A Hotspot of Gene Order Rearrangement by Tandem Duplication and Random Loss in the Vertebrate Mitochondrial Genome

Diego San Mauro, David J. Gower,† Rafael Zardoya,* and Mark Wilkinson†*

*Departamento de Biodiversidad y Biología Evolutiva, Museo Nacional de Ciencias Naturales, CSIC, Madrid, Spain; and †Department of Zoology, The Natural History Museum, London, United Kingdom

Most reported examples of change in vertebrate mitochondrial (mt) gene order could be explained by a tandem duplication followed by random loss of redundant genes (tandem duplication–random loss [TDRL] model). Under this model of evolution, independent loss of genes arising from a single duplication in an ancestral species are predicted, and remnant pseudogenes expected, intermediate states that may remain in rearranged genomes. However, evidence for this is rare and largely scattered across vertebrate lineages. Here, we report new derived mt gene orders in the vertebrate "WANCY" region of four closely related caecilian amphibians. The novel arrangements found in this genomic region (one of them is convergent with the derived arrangement of marsupials), presence of pseudogenes, and positions of intergenic spacers fully satisfy predictions from the TDRL model. Our results, together with comparative data for the available vertebrate complete mt genomes, provide further evidence that the WANCY genomic region is a hotspot for gene order rearrangements and support the view that TDRL is the dominant mechanism of gene order rearrangement in vertebrate mt genomes. Convergent gene rearrangements are not unlikely in hotspots of gene order rearrangement by TDRL.

Introduction

Most animal mitochondrial (mt) genomes studied contain the same 37 genes (Boore 1999; Jameson et al. 2003), but their order is variable among and, to a lesser extent, within major groups. Of the several mechanisms proposed to explain gene order rearrangements (e.g., Moritz and Brown 1986; Pääbo et al. 1991; Macey et al. 1997), tandem duplication followed by random gene loss is generally considered the most important in vertebrates (e.g., Moritz and Brown 1986, 1987; Moritz, Dowling, and Brown 1987; Pääbo et al. 1991; Arndt and Smith 1998; Boore 2000; Inoue et al. 2003). However, evidence for this in the form of duplicated genes that either remain functional or have become pseudogenes in the process of being eliminated is rather limited (Arndt and Smith 1998; Kumazawa et al. 1998; Macey et al. 1998; Liu, Wang, and Su 2005; Mueller and Boore 2005; Zhang et al. 2005), and most quantitative methods for the phylogenetic analysis of gene order data assume other rearrangement mechanisms (e.g., Sankoff et al. 1992; Blanchette, Kunisawa, and Sankoff 1999; Cosner et al. 2000; Larget, Kadane, and Simon 2005; Larget et al. 2005).

According to the tandem duplication–random loss (TDRL) model, novel gene orders result from random deletion of one of each of the pairs of the redundant paralogs produced by a tandem duplication (Moritz, Dowling, and Brown 1987; Boore 2000). Which gene is lost is determined by the accumulation of random (but see Lavrov, Boore, and Brown 2002) mutations that disrupt normal function and create a pseudogene that is further selected against and eventually lost from the genome. Alternative mechanisms including inversion (Smith et al. 1989), transposition (Macey et al. 1997), and intramolecular recombination (Lunt and Hyman 1997) have been suggested and sometimes invoked to account for mt gene order rearrangements that cannot be explained by TDRL alone (e.g., change in encoding strand requires some inversion), particularly in invertebrates (Dowton, Castro, and Austin 2002). Importantly, none of these alternative mechanisms explains the existence of pseudogenes, which require at least one duplication and that are expected intermediate steps in changing mt gene orders under TDRL (Macey et al. 1998).

We here report new data for the "WANCY" genomic region (including one new complete mt genome) of four closely related South American caecilian amphibians (Gymnophiona), three of the five nominate species of *Siphonops* and the closely related (Taylor 1968; Wilkinson and Nussbaum 1992) monotypic *Lutkenotyphlus*. These caecilians present novel arrangements of this region, presence of pseudogenes, and positions of intergenic spacers that fully satisfy predictions from the TDRL model. Our results and comparisons across the available mt gene order data for 453 vertebrates provide further evidence that the WANCY region is a hotspot for gene order rearrangements by TDRL (Boore and Brown 1998) and suggest that TDRL has been the principal mechanism of gene order rearrangement operating in the history of the vertebrate mt genome.

Materials and Methods

Taxon Sampling and DNA Sequencing

We determined the nucleotide sequence of the complete mt genome of the caecilian amphibian *Siphonops annulatus* and an mtDNA fragment that covered the WANCY region and part of flanking genes in two other species of *Siphonops* (*Siphonops paulensis* and *Siphonops hardyi*) and in *Lutkenotyphlus brasiliensis*. The WANCY region is a cluster of five tRNA genes ($tRNA^{Trp}$, $tRNA^{Ala}$, $tRNA^{Asn}$, $tRNA^{Cys}$, and $tRNA^{Tyr}$) surrounding the origin of light-strand replication (O_L) that is located between the genes for nicotinamide adenine dinucleotide dehydrogenase subunit 2 (*ND2*) and cytochrome c oxidase subunit 1 (*COX1*) in almost all vertebrate mt genomes (Seutin et al. 1994; Boore 1999; Jameson et al. 2003).

Caecilians (order Gymnophiona) are limbless, elongate amphibians distributed throughout mostly tropical habitats in Africa, America, and Asia (Taylor 1968; Duellman and Trueb 1994). All caecilian species examined

Key words: gene rearrangement, tandem duplication, gene loss, mitochondrial genome, convergent rearrangement, Gymnophiona.

E-mail: diegos@mncn.csic.es.

Mol. Biol. Evol. 23(1):227–234. 2006
doi:10.1093/molbev/msj025
Advance Access publication September 21, 2005

Table 1
Data for Caecilian Samples Employed in This Study

Species	Region Determined	Voucher Number	Collection Locality	GenBank Accession Number
Siphonops annulatus	Complete mt genome	BMNH 2005.9	Dominguez Martins, ES, Brazil	AY954506
Siphonops paulensis	WANCY region	CHUNB 39114	Formosa, GO, Brazil	AY954507
Siphonops hardyi	WANCY region	BMNH 2005.6	Dominguez Martins, ES, Brazil	AY954508
Lutkenotyphlus brasiliensis	WANCY region	BMNH 2005.3	Sao Paulo, SP, Brazil	AY954509

NOTE.—BMNH, The Natural History Museum, London; CHUNB, Departamento de Ciências Fisiológicas, Universidade de Brasilia.

in this study belong to the so-called "higher" caecilians (Nussbaum 1991), a well-defined clade that comprises three families (Caeciliidae, Scolecomorphidae, and Typhlonectidae) of still poorly known inter- and intrafamilial phylogenetic relationships (Wilkinson 1997; Wilkinson et al. 2003; San Mauro et al. 2004, 2005).

In all cases, total DNA was purified from ethanol-preserved liver or muscle with standard phenol/chloroform extraction procedures (Sambrook, Fritsch, and Maniatis 1989), and nucleotide sequences were determined using the primers, conditions, and methods reported by San Mauro et al. (2004). Details of the employed taxa, region sequenced, voucher specimens, collection localities, and GenBank accession numbers can be found in table 1.

The sequences of other available higher caecilians (Zardoya and Meyer 2000; San Mauro et al. 2004) were used as outgroups in phylogenetic analyses (GenBank accession numbers in parentheses): *Gegeneophis ramaswamii* (AY456250), *Scolecomorphus vittatus* (AY456253), and *Typhlonectes natans* (AF154051).

Molecular and Phylogenetic Analyses

Gene boundaries were determined from sequence data by comparison with other available caecilian mt genomes using MacClade version 4.05 (W. P. Maddison and D. R. Maddison 1992) and PAUP* version 4.0b10 (Swofford 1998).

The phylogenetic relationships of the three *Siphonops* and *Lutkenotyphlus* were inferred using a concatenated data set that included all five tRNA genes of the WANCY region and fragments of the two flanking protein-coding genes (3′-end of the *ND2* gene and 5′-end of *COX1*). Sequences were manually aligned against a previous database (San Mauro et al. 2004), and gaps and ambiguous alignments (42 positions) were excluded from the data using GBLOCKS version 0.91b (Castresana 2000) with default parameters. The final alignment is 572 bp, of which 185 are parsimony informative. The sequences of all other available higher caecilians (*Gegeneophis*, *Scolecomorphus*, and *Typhlonectes*) were used as outgroups. The concatenated alignment was subjected to Bayesian inference (BI; Huelsenbeck et al. 2001), maximum likelihood (ML; Felsenstein 1981), and minimum evolution (ME; Rzhetsky and Nei 1992). All methods were executed using the General Time Reversible (Rodríguez et al. 1990) + Γ model of nucleotide substitution as selected using the Akaike information criterion (Akaike 1973) in Modeltest version 3.6 (Posada and Crandall 1998). BI analysis was conducted with MrBayes version 3.0b4 (Huelsenbeck and Ronquist 2001) simulating four simultaneous chains, for a million generations, sampling every 100 generations, and discarding generations sampled before the chain reached stationarity (100,000) as "burn-in." Statistical support for clades obtained by BI was measured by Bayesian posterior probability. Two independent BI runs were performed to verify congruence of resulting topologies and support. ML and ME analyses were carried out with PAUP*, using heuristic searches with Tree Bisection-Reconnection branch swapping and 10 random stepwise additions of taxa. Support of the resulting ML and ME trees was evaluated by nonparametric bootstrapping with 1,000 pseudoreplicates. The reconstructed phylogeny indicates that *Lutkenotyphlus* is the sister taxon of a monophyletic *Siphonops* (fig. 1). *Siphonops* monophyly is not overwhelmingly robust but receives additional support from the uniquely shared gene order.

To investigate divergence and substitution rates among $tRNA^{Asn}$ genes and pseudogenes found in the three *Siphonops* and *Lutkenotyphlus*, their nucleotide sequences were aligned, together with that of the $tRNA^{Asn}$ gene of *Gegeneophis* (as outgroup), yielding an alignment of 79 bp. Gapped positions were excluded from the alignment, and the resulting 56 sites (33 parsimony informative) were employed to reconstruct a distance phylogeny by ME using JC (Jukes and Cantor 1969) distances (no parameter-richer model was assumed because of the low number of positions analyzed). Relative-rate tests (Robinson et al. 1998) were employed to assess variations in substitution rates using RRTree version 1.1.11 (Robinson-Rechavi and Huchon 2000) assuming JC distances. Base frequencies were compared between $tRNA^{Asn}$ genes and pseudogenes of the three *Siphonops* and *Lutkenotyphlus* using analyses of variance as implemented in STATISTICA version 6.0 (StatSoft Inc. 2001).

The Department of Energy (DOE) Joint Genome Institute database (http://evogen.jgi.doe.gov/) was used to provide comparative information on the 453 complete vertebrate gene mt gene orders included as of April 2005.

Results and Discussion

Rearrangement of the WANCY Region

Our sequencing revealed two different WANCY region gene orders, both of which depart from the consensus order of vertebrates (Seutin et al. 1994; Boore 1999; Jameson et al. 2003) and other analyzed caecilians (Zardoya and Meyer 2000; San Mauro et al. 2004) (fig. 1). The WANCY gene orders of *Siphonops* and *Lutkenotyphlus* are clearly derived. Given that duplications of genes appear to be infrequent among mt genomes (Boore 2000), independent duplications of the WANCY region in

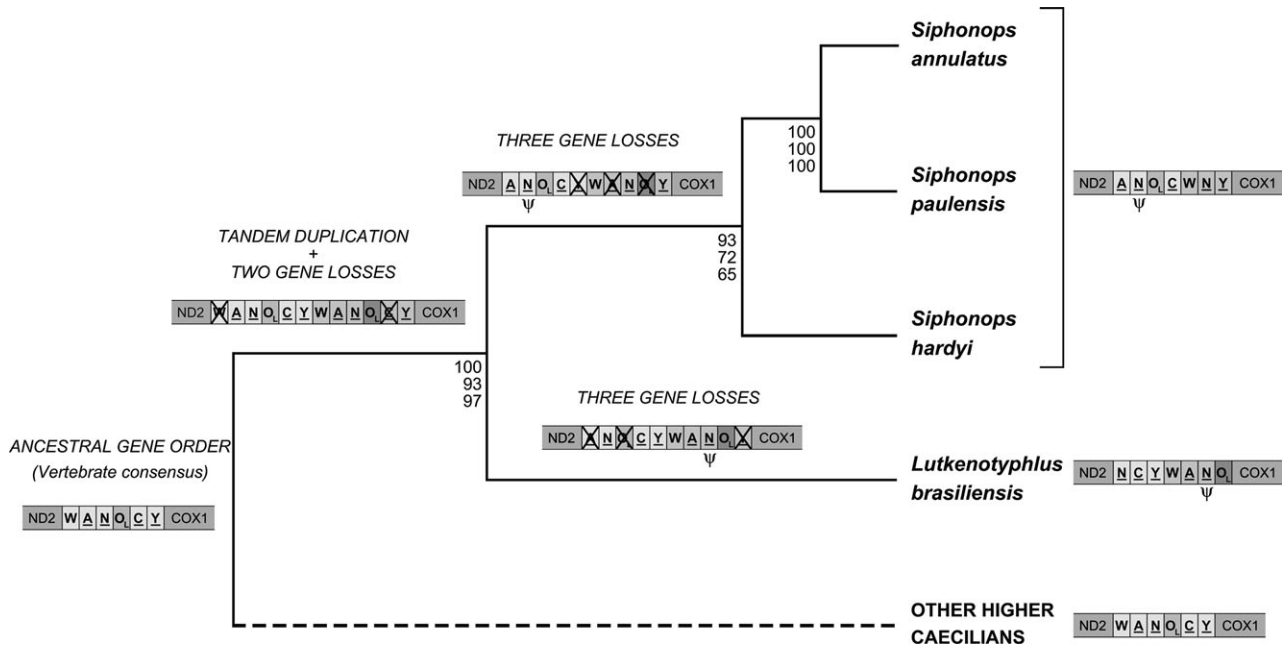

FIG. 1.—Most parsimonious reconstruction of changes producing derived mt gene orders in the WANCY region of three *Siphonops* and *Lutkenotyphlus*. tRNA genes are abbreviated by the corresponding one-letter amino acid code, and genes encoded by the light strand are underlined. ψ indicates the pseudogene. The phylogeny was inferred from a single concatenated data set with all five tRNA genes of the WANCY region and fragments of the two flanking protein-coding genes (*ND2* and *COX1*). Numbers below branches represent support for (from top to bottom) BI, ML, and ME. The derived gene order in *Lutkenotyphlus* dictates that the entire WANCY region must have been involved in the initial tandem duplication, whereas losses reconstructed parsimoniously as occurring before the divergence of *Lutkenotyphlus* and *Siphonops* might plausibly have occurred independently in these lineages.

Lutkentoyphlus and in *Siphonops* provide a less plausible explanation of the derived gene orders of these closely related caecilians than their resulting from a single ancestral tandem duplication of the entire WANCY region followed by almost instant loss of two redundant gene duplicates ($tRNA^{Trp}$, $tRNA^{Cys}$), and independent, random loss of three ($tRNA^{Ala}$, $tRNA^{Asn}$, $tRNA^{Tyr}$) redundant gene duplicates in *Siphonops* and *Lutkenotyphlus* (fig. 1). An alternative reconstruction in which all redundant duplicates are independently lost after the first speciation event (the split between *Siphonops* and *Lutkenotyphlus*) seems equally plausible (not shown).

All rearranged tRNA genes retain the ancestral strand-coding polarity, providing no evidence for inversion. In all *Siphonops* and *Lutkenotyphlus*, there are five intergenic spacers. Most of these range between 4 and 13 nt, and all are in positions expected of pseudogenes under the TDRL model (fig. 2). A more substantial intergenic spacer between the $tRNA^{Ala}$ gene and the O_L is similar to the known, functional $tRNA^{Asn}$ genes of caecilians (fig. 3), but with substantial length and substitution mutations, and can be more confidently identified as the $tRNA^{Asn}$ pseudogene predicted by the TDRL model. All other sequenced caecilian mt genomes (Zardoya and Meyer 2000; San Mauro et al. 2004) typically possess one single intergenic spacer between the WANCY genes, located between $tRNA^{Trp}$ and $tRNA^{Ala}$ (in *T. natans* the spacer is located between $tRNA^{Ala}$ and $tRNA^{Asn}$, and in *S. vittatus* there are no spacers at all between the WANCY genes). In all cases, these spacers comprise a single nucleotide.

Evolution of $tRNA^{Asn}$ Pseudogenes

Although their anticodon sequences are conserved (fig. 3*A*), the *Siphonops* and *Lutkenotyphlus* $tRNA^{Asn}$ pseudogenes have all lost the potential to fold into stable cloverleaf structures, indicating loss of primary function. Moreover, divergence among the pseudogene sequences is far greater than that for their functional paralogs (fig. 3*B*). The phylogeny of all the $tRNA^{Asn}$ paralogs (fig. 3*B*)

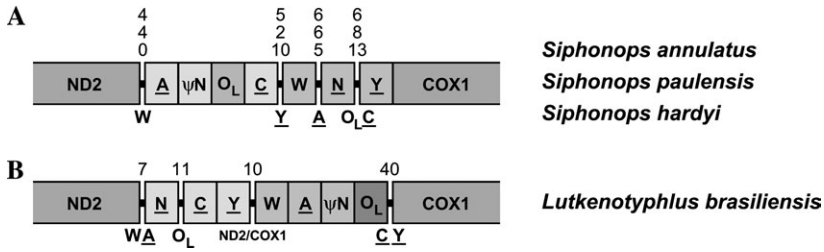

FIG. 2.—Intergenic spacers occurring around the tRNA genes (marked in black) in the WANCY region of the three *Siphonops* (*A*) and *Lutkenotyphlus* (*B*). For every intergenic spacer, its length (in bp, above) in each species and the likely lost gene (below) are shown. tRNA genes are abbreviated by the corresponding one-letter amino acid code, and genes encoded by the light strand are underlined. ψ indicates the pseudogene.

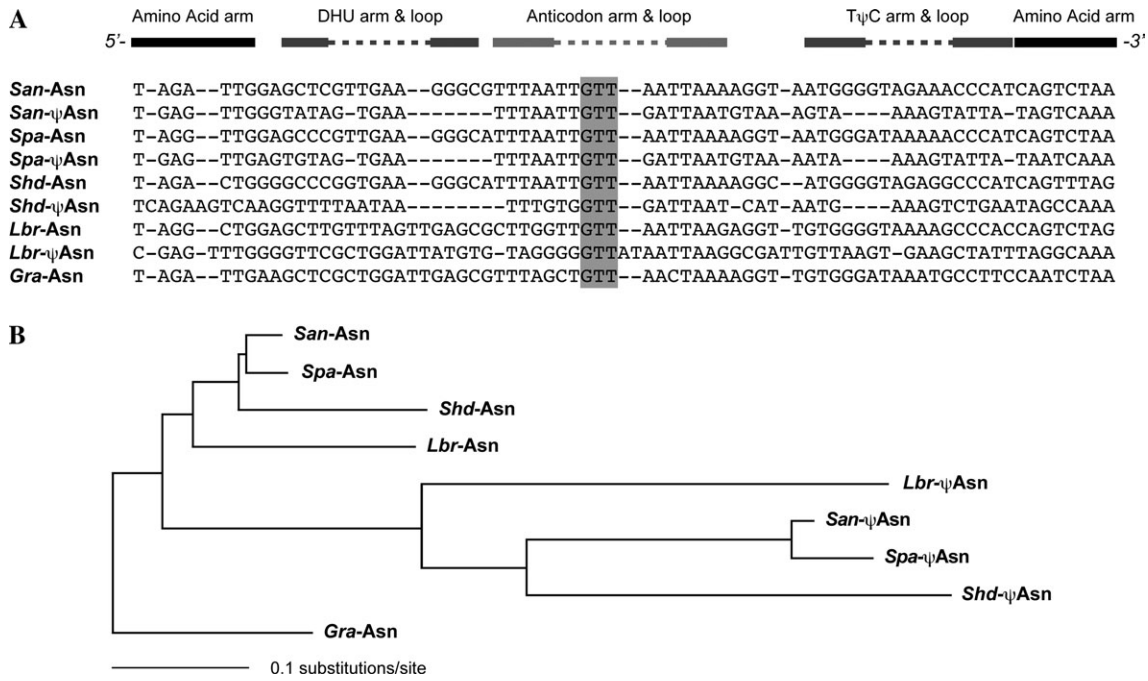

FIG. 3.—Alignment of the $tRNA^{Asn}$ genes and pseudogenes of the three *Siphonops* and *Lutkenotyphlus* (*A*), and ME phylogram inferred from the alignment (after excluding gapped positions) (*B*). tRNA secondary structure is designated above the alignment (full bar indicates arm, dashed bar indicates loop), and the position of the anticodon in highlighted. Species and gene codes are as in table 2. *Gra*, *Gegeneophis ramaswamii* (used as outgroup).

is not as expected because, although their relationships mirror the species phylogeny (fig. 1), duplicates are clustered by functionality instead of homology. Our "incorrect" gene tree probably results from the few data, long branches associated with pseudogenes, and marked base composition differences (see below) between functional and nonfunctional paralogs.

Relative-rate tests show that $tRNA^{Asn}$ pseudogenes evolved more than twice as fast as their functional paralogs (table 2). The contrast of all genes versus all pseudogenes is highly significant, and contrasts of each gene versus its pseudogene paralog are significant with the exception of *Lutkenotyphlus* (table 2). These results suggest that, following duplication, the redundant $tRNA^{Asn}$ paralogs have experienced more relaxed selective constraints (Moritz and Brown 1987). The $tRNA^{Asn}$ pseudogenes have a lower and higher frequency of C and T, respectively (1.7%–9.5% vs. 11.3%–14.3%, $F_{1,6} = 12.760$, $P = 0.012$; 33.3%–38.3% vs. 25.7%–31.5%, $F_{1,6} = 14.086$, $P = 0.010$), than their functional paralogs. Assuming relaxed selection, these biases provide further evidence for asymmetric mutation pressures in mt genomes (Jermiin, Graur, and Crozier 1995).

The pseudogene remnants predicted by TDRL are uncommon in known mt genomes (e.g., Macey et al. 1998; Mueller and Boore 2005; Zhang et al. 2005), consistent with the idea that they are lost rapidly under strong selective pressure to constrain mt genome size and gene number (Wolstenholme 1992). Persistence of an ancestral tandem duplication through a speciation event with subsequent independent random loss of paralogs is a predicted rare event under the TDRL model (Boore 2000) for which our caecilian data may provide the first evidence. Similarly, the $tRNA^{Asn}$ pseudogenes of multiple caecilian lineages provide powerful evidence for TDRL while simultaneously prompting questions about their persistence. $tRNA^{Asn}$ is not distinct from the other four tRNAs in its length and usage, and the $tRNA^{Asn}$ gene is no more or less variable than other caecilian tRNA genes. Their adjacency to O_L is the only obvious variable that correlates with the persistence of these pseudogenes. It may be possible that they (or part of them) have acquired some type of functional role perhaps related to the O_L. Zardoya and Meyer (2000) reported that the O_L of another caecilian, *T. natans*, has the potential to fold into alternative secondary structures with the adjacent $tRNA^{Cys}$. However, similar alternative stem-loop structures have not been found in the caecilian sequences reported here, and the persistence of the pseudogenes is somewhat enigmatic.

Table 2
Results of the Relative-Rate Test for Contrasts Between $tRNA^{Asn}$ Genes and Pseudogenes

Contrast	Rates	SD	*P* Value
All Asn versus all ψAsn	0.320 versus 0.703	0.124	0.002*
San-Asn versus *San*-ψAsn	0.252 versus 0.724	0.165	0.004*
Spa-Asn versus *Spa*-ψAsn	0.278 versus 0.678	0.158	0.011*
Sha-Asn versus *Sha*-ψAsn	0.389 versus 0.772	0.193	0.047*
Lbr-Asn versus *Lbr*-ψAsn	0.360 versus 0.636	0.160	0.085

NOTE.—Results of all possible pairwise contrasts among $tRNA^{Asn}$ genes and among $tRNA^{Asn}$ pseudogenes are nonsignificant ($P > 0.05$). SD, standard deviation; *San*, *Siphonops annulatus*; *Spa*, *Siphonops paulensis*; *Sha*, *Siphonops hardyi*; *Lbr*, *Lutkenotyphlus brasiliensis*; Asn, $tRNA^{Asn}$ gene; and ψAsn, $tRNA^{Asn}$ pseudogene.

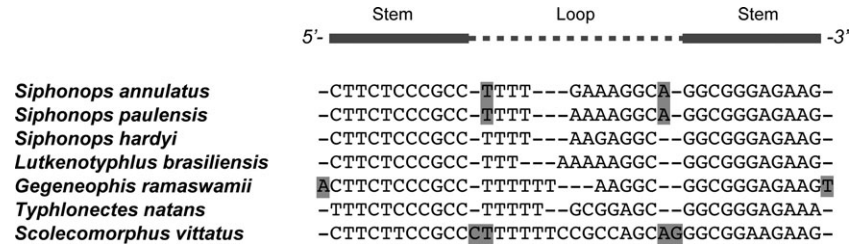

FIG. 4.—Alignment of the O_L's of the three *Siphonops* and *Lutkenotyphlus* (and other higher caecilians). Consensus stem-loop secondary structure is designated above the alignment. Positions highlighted, although outside the indicated consensus stem region, are also part of the stem in those species having them.

Convergence in a Hotspot of Gene Rearrangement

Gene order arrangements may provide exceptionally useful data for phylogenetic inference because of both the relative rarity of rearrangements and the potential complexity of the characters and consequent large character state space (Macey et al. 1997; Boore and Brown 1998; Dowton, Castro, and Austin 2002). These features reduce the chances of homoplasy, and only four convergent derived gene orders among metazoan mitochondria have been previously reported (Flook, Rowell, and Gellissen 1995; Mindell, Sorenson, and Dimcheff 1998; Dowton and Austin 1999; Macey et al. 2004). The derived WANCY region of *Siphonops* is exceptionally similar to that of marsupials (Pääbo et al. 1991), and the order of functional tRNA genes is identical, providing a fifth example of such convergence. The three *Siphonops* differ from marsupials in having a complete O_L in the ancestral vertebrate position relative to the WANCY tRNAs (fig. 1), with nucleotides and stem-loop structures similar to those of other caecilians (fig. 4). This suggests that their O_L's are not secondarily derived, "drifted" duplicated tRNA genes like that found in the derived WANCY region of marsupials (Pääbo et al. 1991). Thus, contrary to previous proposals (Macey et al. 1997, 1998), displacement or loss of the O_L does not always precede vertebrate mt gene order change by tandem duplication.

Tandem duplication can occur during replication by slipped-strand mispairing (Levinson and Gutman 1987) or by illicit priming of replication by tRNAs (Cantatore et al. 1987). In mt genomes, these are thought to particularly involve stem-loop structures and thus to most commonly involve regions including tRNA genes and/or near the origins of replication of the light (O_L in the vertebrate WANCY region) and heavy (O_H in the vertebrate control region) strands (e.g., Moritz and Brown 1987; Pääbo et al. 1991; Stanton et al. 1994; Macey et al. 1997, 1998; Kumazawa et al. 1998; Mindell, Sorenson, and Dimcheff 1998; Boore 1999). Previous studies have cautioned that tandem duplications and gene deletions may be subject to mechanistic constraints such that genes flanking the origins of strand replication are more likely to be duplicated, forming "hotspots" that make convergent gene order rearrangement more probable (Boore and Brown 1998; Mindell, Sorenson, and Dimcheff 1998; Dowton and Austin 1999; Boore 2000).

Ignoring deleted genes and random duplicates, the 453 vertebrate mt genomes in the DOE Joint Genome Institute database display 31 distinct gene orders, with most (368) conforming to the vertebrate consensus. Of the 30 derived gene orders, 4 involve the WANCY region and 26 are found elsewhere in the mt genome. For simplicity, we do not consider rearrangements that involve both the WANCY and other adjacent genomic regions, those evidenced by the genomes of the worm snake *Leptotyphlops dulcis* (Kumazawa and Nishida 1995) and the gluper eels *Eurypharynx pelecanoides* and *Saccopharynx lavenbergi* (Inoue et al. 2003). The four derived gene orders of the WANCY region can be explained by a single TDLR (table 3). Our

Table 3
The 32 Possible Outcomes from Deleting Redundant Gene Copies Subsequent to a Single Tandem Duplication of the Entire Ancestral $W_1A_1N_1C_1Y_1$ Region to Produce $W_1A_1N_1C_1Y_1W_2A_2N_2C_2Y_2$

$W_1A_1N_1C_1Y_1$	Vertebrate consensus
$W_1A_1N_1C_1Y_2$	Vertebrate consensus
$W_1A_1N_1Y_1C_2$	*Chauliodus sloani* (viperfish; Miya, Kawaguchi, and Nishida 2001)
$W_1A_1C_1Y_1N_2$	
$W_1N_1C_1Y_1A_2$	*Hydromantes brunus* (salamander; Mueller et al. 2004)
$A_1N_1C_1Y_1W_2$	
$W_1A_1N_1C_2Y_2$	Vertebrate consensus
$W_1A_1C_1N_2Y_2$	
$W_1A_1Y_1N_2C_2$	*Batrachoseps attenuatus* (salamander; Mueller et al. 2004)
$W_1N_1C_1A_2Y_2$	
$W_1N_1Y_1A_2C_2$[a]	
$W_1C_1Y_1A_2N_2$	
$A_1N_1C_1W_2Y_2$	
$A_1N_1Y_1W_2C_2$[a]	
$A_1C_1Y_1W_2N_2$[a]	
$N_1C_1Y_1W_2A_2$	*Lutkenotyphlus brasiliensis* (caecilian, this study)
$W_1A_1N_2C_2Y_2$	Vertebrate consensus
$W_1N_1A_2C_2Y_2$	
$W_1C_1A_2N_2Y_2$	
$W_1Y_1A_2N_2C_2$	
$A_1N_1W_2C_2Y_2$	
$A_1C_1W_2N_2Y_2$[a]	*Siphonops* species (caecilian, this study; marsupials, Pääbo et al. 1991)
$A_1Y_1W_2N_2C_2$[a]	
$N_1C_1W_2A_2Y_2$	
$N_1Y_1W_2A_2C_2$[a]	
$C_1Y_1W_2A_2N_2$	
$W_1A_2N_2C_2Y_2$	Vertebrate consensus
$A_1W_2N_2C_2Y_2$	
$N_1W_2A_2C_2Y_2$	
$C_1W_2A_2N_2Y_2$	
$Y_1W_2A_2N_2C_2$	
$W_2A_2N_2C_2Y_2$	Vertebrate consensus

[a] A gene order that cannot be explained by a single transposition affected the vertebrate consensus. Although a part of most WANCY regions, the O_L is not considered here. In all cases, it is possible that pseudogenes may remain.

new data for caecilians provide evidence of two derived arrangements of the WANCY region that are also readily explained by the TDRL model of gene order rearrangement. Approximately 15% of all known derived arrangements of the vertebrate mt gene order are explicable in terms of TDRLs of the WANCY region, consistent with the hypothesis that this region may be a mechanistic hotspot of gene duplication by virtue of its association with the O_L.

However, for a tandem duplication to produce a gene order rearrangement, the duplicated region must include more than one complete gene, and the chances of rearrangement are increased with the number of duplicated genes. Thus, tandem duplications are more likely to be detectable in regions, such as the WANCY cluster, with relatively many, small genes, making such regions potential epistemic hotspots. With additional data, it may be possible to address whether rearrangements of the WANCY cluster are significantly more common than expected for any cluster of five small genes and thus better test the hypothesis that the region is a mechanistic hotspot of gene order rearrangement.

Likelihood of Gene Order Change

Dowton, Castro, and Austin (2002) have discussed the probability of convergence in mt gene orders under a "cut and paste" model of gene transposition and inversion. Here, we consider the probability of the observed convergence in the order of tRNAs in the WANCY regions of *Siphonops* and marsupials under the TDRL model. Although which of each of a pair of paralogs is lost or retained subsequent to a single duplication is in principle random (but see Lavrov, Boore, and Brown 2002), those retained from the same duplicate must preserve the original relative order. This leads to some differences in expectations for the TDRL and transposition models. In particular, whereas any derived gene order arrangements that can be explained by a single transposition can also be explained by a single TDRL, some TDRLs produce arrangements that cannot be explained by single transpositions (table 3).

Ignoring changes in the coding strand, there are 120 (5!) possible orders of the five tRNA genes of the WANCY region, suggesting a large character space and low probability of convergence. However, less than a quarter of the arrangements can be produced from the vertebrate consensus by a single TDRL, constraining the character state space and increasing the chance of convergence. There are 32 (2^5) possible random selections of one from each pair of paralogs of a tandemly duplicated WANCY region (not including the O_L) that yield 27 distinct gene orders (table 3). TDRLs of smaller parts of the WANCY region would not add to these 27 different gene orders. Note that six of the 26 derived gene orders cannot be explained by a single transposition (table 3). Note also that six random selections return the original order, so that approximately one-fifth of all WANCY region TDRLs are expected to be undetectable (table 3). In general, for n genes, the probability of undetected TDRLs is $(n + 1)/2^n$. Thus, with fewer genes, the chances of TDRLs being undetected are higher. For example, only one in four TDRLs of two genes yield rearrangements.

The majority, 93 of 119, possible derived gene orders of the vertebrate WANCY region are prohibited by a single TDRL (i.e., require either multiple TDRLs and/or alternative mechanisms of gene order change), but all six currently documented independently derived gene orders found in the WANCY clusters of vertebrates are ones that are permitted by a single TDRL (table 3). Of these six, the convergent WANCY gene orders of marsupial and *Siphonops* cannot be explained by single transpositions, providing further evidence that they have arisen through TDRL. In fact, the conditional probability of at least one convergence given six independent rearrangements produced by single TDRLs of the WANCY region is 0.463 $(1 - ((25/26)\cdot(24/26)\cdot(23/26)\cdot(22/26)\cdot(21/26)))$, so that the observed convergence is hardly surprising given the probable mode of origin.

The Importance of TDRL in Vertebrate mt Evolution

Our data provide compelling evidence, both from the pattern of gene orders and the presence of pseudogenes and intergenic spacers in the positions predicted by the model, that derived caecilian mt gene orders in the WANCY region have evolved through TDRL. Comparing published vertebrate mt gene orders (of 453 complete mt genomes), we find that 24 of the 30 derived arrangements can each be explained by a single TDRL and that the six exceptions can each be explained by two TDRLs. Several of these derived gene orders, like those of *Siphonops* and marsupials, can be explained by a single TDRL but alternatively require multiple transpositions. For example, the highly divergent mt gene order of the gulper eels *E. pelecanoides* and *S. lavenbergi* can be derived from the vertebrate consensus by a single TDRL (Inoue et al. 2003) or by five transpositions. These observations are consistent with the view that TDRL is the dominant mechanism of gene rearrangement in vertebrate mt genomes (e.g., Boore 2000).

Rare genomic changes have attracted great interest because of their potential to provide homoplasy-free evidence of phylogenetic relationships (e.g., Rokas and Holland 2000). Of course, the likelihood of convergence depends on just how rare such changes are, and changes in gene order are not so infrequent that homoplasy is nonexistent (Dowton and Austin 1999; Inoue et al. 2003; Mueller and Boore 2005). The above considerations suggest that convergence in gene order may be more or less common depending also on the mechanism of rearrangement and the mt genomic region considered (Dowton and Austin 1999; Boore 2000; Dowton, Castro, and Austin 2002) and that duplication events may be more or less detectable. In particular, it may be unsurprising if hotspots of tandem duplication coincide with clusters of small genes within which gene order rearrangement is more likely to accompany tandem duplications (Boore 1999). As Darwin (1859) cautioned in *On the Origin of Species*, classifications based on single characters have always failed. Empirical evidence on the relative importance of different mechanisms of gene order rearrangement should provide a basis for more realistic models of gene order rearrangements and best use of comparative gene order data for phylogenetic inference.

Acknowledgments

We are grateful to C. Jared, E. F. Schwartz, C. Schwartz, A. Sebben, and the late C. Zamprogno for invaluable help in obtaining tissue samples. J. A. Cotton and two anonymous reviewers made insightful comments on an earlier version of the manuscript. D.S.M. was sponsored by a predoctoral fellowship of the Ministerio de Educación y Ciencia of Spain and was granted a SYNTHESYS award (Sixth Framework Programme of the European Union) to visit the Natural History Museum. D.J.G. and M.W. were granted BIODIBERIA awards (Fifth Framework Programme of the European Union) to visit the Museo Nacional de Ciencias Naturales. This work received financial support from a project of the Ministerio de Educación y Ciencia of Spain (CGL2004-00401). Fieldwork was supported by National Environmental Research Council GST/02/832.

Literature Cited

Akaike, H. 1973. Information theory as an extension of the maximum likelihood principle. Pp. 267–281 *in* B. N. Petrov and F. Csaki, eds. Second International Symposium of Information Theory. Akademiai Kiado, Budapest, Hungary.

Arndt, A., and M. J. Smith. 1998. Mitochondrial gene rearrangement in the sea cucumber genus *Cucumaria*. Mol. Biol. Evol. **15**:1009–1016.

Blanchette, M., T. Kunisawa, and D. Sankoff. 1999. Gene order breakpoint in animal mitochondrial phylogeny. J. Mol. Evol. **49**:193–203.

Boore, J. L. 1999. Animal mitochondrial genomes. Nucleic Acids Res. **27**:1767–1780.

———. 2000. The duplication/random loss model for gene rearrangement exemplified by mitochondrial genomes of deuterostome animals. Pp. 133–147 *in* D. Sankoff and J. Nadeau, eds. Comparative genomics, computational biology series. Volume 1. Kluwer Academic Publishers, Dordrecht, Netherlands.

Boore, J. L., and W. M. Brown. 1998. Big trees from little genomes: mitochondrial gene order as a phylogenetic tool. Curr. Opin. Genet. Dev. **8**:668–674.

Cantatore, P., M. N. Gadaleta, M. Roberti, C. Saccone, and A. C. Wilson. 1987. Duplication and remoulding of tRNA genes during the evolutionary rearrangement of mitochondrial genomes. Nature **329**:853–855.

Castresana, J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol. Biol. Evol. **17**:540–552.

Cosner, M. E., R. K. Jansen, B. M. E. Moret, L.-S. Wang, T. Warnow, and S. Wyman. 2000. An empirical comparison of phylogenetic methods on chloroplast gene order data in Campanulaceae. Pp. 99–121 *in* D. Sankoff and J. Nadeau, eds. Comparative genomics, computational biology series. Volume 1. Kluwer Academic Publishers, Dordrecht, Netherlands.

Darwin, C. 1859. On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life. John Murray, London.

Dowton, M., and A. D. Austin. 1999. Evolutionary dynamics of a mitochondrial rearrangement "hot spot" in the Hymenoptera. Mol. Biol. Evol. **16**:298–309.

Dowton, M., L. R. Castro, and A. D. Austin. 2002. Mitochondrial gene rearrangements as phylogenetic characters in the invertebrates: the examination of the genome "morphology". Invertebr. Syst. **16**:345–356.

Duellman, W. E., and L. Trueb. 1994. Biology of amphibians. Johns Hopkins University Press, Baltimore, Md.

Felsenstein, J. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. J. Mol. Evol. **17**:368–376.

Flook, P., H. Rowell, and G. Gellissen. 1995. Homoplastic rearrangements of insect mitochondrial tRNA genes. Naturwissenschaften **82**:336–337.

Huelsenbeck, J. P., and F. R. Ronquist. 2001. MrBayes: Bayesian inference of phylogeny. Bioinformatics **17**:754–755.

Huelsenbeck, J. P., F. R. Ronquist, R. Nielsen, and J. P. Bollback. 2001. Bayesian inference of phylogeny and its impact on evolutionary biology. Science **294**:2310–2314.

Inoue, J. G., M. Miya, K. Tsukamoto, and M. Nishida. 2003. Evolution of the deep-sea gulper eel mitochondrial genomes: large-scale gene rearrangements originated within the eels. Mol. Biol. Evol. **20**:1917–1924.

Jameson, D., A. P. Gibson, C. Hudelot, and P. G. Higgs. 2003. OGRe: a relational database for comparative analyses of mitochondrial genomes. Nucleic Acids Res. **31**:202–206.

Jermiin, L., D. Graur, and R. H. Crozier. 1995. Evidence from analyses of intergenic regions for strand-specific directional mutation pressure in metazoan mtDNA. Mol. Biol. Evol. **12**:558–563.

Jukes, T. H., and C. R. Cantor. 1969. Evolution of protein molecules. Pp. 21–132 *in* H. N. Munro, ed. Mammalian protein metabolism. Academic Press, New York.

Kumazawa, Y., and M. Nishida. 1995. Variation in mitochondrial tRNA gene organization of reptiles as phylogentic markers. Mol. Biol. Evol. **12**:759–772.

Kumazawa, Y., H. Ota, M. Nishida, and T. Ozawa. 1998. The complete nucleotide sequence of snake (*Dinodon semicarinatus*) mitochondrial genome with two identical control regions. Genetics **150**:313–329.

Larget, B., J. B. Kadane, and D. L. Simon. 2005. A Bayesian approach to the estimation of ancestral genome arrangements. Mol. Phylogenet. Evol. **36**:214–223.

Larget, B., D. L. Simon, J. B. Kadane, and D. Sweet. 2005. A Bayesian analysis of metazoan mitochondrial genome arrangements. Mol. Biol. Evol. **22**:486–495.

Lavrov, D. V., J. L. Boore, and W. M. Brown. 2002. Complete mtDNA sequences of two millipedes suggest a new model for mitochondrial gene rearrangements: duplication and nonrandom loss. Mol. Biol. Evol. **19**:163–169.

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

Liu, Z. Q., Y. Q. Wang, and B. Su. 2005. The mitochondrial genome organization of the rice frog, *Fejervarya limnocharis* (Amphibia: Anura): a new gene order in the vertebrate mtDNA. Gene **346**:145–151.

Lunt, D. H., and B. C. Hyman. 1997. Animal mitochondrial DNA recombination. Nature **387**:247.

Macey, J. R., A. Larson, N. B. Ananjeva, Z. Fang, and T. J. Papenfuss. 1997. Two novel gene orders and the role of light-strand replication in rearrangement of the vertebrate mitochondrial genome. Mol. Biol. Evol. **14**:91–104.

Macey, J. R., T. J. Papenfuss, J. V. Kuehl, H. M. Fourcade, and J. L. Boore. 2004. Phylogenetic relationships among amphisbaenian reptiles based on complete mitochondrial genomic sequences. Mol. Phylogenet. Evol. **33**:22–31.

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

Maddison, W. P., and D. R. Maddison. 1992. MacClade: analysis of phylogeny and character evolution. Sinauer Associates Inc., Sunderland, Mass.

Mindell, D. P., M. D. Sorenson, and D. E. Dimcheff. 1998. Multiple independent origins of mitochondrial gene order in birds. Proc. Natl. Acad. Sci. USA **95**:10693–10697.

Miya, M., A. Kawaguchi, and M. Nishida. 2001. Mitogenomic exploration of higher teleostean phylogenies: a case study for moderate-scale evolutionary genomics with 38 newly determined complete mitochondrial DNA sequences. Mol. Biol. Evol. **18**:1993–2009.

Moritz, C., and W. M. Brown. 1986. Tandem duplications of D-loop and ribosomal RNA sequences in lizard mitochondrial DNA. Science **233**:1425–1427.

———. 1987. Tandem duplications in animal mitochondrial DNAs: variation in incidence and gene content among lizards. Proc. Natl. Acad. Sci. USA **84**:7183–7187.

Moritz, C., T. E. Dowling, and W. M. Brown. 1987. Evolution of animal mitochondrial DNA: relevance for population biology and systematics. Annu. Rev. Ecol. Syst. **18**:269–292.

Mueller, R. L., and J. L. Boore. 2005. Molecular mechanisms of extensive mitochondrial gene rearrangement in plethodontid salamanders. Mol. Biol. Evol. **22**:2104–2112.

Mueller, R. L., J. R. Macey, M. Jaekel, D. B. Wake, and J. L. Boore. 2004. Morphological homoplasy, life history evolution, and historical biogeography of plethodontid salamanders inferred from complete mitochondrial genomes. Proc. Natl. Acad. Sci. USA **101**:13820–13825.

Nussbaum, R. A. 1991. Cytotaxonomy of caecilians. Pp. 22–76 *in* S. K. Sessions and D. M. Green, eds. Amphibian cytogenetics and evolution. Academic Press, San Diego, Calif.

Pääbo, S., W. K. Thomas, K. M. Whitfield, and Y. Kumazawa. 1991. Rearrangements of mitochondrial transfer RNA genes in marsupials. J. Mol. Evol. **33**:426–430.

Posada, D., and K. A. Crandall. 1998. MODELTEST: testing the model of DNA substitution. Bioinformatics **14**:817–818.

Robinson, M., M. Gouy, C. Gautier, and D. Mouchiroud. 1998. Sensitivity of the relative-rate test to taxonomic sampling. Mol. Biol. Evol. **15**:1091–1098.

Robinson-Rechavi, M., and D. Huchon. 2000. RRTree: relative-rate tests between groups of sequences on a phylogenetic tree. Bioinformatics **16**:296–297.

Rodríguez, F., J. F. Oliver, A. Marín, and J. R. Medina. 1990. The general stochastic model of nucleotide substitution. J. Theor. Biol. **142**:485–501.

Rokas, A., and P. W. H. Holland. 2000. Rare genomic changes as a tool for phylogenetics. Trends Ecol. Evol. **15**:454–459.

Rzhetsky, A., and M. Nei. 1992. A simple method for estimating and testing minimum-evolution trees. Mol. Biol. Evol. **9**: 945–967.

Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989. Molecular cloning. A laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

San Mauro, D., D. J. Gower, O. V. Oommen, M. Wilkinson, and R. Zardoya. 2004. Phylogeny of caecilian amphibians (Gymnophiona) based on complete mitochondrial genomes and nuclear RAG1. Mol. Phylogenet. Evol. **33**:413–427.

San Mauro, D., M. Vences, M. Alcobendas, R. Zardoya, and A. Meyer. 2005. Initial diversification of living amphibians predated the breakup of Pangaea. Am. Nat. **165**:590–599.

Sankoff, D., G. Leduc, N. Antoine, B. Paquin, B. F. Lang, and R. Cedergren. 1992. Gene order comparisons for phylogenetic inference: evolution of the mitochondrial genome. Proc. Natl. Acad. Sci. USA **89**:6575–6579.

Seutin, G., B. F. Lang, D. P. Mindell, and R. Morais. 1994. Evolution of the WANCY region in amniote mitochondrial DNA. Mol. Biol. Evol. **11**:329–340.

Smith, M. J., D. K. Banfield, K. Doteval, S. Gorski, and D. J. Kowbel. 1989. Gene arrangement in sea star mitochondrial DNA demonstrates a major inversion event during echinoderm evolution. Gene **76**:181–185.

Stanton, D. J., L. L. Daehler, C. C. Moritz, and W. M. Brown. 1994. Sequences with the potential to form stem-and-loop structures are associated with coding-region duplications in animal mitochondrial DNA. Genetics **137**:233–241.

StatSoft Inc. 2001. STATISTICA (data analysis software system). (http://www.statsoft.com).

Swofford, D. L. 1998. PAUP*: phylogenetic analysis using parsimony (*and other methods). Version 4.0. Sinauer Associates, Inc., Sunderland, Mass.

Taylor, E. H. 1968. The caecilians of the world: a taxonomic analysis. University of Kansas Press, Lawrence, Kans.

Wilkinson, M. 1997. Characters, congruence and quality: a study of neuroanatomical and traditional data in caecilian phylogeny. Biol. Rev. **72**:423–470.

Wilkinson, M., S. P. Loader, D. J. Gower, J. A. Sheps, and B. L. Cohen. 2003. Phylogenetic relationships of African caecilians (Amphibia: Gymnophiona): insights from mitochondrial rRNA gene sequences. Afr. J. Herpetol. **52**:83–92.

Wilkinson, M., and R. A. Nussbaum. 1992. Taxonomic status of *Pseudosiphonops ptychodermis* Taylor and *Mimosiphonops vermiculatus* Taylor (Amphibia: Gymnophiona: Caeciliaidae) with a description of a new species. J. Nat. Hist. **26**: 675–688.

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

Zardoya, R., and A. Meyer. 2000. Mitochondrial evidence on the phylogenetic position of caecilians (Amphibia: Gymnophiona). Genetics **155**:765–775.

Zhang, P., H. Zhou, D. Liang, Y. F. Liu, Y. Q. Chen, and L. H. Qu. 2005. The complete mitochondrial genome of a tree frog, *Polypedates megacephalus* (Amphibia: Anura: Rhacophoridae), and a novel gene organization in living amphibians. Gene **346**:133–143.

Franz Lang, Associate Editor

Accepted September 15, 2005