

## Mitochondrial genomic divergence in coelacanths (*Latimeria*): slow rate of evolution or recent speciation?

Sudarto · Xandramaya C. Lalu · Janny D. Kosen · Agus H. Tjakrawidjaja · Ruby Vidia Kusumah · Bambang Sadhotomo · Kadarusman · Laurent Pouyaud · Jacques Slembrouck · Emmanuel Paradis

Received: 21 November 2009 / Accepted: 2 June 2010 / Published online: 18 June 2010  
© Springer-Verlag 2010

**Abstract** Dating the divergence between the two known living species of coelacanths has remained a difficult issue because of the very ancient origin of this lineage of fish, which is more closely related to tetrapods than to other fishes. We sequenced the complete mitochondrial genome of a recently captured individual of the Indonesian coelacanth in order to solve this issue. Using an approach based on loglinear models, we studied the molecular divergence between the two species of coelacanths and three other pairs of species, one that has diverged recently (*Pan*) and two that have diverged more distantly in the past. The

loglinear models showed that the divergence between the two species of coelacanths is not significantly different from the two species of *Pan*. A detailed gene by gene analysis of the patterns of nucleotide and amino acid substitutions between these two pairs of species further supports the similarity of these divergences. On the other hand, a molecular dating analysis suggested a much older origin of the two coelacanth species (between 20 and 30 million years ago). We discuss the potential reasons for this discrepancy. The analysis of new individuals of the Indonesian coelacanth will certainly help to solve this issue.

---

Communicated by M. I. Taylor.

---

Sudarto  
Loka Riset Budidaya Ikan Hias Air Tawar,  
Jl. Perikanan No. 13 RT, 001 RW 02, Pancoranmas Depok,  
Jawa Barat 16436, Indonesia

X. C. Lalu  
Dinas Kelautan dan Perikanan, Sulawesi Utara, Jl. Komp.  
Pertanian Kalasey, PO BOX 1039, Manado 5013, Indonesia

J. D. Kosen  
Centre of Research for Coastal Management,  
Small Islands and Pacific Affair, Sam Ratulangi University,  
Manado, North Sulawesi, Indonesia

A. H. Tjakrawidjaja  
Zoology Division, Research Center for Biology,  
Indonesian Institute of Sciences (LIPI), Cibinong,  
Bogor 16911, Indonesia

R. V. Kusumah  
Faculty of Fisheries and Marine Science,  
Bogor Agricultural University (IPB), Bogor,  
Jawa Barat, Indonesia

B. Sadhotomo  
Pusat Riset Perikanan Tangkap, BRKP-DKP,  
Jl. Pasir Putih, Ancol Jakarta Utara, Indonesia

Kadarusman  
Akademi Perikanan Sorong (APSOR-BPSDM.KP),  
Jl. Kapitan Pattimura, Tanjung Kasuari,  
Kota Sorong, Papua Barat, Indonesia

L. Pouyaud · E. Paradis (✉)  
Institut de Recherche pour le Développement,  
UR175 CAVIAR, GAMET, 361 rue Jean-François Breton,  
BP 5095, 34196 Montpellier Cédex 5, France  
e-mail: Emmanuel.Paradis@ird.fr

J. Slembrouck  
Institut de Recherche pour le Développement,  
Wisma Anugraha, Jl. Taman Kemang 32B,  
Jakarta 12730, Indonesia

## Introduction

Molecular sequences are a rich source of information for dating the divergence of species, especially when paleontological or biogeographical evidence are not available. However, molecular dating remains a difficult methodological field, and many molecular date estimates have remained contentious (Pulquério and Nichols 2007; Miller et al. 2009; Subramanian et al. 2009). One of the difficulties is the interpretation of fossils used as calibration points in molecular phylogenies (Parham and Irmis 2008). Benton and Donoghue (2007) reviewed a large literature on fossils in order to provide a set of standard calibration points for metazoans. Santini et al. (2009) complemented this work by providing a list of calibration points for fishes. However, the use of molecular dating methods remains difficult if one is interested in lineages where no calibration point is available.

The Coelacanthimorpha is a subclass of vertebrates that has flourished from the Paleozoic (Upper Devonian) until the end of the Mesozoic (Upper Cretaceous). Among the nine recognized families (Nelson 2006), only one, Latimeriidae, has known living representatives with two currently described species: *Latimeria chalumnae* (Smith 1939) and *L. menadoensis* (Pouyaud et al. 1999). Since the Coelacanthimorpha originated ca. 400 million years ago (Myr), the study of the divergence among living coelacanths is a unique opportunity to investigate the evolutionary dynamics of this lineage considered as “living fossils”. During about a decade, only a single specimen of *L. menadoensis* was available for scientific investigation since the first specimen was not conserved (Erdmann 1999; Erdmann et al. 1999). The capture of a third coelacanth off Manado (North Sulawesi, Indonesia) in May 2007 gave a new opportunity to investigate the divergence between the two species of *Latimeria*.

There are two main hypotheses on the origin of the living species of coelacanths. Springer (1999) hypothesized that a large population of coelacanths present in the Indian Ocean was fragmented after the collision of the Indian subcontinent with the Eurasian plate 50 Myr. From this hypothesis, the date of the divergence between *L. chalumnae* and *L. menadoensis* is expected to be around 40–50 Myr. On the other hand, Gordon (1998) pointed out the existence of a transoceanic flow from the West Pacific up to the Southwest Indian Ocean, so that the divergence between both species may be much more recent because coelacanths may disperse, more or less passively, along this transoceanic current. However, dating this divergence is a difficult exercise because this speciation event could have occurred, at least in theory, anytime between 400 Myr and the present. Thus,

molecular dating methods that use calibration points are difficult to apply because these points will be far from the focus event. Pouyaud et al. (1999) used a simple molecular clock method with two mitochondrial genes (12S and cytochrome *b*) and estimated the divergence date between both species between 1.22 and 1.42 Myr. Holder et al. (1999), using a larger set of mtDNA sequences (4,823 bp), found older estimates: 4.7–6.3 Myr. Inoue et al. (2005) used the complete sequence of the mitochondrial genome (mtGenome) and Bayesian methods to estimate the same divergence at around 40 Myr. It should be noted that in these three studies, the DNA of the same individual of *L. menadoensis* was used. So, the present evidence is not definitive, and it does not seem possible to reject either hypothesis.

In the present paper, we report a new assessment of this issue using the data of the recently captured individual of *L. menadoensis*. Specifically, we studied the divergence between *L. chalumnae* and *L. menadoensis* by investigating in detail the mitochondrial genomic divergence between these two species. We compared the divergence between the two species of *Latimeria* and the divergence between other pairs of species for which complete mtGenome sequences are available and the divergence date between them is known. We selected the two species of chimpanzees: *Pan troglodytes* and *P. paniscus*. Various molecular studies have dated the divergence between these two species between 0.86 and 2.4 Myr (Raaum et al. 2005; Won and Hey 2005; Perry et al. 2007). Furthermore, hominid fossils give an older bound for this date since the species of chimpanzees have speciated after the common ancestor of *Homo* and *Pan*. Benton and Donoghue (2007) recognized *Sahelanthropus*, dated between 7 and 10 Myr, as this common ancestor. Consequently, the most recent common ancestor of *P. troglodytes* and *P. paniscus* cannot be older than 10 Myr. Therefore, if the divergence between *L. chalumnae* and *L. menadoensis* is not significantly greater than the divergence between *P. troglodytes* and *P. paniscus*, we may reject the hypothesis that the two species of coelacanths have diverged through vicariance and conclude that Gordon's (1998) hypothesis is the most likely one.

The complementary test is to compare the divergence between the two species of coelacanths with a pair of species that diverged 40–50 Myr: if these divergences are not significantly different, then Springer's (1999) hypothesis could be accepted. We selected two pairs of species: *Bos taurus* and *Sus scrofa*, and *Takifugu rubripes* and *Tetraodon nigroviridis*. Benton and Donoghue (2007) give the divergence of these two pairs as 48.3–53.5 and 32.25–56 Myr, respectively. We also conducted a relaxed-clock molecular analysis using a Bayesian method.

## Materials and methods

### DNA extraction

The genomic DNA of the second specimen of *L. menadoensis* was extracted from the pieces of gills that were preserved in absolute ethanol at  $-20^{\circ}\text{C}$  using the Genomic DNA Mini Kit developed by Geneaid Biotech, Ltd. (Taipei, Taiwan).

### PCR and sequencing

The whole mtGenome was amplified using 24 couples of primers defined on the sequence of the holotype of *L. menadoensis* (GenBank accession number NC 006921; Inoue et al. 2005). The primer sequences are given in Table 1. The PCRs were done using Illustra Puretaq ready-to-go PCR Beads GE Healthcare (Chalfont St Giles, UK) and with the following amplification program: an initial denaturation step at  $95^{\circ}\text{C}$  for 2 min, a cycle of three steps (denaturation at  $94^{\circ}\text{C}$  for 1 min; annealing at  $60^{\circ}\text{C}$  for 2 min, elongation at  $72^{\circ}\text{C}$  for 1 min) repeated 30 times, and a final extension of 10 min at  $72^{\circ}\text{C}$ . Double-stranded PCR products were purified with a Gel/PCR DNA

Fragments Extraction Kit developed by Geneaid Biotech and were subsequently used for direct cycle sequencing with dye-labeled terminators by MacroGen, Inc. (Rockville, MD) using the same primers as those for PCR. The sequence has been deposited in GenBank (accession number: GQ911586).

Additional complete mtGenome sequences were obtained from GenBank (Table 2). Sasaki et al. (2007) showed convincingly that the sequence published by Zardoya and Meyer (1997, NC 001804) has sequencing errors so that it is not usable.

### Statistical analyses

Sequences were aligned with clustal (Thompson et al. 1997). All nucleotide sequence manipulations and analyses were done with R 2.10.1 (R Development Core Team 2009) and the package ape 2.5 (Paradis et al. 2004). Nucleotide sequences were translated into amino acid sequences with the package seqinr 2.0-8 (Charif and Lobry 2007) in order to examine the patterns of amino acid changes. The patterns of nucleotide substitutions between both pairs of species of *Latimeria* and *Pan* were compared using loglinear regression models assuming a multinomial

**Table 1** Primers used for PCR of *Latimeria menadoensis* mtDNA

Primer name	Sequence	Primer name	Sequence
ND2F1	AGAAGTGCTACAAGGACTCC	ALAR1	GAATATGTGGGGTAGTGTCC
ALAF1	AAGAGCCTTCAAAGCCCTCA	COX1R1	GTATACTGTTTCATCCTGTGC
COX1F1	CTACTCCTACTAGCATCCTC	COX1R2	GAGTTTGCCAGTACGATGCC
COX1F2	ACCCCTGCTATGAGCACTAG	COX2R1	GTGAAAGTGGAGGAGTTCTTC
COX2F1	CCTTAACCAAATGGCACACC	COX2R2	TTGTGGCATGGCACTGAGG
COX2F2	GCTAAATAGAAAGCGTAGCC	ATP6R1	CTGTTGGTATTATGGGTAGGA
ATP6F1	CGACTAACAGCCAATCTCAC	COX3R1	AAAGGTTGAGCCATAGGCTC
COX3F1	ACTTTACAACACTGCAAGCC	ND4LR1	AGCCTAGTGTGAATGCAGTG
ND4LF1	TAACGTGGTGAAACCCAC	ND4R2	GCAATTAATAGTGGTAGTGACC
ND4F2	CCGATGGGGTAATCAAACAG	ND4R3	CCTGTTATAGTTAGTGTCCAG
ND4F3	CTAGCTAACTTGGCCTTACC	ND5R1	TGACATATAGTGCATTGGG
ND5F1	CTATCTTGACCAGGGAATCG	ND5R2	TGGAGGAATGCTAGTTGTGG
ND5F2	GCACTAACCACCCTATTAC	ND5R3	GAAATATCCTAGTATGTTGGATG
ND5F3	AACAGCCGCAATCATAGTTAC	CYTBR1	ATTTTGCTGCCTTTGGGTTAG
CYTBF1	AAGAGCCACCATATCCAACC	CYTBR2	CCTGAGATTACGAATGGCAG
CYTBF2	GAGGAGGTTTTTCCGTAGAC	CYTBR3	TGAGAGCTTTAGGAGAATGGT
CYTBF3	TGGATGGCTGAAAACAAGG	DLOOPR1	AAGGGTACTTGAGTGTCCGAC
DLOOPF1	TTAACTCGCGTCAAACGTCG	12SR1	TCAATGTCTATCACTGCTGAG
12SF1	CCACTGACACAAAGGTTTGG	12SR2	TTTTGGGTGAGGAGAGTGAC
12SF2	TACAGCGAAACCCGTACTTC	16SR1	CACTTATGCAGTTAGAGAAGG
16SF1	GCGTTAAAGCTTAAGCAGCAC	16SR2	AAGGATTGCGCTGTTATCCC
16SF2	TAATCCTCCAAGCCGATTGG	ND1R1	TGTTATTGGATGTGGGAGGG
ND1F1	ATACTAGCACTCACCTAGC	TRILER1	GAGGTTTTAACTCCTATAGTCC
ND1F2	TGATCCTATGACACACCTCC	ND2R1	CTTTTCGCAGTTGAGTTTGG

**Table 2** Sequences from GenBank used in the present study

Species	GenBank no.
<i>Latimeria chalumnae</i>	AB257296
<i>Latimeria chalumnae</i>	AB257297
<i>Latimeria menadoensis</i>	NC_006921
<i>Pan paniscus</i>	NC_001643
<i>Pan troglodytes</i>	NC_001644
<i>Bos taurus</i>	NC_006853
<i>Sus scrofa</i>	NC_000845
<i>Takifugu rubripes</i>	AP006045
<i>Tetraodon nigroviridis</i>	AP006046
<i>Scyliorhinus canicula</i>	NC_001950
<i>Polypterus ornatipinnis</i>	NC_001778
<i>Amia calva</i>	AB042952
<i>Hiodon alosoides</i>	AP004356
<i>Engraulis japonicus</i>	NC_003097
<i>Carassius carassius</i>	NC_006291
<i>Danio rerio</i>	NC_002333

distribution for the observations. The principle of this method is to generalize the idea of the logistic regression, which applies to data that follow a binomial distribution, i.e. data with two possible outcomes. The multinomial regression considers variables with three or more outcomes. Therefore, it may be applied to the patterns of nucleotide differences between several pairs of sequences where the outcomes are no change, transition, and transversion. It is also possible to consider more categories of possible outcomes. With the four DNA bases, a comparison of two sequences may lead to 16 possible categories. If we consider that changes are time-reversible, then 10 categories will be considered. However, in the present analyses, we found that considering the three above categories was sufficient.

For a multinomial variable  $Y$  with  $k$  categories, we have to consider  $k - 1$  probabilities (because of the constraint that they must sum to one). The loglinear regression models the logarithmic ratio of these probabilities on the probability of the last category (Agresti 2002):

$$\ln \frac{\Pr(Y = y_j)}{\Pr(Y = y_k)} \quad j = 1, \dots, k - 1 \quad (1)$$

where  $y_1, \dots, y_k$  are the values taken by  $Y$ . These  $k - 1$  ratios are then modeled with a standard regression model with predictors  $x_1, x_2, \dots$ , which may be continuous and/or categorical like in any regression:

$$\ln \frac{\Pr(Y = y_j)}{\Pr(Y = y_k)} = \beta_{j1}x_1 + \beta_{j2}x_2 + \dots + \alpha_j \quad (2)$$

where  $\beta_{j1}, \beta_{j2}, \dots, \alpha_j$  are the parameters for the  $j$ th regression.

We expected that the number of nucleotide changes are affected by the time separating each species within a pair but also by the position along the sequence of the mtGenome. So, the predictors  $x$  in Eq. 2 were ‘gene’, ‘codon position’, and ‘pair of species’ (or ‘lineage’). The data were prepared as follows: for each gene, codon position, and pair of species, the number of transitions, transversions, and no change ( $y_1, y_2$ , and  $y_3$ , respectively) were counted and associated with these predictors. The last category was used as the reference (since here  $k = 3$ ). All predictors were considered as categorical. For non-coding loci, codon position was coded as a fourth category. Several models were fitted using these predictors separately or in combination, including a null model (i.e., the probabilities  $\Pr(Y = y_j)$  are the same along the mtGenome and among all lineages).

Since there are 38 genes in the mtGenome, a model with all three predictors cannot be statistically identified because not all categories of nucleotide change were represented for all combinations of predictors: there was no transversion for ND4L and ATP8 and on the second codon position for all protein-coding genes but ND2. This codon position was monomorphic for three genes (ND4L, COX1, ATP8). Even the model gene + lineage has 82 parameters to be estimated that would be difficult to interpret. Furthermore, substitution rates may not be significantly different between different genes. To test whether variation within different categories of genes is significant, we pooled the data for the protein-coding, for the rRNA, and for the tRNA genes, respectively, thus creating a new predictor labeled ‘gene2’ with four categories (D-loop was kept apart). Since gene and codon position are partly confounded, we pooled together the data from the different positions as if they were from three genes, thus creating a new variable ‘gene3’ with six categories (the three codon positions, rRNA, tRNA, and D-loop). Finally, we tested whether divergence among lineages of the same approximate age was different by creating a variable ‘lineage2’ pooling *Latimeria* and *Pan* together on one side, and the pairs *Bos* + *Sus* and *Takifugu* + *Tetraodon* on the other.

The models were fitted by maximum likelihood, so that they can be compared with their respective Akaike information criterion (AIC; Akaike 1973), which is equal to minus twice the log-likelihood plus twice the number of estimated parameters of the model. The model with the lowest AIC was selected. This procedure can thus be used for hypothesis testing. Additionally, we performed likelihood-ratio tests (LRT) to compare some pairs of models when relevant. After selecting a model, the probabilities for each predictor can be calculated by inverting Eq. 2. The loglinear modeling analysis was done with the R package VGAM (Yee 2009).

**Table 3** The six positions where the two mtGenome sequences of *L. menadoensis* differ

	Sequence	Position					
		4950	5346	5547	5669	9878	9882
Aligned positions <sup>a</sup> :		4950	5346	5547	5669	9878	9882
Unaligned positions <sup>b</sup> :		4947	5341	5542	5664	9873	9877
Gene:		ND2	tRNA-Cys	CO1	CO1	ND3	ND3
<i>L. menadoensis</i>	NC_006921	C	T	G	G	G	G
	New sample	T	C	T	T	A	A
<i>L. chalumnae</i>	AB257297	T	C	T	T	A	A
	AB257296	T	C	T	T	A	A

<sup>a</sup> Obtained after alignment with *L. chalumnae*

<sup>b</sup> Refers to the *L. menadoensis* sequences without alignment gaps

Molecular dating was done with the Bayesian method developed by Thorne et al. (1998) and Thorne and Kishino (2002) where rates of molecular evolution are assumed to vary among branches in a phylogenetic tree but in a correlated way between adjacent branches (relaxed molecular clock). Two data sets were used: the first one used the same data than above, and the second one used additional fish calibration points taken from Santini et al. (2009). In both data sets, the phylogeny was unambiguous. The branch lengths were estimated with PAML 4.3 (Yang 2007). These estimates were used by Multidivtime that runs a Markov chain Monte Carlo (MCMC) to calculate the posterior distribution of the node times of the phylogeny. The MCMC was sampled  $10^4$  times every 100 cycles, and  $10^5$  burn-in cycles were discarded. The sequences were dispatched in five partitions: the three codon positions of protein-coding genes, t-RNA, and rRNA. The D-loop was discarded because of its fast evolving rate and many insertions–deletions, as well as ND6 (see Inoue et al. 2005). All genes were aligned separately prior to building the partitions. This analysis was run from R with the LAGOPUS package (Heibl 2010).

## Results

### Sequence alignment

The alignment of the four complete mtGenomes of *Latimeria* using clustal with its default parameters (gap opening = 15, gap extension = 6.66) resulted in the insertion of six gaps in *L. menadoensis* and seven in *L. chalumnae*. An alternative alignment was obtained by setting gap opening = 60 and gap extension = 60: it has two fewer indels (one in each sequence) located in the D-loop, but this results in eight additional base mismatches. Additionally, a gap has been placed differently within the sequence of the tRNA-Trp (certainly because of the increased penalty for gap opening) resulting in three new

mismatches. Therefore, we considered the first alignment in the subsequent analyses.

The comparison of the first sequenced coelacanth from Manado with the new individual showed six differences (Table 3). Referring to the sequences of *L. chalumnae* after alignment, it appeared that these six sites were identical between this species and the new individual. Additionally, none of these sites were present in the fastest evolving portions of the mtGenome (e.g., D-loop), and of the five differences observed in protein-coding genes, two occurred on the second codon position, which is the slowest evolving position; the three others occurred on the third codon position. Thus, we suggest that these six differences are the results of sequencing errors in NC\_006921, and, therefore, that both individuals have identical mtGenomes.

### Loglinear modeling

The best model selected by AIC is the one with the predictors ‘gene3’ and ‘age’ (Table 4). The model that comes in second in terms of AIC has ‘lineage’ instead of ‘lineage2’. Both models have very close values of log-likelihood meaning that the four additional parameters involved in ‘lineage’ do not significantly improve the fit. Indeed, the LRT comparing the two models was not significant ( $\chi^2_4 = 1.22$ ,  $P = 0.875$ ).

Interestingly, the model assuming four categories of substitution rates (for each codon position and the non-coding loci, ‘pos’) fitted better than the model with different substitution rates for each gene (‘gene’), which is clearly over-parameterized.

Table 5 shows the predicted probabilities of substitution from the selected model. These probabilities clearly show the contrast between the two groups of lineages. Overall, they conform to what is known on the rates of molecular evolution. It is interesting to note that for fastest evolving portions of the mtGenome, namely third codon positions and D-loop, the transition (Ts) and transversion (Tv) probabilities are markedly different for the *Latimeria* and

**Table 4** Results of loglinear modeling

Effect(s)	AIC	Log-likelihood	np	df
Gene3 + lineage2	52219.69	−26095.85	14	476
Gene3 + lineage	52226.48	−26095.24	18	472
Pos + lineage	52387.28	−26179.64	14	476
Gene + lineage	54344.04	−27090.02	82	408
Gene2 + lineage	54409.41	−27190.71	14	476
Lineage2	54706.44	−27349.22	4	486
Lineage	54713.10	−27348.55	8	482
Gene3	55850.90	−27913.45	12	478
Pos	56013.21	−27998.61	8	482
Gene	58021.10	−28934.55	76	414
Gene2	58092.48	−29038.24	8	482
Null model	58401.38	−29198.69	2	488

Models are ordered with the lowest AIC on top (*np* number of parameters, *df* degrees of freedom)

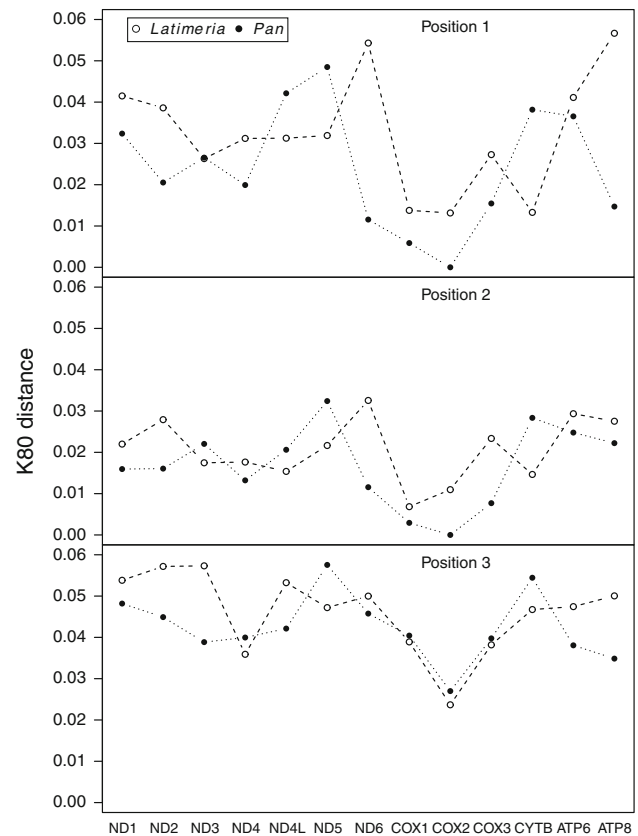
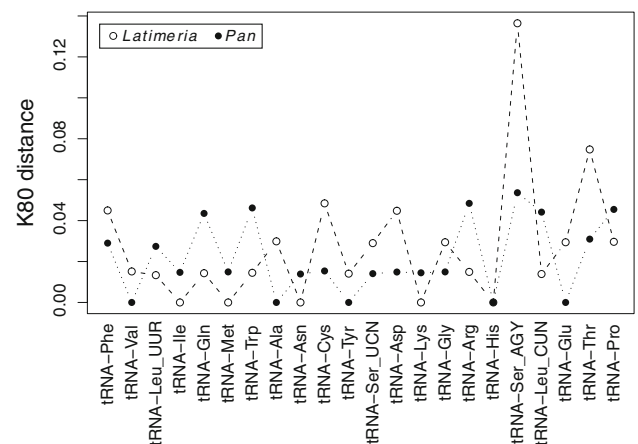
**Table 5** Probabilities of change predicted from the model selected with the lowest AIC

Lineages	Gene	Ts	Tv	No change
<i>Latimeria</i>	Protein (pos 1)	0.0323	0.0035	0.9642
<i>Pan</i>	Protein (pos 2)	0.0125	0.0010	0.9866
	Protein (pos 3)	0.0657	0.0079	0.9265
	rRNA	0.0258	0.0025	0.9717
	D-loop	0.0404	0.0052	0.9544
	tRNA	0.0263	0.0014	0.9723
<i>Bos–Sus</i>	Protein (pos 1)	0.0826	0.0656	0.8518
<i>Takifugu–</i>	Protein (pos 2)	0.0346	0.0196	0.9458
<i>Tetraodon</i>	Protein (pos 3)	0.1481	0.1302	0.7217
	rRNA	0.0678	0.0491	0.8831
	D-loop	0.0988	0.0940	0.8072
	tRNA	0.0706	0.0278	0.9016

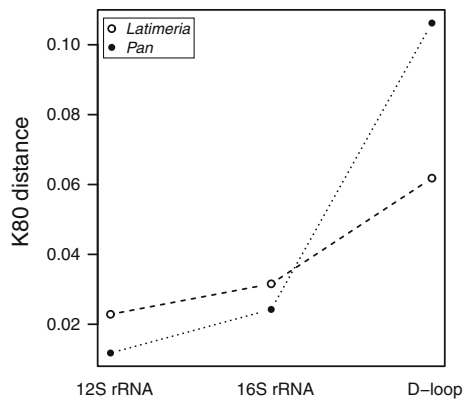
*Pan* lineages, whereas they were close for the two others. The values of the two types of RNA were also very similar for each category of lineages.

#### Nucleotide differences

In the following analyses, we considered only the two species of *Latimeria* and of *Pan*. Figure 1 shows the Kimura (1980) evolutionary distances between each pair of species computed separately for each protein-coding gene and the three codon positions. Overall, the distances are similar between both pairs of species. The similarity is stronger for the third position, which is the most selectively neutral one within the codon. Interestingly, the three shortest sequences (ND3, ND4L, ATP8, all less than 500 bp) show substantial differences in these distances.

**Fig. 1** Evolutionary distances within *Latimeria* and *Pan* for the 13 mitochondrial protein-coding genes and the three codon positions separately**Fig. 2** Evolutionary distances within *Latimeria* and *Pan* for the 22 tRNA genes

The evolutionary distances among the 22 tRNA sequences (Fig. 2) were also overall similar between both pairs of species, though with more variation than for the protein-coding genes, which is certainly due to the short sequences of the tRNAs (all less than 80 bp).



**Fig. 3** Evolutionary distances within *Latimeria* and *Pan* for the two rRNA genes and D-loop

The evolutionary distances for the rRNA sequences are slightly larger between the two *Latimeria* species than for the *Pan* ones (Fig. 3). On the other hand, this was the opposite for the D-loop sequences where the distance among *Latimeria* was about half that among *Pan* (Fig. 3).

#### Amino acid differences

Each protein-coding gene sequence was translated into its amino acid sequence using the vertebrate mitochondrial genetic code, and after eventually inverting the sequence of the ND6 gene, which is coded on the opposite strand of the DNA molecule (see Zardoya and Meyer 1997). The amino acid sequences were then analyzed with the same procedure as for the nucleotide sequences, counting each substitution type for each molecule for each genus.

Between *L. menadoensis* and *L. chalumnae*, 112 amino acid substitutions were observed, while there were 95 between *P. troglodytes* and *P. paniscus*. Among the 190 possible types of amino acid changes, 29 were observed between the *Latimeria* species, while 30 were observed between the *Pan* ones (21 were common to both pairs). The Ala–Thr substitution was the most frequent one in both genera. The second most frequent substitution (Ile–Val) was also the same for both genera. Overall, both matrices of amino acid substitution frequencies are consistent between both genera in the sense that substitutions in middle frequencies (5–10 occurrences) were observed in both sides.

Figure 4 shows how the types of amino acid substitutions were observed for each of the 13 genes. Both genera yielded similar results, with most substitutions in the ND5 gene with several Ile–Val changes.

#### Bayesian molecular dating

The credibility intervals of the divergence dates calculated from their posterior distributions are indicated in Fig. 5.

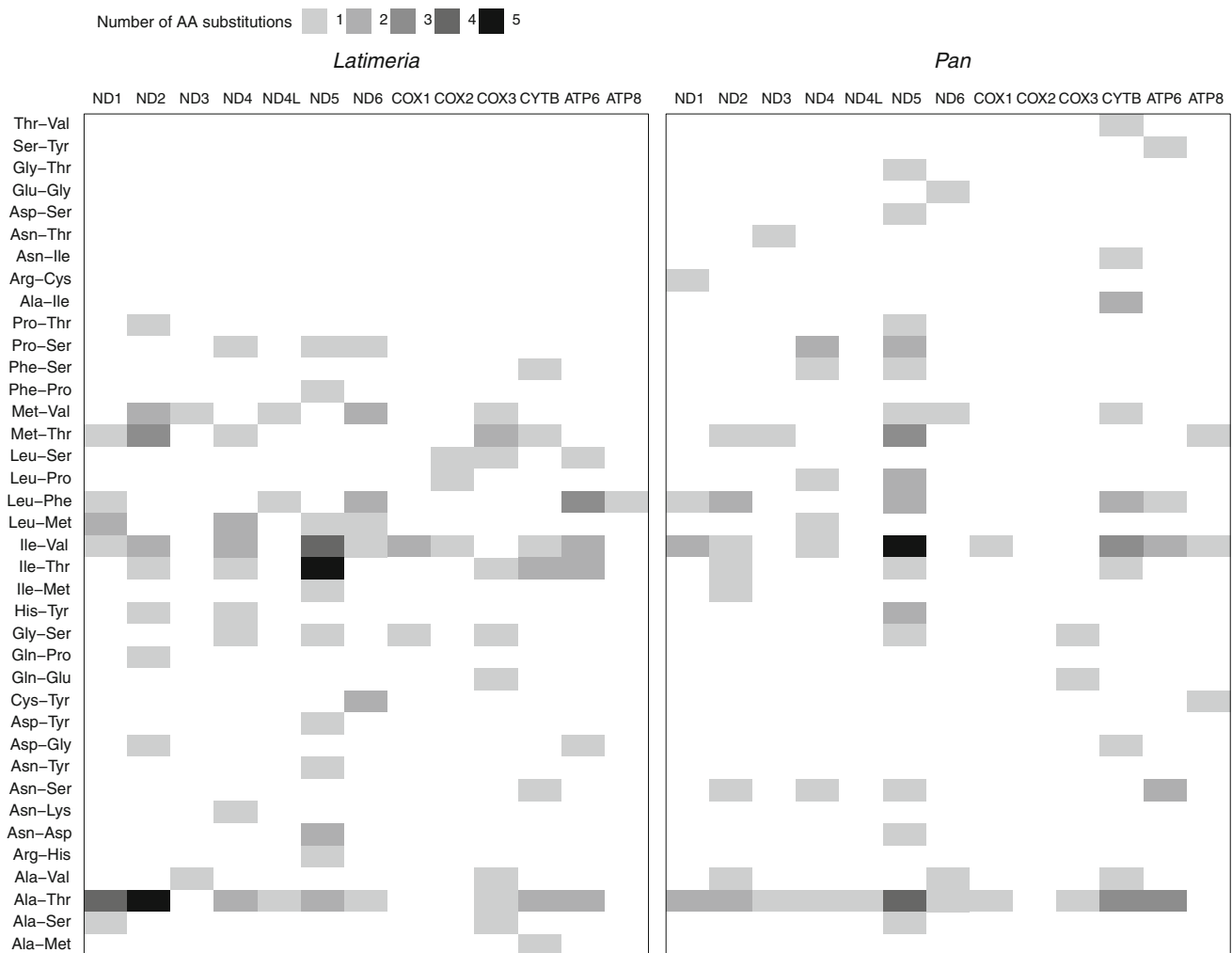
Except for the divergence between the two species of *Latimeria*, all dates were constrained with upper and lower bounds taken from Benton and Donoghue (2007) and Santini et al. (2009). The credibility interval for the *Latimeria* divergence was  $21 \pm 4$  when using the same species than above. However, it was  $28 \pm 6$  after including the fish sequence and calibration data.

#### Discussion

The sequence of the mtGenome of a second individual of *L. menadoensis* showed that it is similar to the previously sequenced one. The six observed differences between these two individuals are most likely attributable to sequencing errors from Inoue et al. (2005). This observed genetic homogeneity is not a surprising result. Sasaki et al. (2007) sequenced the complete mtGenome of two individuals of *L. chalumnae* from Tanzania and found four differences (one on a third codon position of ND5, one on tRNA–Gln, and two on D-loop). Scharl et al. (2005) found no variation in the cytochrome *b* sequences of 47 individuals of *L. chalumnae* from a much larger geographic area and only six variable sites in the D-loop allowing to define six haplotypes. All these results provide evidence for a very low mitochondrial genetic diversity in the two species of *Latimeria*. The low population numbers, the long sexual maturation time, and low fecundity of females are factors that decrease the effective population size of females, and therefore lead to lower variability of the matrilineally inherited mtGenome. Scharl et al. (2005) also analyzed microsatellite data and found low levels of heterozygosity further suggesting that effective population sizes are low in *L. chalumnae*.

The patterns of nucleotide substitutions between *L. chalumnae* and *L. menadoensis* were very similar to those observed between *P. troglodytes* and *P. paniscus*. This was particularly the case for the third codon position of the protein-coding genes. It is generally admitted that these nucleotide sites are the most neutral among all in the genome because of the degeneracy of the genetic code (Kimura 1983; Nei and Kumar 2000). Furthermore, there was congruence between the difference in divergence between both pairs of species and the length of gene sequences: the longer the sequence, the closer the distance between *Latimeria* and *Pan*. This can be explained by the variance of the substitution process that leads to higher observed variance for the shortest sequences (Nei and Kumar 2000). These results support the hypothesis of similar divergence for both pairs of species.

Yokoyama and Tada (2000) studied the sequences of two rhodopsin genes (2,130 bp) in both species of *Latimeria* and found twelve nucleotide differences; ten of



**Fig. 4** Patterns of amino acid substitutions within *Latimeria* and *Pan* for the 13 mitochondrial protein-coding genes. The gray level indicates how many each substitution type (as rows) occurred between each pair of species

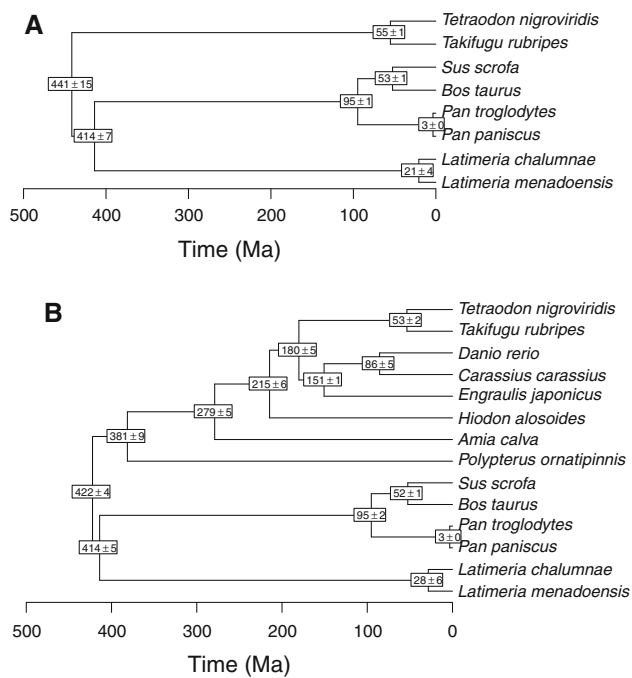
them occurred on third codon positions and the two others on the first codon position resulting in one amino acid difference in each gene. This result agrees with ours showing a recent divergence between *L. chalumnae* and *L. menadoensis*.

The molecular dating analysis shows contrasting results compared to those discussed above: we found a relatively old date of divergence between 20 and 30 Ma. Inoue et al. (2005) estimated the same divergence date at ca. 40 Ma. These authors used the same statistical method employed here, and it is very likely that the difference with our results arises from the selection of species. The date of 20 Ma was obtained with a limited number of species: four mammals, two fishes, and the two coelacanths. After adding six other fish species, the same date estimate was pushed 10 Myr back into the past. This is certainly due to the autocorrelation of rates as assumed in the Multidivtime model (Thorne et al. 1998). This also

explains why the divergence among coelacanths is estimated to be older than the divergence between the chimpanzees in spite of our results discussed in the previous paragraph. Considering that the Sarcopterygii and other tetrapods diverged very far in the past (more than 400 Ma), the molecular dating method will tend to spread the amount of change along the branch leading to the coelacanths in order to increase the correlation in molecular substitution rates.

It is remarkable that the patterns of amino acid changes between both pairs of species of *Latimeria* and *Pan* show several similarities. The ND-series of proteins show many more amino acid changes than the cytochrome ones. For the former proteins, the number of changes were found to be proportional to the length of the protein. On the other hand, for the cytochromes, the longest protein (COX1) showed the smallest number of changes. It will be interesting in future studies to assess whether these amino





**Fig. 5** Bayesian molecular dating of the divergence between the two species of *Latimeria* **a** using calibration points from Benton and Donoghue (2007), **b** using calibration points from Benton and Donoghue (2007) and Santini et al. (2009). The boxes give the credibility intervals for the divergence dates

acids' changes have similar impacts on the secondary structure of the proteins.

Our results do not support the hypothesis that homeotherms (here mammals) have higher molecular evolutionary rates than poikilotherms (fishes). The molecular divergences between *Bos–Sus* and *Takifugu–Tetraodon* were not statistically different, whereas they have widely overlapping divergence dates according to the fossil record (Benton and Donoghue 2007). This is consistent with recent comparative studies that failed to find a relationship between molecular rates of evolution and metabolic rates across a wide range of animal taxa (Thomas et al. 2006; Lanfear et al. 2007).

Several recent studies have questioned the adequacy of mitochondrial DNA in quantifying population genetic variability (Bazin et al. 2006; Balloux 2009). Balloux et al. (2009) showed that some protein-coding genes are under selective pressure in the human mtGenome. Nabholz et al. (2009) showed that mutation rates of cytochrome *b* are on average three times higher in mammals than in birds. They also provided evidence for substantial variation among species of the same group. On the other hand, Piganeau and Eyre-Walker (2009) showed a relationship between effective population size and mitochondrial genetic diversity using a different measure based on synonymous polymorphism. All these studies indicate clearly that mtDNA data should be interpreted with care in population studies. In

phylogenetic studies, it is clear that mtDNA exhibits significant rate heterogeneity in many cases, a factor that should be taken into account (Pereira and Baker 2006).

There seems to be two alternative explanations in order to resolve the apparent paradox in our results: either the divergence between both species of coelacanths is recent and the dating analysis is biased as described above, or the divergence is old and the rate of molecular evolution is remarkably slow in coelacanths. It seems difficult to evoke selective pressures that would be different in these two species because of the similarities of molecular substitutions with chimpanzees, even on third codon positions. Nabholz et al. (2009) evoked the “aging hypothesis” to explain why mutation rate is higher in mammals compared to birds. However, it is not clear how this hypothesis could apply here. Coelacanths are long-lived animals and if mutations accumulate at a low rate in their mtGenome that would be interesting since the metabolic rate hypothesis has been refuted in previous studies (Thomas et al. 2006; Lanfear et al. 2007; Nabholz et al. 2009), and thus would imply the existence of repair mechanisms as suggested for birds (Nabholz et al. 2009).

Our results suggest it is unlikely that the question of the divergence date of the two living coelacanth species could be answered using a traditional approach for the estimation of molecular divergence. The possible future captures of other specimens of *L. menadoensis* would be particularly welcome in order to quantify the variation in mtGenome in this species. These new data will certainly be essential to solve the issue of the evolutionary dynamics of these fascinating tetrapods.

**Acknowledgments** We are grateful to Thibaut Jombart and an anonymous reviewer for their constructive comments on a previous version of this paper. This research was financially supported by an Action Thématique Incitative of the Institut de Recherche pour le Développement.

## References

- Agresti A (2002) Categorical data analysis, 2nd edn. Wiley, New York
- Akaike H (1973) Information theory and an extension of the maximum likelihood principle. In: Petrov BN, Csaki F (eds) Proceedings of the second international symposium on information theory, Akadémia Kiado, Budapest, pp 267–281
- Balloux F (2009) The worm in the fruit of the mitochondrial DNA trees. Heredity doi:10.1038/hdy.2009.122
- Balloux F, Handley LJ, Jombart T, Liu H, Manica A (2009) Climate shaped the worldwide distribution of human mitochondrial DNA sequence variation. Proc R Soc Lond B 276:3447–3455
- Bazin E, Glémin S, Galtier N (2006) Population size does not influence mitochondrial genetic diversity in animals. Science 312:570–572
- Benton MJ, Donoghue PCJ (2007) Paleontological evidence to date the tree of life. Mol Biol Evol 24(1):26–53. Erratum 24:889–891

- Charif D, Lobry J (2007) SeqinR 1.0-2: A contributed package to the R project for statistical computing devoted to biological sequences retrieval and analysis. In: Bastolla U, Porto HRM, Vendruscolo M (eds) Structural approaches to sequence evolution: molecules, networks, populations. Springer, New York, pp 207–232
- Erdmann MV (1999) An account of the first living coelacanth known to scientists from Indonesian waters. *Environ Biol Fish* 54:439–443
- Erdmann MV, Caldwell RL, Jewett SL, Tjakrawidjaja A (1999) The second recorded living coelacanth from north Sulawesi. *Environ Biol Fish* 54(4):445–451
- Gordon AL (1998) Coelacanth populations may go with the flow. *Nature* 395:634
- Heibl C (2010) LAGOPUS: R-package for Bayesian relaxed-clock molecular dating. <http://www.christophheibl.de>
- Holder MT, Erdmann MV, Wilcox TP, Caldwell RL, Hillis DM (1999) Two living species of coelacanths? *Proc Natl Acad Sci USA* 96(22):12616–12620
- Inoue JG, Miya M, Venkatesh B, Nishida M (2005) The mitochondrial genome of Indonesian coelacanth *Latimeria menadoensis* (Sarcopterygii: Coelacanthiformes) and divergence time estimation between the two coelacanths. *Gene* 349:227–235
- Kimura M (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol* 16:111–120
- Kimura M (1983) The neutral theory of molecular evolution. Cambridge University Press, Cambridge
- Lanfear R, Thomas JA, Welch JJ, Brey T, Bromham L (2007) Metabolic rate does not calibrate the molecular clock. *Proc Natl Acad Sci USA* 104(39):15388–15393
- Miller HC, Moore JA, Allendorf FW, Daugherty CH (2009) The evolutionary rate of tuatara revisited. *Trends Genet* 25(1):13–15
- Nabholz B, Glémin S, Galtier N (2009) The erratic mitochondrial clock: variations of mutation rate, not population size, affect mtDNA diversity across birds and mammals. *BMC Evol Biol* 9:54
- Nei M, Kumar S (2000) Molecular evolution and phylogenetics. Oxford University Press, Oxford
- Nelson JS (2006) Fishes of the world, 4th edn. Wiley, Hoboken
- Paradis E, Claude J, Strimmer K (2004) APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* 20(2):289–290
- Parham JF, Irmis RB (2008) Caveats on the use of fossil calibrations for molecular dating: a comment on Near et al. *Am Nat* 171(1):132–136
- Pereira SL, Baker AJ (2006) A mitogenomic timescale for birds detects variable phylogenetic rates of molecular evolution and refutes the standard molecular clock. *Mol Biol Evol* 23:1731–1740
- Perry GH, Tito RY, Verrelli BC (2007) The evolutionary history of human and chimpanzee Y-chromosome gene loss. *Mol Biol Evol* 24(3):853–859
- Piganeau G, Eyre-Walker A (2009) Evidence for variation in the effective population size of animal mitochondrial DNA. *Plos One* 4(2):e4396
- Pouyaud L, Wirjoatmodjo S, Rachmatika I, Tjakrawidjaja A, Hadiaty R, Hadie W (1999) A new species of coelacanth. *C R Acad Sci III* 322:261–267
- Pulquério MJF, Nichols RA (2007) Dates from the molecular clock: how wrong can we be? *Trends Ecol Evol* 22(4):180–184
- R Development Core Team (2009) R: a language and environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org>
- Raaum RL, Sterner KN, Noviello CM, Stewart CB, Disotell TR (2005) Catarrhine primate divergence dates estimated from complete mitochondrial genomes: concordance with fossil and nuclear DNA evidence. *J Hum Evol* 48(3):237–257
- Santini F, Harmon LJ, Carnevale G, Alfaro ME (2009) Did genome duplication drive the origin of teleosts? A comparative study of diversification in ray-finned fishes. *BMC Evol Biol* 9:194
- Sasaki T, Sato T, Miura S, Bwathondi POJ, Ngatunga BP, Okada N (2007) Mitogenomic analysis for coelacanths (*Latimeria chalumnae*) caught in Tanzania. *Gene* 389(1):73–79
- Schartl M, Hornung U, Hissmann K, Schauer J, Fricke H (2005) Relatedness among east African coelacanths. *Nature* 435:901
- Smith JLB (1939) A living fish of Mesozoic type. *Nature* 143:455–456
- Springer VG (1999) Are the Indonesian and western Indian Ocean coelacanths conspecific: a prediction. *Environ Biol Fish* 54(4):453–456
- Subramanian S, Hay JM, Mohandesan E, Millar CD, Lambert DM (2009) Molecular and morphological evolution in tuatara are decoupled. *Trends Genet* 25(1):16–18
- Thomas JA, Welch JJ, Woolfit M, Bromham L (2006) There is no universal molecular clock for invertebrates, but rate variation does not scale with body size. *Proc Natl Acad Sci USA* 103(19):7366–7371
- Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL\_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* 25(24):4876–4882
- Thorne JL, Kishino H (2002) Divergence time and evolutionary rate estimation with multilocus data. *Syst Biol* 51(5):689–702
- Thorne JL, Kishino H, Painter IS (1998) Estimating the rate of evolution of the rate of molecular evolution. *Mol Biol Evol* 15(12):1647–1657
- Won YJ, Hey J (2005) Divergence population genetics of chimpanzees. *Mol Biol Evol* 22(2):297–307
- Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* 24(8):1586–1591
- Yee TW (2009) VGAM: vector generalized linear and additive models. R package version 0.7-8. <http://www.stat.auckland.ac.nz/~yee/VGAM>
- Yokoyama S, Tada T (2000) Adaptive evolution of the African and Indonesian coelacanths to deep-sea environments. *Gene* 261(1):35–42
- Zardoya R, Meyer A (1997) The complete DNA sequence of the mitochondrial genome of a “living fossil,” the coelacanth (*Latimeria chalumnae*). *Genetics* 146:995–1010