Evolution of the Deep-Sea Gulper Eel Mitochondrial Genomes: Large-Scale Gene Rearrangements Originated Within the Eels

Jun G. Inoue,* Masaki Miya,† Katsumi Tsukamoto,*, and Mutsumi Nishida*

*Ocean Research Institute, University of Tokyo, Tokyo, Japan; and †Department of Zoology, Natural History Museum & Institute, Chiba, Japan

Recent studies have demonstrated that deviations from the typical vertebrate mitochondrial gene order are more frequent than initially thought. Such deviations, however, are minor, with inversions and/or translocations of a few genes being involved and tandem duplication of the gene regions followed by deletions of genes having been invoked as mechanisms originating in such novel gene order. During the course of molecular phylogenetic studies on the Elopomorpha (eels and their allies), we found that mitochondrial genomes (mitogenomes) from the two deep-sea gulper eels, *Eurypharynx pelecanoides* (Eurypharyngidae) and *Saccopharynx lavenbergi* (Saccopharyngidae), exhibit an identical gene order which greatly differs from that of any other vertebrates. Phylogenetic analysis using the mitogenomic data from 59 species of fish not only confirmed a single origin of such a gene order with confidence but also indicated that it had been derived from the typical vertebrate gene order. Detailed comparisons of the gulper eel gene order with that of typical vertebrates suggested that occurrence of a single step, large-scale duplication of gene region extending >12 kb, followed by deletions of genes in a common ancestor of the two species, most parsimoniously accounts for this unusual gene arrangement.

Introduction

Vertebrate mitochondrial gene order was initially considered conservative because the complete nucleotide sequences of the mitochondrial genomes (mitogenomes) from mammals (Anderson et al. 1981, 1982; Bibb et al. 1981) and the African clawed frog (Roe et al. 1985) showed a common gene order. As the complete mitochondrial DNA (mtDNA) sequences have been determined for a number of vertebrates (269 species as of 11 June 2003; National Center for Biotechnology Information [NCBI] Web site: http://www.ncbi.nlm.nih.gov/PMGifs/Genomes/7742.html), it has been recognized that deviations from the typical gene order are not so rare in vertebrates (fig. 1; Boore 1999). With the exception of placental mammals, examples include all major lineages of vertebrates such as lampreys (Lee and Kocher 1995), amphibians (Macey et al. 1997; Sumida et al. 2001), reptiles (Kumazawa and Nishida 1995; Quinn and Mindell 1996; Macey et al. 1997), birds (Desjardins and Morais 1990, 1991; Quinn and Wilson 1993; Mindell, Sorensen, and Dimcheff 1998), marsupials (Pääbo et al. 1991; Janke et al. 1994), and fish (Miya and Nishida 1999; Inoue et al. 2001b; Miya, Kawaguchi, and Nishida 2001). Such gene rearrangements, however, are local (most notably within a cluster of tRNA genes), and no example has been found that involves genomic-scale rearrangements.

Gulper eels are well-known deep-sea creatures, collectively including four families placed in the order Saccopharyngiformes, that occur at bathypelagic depths (>1,000 m) throughout the world’s oceans (Bertelsen, Nielsen, and Smith 1989). Their bizarre appearance (fig. 1), which results from extremely enlarged, deformed gapes at the tip of the eel’s snake-like head and body, has attracted much attention from many researchers (e.g., Bertelsen, Nielsen, and Smith 1989; Robins 1989). Such specialized anatomical features have led investigators to suppose that the phylogenetic positions of these animals are outside the bony fish (e.g., Tchernavin 1946), although they have generally been considered to be close relatives of the true eels (order Anguilliformes), because they have a leaf-like leptocephalous larval stage (Greenwood et al. 1966; Inoue and Miya 2001).

During the course of molecular phylogenetic studies of the Elopomorpha (eels and their allies) using mitogenomic data, we found an unusual mitochondrial gene order for one of the gulper eels, *Eurypharynx pelecanoides* (fig. 1). This article describes novel gene organization in the mitogenomes of two species of gulper eels, *E. pelecanoides* and *Saccopharynx lavenbergi*. We employed a polymerase chain reaction (PCR)-based approach developed by Miya and Nishida (1999) for sequencing the complete mitogenomes of these fishes. To explore the origin of such unique mitogenomes, we subjected the mitogenomic data to phylogenetic analysis, and we discuss here the possible mechanisms generating such a gene arrangement in the two gulper eels.

Materials and Methods

Specimens

Mitochondrial DNA sequences were obtained from six individuals from two species representing two families in the order Saccopharyngiformes (Robins 1989). For the monotypic family Eurypharyngidae, one individual of *E. pelecanoides* was used as a specimen for determination of the complete mtDNA sequence, and the other four individuals, including two leptocephalous larvae, were used for determination of partial sequences of four major gene junctions (fig. 2) to confirm the gene order in the *E. pelecanoides* mitogenome. For the family Saccopharyngidae, a single individual of *S. lavenbergi* was used as a specimen for determination of the complete mtDNA sequence. Voucher specimens were deposited in the fish collections at the Natural History Museum & Institute, Chiba, Japan (CBM-ZF), and at the University of Washington (UW): *E. pelecanoides* (CBM-ZF 10311: Hyuga-Nada,
southern Japan; CBM-ZF 10312, 10313, 10316, and 10317: equatorial central Pacific Ocean) and *S. lavenbergi* (UW 045633: Eastern Pacific).

**DNA Extraction**

A portion of the epaxial musculature (ca. 0.25 g) was excised from fresh specimens of each species and immediately preserved in 99.5% ethanol. Total genomic DNA was extracted using the Qiagen QIAamp tissue kit following the manufacturer’s protocol.

**Polymerase Chain Reaction and Sequencing**

The entire mitogenomes of two gulper eels were amplified using a long-PCR technique (Cheng, Higuchi, and Stoneking 1994). Five fish-universal and three species-specific long-PCR primers (S-LA-16S-L + H12293-Leu and L12321-Leu + S-LA-16S-H for *E. pelecanoides*; and L2508–16S + Sala-ND4-H and Sala-COI-L + Sala-ND1-H for *S. lavenbergi* (fig. 1) were used to amplify the entire mitogenome of each eel in two reactions. The entire mitogenomes of the other four *E. pelecanoides* individuals
were amplified with three fish-universal primers and one species-specific long-PCR primer (L2508–16S + Eupe-
ND2-H and L12321-Leu + S-LA-16S-H). Four species-
specific primers (Eupe-ND2-H for E. pelecanoides; Sala-COI-L, Sala-ND1-H, and Sala-ND4-H for S. laven-
bergi) were alternatively designed with reference to the partial nucleotide sequences from the ND2 gene of E. pele-
canooides and the COI, ND1, and ND4 genes of S. lavenbergi determined from each total DNA using fish-universal
primers.

The long-PCR products were diluted with TE buffer (1:20) for subsequent use as PCR templates, except for a region intervening between the two long-PCR primers (between S-LA-16S-H and S-LA-16S-L, and H12293-Leu and L12321-Leu for E. pelecanoides; fig. 1), in which total
genomic DNA was used alternatively.

A total of 82 fish-universal primers (Miya and Nishida 1999, 2000; Inoue et al. 2000, 2001a, 2001b, 2001c, 2001d; Ishiguro, Miya, and Nishida 2001; Kawaguchi, Miya, and Nishida 2001) were used in various combinations to
amplify contiguous, overlapping segments of the entire
mitogenome for each of the two species. Species-specific
primers were designed in cases where no appropriate fish-
universal primers were available. A list of PCR primers
used for a specific species is available from J.G.I. upon
request. Long-PCR and subsequent PCR reactions were
carried out as previously described (e.g., Miya and Nishida

Double-stranded PCR products, purified using a Pre-
Sequencing Kit (USB), were subsequently used for direct
cycle sequencing using dye-labeled terminators (Applied
Biosystems). The primers used were the same as those for
PCR. All sequencing reactions were performed according to
the manufacturer’s instructions. Labeled fragments were
analyzed by means of a Model 377 DNA sequencer
(Applied Biosystems).

Sequence Analysis

DNA sequences were analyzed using the computer
software package program DNASIS version 3.2 (Hitachi
Software Engineering). Locations of the 13 protein-coding
genes were determined by comparisons of DNA or amino
acid sequences of bony fish mitogenomes. The 22 tRNA
genes were identified by their cloverleaf secondary
acid sequences of bony fish mitogenomes. The 22 tRNA
genes were determined by comparisons of DNA or amino
groups were excluded from the analyses. In addition, 3rd codon positions in the protein-coding genes that could positively mislead an
analysis of higher-level relationships of actinopterygians (Miya and Nishida 2000) were excluded from the analyses,
leaving 7,984 available nucleotide positions from the 12
protein-coding genes and 21 tRNA genes.

Bayesian phylogenetic analyses were conducted using
MrBayes 2.01 (Huelsenbeck and Ronquist 2001). The
general time reversible model with some sites assumed to
be invariable and with variable sites assumed to follow
a discrete gamma distribution (GTR + I + Γ; Yang 1994)
was selected as the best-fit model of nucleotide substitution
(ModelTest version 3.06; Posada and Crandall 1998). We
set the maximum likelihood parameters in MrBayes as
follows: “iset nst = 6” (GTR), “rates = invgamma” (I + Γ),
and “basefreq = estimate” (estimated proportion of base
types from the data). The Markov chain Monte Carlo
process was set so that four chains (three heated and one
cold) ran simultaneously. We conducted two independent
runs for 1 million generations, with trees being sampled
every 100 generations, each of which started from a random
tree. Two independent analyses converged on similar log-
likelihood scores and reached “stationarity” (lack of
improvement in ML scores) at no later than 120,000
generations. Thus, the first 1,200 trees were discarded from
each analysis, and the remaining 17,602 combined samples
from two independent analyses were used to generate a 50% majority rule consensus tree, with the percentage of samples
recovering any particular clade representing that clade’s
posterior probability (Huelsenbeck and Ronquist 2001).

Results

Phylogenetic Analysis

The data matrices from Inoue et al. (2001d) and Miya,
Kawaguchi, and Nishida (2001) were combined and redundant taxa were eliminated. Mitogenic sequences from the two gulper eels and two teleosts (Gymnothorax kidako, AP002976 [Inoue et al. 2003] and Salmo salar,
U12143 [Hurst et al. 1999]) were added. Because unam-
biguous alignments of the two rRNA genes on the basis of
secondary structure models were not feasible (Miya and
Nishida 2000), they were not used in the analyses. The ND6
gene was not used in the phylogenetic analyses because of
its heterogeneous base composition and consistently poor
phylogenetic performance (e.g., Zardoya and Meyer 1996;
Miya and Nishida 2000). Also, tRNA loops, tRNA^Ser(AGY)
(unstable inferred secondary structures in some species),
and other ambiguously aligned regions, such as the 5’ and
3’ ends of several protein-coding genes, were excluded
from the analyses. In addition, 3rd codon positions in the
protein-coding genes that could positively mislead an
analysis of higher-level relationships of actinopterygians
(Miya and Nishida 2000) were excluded from the analyses,
leaving 7,984 available nucleotide positions from the 12
protein-coding genes and 21 tRNA genes.

Bayesian phylogenetic analyses were conducted using
MrBayes 2.01 (Huelsenbeck and Ronquist 2001). The
general time reversible model with some sites assumed to
be invariable and with variable sites assumed to follow
a discrete gamma distribution (GTR + I + Γ; Yang 1994)
was selected as the best-fit model of nucleotide substitution
(ModelTest version 3.06; Posada and Crandall 1998). We
set the maximum likelihood parameters in MrBayes as
follows: “iset nst = 6” (GTR), “rates = invgamma” (I + Γ),
and “basefreq = estimate” (estimated proportion of base
types from the data). The Markov chain Monte Carlo
process was set so that four chains (three heated and one
cold) ran simultaneously. We conducted two independent
runs for 1 million generations, with trees being sampled
every 100 generations, each of which started from a random
tree. Two independent analyses converged on similar log-
likelihood scores and reached “stationarity” (lack of
improvement in ML scores) at no later than 120,000
generations. Thus, the first 1,200 trees were discarded from
each analysis, and the remaining 17,602 combined samples
from two independent analyses were used to generate a 50% majority rule consensus tree, with the percentage of samples
recovering any particular clade representing that clade’s
posterior probability (Huelsenbeck and Ronquist 2001).

Results

Phylogenetic Analysis

The data matrices from Inoue et al. (2001d) and Miya,
Kawaguchi, and Nishida (2001) were combined and redundant taxa were eliminated. Mitogenic sequences from the two gulper eels and two teleosts (Gymnothorax kidako, AP002976 [Inoue et al. 2003] and Salmo salar,
U12143 [Hurst et al. 1999]) were added. Because unam-
biguous alignments of the two rRNA genes on the basis of
secondary structure models were not feasible (Miya and
Nishida 2000), they were not used in the analyses. The ND6
gene was not used in the phylogenetic analyses because of
its heterogeneous base composition and consistently poor
phylogenetic performance (e.g., Zardoya and Meyer 1996;
Miya and Nishida 2000). Also, tRNA loops, tRNA^Ser(AGY)
(unstable inferred secondary structures in some species),
and other ambiguously aligned regions, such as the 5’ and
3’ ends of several protein-coding genes, were excluded
from the analyses. In addition, 3rd codon positions in the
protein-coding genes that could positively mislead an
analysis of higher-level relationships of actinopterygians
(Miya and Nishida 2000) were excluded from the analyses,
leaving 7,984 available nucleotide positions from the 12
protein-coding genes and 21 tRNA genes.

Bayesian phylogenetic analyses were conducted using
MrBayes 2.01 (Huelsenbeck and Ronquist 2001). The
general time reversible model with some sites assumed to
be invariable and with variable sites assumed to follow
a discrete gamma distribution (GTR + I + Γ; Yang 1994)
was selected as the best-fit model of nucleotide substitution
(ModelTest version 3.06; Posada and Crandall 1998). We
set the maximum likelihood parameters in MrBayes as
follows: “iset nst = 6” (GTR), “rates = invgamma” (I + Γ),
and “basefreq = estimate” (estimated proportion of base
types from the data). The Markov chain Monte Carlo
process was set so that four chains (three heated and one
cold) ran simultaneously. We conducted two independent
runs for 1 million generations, with trees being sampled
every 100 generations, each of which started from a random
tree. Two independent analyses converged on similar log-
likelihood scores and reached “stationarity” (lack of
improvement in ML scores) at no later than 120,000
generations. Thus, the first 1,200 trees were discarded from
each analysis, and the remaining 17,602 combined samples
from two independent analyses were used to generate a 50% majority rule consensus tree, with the percentage of samples
recovering any particular clade representing that clade’s
posterior probability (Huelsenbeck and Ronquist 2001).

Results

Genome Organizations

The total lengths of the E. pelecanoides and the S.
lavenbergi mitogenomes are 18,978 bp and 18,495 bp
(except for a portion of the putative control region for S.
lavenbergi), respectively (see Supplementary Material
online). The genome content of these two gulper eels
includes two rRNA, 22 tRNA, and 13 protein-coding
genes, as found in other vertebrates (fig. 2). Also as in
other vertebrates, most genes of these two species are
encoded on the H-strand, except for the ND6 and eight
tRNA genes. All features are similar in length to those in
other bony fishes with the exception of the COI gene and
the control region for E. pelecanoides and the COI and cyt
b genes for S. lavenbergi (approximately 100, 800, 150,
and 30 bp longer than those in other bony fishes,
respectively).

The E. pelecanoides mitogenome contains three noncoding regions longer than 200 bp (fig. 2; see also
Supplementary Materials online; noncoding region [nc1 and 2] and control region [CR]). The nc1, located between tRNASer(UCN) and tRNAAsp genes, and the nc2, located between the tRNAAsp and COII genes, contain two and four pseudogenes corresponding to degenerating copies of tRNAAsp gene, respectively. A noncoding region (1,818 bp) located between the cyt b and tRNAPhe genes appears to correspond to the control region. The S. lavenbergi mitogenome also contains three noncoding regions longer than 200 bp (noncoding region [nc] and control regions [CR1 and CR2]). The nc, located between the tRNALeu(CUN) and ND5 genes, contains at least two pseudogenes corresponding to degenerating copies of the tRNALeu(CUN) gene. CR1 (992 bp), located between two tRNA gene clusters (TP and IM regions), and CR2 (> 837 bp), located between cyt b and tRNAPhe genes, appear to correspond to the control region, and they have completely identical sequences over 800 bp, indicating that they are undergoing concerted evolution (see Kumazawa et al. 1998).

With the exception of the two duplicated control regions (CR1 and CR2) in the S. lavenbergi mitogenome, the mitogenomes from the two deep-sea gulper eels exhibit an identical gene order (fig. 2). However, the mitochondrial gene order of the two gulper eels differs greatly from that of all other vertebrates known to date. When some tRNA genes were excluded from the comparisons (fig. 2), we identified five gene blocks (A, tRNA^Glu^-tRNA^Gln^, B, ATP8-ND3; C, tRNA^Leu(CUN)^-ND6; D, tRNA^Ile^-tRNA^Lys^; and E, tRNA^Arg^-tRNA^Ser(AGY)^ gene regions), the gene order of which is identical to that of typical vertebrates.

Phylogenetic Positions of Two Gulper Eels

Figure 3 is a 50% majority rule consensus tree of the 17,602 combined samples from two independent Bayesian analyses of the 59 mitochondrial nucleotide sequences from the concatenated 12 protein-coding (no 3rd codon positions), plus 21 tRNA genes (stem regions only) using the GTR + I + C model of sequence evolution (Yang 1994). With the exception of a few nodes within the Neoteleostei, most internal branches were supported by high Bayesian posterior probabilities (≥95%). It should be noted that the resultant tree explicitly demonstrates not only that the two gulper eels form a sister-group relationship with a high posterior probability (100%), but also that they are deeply nested within the Anguilliformes (true eels), strongly suggesting that they have been derived from an eel-like ancestor.

Discussion

Recently, more than 140 complete mtDNA sequences from the actinopterygian fishes have been determined (e.g., Miya and Nishida 1999, 2000; Inoue et al. 2000, 2001a, 2001b, 2001c, 2001d, 2003; Ishiguro, Miya, and Nishida 2001, 2003; Kawaguchi, Miya, and Nishida 2001; Miya, Kawaguchi, and Nishida 2001; Miya et al. 2003; Saitoh et al. 2003), and several examples of gene rearrangements have been reported (fig. 1). It should be noted that such gene rearrangement events involve few genes and that the gene arrangement of the gulper eel mitogenomes differs greatly from that of other vertebrates.

Phylogenetic Implications on the Origin of Novel Gene Order

To deduce the evolution of the gene arrangements in the two gulper eels, inference for the ancestral organization based on a reliable phylogenetic framework is required. Although the gulper eels have been included in the
FIG. 3.—The 50% majority rule consensus tree derived from Bayesian analysis of mitogenomic data from 57 teleosts and two outgroup species using MrBayes 2.01 (Huelsenbeck and Ronquist 2001) with GTR + I + C model (Yang 1994) of sequence evolution. Numerals indicate Bayesian posterior probabilities (shown as percentages). Branch lengths were estimated by constrained ML analysis using PAUP*4.0b10 (Swofford 2002) with the GTR + I + C model of sequence evolution. The scale indicates expected nucleotide substitutions per site. Thick branches indicate occurrences of the gene rearrangements.
Elopomorpha based solely on a distinct pelagic larval form, termed the leptocephalus, no one has corroborated their phylogenetic position using character matrices derived from morphological or molecular data (Inoue and Miya 2001).

Bayesian analysis using the mitogenomic data from 59 species which fully represent teleostean fish diversity (fig. 3) not only corroborated that the novel gene order of the two gulper eels originated in a common ancestor of the two species, but also demonstrated that their origin was independent from those of the other novel gene orders. It also appears that the novel gene order of Conger myriaster, another member of the Elopomorpha, has an origin independent from that of gulper eels. It should be noted that Anguilla japonica, which has the typical vertebrate gene order, was confidently placed as a sister species of the two gulper eels. The most parsimonious reconstruction of the gene rearrangement events on the phylogenetic tree indicated that the aberrant gene order of the two gulper eels is derived from that of typical vertebrates.

### Possible Mechanism for the Gene Rearrangement in Gulper Eels

Two major mechanisms have been proposed to explain the gene rearrangements in vertebrate mitogenomes (Lee and Kocher 1995; Kumazawa et al. 1996). One is the tandem duplication of gene regions as a result of slipped strand mispairing, followed by the deletions of genes (Moritz and Brown 1986; Levinson and Gutman 1987). Although the dynamics of mitochondrial gene arrangements have not been explained in some invertebrates (e.g., Kurabayashi and Ueshima 2000; Machida et al. 2002; Tomita et al. 2002), vertebrate mitochondrial gene rearrangements may well be explained by such a mechanism (Desjardins and Morais 1990; Quinn and Wilson 1993; Kumazawa and Nishida 1995; Kumazawa et al. 1996; Mindell, Sorenson, and Dimcheff 1998; Miya and Nishida 1999; Inoue et al. 2001b), and its feasibility is also supported by the frequent polymorphic duplications of mtDNA sequences (e.g., Stanton et al. 1994; Gach and Brown 1997; Mindell, Sorenson, and Dimcheff 1998; Miya and Nishida 1999). The other suggested mechanisms invoke the illicit priming of replication by tRNAs and the resultant integration of tRNA genes around the control region (Cantatore et al. 1987; Jacobs et al. 1989).

Although the second model is not excluded at this point, the first model is favored as the mechanism of gene rearrangement in the gulper eel mitogenomes. Assuming that a long DNA fragment in the tRNA\textsuperscript{The}–CR region of the typical organization is duplicated (fig. 4), subsequent deletions of redundant genes would give rise to the rearranged gene organization in gulper eels. Such tandem duplication and subsequent deletions most parsimoniously resulted in the observed gene order and associated intergenic spacers in these two gulper eels.

Multiple deletions of redundant genes seemed to be incomplete in the two gulper eels, owing to several stretches of noncoding sequences (fig. 2) occurring around the genes involved in the rearrangement (fig. 4). Although nc1 and nc2 in E. pelecanoides and nc in S. lavenbergi are composed of repetitive pseudogenes corresponding to degenerating copies of adjacent tRNA gene, CR1 and CR2 in S. lavenbergi mitogenome have identical sequences over 800 bp, and CR1 is located at the putative duplicated position while CR2 is located at the original position of the control region. This observation in the S. lavenbergi...
mitogenome supports the concept of the tandem duplication and subsequent deletion events as having occurred in the tRNA\textsubscript{Ile}–CR region (fig. 4).

If four of five contiguous gene blocks were duplicated together in a common ancestor of the two species, and if subsequent deletion of genes occurred before the speciation, the assumed duplicated portions of gulper eels, ranging from the tRNA\textsubscript{Ile} gene to CR of the typical organization, were extending beyond 12 kb (fig. 1). Prior to this study, the putative duplicated regions of ever-known vertebrate rearrangement were 4 kb at best. Moreover, all the known tandemly repeated coding sequences were restricted to regions adjacent to the CR, IQM, or WANCY regions, possibly reflecting occasional imprecise termination of duplications beyond their origins. The duplicated region in the gulper eel mitogenome was much greater than that of ever-known vertebrate rearrangements and involved almost the entire mitogenome (fig. 1).

Acknowledgments

We thank the captains, officers, crew, scientists, and students on board during the KT97-10 cruise of the R/V Tansei Maru and the KH00-1 cruise of the R/V Hakuyo Maru for their assistance in collecting samples. This study could not have been possible without the generous donation of S. lavenbergi tissue material, for which we sincerely thank E. O. Wiley. We also thank two anonymous reviewers for their valuable comments on the manuscript. Thanks are also due to Y. Fukuyo, K. Furuya, K. Nanba, and graduate students at the Fisheries Oceanography Laboratory, University of Tokyo, for generously allowing us to use their experimental facilities, and to Marshall Editions Ltd.; Andromeda Oxford Ltd.; and Tokai University Press for permission to use illustrations and photographs. K. Pearson kindly provided relevant information about the identity of the “S. lavenbergi” specimen. This study was supported in part by grants-in-aid from the Ministry of Education, Science and Culture of Japan (Nos. 09740644, 10460081, 10660189, 11640705, 11691177, 12460083, and 12NP0201) and “Research for the Future” Program No. JSPS-RFTF 97L00901 from the Japan Society for the Promotion of Science. K.T. was supported by the Research Foundation from Touwa Shokuhin Shinkoukai and the Eel Research Foundation from Nobori-kai.

Literature Cited


Franz Lang, Associate Editor

Accepted June 22, 2003