

The relationships of the coelacanth *Latimeria chalumnae*: evidence from sequences of vertebrate 28S ribosomal RNA genes

David M. Hillis, Michael T. Dixon & Loren K. Ammerman
Department of Zoology, The University of Texas, Austin, TX 78712, U.S.A.

Received 1.8.1989 Accepted 30.7.1990

Key words: Sarcopterygians, Phylogeny, Outgroup comparison

Synopsis

A subgenomic library created from genomic DNA of *Latimeria chalumnae* was screened for 28S ribosomal RNA (rRNA) clones. The resulting clone was subcloned into a plasmid vector, and over 2 kb of the 28S rRNA region was sequenced. Sequences of 28S rRNA genes were also obtained for *Rhineura floridana* (Squamata), *Cyprinella lutrensis* (Actinopterygii), and *Lampetra aepyptera* (Petromyzontiformes) by cloning and/or amplification by the polymerase chain reaction. The 28S rDNA sequences were aligned for all the above species as well as for the previously published 28S rDNA sequences of the genera *Mus*, *Rattus*, and *Homo* (Mammalia), *Xenopus* (Amphibia), and *Drosophila* (Insecta). Phylogenetic analysis of these species (using both the insect and lamprey sequences for outgroup comparison, or using only the lamprey sequence in the outgroup) produced a single optimal solution: (Outgroup(*Cyprinella*(*Latimeria*(*Xenopus*(*Rhineura*(*Homo*(*Rattus*(*Mus*))))))). Bootstrap analysis indicated that the placement of *L. chalumnae* on this tree was significant at $p < 0.01$. Previously published alternative hypotheses of relationships of *Latimeria* require at least 19 additional steps compared to the optimal solution; the rDNA data are sufficient to reject the hypotheses that place *Latimeria* in groups other than the sarcopterygians.

Introduction

There have been more diverse hypotheses about the relationships of coelacanths (Actinistia) than virtually any other major group of vertebrates. Interest in actinistians has been great not only because of the key position they hold in the problem of vertebrate transition from water to land, but also because of the persistence to the present of but a single living member of this group, *Latimeria chalumnae*. The existence of a living species (in addition to the fairly extensive fossil record of coelacanths) permits the collection of a diversity of biological data that can be used to infer relationships. However, past studies on coelacanth phylogeny have generated virtually every conceivable hypoth-

esis for the relationship of this group to lungfishes, tetrapods, actinopterygians, and chondrichthyans (see Forey 1988).

All methods of phylogenetic inference depend on two primary assumptions: (1) the characters examined are heritable; and (2) the characters examined are historically independent. The largest possible set of independent, heritable characters for any organism consists of a subset of its DNA sequences; only a subset of DNA sequences are independent because many repeated sequences evolve in concert (see Hillis 1987). However, molecular biology has tapped a miniscule portion of this reservoir of phylogenetic information to date, and virtually nothing is known about the DNA sequences of *Latimeria chalumnae*. We describe in

this paper the creation of a gene library for *Latimeria*; we also describe the first gene that we have screened and partially sequenced from this library. We chose to examine the 28S ribosomal RNA gene because of the conserved nature of this gene among vertebrates (Hillis & Davis 1987) and its potential for elucidating the relationships of coelacanths. Preliminary analysis of a portion of the 28S rRNA gene (Hillis & Dixon 1989) suggested that this sequence could be used to resolve relationships among major groups of vertebrates. We here report the sequence of over 2 kb of the 28S rRNA gene of *L. chalumnae* (along with the homologous genes of other major vertebrate groups) and examine the relationships suggested by these sequences.

Materials and methods

Frozen muscle tissue from *Latimeria chalumnae* was obtained from a specimen procured by the Explorer's Club from Grand Comoro island in November 1986 (VIMS 8118). High molecular weight DNA was isolated from this tissue following the protocol of Hillis & Davis (1986). *L. chalumnae* DNA was cleaved with the restriction enzyme *Eco*RI and ligated into the lambda vector Lambda Zap II (Stratagene) to produce a subgenomic library (see Hillis et al. 1990). This library was screened (via filter-lift hybridization) with a cloned 28S rDNA gene of *Rana catesbeiana* (pE2528, Hillis & Davis 1987). Positive plaques were selected and the inserts were subcloned in the vector pBluescript (Stratagene). Subclones were verified by restriction digestion, Southern blotting, and sequencing. A verified 28S clone was designated pDH8804.

In addition to the 28S rRNA gene of *L. chalumnae* we also examined the 28S rRNA genes of *Cyprinella* (*Notropis sensu lato*) *lutrensis* (Actinopterygii) and *Rhineura floridana* (Squamata) (see Hillis & Dixon 1989). Furthermore, we compared these sequences to the published 28S DNA sequences of *Mus musculus*, *Rattus norvegicus*, and *Homo sapiens* (Mammalia; Hassouna et al. 1984, Hadjiolov et al. 1984, Gonzalez et al. 1985), *Xenopus laevis* (Amphibia; Ware et al. 1983), and *Drosophila melanogaster* (Insecta; Tautz et al. 1988).

We also amplified (via the polymerase chain reaction, or PCR; Mullis & Faloona 1987) and sequenced a portion of the 28S rRNA gene of *Lampetra aepyptera* (Petromyzontiformes). PCR amplification was also used to verify pDH8804 as a 28S rDNA clone of *L. chalumnae*. A 700 bp fragment of 28S rDNA was amplified from genomic DNA with modified *Taq* polymerase (AmpliTaq, Cetus) using primers 28v and 28x (Table 1, Fig. 1.). Thirty cycles of melting/annealing/polymerization were performed to produce double-stranded DNA. This product was used as the starting template for asymmetrical amplification to generate single-stranded DNA which was used for direct sequencing (Gyllensten & Erlich 1988). Both an amplification primer (28v) and an internal primer (28w) were used to sequence the amplified DNA.

Plasmid DNA was purified as described by Hillis et al. (1990), denatured in alkali, and sequenced by the base-specific dideoxynucleotide chain termination method (Sanger et al. 1977) using modified T7 DNA polymerase (Tabor & Richardson 1987). DNA amplified via PCR was sequenced in the same manner, except that the single-stranded DNA did not require the treatment with alkali. Primers used are shown in Table 1 and Figure 1. Reaction products were separated on 55 cm 4–6% polyacrylamide field gradient gels (Ansorge & Laibit 1984) and visualized by autoradiography. DNA sequences were aligned against the *Mus* 28S rDNA sequence with the alignment subroutines described by Pustell & Kafatos (1982, 1984, 1986), with adjustments made manually to increase similarity.

All possible tree topologies were examined using the exhaustive search routine of the Phylogenetic Analysis Using Parsimony (PAUP) software package (Swofford 1985 et seq.), with all characters coded as unordered. The insect and lamprey sequences were used for outgroup comparison. The distribution of lengths of all possible trees provides a means of evaluating the relative information to noise ratio in a data set (Fitch 1984, Hillis & Dixon 1989). Confidence limits of nodes on the most parsimonious tree were estimated using bootstrap analysis (Felsenstein 1985). Briefly, this involves sampling characters with replacement from the

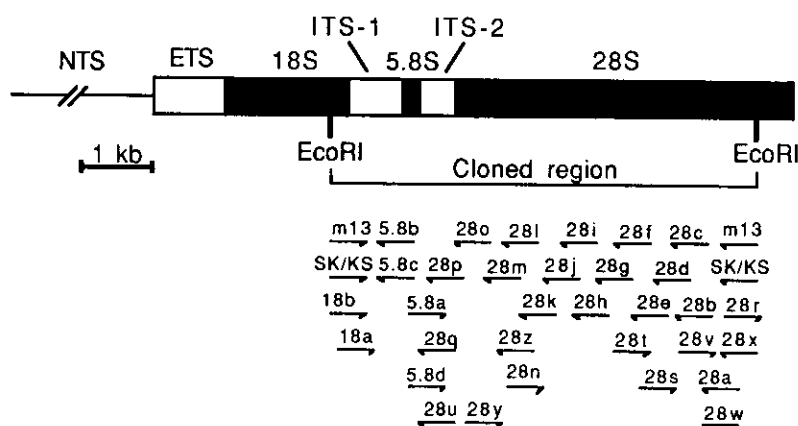


Fig. 1. Cloning and sequencing strategy of ribosomal DNA. Abbreviations: ETS = external transcribed spacer; ITS = internal transcribed spacer; NTS = non-transcribed spacer.

original data matrix to produce new data matrices of equal size to the original, each of which is then analyzed to find the most parsimonious solution. The branch and bound algorithm of PAUP was

used to find optimal solutions during bootstrapping. Bootstrap sampling was replicated 1000 times. Another means of assessing relative stability of nodes on a most parsimonious cladogram is con-

Table 1. Primers used to obtain the sequences in Appendix 1. Positions of the primers in the *Mus* sequence correspond to the positions of the RNA nucleotides in the mature transcripts. 'S' stands for the strand synonymous to RNA; 'C' stands for the complementary strand.

Primer	<i>Mus</i> position	Strand	Sequence
28a	3913-3927	S	5'-CCTTCTGCTCCACGG-3'
28b	3695-3709	S	5'-AGAGTAGTGGTATTT-3'
28c	3481-3495	S	5'-ACAGTGGAATCTCG-3'
28d	3284-3298	S	5'-TTAAACAGTCGGATT-3'
28e	2976-2990	S	5'-GTCCAGAGTCGCCGC-3'
28f	2617-2631	S	5'-TCCCGAAGTTACGGA-3'
28g	2386-2400	S	5'-CTGCCCTTCAAAAG-3'
28h	2101-2115	S	5'-CTACCACCAAGATCT-3'
28i	1840-1854	S	5'-GCGCCATCCATTTTC-3'
28j	1665-1679	S	5'-CCAGTTCTGCTTACC-3'
28k	1402-1416	S	5'-CGATTTGCACGTCAG-3'
28l	1131-1145	S	5'-GGTCCGTGTTTCAAG-3'
28r	3828-3842	C	5'-CAGGTGGGAGTTTG-3'
28s	2559-2573	C	5'-AGGTGAACAGCCTCT-3'
28t	2338-2352	C	5'-ACCGATCCCGGAGAA-3'
28v	3429-3452	C	5'-AAGGTAGCCAAATG(T,C)CTCGTCATC-3'
28w	3565-3588	C	5'-CCTGTTGAGCTTGACTCTAGTCTG-3'
28x	4106-4137	S	5'-GTGAATTCTGCTTCAATGATAGGAAGAGCC-3'
m13F	NA	S/C	5'-GTTTTCCAGTCACGAC-3'
m13R	NA	S/C	5'-CAGGAAACAGCTATGAC-3'
SK	NA	S/C	5'-TCTAGAACTAGTGGATC-3'
KS	NA	S/C	5'-CGAGGTGACGGTATCG-3'

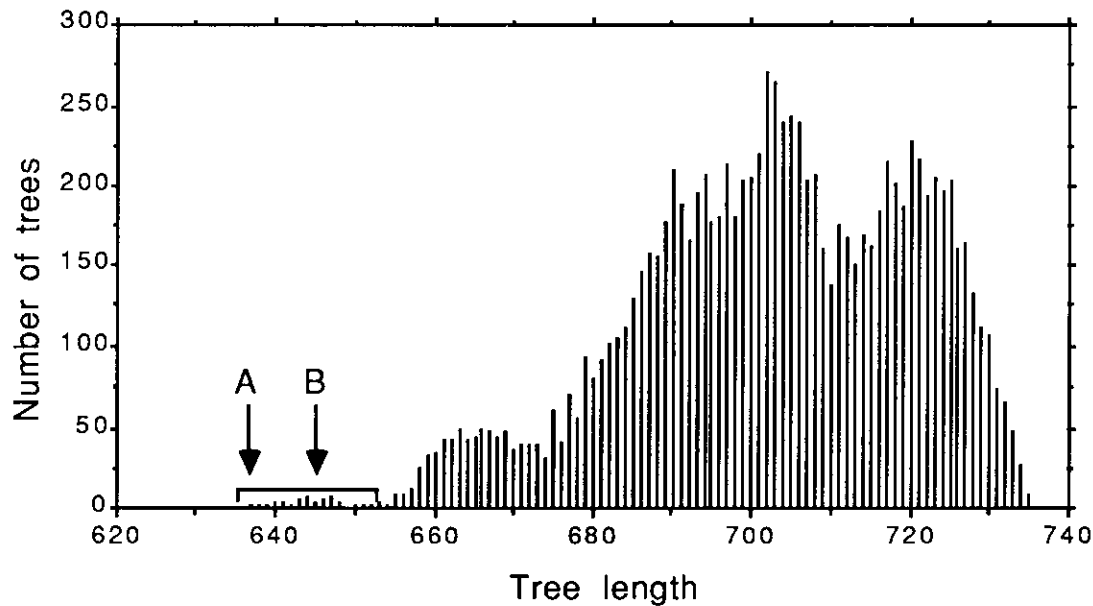


Fig. 2. Distribution of lengths of all possible tree topologies fit to the 28S rDNA sequence data. A = Shortest tree (637 steps); see Figure 3A. B = Distribution of 45 shortest trees used to construct consensus tree (see Figure 3B).

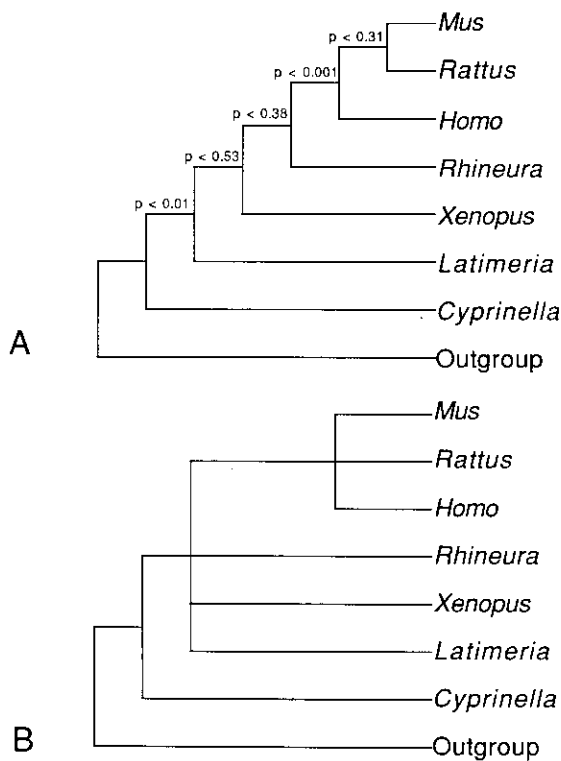


Fig. 3. A = Shortest tree, showing bootstrap probabilities (based on the number of bootstrap trees that do not contain the corresponding node). B = Strict consensus tree of the 45 shortest trees.

sensus analysis of near-parsimonious trees. Strict consensus trees were calculated for all solutions near the most parsimonious tree in order to determine the minimum number of steps needed to break the branch that united *L. chalumnae* to other vertebrates.

Results and discussion

The sequences from over 2 kb of the 28S rRNA genes of the study species are given in Appendix 1. We aligned 1989 nucleotide positions among the species; unambiguous alignment among all the species was not possible in six divergent domains; these regions were not used in the phylogenetic analysis. These divergent domains are characterized by high G/C content and major length variation among the vertebrates, which makes unambiguous alignment between distantly related species difficult or impossible (Appendix 1). Sequence similarities within divergent domains do exist among the more closely related species (e.g. among the mammals), but the similarity among

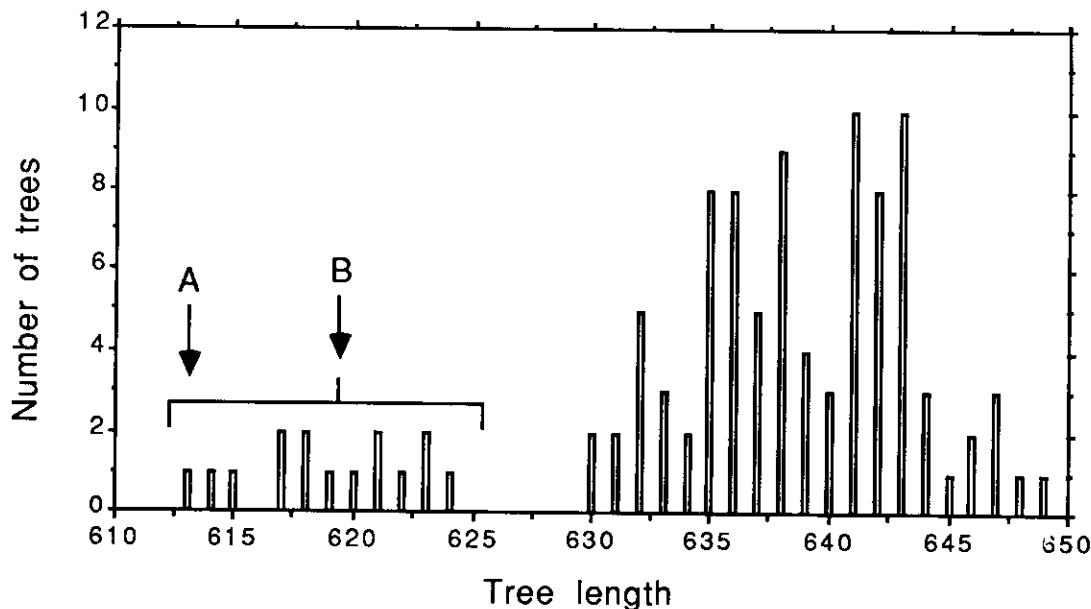


Fig. 4. Distribution of lengths of all possible tree topologies fit to the 28S rDNA sequence data with only one mammal (*Homo*) included. A = shortest tree (613 steps); this tree is the same as Figure 3A with the two rodent lineages removed. B = Distribution of the 15 shortest trees, which are all the possible groupings of *Latimeria* with the tetrapods to the exclusion of the actinopterygian.

distantly related taxa in these regions is not significantly better than random.

The distribution of all possible trees in the phylogenetic analysis was strongly skewed with a long left tail (Fig. 2). Such a distribution indicates that the data matrix has a high information content; there are very few trees that explain the data well and many that are poor solutions to the problem. There is a single most parsimonious tree which places *L. chalumnae* as the sister group to the tetrapods (Fig. 2, 3A). All other nodes of the most parsimonious tree are consistent with traditional views of vertebrate phylogeny (e.g. the two rodents, the three mammals, the four amniotes, and the five tetrapods are united as monophyletic groups).

The distribution of all possible trees (Fig. 2) is skewed partly as a result of the inclusion of three species of mammals in the analysis. However, if only one species of mammal is included the distribution is still strongly skewed, and all solutions that place *L. chalumnae* with the tetrapods are considerably shorter than the remaining possibilities (Fig. 4). As with the full data set, the shortest tree in this reduced data set includes the amniotes, the tetra-

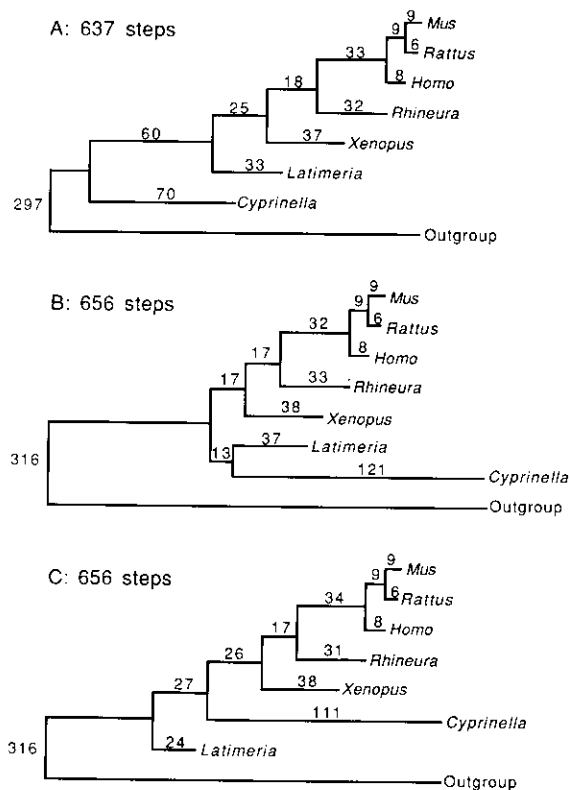


Fig. 5. Branch lengths of the most parsimonious tree (A) and alternative hypotheses of *Latimeria* relationships (B and C).

pods, and the sarcopterygians as monophyletic groups.

Bootstrap analysis of the full data set demonstrates that the branch that connects *Latimeria* with the tetrapods has a 99% confidence interval (Fig. 3A). All of the trees within 15 steps of the most parsimonious solution unite *Latimeria* and the tetrapods; all of these trees also unite the three species of mammals (Fig. 3B). The 45 shortest trees (637 to 652 steps) are the 45 possible resolutions of the consensus tree shown in Figure 3B.

The sequences of 28S rDNA clearly support the relationship of *L. chalumnae* with tetrapods (as suggested by Romer 1966, Rosen et al. 1981, Schultze 1987, Northcutt 1987, and Fritsch 1987, as well as in the preliminary analysis of a portion of these data by Hillis & Dixon 1989) rather than outside of actinopterygians plus tetrapods (as suggested by von Wahlert 1968, Lagios 1979, 1982, Løvtrup 1977, and Wiley 1979). The shortest trees that place *Latimeria* with actinopterygians or outside of actinopterygians plus tetrapods require significantly more steps than the most parsimonious tree (Fig. 5). These alternatives require a large number of reversals in the lineage leading to the actinopterygians. Based on the rDNA sequences, one can reject hypotheses that place *L. chalumnae* apart from tetrapods with a probability of making a Type I error of $p < 0.01$ (Fig. 3A).

One could argue that the rooting of the tree might not be accurate because of the inclusion of the distantly related *Drosophila* sequence in the outgroup. However, if only the partial lamprey sequence is used for the outgroup, then the results are the same. There is a single most parsimonious tree identical to Figure 3A, and all 37 trees within 12 steps of the most parsimonious tree place *Latimeria* with the tetrapods. A bootstrap analysis again indicates a 99% confidence interval for the *Latimeria*-tetrapod clade. The more complete *Drosophila* sequence lends further support to the position of the root, so it is informative as an additional outgroup.

Among the hypotheses that link coelacanths to other sarcopterygians, there is no consensus of opinion concerning the relationships among coelacanths, lungfishes, and tetrapods. Each of the three

possible resolutions has been suggested: coelacanths as the living sister group to lungfishes plus tetrapods (e.g. Rosen et al. 1981); lungfishes as the sister group to coelacanths plus tetrapods (e.g. Schultze 1987, Fritsch 1987); and tetrapods as the sister group to lungfishes plus coelacanths (Northcutt 1987). Because we have not yet sequenced the 28S gene of a lungfish, we cannot address this problem at present. However, the long branch that unites tetrapods and *L. chalumnae* suggests that an abundance of informative sites exist within the 28S gene that should be relevant to this question. In any case, hypotheses that place *Latimeria* outside of the Sarcopterygii are strongly rejected by the 28S rDNA sequences.

Acknowledgements

This work was supported by National Science Foundation grants BSR 8614622 and BSR 8796293 to D.M.H. We thank the Explorer's Club, the Virginia Institute of Marine Science Coelacanth Program (J.A. Musick, P.I.), and E.O. Wiley for providing the *Latimeria* tissue. John Gold and Scott Davis provided the clone of *Cyprinella lutrensis*, and William Ramos and Rafael de Sá provided laboratory assistance.

References cited

- Ansorge, W. & S. Labeit. 1984. Field gradients improve resolution on DNA sequencing gels. *J. Biochem. Biophys. Meth.* 10: 237-243.
- Felsenstein, J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39: 783-791.
- Fitch, W.M. 1984. Cladistic and other methods: problems, pitfalls, and potentials, pp. 221-252. *In*: T. Duncan & T.F. Stuessy (ed.) *Cladistics: Perspectives on the Reconstruction of Evolutionary History*, Columbia University Press, New York.
- Forey, P.L. 1988. Golden jubilee for the coelacanth *Latimeria chalumnae*. *Nature* 336: 727-732.
- Fritsch, B. 1987. Inner ear of the coelacanth fish *Latimeria* has tetrapod affinities. *Nature* 327: 153-154.
- Gonzalez I.L., J.L. Gorski, T.J. Campen, D.J. Dorney, J.M. Erickson, J.E. Sylvester & R.D. Schmickel. 1985. Variation among human 28S ribosomal RNA genes. *Proc. Natl. Acad. Sci. USA* 82: 7666-7670.

- Gyllenstein, V. & H. Erlich. 1988. Generation of single-stranded DNA by the polymerase chain reaction and its applications to direct sequencing of the HLA-DQA locus. *Proc. Natl. Acad. Sci. USA* 85: 7652-7656.
- Hadjiolov, A.A., O.I. Georgiev, V.V. Nosikov & L.P. Yavachev. 1984. Primary and secondary structure of rat 28S ribosomal RNA. *Nucl. Acids Res.* 12: 3677-3693.
- Hassouna, N., B. Michot & J.-P. Bachelierie. 1984. The complete nucleotide sequence of mouse 28S rRNA gene: implications for the process of size increase of the large subunit rRNA in higher eukaryotes. *Nucl. Acids Res.* 12: 3563-3583.
- Hillis, D.M. 1987. Molecular versus morphological approaches to systematics. *Ann. Rev. Ecol. Syst.* 18: 23-42.
- Hillis, D.M. & S.K. Davis. 1986. Evolution of ribosomal DNA: fifty million years of recorded history in the frog genus *Rana*. *Evolution* 40: 1275-1288.
- Hillis, D.M. & S.K. Davis. 1987. Evolution of the 28S ribosomal RNA gene in anurans: regions of variability and their phylogenetic implications. *Mol. Biol. Evol.* 4: 117-125.
- Hillis, D.M. & M.T. Dixon. 1989. Vertebrate phylogeny: evidence from 28S ribosomal DNA sequences, pp. 355-367. *In*: B. Fernholm, K. Bremer & H. Jornvall (ed.) *The Hierarchy of Life*, Elsevier Science Publishers, Amsterdam.
- Hillis, D.M., A. Larson, S.K. Davis & E.A. Zimmer. 1990. Nucleic acids III: sequencing, pp. 318-370. *In*: D.M. Hillis & C. Mortiz (ed.) *Molecular Systematics*, Sinauer Associates, Sunderland.
- Lagios, M.D. 1979. The coelacanth and the Chondrichthyes as sister groups: a review of shared apomorph characters and a cladistic analysis and reinterpretation. *Occ. Pap. California Acad. Sci.* 134: 25-44.
- Lagios, M.D. 1982. *Latimeria* and the Chondrichthyes as sister taxa: a rebuttal to recent attempts at refutation. *Copeia* 1982: 942-948.
- Løvtrup, S. 1977. *The phylogeny of Vertebrata*. John Wiley and Sons, London. 330 pp.
- Mullis, K.B. & F.A. Faloon. 1987. Specific synthesis of DNA *in vitro* via a polymerase catalyzed chain reaction. *Meth. Enzymol.* 155: 335-350.
- Northcutt, R.G. 1987. Lungfish neural characters and their bearing on sarcopterygian phylogeny. *J. Morph. Suppl.* 1: 277-297.
- Pustell, J. & F.C. Kafatos. 1982. A convenient and adaptable package of DNA sequence analysis programs. *Nucl. Acids Res.* 10: 51-59.
- Pustell, J. & F.C. Kafatos. 1984. A convenient and adaptable package of computer programs for DNA and protein sequence management, analysis, and homology determination. *Nucl. Acids Res.* 12: 643-655.
- Pustell, J. & F.C. Kafatos. 1986. A convenient and adaptable microcomputer environment for DNA and protein manipulation and analysis. *Nucl. Acids Res.* 14: 479-488.
- Romer, A.S. 1966. *Vertebrate paleontology*, 3rd ed. University of Chicago Press, Chicago. 468 pp.
- Roscn, D.E., P.L. Forey, B.G. Gardiner & C. Patterson. 1981. Lungfishes, tetrapods, paleontology, and plesiomorphy. *Bull. Amer. Mus. Nat. Hist.* 167: 159-276.
- Sanger, F., S. Nicklen & A.R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* 74: 5463-5467.
- Schultze, H.-P. 1987. Dipnoans as sarcopterygians. *J. Morph. Suppl.* 1: 39-74.
- Swofford, D.L. 1985 et seq. *Phylogenetic analysis using parsimony*. University of Illinois, Urbana. (computer program).
- Tabor, S. & C.C. Richardson. 1987. DNA sequence analysis with a modified bacteriophage T7 DNA polymerase. *Proc. Natl. Acad. Sci. USA* 84: 4767-4771.
- Tautz, D., J.M. Hancock, D.A. Webb, C. Tautz & G.A. Dover. 1988. Complete sequences of the rRNA genes of *Drosophila melanogaster*. *Mol. Biol. Evol.* 5: 366-376.
- von Wahlert, G. 1968. *Latimeria* und die Geschichte der Wirbeltiere: eine evolutionsbiologische Untersuchung. Gustav Fischer Verlag, Stuttgart. 125 pp.
- Ware, V.C., B.W. Tague, C.G. Clark, R.L. Gourse, R.C. Brand & S.A. Gerbi. 1983. Sequence analysis of 28S ribosomal DNA from the amphibian *Xenopus laevis*. *Nucl. Acids Res.* 11: 7795-7817.
- Wiley, E.O. 1979. Ventral gill arch muscles and the interrelationships of gnathostomes, with a new classification of the Vertebrata. *Zool. J. Linnean Soc.* 67: 149-179.

Appendix 1

Aligned sequences of 28S rDNA. The reference numbers correspond to the positions in the *Mus* sequence. Outgroups are designated by OD for outgroup, *Drosophila* and OL for outgroup, *Lampetra*. Positions that vary among the species are marked with an asterisk below the site. Sequences within brackets are not aligned. In the outgroup, positions that could not be aligned with the ingroup taxa are represented by 'N' (five next pages).

	11 00	11 20	11 40	11 60
Mus	CACGGGGTCGGCGCGATGTCGGGTACCCACCCGACCCGCTCTTGAACACGGACCAAGGAGTCTAACCGGTG			
Rattus	CACGGGGTCGGCGCGATGTCGGGTACCCACCCGACCCGCTCTTGAACACGGACCAAGGAGTCTAACCGGTG			
Homo	CACGGGGTCGGCGCGACGTGTCGGGTACCCACCCGACCCGCTCTTGAACACGGACCAAGGAGTCTAACCGGTG			
Rhineura	CCCGGGTCCGCGCGATGTCGGGTACCCACCCGACCCGCTCTTGAACACGGACCAAGGAGTCTAACCGGTG			
Xenopus	CCNAGGGTCCGCGCGATGTCGGGTACCCACCCGACCCGCTCTTGAACACGGACCAAGGAGTCTAACCGGTG			
Latimeria	CA-GGGGTCTGCGCGATGTCGGGTACCCACCCGACCCGCTCTTGAACACGGACCAAGGAGTCTAACCGGTG			
Cyprinella	CTCAGGTCGCGCGATGTCGGGTACCCACCCGACCCGCTCTTGAACACGGACCAAGGAGTCTAACCGGTG			
Outgroup (D)	ATTAACAATGCGAAAGATTGAGTACCTTCGGGACCCGCTCTTGAACACGGACCAAGGAGTCTAACATATG			
	*****	****	***	**

	11 80	12 00	12 20	
Mu	CGCGAGTCAGGGGCTCG-TC--CGAAAGCCGCGCTGGCGCAATGAAGGTGAAGGGC		[CCCGCCCGGGGGCCC]	GA
Ra	CGCGAGTCAGGGGCTCG-TC--CGAAAGCCGCGCTGGCGCAATGAAGGTGAAGGGC		[CCCGTTCGCGGGGGCCC]	GA
Ho	CGCGAGTCAGGGGCTCGCA---CGAAAGCCGCGCTGGCGCAATGAAGGTGAAGGGC		[GGCGCGCTCGCCGGCC]	GA
Rh	CGCGAGTCAGAGGCTCG-ACC-CGAAAGCC-CGCTGGCGCAATGAAGGTGACGGCG		[CGCGCCGGCC]	GA
Xe	CGCGAGTCAGAGGGACTCTGCGCGAAACC--CTGTGGCGCAATGAAGGTGAGGGCC		[GGGGCGCCCCGGCT]	GA
La	CGCGAGTCAGAGGGCAG-AG--CGAAAGCC-CCATGGCGCAATGAAGGTGAGGGCG		[CGGGCCGGCT]	GA
Cy	CGCGAGTCAGAGGGT-G-TC--CACGAGCCCCACGGCGCAATGAAGGTGAGCGGC		[GGCGCGCCGGCCC]	G-
OD	TGCAAGTTATTGGGATA-T-----AAACCTAATA--GCGTAATTAACCTGANNNNN			NN
	* * **** ***** * ***** * * ** *****			*

	12 40	12 60	12 80	13 00
Mu	GGTGGGATCCCGAGGC [CTCTCCAGTCC]	GCC-GAGGGCGCACACCAGCGCCGCTCTCGCCCGCCGCGCGGGGAGGTG		
Ra	GGTGGGATCCCGAGGC [CTCTCCAGTCC]	GCC-GAGGGCGCACACCAGCGCCGCTCTCGCCCGCCGCGCGGGGAGGTG		
Ho	GGTGGGATCCCGAGGC [CTCTCCAGTCC]	GCC-GAGGG-GCACACCAGCGCCGCTCTCGCCCGCCGCGCGGGGAGGTG		
Rh	GGTGGGATCCCGAGGC [GCCGA]	GCG-GAGGGCGCACACCAGCGCCGCTCTCGCCCGCC-CGTCGGGGAGGTG		
Xe	GGTGGGATCCCGCGC [CCCTCCCTCC1]	GCCGGCGGCACCACCAGCGCCGCTCTCGCCCGCCCGCTCGGGNGGTG		
La	GGTGGGATCCCGAGGC [CTCGT]	GCG-GGGGGCGCACACCAGCGCCGCTCTCACCCCGAGCGTCGGGGAGGTG		
Cy	GGTGGGATCCCGTCCG	GAG-GGGGGCGCACACCAGCGCCGCTCTCACCCGGTCCGCGGGGAGGTG		
OD	NNNNNNNNNNNNNNNN	NN		
	*****	*** *		

[1: GCCCCCCCGGGGGCGGGGGGGGGG]

	13 20	13 40	13 60	13 80
Mu	GAGCACGAGCGTACGCG-TTAGGACCCGAAAGATGGTGAACATATGCTTGGGCAGGGCGAAGCCAGAGGAACTCTGGTGG			
Ra	GAGCACGAGCGTACGCG-TTAGGACCCGAAAGATGGTGAACATATGCTTGGGCAGGGCGAAGC-AGAGGAACTCTGGTGG			
Ho	GAGCACGAGCGCACGTG-TTAGGACCCGAAAGATGGTGAACATATGCTTGGGCAGGGCGAAGCCAGAGGAACTCTGGTGG			
Rh	GAGCGTGAGCGCGCGTGC-TAGGACCCGAAAGATGGTGAACATATGCTTGGGCAGGGCGAAGCCAGAGGAACTCTGGTGG			
Xe	GNGCGTGAGCGCGCGGATTAGGACCCGAAAGATGGTGAACATATGCTTGGGCAGNGCGAAGCCAGAGGAACTCTGGTGG			
La	GAGCATGAGCGTGCCTGA-TAGGACCCGAAAGATGGTGAACATATGCTTGGGCAGGGCGAAGCCAGAGGAACTCTGGTGG			
Cy	GAGC--GAGAGCGCGCA-TGGTACCCGAAAGATGGTGAACATATGCTTGGGCAGGGCGAAGCCAGAGGAACTCTGGTGG			
OD	NNNNNNNNNNNNNNNNNTGTGACCCGAAAGATGGTGAACATATGCTTGGGCAGGGCGAAGCCAGAGGAACTCTGGTGG			
	** *			

	14 00	14 20	14 40	14 60
Mu	AGGTCCGTAGCGGTCTGACGTGCAAATCGGTCTGTCGA-CCTGGGTATAGGGGCGAAAGACTAATCGAACCATCTAGTA			
Ra	AGGTCCGTAGCGGTCTGACGTGCAAATCGGTCTGTCGA-CCTGGGTATAGGGGCGAAAGACTAATCGAACCATCTAGTA			
Ho	AGGTCCGTAGCGGTCTGACGTGCAAATCGGTCTGTCGA-CCTGGGTATAGGGGCGAAAGACTAATCGAACCATCTAGTA			
Rh	AGGTCCGTAGCGGTCTGACGTGCAAATCG-TGTCGA-CCTGGGTATAGGGGCGAAAGACTAATCGAACCATCTAGTA			
Xe	AGTCCGTAGCGGTCTGACGTGCAAATCGGTCTGTCGA-CCTNGGTATAGGGGCGAAAGACTAATCGAACCATCTAGTA			
La	AGGTCCGTAGCGGTCTGACGTGCAAATCGGTCTGTCGA-CCTGGGTATAGGGGCGAAAGACTAATCGAACCATCTAGTA			
Cy	AGG-CCGAGCGGTCTGACGTGCAAATCGGTCTGTCGA-CCTGGGTATAGGGGCGAAAGACTAATCGAACCATCTAGTA			
OD	AAGACCGAAACAGTTCTGACGTGCAAATCGATTGTCAGAA-TTGAGTATAGGGGCGAAAGACCAATCGAACCATCTAGTA			
	* *			

	14 80	15 00	15 20	
Mu	GCTGGTTCCTCCGAAGTTCCCTCAGGATAGCTGGCGCTCTCGC [TCCCGACGTA]			CGCAGTTTTA
Ra	GCTGGTTCCTCCGAAGTTCCCTCAGGATAGCTGGCGCTCTCGC [AACCGTTCGCTCGACAACC]			CGCAGTTTTA
Ho	GCTGGTTCCTCCGAAGTTCCCTCAGGATAGCTGGCGCTCTCGC [AGACCCGACGACCCCGCCA]			CGCAGTTTTA
Rh	GCTGGTTCCTCCGAAGTTCCCTCAGGATAGCTGGCGCTCTCGC [CGCGAACC]			CGCAGTTTTA
Xe	GCTGGTTCCTCCGAAGTTCCCTCAGGATAGCTGGCGCTCTCGC [GT]			CGCAGTTTTA
La	GCTGGTTCCTCCGAAGTTCCCTCAGGATAGCTGGTGTCTCGAGC [GAA]			CGCAGTTTTA
Cy	GCTGGTTCCTCCGAAGTTCCCTCAGGATAGCTGGCGCTCGCCA			CGCAGTTTTA
OD	GCTGGTTCCTCCGAAGTTCCCTCAGGATAGCTGGTGCATTTTA [ATATTATATAA]			AATAATCTTA
	*	* *****		*** * *

15|40 15|60 15|80 16|00
 Mu TCCGGTAAAGCGAATGATTAGAGGCTCTGGGGCCGAAA-CGATCTCAACCTATTCTCAAACCTTTAAATGGGTAAGAAGCC
 Ra TCCGGTAAAGCGAATGATTAGAGGCTCTGGGGCCGAAA-CGATCTCAACCTATTCTCAAACCTTTAAATGGGTAAGAAGCC
 Ho TCCGGTAAAGCGAATGATTAGAGGCTCTGGGGCCGAAA-CGATCTCAACCTATTCTCAAACCTTTAAATGGGTAAGAAGCC
 Rh TCTGGTAAAGCGAATGATTAGAGGCTCTGGGGCCGAAA-CGATCTCAACCTATTCTCAAACCTTTAAATGGGTAAGAAGCC
 Xe TCCGGTAAAGCGAATGATTAGAGGCTCTGGGGCCGAAA-CGATCTCAACCTATTCTCAAACCTTTAAATGGGTAAGAAGCC
 La TCTGGTAAAGCGAATGATTAGAGGCTCTGGGGCCGAAA-CGATCTCAACCTATTCTCAAACCTTTAAATGGGTAAGAAGCC
 Cy TCCGGTAAAGCGAATGATTAGAGGCTCTGGGGCCGAAA-CGATCTCAACCTATTCTCAAACCTTTAAATGGGTAAGAAGCC
 OD TCTGGTAAAGCGAATGATTAGAGGCTCTGGGGCCGAAA-CGATCTCAACCTATTCTCAAACCTTTAAATGGGTAAGAAGCC

16|20 16|40 16|60 16|80
 Mu CCGCTCGCTGGCGTGGAGCCGGG-CGTGGAATGCGA-GT--GCC-TAGTGGGCCACTTTTGGTAAGCAGAAGTGGCGCTG
 Ra CCGCTCGCTGGCGTGGAGCCGGG-CGTGGA-TGCGA-GT--GCC-TAGTGGGCCACTTTTGGTAAGCAGAAGTGGCGCTG
 Ho CCGCTCGCTGGCGTGGAGCCGGG--GTGGAATGCGA-GT--GCC-TAGTGGGCCACTTTTGGTAAGCAGAAGTGGCGCTG
 Rh CCGCTCGCTGGCGTGGAGCCGGG-CGTGGAATGCGA-GT--GCC-TAGTGGGCCACTTTTGGTAAGCAGAAGTGGCGCTG
 Xe CCGCTCGCTGGCTTGGAGCCGGGCGTGGGAATGCGNNGCAGCCATAGTGGGCCACTTTTGGTAAGCAGAAGTGGCGCTG
 La CCGCTCGCTGGCTTGGAGCCGGG-CGTGGAATGCGA-GT--GCC-TAGTGGGCCACTTTTGGTAAGCAGAAGTGGCGCTG
 Cy CCGCTCGCTGGCTTGGAGCCGGG-CGTGGAATGCGA-GA-GCCC-CAGTGGGCCACTTTTGGTAAGCAGAAGTGGCGCTG
 OD TAACCTTTGATATGA-CCNNNNNNNNNNNNNNNNNNNNCC-CAGTGGGCCACTTTTGGTAAGCAGAAGTGGCGCTG
 *** ** * **** ** * **** ** * **** ** * **** *

17|00 17|20 17|40 17|60
 Mu CGGGATGAACCGAACGCCGGGTTAAGGCGCCCGATGCCGACGCTCAT-CAGACCCAGAAAAGGTGTTGGTTGATATAGA
 Ra CGGGATGAACCGAACGCCGGGTTAAGGCGCCCGATGCCGACGCTCAT-CAGACCCAGAAAAGGTGTTGGTTGATATAGA
 Ho CGGGATGAACCGAACGCCGGGTTAAGGCGCCCGATGCCGACGCTCAT-CAGACCCAGAAAAGGTGTTGGTTGATATAGA
 Rh CGGGATGAACCGAACGCCGGGTTAAGGCGCCCGATGCCGACGCTCAT-CAGACCCAGAAAAGGTGTTGGTTGATATAGA
 Xe CGGGATGAACCGAACGCCGGGTTAAGGCGCCCGATGCCGACGCTCAT-CAGACCCAGAAAAGGTGTTGGTTGATATAGA
 La CGGGATGAACCGAACGCCGGGTTGTTGGCGCCCGATGCCGACGCTCAT-CAGACCCAGAAAAGGTGTTGGTTGATATAGA
 Cy CGGGATGAACCGAACGCCGGGTTAAGGCGCCCGATGCCGACGCTCAT-CAGACCCAGAAAAGGTGTTGGTTGATATAGA
 OD TGGGAGTAACCAACGTAATGTTACGTGCCAAATTAACAA-CTCATGCAGATACCATGAAAGGCGTTGTTGCTTAAAA
 * ** * **** ** *** ** **** ** * ** ** *

17|80 18|00 18|20 18|40
 Mu CAGCAGGACGGTGGCCATGGAAGTCGGAATCCGCTAAGGAGTGTGTAACAACCTACCTGCCGAATCAACTAGCCCTGAAA
 Ra CAGCAGGACGGTGGCCATGGAAGTCGGAATCCGCTAAGGAGTGTGTAACAACCTACCTGCCGAATCAACTAGCCCTGAAA
 Ho CAGCAGGACGGTGGCCATGGAAGTCGGAATCCGCTAAGGAGTGTGTAACAACCTACCTGCCGAATCAACTAGCCCTGAAA
 Rh CAGCAGGACGGTGGCCATGGAAGTCGGAATCCGCTAAGGAGTGTGTAACAACCTACCTGCCGAATCAACTAGCCCTGAAA
 Xe CAGCAGGACGGTGGCCATGGAAGTCGGAATCCGCTAAGGAGTGTGTAACAACCTACCTGCCGAATCAACTAGCCCTGAAA
 La CAGCAGGACGGTGGCCATGGAAGTCGGAATCCGCTAAGGAGTGTGTAACAACCTACCTGCCGAATCAACTAGCCCTGAAA
 Cy CAGCAGGACGGTGGCCATGGAAGTCGGAATCCGCTAAGGAGTGTGTAACAACCTACCTGCCGAATCAACTAGCCCTGAAA
 OD CAGCAGGACGGTGAAGGAAGTCGGAATCCGCTAAGGAGTGTGTAACAACCTACCTGCCGAAGCAACTACGGGTAAA
 * ** * ** * ** *

18|60 21|00
 Mu ATGGATGGCGCTGGAGC-GTCGGGCCCATACCCGCGCTGCGCG AGGTGCA-GATCTTGGT-GGTAGTAGC
 Ra ATGGATGGCGCTGGAGC-GTCGGGCCCATACCCGCGCTGCGCG AGGTGCA-GATCTTGGT-GGTAGTAGC
 Ho ATGGATGGCGCTGGAGC-GTCGGGCCCATACCCGCGCTGCGCG AGGTGCA-GATCTTGGT-GGTAGTAGC
 Rh ATGGATGGCGCTGGAGC-GTCGGGCCCATACCCGCGCTGCGCG AGGTGCA-TATCTTGGT-GGTAGTAGC
 Xe ATGGATGGCGCTGGAGC-GTCGGGCCCATACCCGCGCTGCGCG GGGTGCA-GATCTTGGT-GGTAGTAGC
 La ATGGATGGCGCTGGAGC-GTCGGGCCCATACCCGCGCTGCGCG GGGTGCA-GATCTTGGT-GGTAGTAGC
 Cy ATGGATGGCGCTGGAGC-GTCGGGCCCATACCCGCGCTGCGAG GGGTGCA-GATCTTGGT-GGTAGTAGC
 OD ATGGATGGCGCTTAAGTTGATA-CCTATACATTACCGCTAAA TGTGTCACGA-GTTAGTGGTCTTAA-
 ** ** **** ** ***** ***** ** ** ** ** * ** ** *

21|20 21|40 21|60 21|80
 Mu AAATATTCAA-CG AGAAC--TTGAAGGCCGAAGTGGAGAAGGGTTCCATGTGAACAGCAGTTGAACATGGGTCCAG
 Ra AAATATTCAA-CG AGAAC--TTGAAGGCCGAAGTGGAGAAGGGTTCCATGTGAACAGCAGTTGAACATGGGTCCAG
 Ho AAATATTCAA-CG AGAAC--TTGAAGGCCGAAGTGGAGAAGGGTTCCATGTGAACAGCAGTTGAACATGGGTCCAG
 Rh AAATATTCAA-CG AGAAC--TTGAAGGCCGAAGTGGAGAAGGGTTCCATGTGAACAGCAGTTGAACATGGGTCCAG
 Xe AAATATTCAA-CG AGAAC--TTGAAGGCCGAAGTGGAGAAGGGTTCCATGTGAACAGCAGTTGAACATGGGTCCAG
 La AAATATTCAA-CG AGAAC--TTGAAGGCCGAAGTGGAGAAGGGTTCCATGTGAACAGCAGTTGAACATGGGTCCAG
 Cy AAATATTCAAAGCG AGAGCG-TTGAAGGCCGAAGTGGAGAAGGGTTCCATGTGAACAGCAGTTGAACATGGGTCCAG
 OD ---AGTTCAAGCG [2] AGAAGGGTTTNN
 ***** ** ****

[2: CACTGAAGTGG]

	22 00	22 20	22 40	22 60
Mu	TC-GGTCCTGAGAGATGGGCGAGTGCC-GTTCCGAAGGGACGGGCGATGGCTCCGTTGCCCTCGGCCGATCGAAAGGGA			
Ra	TC-GGTCCTGAGAGATGGGCGAGTGCC-GTTCCGAAGGGACGGGCGATGGCTCCGTTGCCCTCAGCCGATCGAAAGGGA			
Ho	TC-GGTCCTGAGAGATGGGCGAGCGCC-GTTCCGAAGGGACGGGCGATGGCTCCGTTGCCCTCAGCCGATCGAAAGGGA			
Rh	TC-GGTCCTAAGAGATAGGCGAGCGCC-GTTCCGAAGGGACGGGCGATGGCTCCGTTGCCCTCGGCCGATCGAAAGGGA			
Xe	TC-GGTCCTAAGAGATGGGCGAGCGCC-GTTCCGAAGGGACGGGCGATGGCTCCGTTGCCCTCGGCCGATCGAAAGGGA			
La	TC-GGTCCTAAGAGATGGGCGAAGCGCC-GTTCCGAAGGGACGGGCGATGGCTCCGTTGCCCTCAGCCGATCGAAAGGGA			
Cy	TCAGGTCTAAG-GATGGGCGACCCGCTTCGGAAGGGA-GGCCGATGGCTCCGTTGCCCTCGGCCGATCGAAAGGGA			
OD	NN			

	22 80	23 00		23 20
Mu	GTCGGGTTTCAGATCCCCGAATCCGGAGTGGCGGAGATGGGCGCC		GCGA	GG---CCAGTG-CGGTAACGGC
Ra	GTCGGGTTTCAGATCCCCGAATCCGGAGTGGCGGAGATGGGCGCC		GCGA	GGCGTCCAGTGCCGGTAACGGC
Ho	GTCGGGTTTCAGATCCCCGAATCCGGAGTGGCGGAGATGGGCGCC		GCGA	GGCGTCCAGTG-CGGTAACGGC
Rh	GTCGGGTTTCAGATCCCCGAATCCGGAGTGGCGGAGACGGGCGCG		-CGA	GGCGTCCAGTG-CGGTAACGGC
Xe	GTCGGGTTTCAGATCCCCGAACCCGAGTGGCGGAGACGGGCGCC	[3]	CCGA	[4] GGCGTCCAGTG-CGGGACGGC
La	GTTGGGTTTCAGATCCCCGAATCCGGATGGCGGAGAGGGGCGCC		GCTG	[5] GGCGTCCAGTG-CGGCAACGCCA
Cy	GTCGGGTTTCAGATCCNCGAACCCGAGTGGCGNAGACCCGCGC-		GCGA	GGCGCCAGTG-CGGTAACGGC
OD	NN		NNNN	NNNNNCAATCC-TGGCAACAGG

[3: CGCGGCCCCCC]

[4: CGCCTCGCGCGCGGGGGGGCGGG]

[5: TT]

	23 40	23 60		29 80	30 00
Mu	ACCGATCCCGGAGAAGCCG-GCGGGAG			GCGGCGCGCGGACTCTGGACGCGAGCCGGGCCCTTCCCG	
Ra	ACCGATCCCGGAGAAGCCG-GCGGGAG			GCGGCGCGCGGACTCTGGACGCGAGCCGGGCCCTTCCCG	
Ho	ACCGATCCCGGAGAAGCCG-GCGGGAG			GCGGCGCGCGGACTCTGGACGCGAGCCGGGCCCTTCCCG	
Rh	ACCGATCCCGGAGAAGCCGCGGGGAG			GCGGCGCGCGGACTCTGGACGCGAGCCGGGCCCTTCCCG	
Xe	ACCGATCCCGGAGAAGCCGNGNGGAG			CCGGCGCGCGGACTCTGGACGCGAGCCGGGCCCTTCCCG	
La	AC-GATCCCGGAGAAGCCG-GCGGGAG			CGGGCGCGCGGACTCTGGACGCGAGCCGGGCCCTTCCCG	
Cy	AACGAACCTGGAGAAGCTG-GCGAGAG			TGCGGCGCGGCGACTCTGGACGCGAGCCGGGCCCTTCCCG	
OD	AACGACCATAAAGAAGCCG-TCGAGAG			NN	

	30 20			
Mu	TGGATCGCCTCAGCTGCGGCGGGCGTCGC [6]			
Ra	TGGATCGCCCCAGCTGCGGCGGGCGTCGC [7]			
Ho	TGGATCGCCCCAGCTGCGGCGGGCGTCGC [8]			
Rh	TGGATCGCCCCAGCTGCGGCGGCTCGCCC [9]			
Xe	TGGATCGCCCCAGCTGCGGCGCGCGCC [10]			
La	TGGATCGCCTCAGCTGCGGCGCGCGGG [11]			
Cy	CGGATCTCCGAGCTACGGCTCGGTCG [12]			
OD	NN			

[6: 195 bases]

[7: 166 bases]

[8: 283 bases]

[9: GCCCTCCTCTCGCGGGACGGGCGGGCGGTT]

[10: 79 bases]

[11: CGTGGCAGGTCNNT]

[12: ACCTCCGTCGCGCGGCGGCGGCTCTCCTCGCGGGGGGGGGGCTGCGGCGGGGCCACCCGCGGG]

	32 80	33 00	33 20	33 40
Mu	GCGGACCAAGGGAATCCGACTGTTTAATTA AAACAAGCATCGGAAGGCCCGCGGCGGTTGACGCGATGTGATTTCT			
Ra	GCGGACTAGGGGAATCCGACTGTTTAATTA AAACAAGCATCGGAAGGCCCGCGGCGGTTGACGCGATGTGATTTCT			
Ho	GCGGACTAGGGGAATCCGACTGTTTAATTA AAACAAGCATCGGAAGGCCCGCGGCGGTTGACGCGATGTGATTTCT			
Rh	GCGGACCAAGGGAATCCGACTGTTTAATTA AAACAAGCATCGGAAGGCCCGAGGCGGTTGACGCGATGTGATTTCT			
Xe	GCGGACNANGGGAATCCGACTGTTTAATTA AAACAAGCATCGGAAGGCCCGAGGCGGTTGACGCGATGTGATTTCT			
La	GCGGACCAAGGGAATCCGACTGTTTAATTA AAACAAGCATCGGAAGGCCCGCGGCGGTTGACGCGATGTGATTTCT			
Cy	GCGGACCAAGGGAATCCGACTGTTTAATTA AAACAAGCATCGGAAGGCCCGCGGCGGTTGACGCGATGTGATTTCT			
OD	ACGACTTGGGGAATCCGACTGTTTAATTA AAACAAGCATGTTGATGGGCTA-GCGGGTGTGACACAATGTGATTTCT			

* * * * *

33|60 33|80 34|00 34|20
 Mu TGCCAGTGCTCTGAATGTCAAAGTGAAGAAATTCATGAAGCGCGGGTAAACGGCGGGAGTAACTATGACTCTCTTAAAG
 Ra TGCCAGTGCTCTGAATGTCAAAGTGAAGAAATTCATGAAGCGCGGGTAAACGGCGGGAGTAACTATGACTCTCTTAAAG
 Ho TGCCAGTGCTCTGAATGTCAAAGTGAAGAAATTCATGAAGCGCGGGTAAACGGCGGGAGTAACTATGACTCTCTTAAAG
 Rh TGCCAGTGCTCTGAATGTCAAAGTGAAGAAATTCATGAAGCGCGGGTAAACGGCGGGAGTAACTATGACTCTCTTAAAG
 Xe TGCCAGTGCTCTGAATGTCAAAGTGAAGAAATTCATGAAGCGCGGGTAAACGGCGGGAGTAACTATGACTCTCTTAAAG
 La TGCCAGTGCTCTGAATGTCAAAGTGAAGAAATTCATGAAGCGCGGGTAAACGGCGGGAGTAACTATGACTCTCTTAAAG
 Cy TGCCAGTGCTCTGAATGTCAAAGTGAAGAAATTCATGAAGCGCGGGTAAACGGCGGGAGTAACTATGACTCTCTTAAAG
 OD TGCCAGTGCTCTGAATGTCAAAGTGAAGAAATTCAGTAAGCGCGGGTCAACGGCGGGAGTAACTATGACTCTCTTAAAG

34|40 34|60 34|80 35|00
 Mu GTAGCCAAATGCCTCGTCATCTAATTAGTGACGCGCATGAATGGATGAACGAGATT Mu CCCACTGTCCCTACCT
 Ra GTAGCCAAATGCCTCGTCATCTAATTAGTGACGCGCATGAATGGATGAACGAGATT Ra CCCACTGTCCCTACCT
 Ho GTAGCCAAATGCCTCGTCATCTAATTAGTGACGCGCATGAATGGATGAACGAGATT Ho CCCACTGTCCCTACCT
 Rh GTAGCCAAATGCCTCGTCATCTAATTAGTGACGCGCATGAATGGATGAACGAGATT Rh CCCACTGTCCCTACCT
 Xe GTAGCCAAATGCCTCGTCATCTAATTAGTGACGCGCATGAATGGATGAACGAGATT Xe CCCACTGTCCCTACCT
 La GTAGCCAAATGCCTCGTCATCTAATTAGTGACNCGCATGAATGGATGAACGAGATT La CCCACTGTCCCTACCT
 Cy GTAGCCAAATGCCTCGTCATCTAATTAGTGACGCGCATGAATGGATGAACGAGATT Cy CCCACTGTCCCTACCT
 OD GTAGCCAAATGCCTCGTCATCTAATTAGTGACGCGCATGAATGGATTAACGAGATT OL CCCACTGTCCCTATNT

35|20 35|40 35|60 35|80
 Mu ACTATCCAGCGAAACCACAGCCAAGGGAACGGGCTTGGCGGAATCAGCGGGGAAAG-AAGACCCTGTTGAGCTTGACTCT
 Ra ACTATCCAGCGAAACCACAGCCAAGGGAACGGGCTTGGCGGAATCAGCGGGGAAAG-AAGACCCTGTTGAGCTTGACTCT
 Ho ACTATCCAGCGAAACCACAGCCAAGGGAACGGGCTTGGCGGAATCAGCGGGGAAAG-AAGACCCTGTTGAGCTTGACTCT
 Rh ACTATCCAGCGAAACCACAGCCAAGGGAACGGGCTTGGCGGAATCAGCGGGGAAAG-AAGACCCTGTTGAGCTTGACTCT
 Xe ACTATCCAGCGAAACCACAGCCAAGGGAACGGGCTTGGCGGAATCAGCGGGGAAAG-AAGACCCTGTTGAGCTTGACTCT
 La ACTATCCAGCGAAACCACAGCCAAGGGAACGGGCTTGGCGGAATCAGCGGGGAAAG-AAGACCCTGTTGAGCTTGACTCT
 Cy GCTATCCAGCGAAACCACAGCCAAGGGAACGGGCTTGGCGGAATCAGCGGGGAAAG-AAGACCCTGTTGAGCTTGACTCT
 OL ACTATCCAGCGAAACCACAGCCAAGGGAACGGGCTTNGCCGATCAGCGGGNAAAG-AAGACCCTGTTGAGCTTGACTCT

36|00 36|20 37|00
 Mu AGTCTGGCACGGTGAAGAGACATGAGAGGTGTAGAATAAGTGGGAGGCC [13] GCCGCCGGTGAAATACCACTACT
 Ra AGTCTGGCACGGTGAAGAGACATGAGAGGTGTAGAATAAGTGGGAGGCC [14] GCCGCCGGTGAAATACCACTACT
 Ho AGTCTGGCACGGTGAAGAGACATGAGAGGTGTAGAATAAGTGGGAGGCC [15] GCCGCCGGTGAAATACCACTACT
 Rh AGTCTGGCCCTGTGAAGAGACATGAGAGGTGTAGAATAAGTGGGAG-CCC [16] GCCGCCGGTGAAATACCACTACT
 Xe AGTCTGCAACTGTGAAGAGACATGAGAGGTGTAGGATAAGTGGGAGGCC [17] GCCGCCGGTGAAATACCACTACT
 La AGTCTGGCACTGTGAAGAGACATGAGAGGTGTAGGATAAGTGGGAGGCC [CG] GCCGCCGGTGAAATACCACTACT
 Cy AGTCTGGCACTGTGAAGAGACATGAGGGGTGTAGAATAAGTGGGAGGCC [18] GCCGCCGGTGAAATACCACTACT
 OL AGCCTGGCACGGTGAAGAGACATGAGAGGTGTAGGATAAGTGGGAGGTGC [19] GTGACCNGTGAAATACCACTACT

- [13: 52 bases]
- [14: 55 bases]
- [15: 57 bases]
- [16: ACGCGG]
- [17: 21 bases]
- [18: GGGTACCTGGGTCCACGGCGC]
- [19: GTCGCTCCCTCGCTAGCTCTNATGATGCAC]

37|20 37|40 37|60
 Mu CTCATCGTTTTTTCCTACTGACCCGGTGAGGCGGG--GGGGCGAGCCCC---GAGGGG--CTC---TCGCTTCTGGCGCC-
 Ra CTCATCGTTTTTTCCTACTGACCCGGTGAGGCGGG--GGGGCGAGCCCC---GAGGGG--CTC---TCGCTTCTGGCTCCG
 Ho CTGATCGTTTTTTCCTACTGACCCGGTGAGGCGGG--GGGGCGAGCCCC---GAGGGG--CTC---TCGCTTCTGGCTCC-
 Rh CTGATCGTTTTTTCCTACTTACCCGGTGAGGCGGG--GGGGCGAGCCCC---GAGTGG--CTC---TCGCTTCTGGCTCC-
 Xe CTTATCGTTTTTTCCTACTTACCCGGTGAGGCGNN--GGGGCGAGCCCC---GAGGGG--CTC---TCGCTTCTGGACCC-
 La CTTATCGTTTTTTCCTACTTACCCGGTGAGGCGGG--GGGGCGAGTCCC---GAGGGG--CTC---TCGATTCTGGTGAC-
 Cy CTTATCGTTTTTTCCTACTTACCCGGTGAGGCGGG--AGGCC--AGCCCCG--GGCGGG--CTA---GCGCTTCTGGTGTC-
 OL CTGATCGTTTTTTCCTACTTACCCGGTGAGGCTCGGGGAGGCTGAGCTCTCGCTGGCGGGTGCTACATGCGCTTNTGGCTCC-

	37 80		38 00		38 20
Mu	AA--GCG [TCCGTCCC	CGCGTGGC]	GGCGGGCGCGACCCGCTCCGGGGACA	-GTGCCAGGTGGGG	
Ra	AA---CG [CGTCCGCGCGCGC]	GGCGGGCGCGACCCGCTCCGGGGACA	-GTGCCAGGTGGGG	
Ho	AA--GCG [CCCGCCCC]	GCCGGGCGCGACCCGCTCCGGGGACA	-GTGCCAGGTGGGG	
Rh	AA--GCG [TCGGCGCGG]	GCCGGGCGCGACCCGCTCCGGGGACA	-GCGTCAGGTGGGG	
Xe	AA--GCG [CNCGGCCCCCGC]	GCCGGGCGCGACCCGCTCCGAGGACA	-GTGGCAGGTGGGG	
La	AA--GCG [CCGGCTCG]	GCCGGGTGCGACCCGCTCCGGGGACA	-GTGGCAGGTGGGG	
Cy	AA--GCC [GGGGGGTCCCTCTCCGGAGGTTCCC]	CCCCCGGGAC--G-TCCC-GGACA	-GTGGCAGGTGAGG	
OL	AAGTGAC [TCTGGCGCCCGAACACTGTTTGACGGT]]	GCAGAGCACGATCCGCTCCGGGAACATGTGTATAGTGGGG		
	*** *		*****	*** *	*** * * * * * *

	38 40		38 60		38 80		39 00
Mu	AGTTTGACT	Mu	GGGGCGGTACACCTGTCAAACGGTAACGCAGGTGTCTTAAGGCGAGCTCAG-G-GAGG-ACAGA				
Ra	AGTTTGACT	Ra	GGGGCGGTACACCTGTCAAACGGTAACGCAGGTGTCTTAAGGCGAGCTCAG-G-GAGG-ACAGA				
Ho	AGTTTGACT	Ho	GGGGCGGTACACCTGTCAAACGGTAACGCAGGTGTCTTAAGGCGAGCTCAG-G-GAGG-ACAGA				
Rh	AGTTTGACT	Rh	GGGGCGGTACACCTGTCAAACGGTAACGCAGGTGTCTTAAGGCGAGCTCAG-G-GAGG-ACAGA				
Xe	AGTTTGACT	Xe	GGGGCGGTACACCTGTCAAACGGTAACGCAGGTGTCTTAAGGCGAGCTCAG-G-GAGG-ACAGA				
La	AGTTTGACT	La	GGGGCGGTACACCTGTCAAACGGTAACGCAGGTGTCTTAAGGCGAGCTCAG-G-GAGG-ACAGA				
Cy	AGTTTGACT	Cy	GGGGCGGTACACCTGTCAAACGGTAACGCAGGTGTCTTAAGGCGAGCTCAG-G-GAG--CAGA				
OL	AGTTTGACT	OD	GGGGCGGTACATCTCTCAAATAACGGAGGTGTCCCAAGGCCAGCTCAGTGGC-G--ACAGA				
			* * * * *				* * * * *

	39 20		39 40		39 60		39 80
Mu	AACCTCCC	GTGGAGCAGAAGGGCAAAGCTCGCTTGATCTTGATTTT	CAGTACGAATACAGACCGTGAAA--GCGGG-GC				
Ra	AACCTCCC	GTGGAGCAGAAGGGCAAAGCTCGCTTGATCTTGATTTT	CAGTACGAATACAGACCGTGAAA--GCGGG-GC				
Ho	AACCTCCC	GTGGAGCAGAAGGGCAAAGCTCGCTTGATCTTGATTTT	CAGTACGAATACAGACCGTGAAA--GCGGG-GC				
Rh	AACCTCCC	GTGGAGCAGAAGGGCAAAGCTCGCTTGATCTTGATTTT	CAGTACGAATACAGACCGTGAAA--GCGGG-GC				
Xe	AACCTCCC	GTGGAGCAGAAGGGCAAAGCTCGCTTGATCTTGATTTT	CAGTACGAATACAGACCGTGAAA--GCGGG-GC				
La	AACCTCCC	GTGGAGCAGAAGGGCAAAGCTCGCTTGATCTTGATTTT	CAGTACGAATACAGACCGTGAAA--GCGGG-GC				
Cy	AACCTCCC	GTGGAGCAGAAGGGCAAAGCTCGCTTGATCTTGATTTT	CAGTACGAATACAGACCGTGAAA--GCGGG-GC				
OD	AACCAACATAGAGCAAAGGGCAAATGCTGACTTGATCTCGGTGTT	CAGTACACAGGGACAGCAAAGC-CG--GC					
	* * * * *		* * * * *				* * * * *

	40 00		40 20		40 40
Mu	CTCA-CGA-TCCTTCTGACCTTTTGGGTTTT	-AAGCAGGAGG-T-GTCAGAAAAGTTACCACAGGGATAACTGGCTTGTG			
Ra	CTCA-CGA-TCCTTCTGACCTTTTGGGTTTT	-AAGCAGGAGG-T-GTCAGAAAAGTTACCACAGGGATAACTGGCTTGTG			
Ho	CTCA-CGA-TCCTTCTGACCTTTTGGGTTTT	-AAGCAGGAGG-T-GTCAGAAAAGTTACCACAGGGATAACTGGCTTGTG			
Rh	CTCA-CGA-TCCTTCTGACTTTTGGGTTTT	-AAGCAGGAGG-T-GTCAGAAAAGTTACCACAGGGATAACTGGCTTGTG			
Xe	CTCA-CGA-TCCTTCTGACTTTTGGGTTTT	-AAGCAGGAGG-T-GTCAGAAAAGTTACCACAGGGATAACTGGCTTGTG			
La	CTCA-CGA-TCCTTCTGACTTTTGGGTTTT	-AAGCAGGAGG-T-GTCAGAAAAGTTACCACAGGGATAACTGGCTTGTG			
Cy	CTCA-CGA-TCCTTCTGACTTTTGGGTTTT	-AAGCAGGAGG-T-GTCAGAAAAGTTACCACAGGGATAACTGGCTTGTG			
OD	CT-ATCGAATCCTTTTGGTTAAAGAGTTTTTAA-CAAGAGG-T-GTCAGAAAAGTTACCATAGGGATAACTGGCTTGTG				
	* * * * *		* * * * *		

	40 60		40 80		41 00		41 20
Mu	GC-GGCCAAGCGTTCATAG--CGA--CGTCGCTTTTTGATCCTTCGATGTCGGCTCTTCCTATCATTGTGAAGCAGAATTC						
Ra	GC-GGCCAAGCGTTCATAG--CGA--CGTCGCTTTTTGATCCTTCGATGTCGGCTCTTCCTATCATTGTGAAGCAGAATTC						
Ho	GC-GGCCAAGCGTTCATAG--CGA--CGTCGCTTTTTGATCCTTCGATGTCGGCTCTTCCTATCATTGTGAAGCAGAATTC						
Rh	GC-GGCCAAGCGTTCATAG--CGA--CGTCGCTTTTTGATCCTTCGATGTCGGCTCTTCCTATCATTGTGAAGCAGAATTC						
Xe	GC-GGCCAAGCGTTCATAG--CGA--CGTCGCTTTTTGATCCTTCGATGTCGGCTCTTCCTATCATTGTGAAGCAGAATTC						
La	GC-GGCCAAGCGTTCATAG--CGA--CGTCGCTTTTTGATCCTTCGATGTCGGCTCTTCCTATCATTGTGAAGCAGAATTC						
Cy	GC-GGCCAAGCGTTCATAGATCGAATCGTCGCTTTTTGATCCTTCGATGTCGGCTCTTCCTATCATTGTGAAGCAGAATTC						
OD	GC-GGCCAAGCGTTCATAG--CGA--CGTCGCTTTTTGATCCTTCGATGTCGGCTCTTCCTATCATTGTGAAGCAAATTC						
	* * * * *		* * * * *				*