KAWAI, T. OHTA, I. KUMAGAI, K. MIURA, and K. WATANABE. 1987. Unusual genetic codes and a novel gene structure for Ser-AGY-tRNA in starfish mitochondrial DNA. *Gene* **56**:219–230.
- JACOBS, H. T., S. ASAKAWA, T. ARAKI, K. MIURA, M. J. SMITH, and K. WATANABE. 1989. Conserved tRNA gene cluster in starfish mitochondrial DNA. *Curr. Genet.* **15**:193–209.
- JACOBS, H. T., D. ELLIOT, B. M. VEERABHADRACHARYA, and A. FARQUHARSON. 1988. Nucleotide sequence and gene organization of sea urchin mitochondrial DNA. *J. Mol. Biol.* **202**:185–217.
- LANAVE, C., G. PREPARATA, C. SACCONI, and G. SERIO. 1984. A new method for calculating evolutionary substitution rate. *J. Mol. Evol.* **20**:86–93.
- MANIATIS, T., E. F. FRITSCH, and J. SAMBROOK. 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
- SACCONI, C. C. LANAVE, G. PESOLE, and G. PREPARATA. 1990. Influence of base composition on quantitative estimates of gene evolution. *Methods Enzymol.* **183**:570–583.
- SANGER, F., A. R. COULSON, B. G. BARRELL, A. J. H. SMITH, and B. A. ROE. 1980. Cloning in single-stranded bacteriophage as an aid to rapid DNA sequencing. *J. Mol. Biol.* **143**:161–178.
- SMITH, A. 1984. Echinoid paleobiology. George Allen & Unwin, London.
- SMITH, A. B. 1988. Phylogenetic relationship, divergence times, and rates of molecular evolution for camarodont sea urchins. *Mol. Biol. Evol.* **5**:345–365.
- SMITH, M. J., D. K. BANFIELD, K. DOTEVAL, S. GORSKI, and D. J. KOWBEL. 1989. Gene arrangement in sea star mtDNA demonstrates a major inversion event during echinoderm evolution. *Gene* **76**:181–185.
- SUEOKA, N. 1962. On the genetic basis of variation and heterogeneity of DNA base composition. *Proc. Natl. Acad. Sci. USA* **48**:582–592.
- THOMAS, W. K., J. MAA, and A. C. WILSON. 1989. Shifting constraints on tRNA genes during mitochondrial DNA evolution in animals. *New Biol.* **1**:93–100.

WESLEY M. BROWN, reviewing editor

Received January 11, 1990; revision received November 8, 1990

Accepted November 8, 1990