Mitochondrial L-rRNA from Aspergillus nidulans: potential secondary structure and evolution

Heinrich G.Köchel and Hans Küntzel

Max-Planck-Institut für experimentelle Medizin, Abteilung Chemie, Hermann-Rein-Str. 3, D-3400 Göttingen, FRG

Received 12 May 1982; Accepted 14 June 1982

ABSTRACT

The alignment of gene sequences coding for <u>A</u>. <u>nidulans</u> mitochondrial L-rRNA and <u>E</u>. <u>coli</u> 23S rRNA indicates a strong conservation of primary and potential secondary structure of both rRNA molecules, except that homologies to the 5'-terminal 5.8S-like region and the 3'-terminal 4.5S-like region of bacterial rRNA are not detectable on mtDNA. The structural organization of the <u>A</u>. <u>nidulans</u> mt L-rRNA gene corresponds to that of yeast ω^+ strains: both genes are interrupted by a large intron sequence (1678 and 1143 bp, respectively) and by another smaller insert (91 and 66 bp) at homologous positions within domain V. An evolutionary tree derived from conserved L-rRNA gene sequences of yeast nuclei, <u>E</u>. <u>coli</u>, maize chloroplasts and six mitochondrial species exhibits a common root of organelle and bacterial sequences separating early from the nuclear branch.

INTRODUCTION

The comparative analysis of L-rRNA (large ribosomal subunit RNA) gene sequences from <u>E</u>. <u>coli</u> (1), maize chloroplast (2) and yeast nuclei (3,4) has led to a general secondary structure model of prokaryotic and eukaryotic L-rRNA containing six more or less conserved domains (5,6). Only few regions of this model are retained in the highly deleted L-rRNA gene sequences of mammalian mitochondria (5,6), and a complete secondary structure model of a mitochondrial L-rRNA has not yet been presented.

In this study we demonstrate that the <u>A</u>. <u>nidulans</u> mitochondria L-rRNA gene sequence (7) is homologous to the <u>E</u>. <u>coli</u> 23S rRNA sequence, and that both L-rRNA molecules can be folded in a very similar way. A phylogenetic tree analysis of a region common to all sequenced L-rRNA genes supports the previous notion (8,9) that fungal and mammalian mitochondrial rRNAs are more related to eubacterial than to nuclear rRNAs.

MATERIALS AND METHODS

The EcoRI fragment E3 containing the split L-rRNA gene (10) was isolated from recombinant plasmid panE3 (11) and sequenced by chemical cleavage methods (7). The alignment of L-rRNA sequences and the phylogenetic tree analysis of conserved sequences was performed as previously described (12,9).

RESULTS AND DISCUSSION

Fig. 1 shows an alignment of the two exon sequences of the <u>A</u>. <u>nidulans</u> mitochondrial L-rRNA gene (7) with the <u>E</u>. <u>coli</u> 23S rRNA sequence (1). The termini of the mitochondrial sequence are operationally defined by the start and end of sequence homology to the bacterial molecule.

		12	12'	13	13	•		_	11'	14 15	
E.c.	CAGTGTGTGTGTGTTAGTGGA	GCGTCTGGA	AGGCGCGCGCG	ATACAGGO	TGACAGOCC	GTACACAA	AAATGCAC	ATGCTGTG	AGCTOGAT	GAGTAGGGCGG	380
mt								<u>r</u>	TGTTA	-AGTAAAATGO	7212
	15'	16	16		14 '	3'					
	CAR ACCTOCA APCOTOT	A AT ATTOCCC	CACC ATCCT	CC ALCOCT	AAATACTOC	TGACTGACO	GATAGTGA	ACCAGTAC	CGTGAGGG	AAAGGCGAAA	480
E.C.	ATTCAAATC-TGTTT	GAATAAAGGGG	AACCTICCT	CTAAGGCT	AAATA-TAA	TATATAAGO	GATAGTGA	AA-AGTAC	CGTGAGGT	AAAATTTGAA	7305
											•
					••				181	ചില	
	17 17			2.	<u> </u>	~		~			
E.c.	GAACCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	GAAAAAGAAC	TGANACCGT	GTACOTAC	AAGCAGTGG	GAGCACGO	TAQGCGTG	TGACTGCG	TACCTTTT	GTATAATGGG	7374
mt	TAGTAGTTTTTATAAGGAGC				-					down with day	
	1 2	3		_3'	4		4	•			
E.c.	CAGCGACTTATATTCTGTA	COALCOTTAN	COGAATAGG-	GGAGCO	GAAGGG		CCGAG-TC	TTAACTGG	G		- 649
mt	CAGCAAGTTAATTTTAGAT	GCAAGTAAAG.	ATAACCAATI	TAAATTAT	AAAAAGTTA	TGATTAAAA	AGGTTATT	THAACTTA	GATAAACO	AATTATGAAA	1 7474
		2							-		
	2'			5		6			61	7	
	COTTA ACTIVICACIÓN	ATACHCCCCA	A PECCOPTO	TCTAGCC	TOCOLACOT	TGAAGGTT	GGTAACAC	TAACTGGA	GGACCIGA	CCGACTAATG	r 746
mt.	AATGAGTTAGTATCTAAAA	TTAGACCCGA	AGGCTAGTG	TCTTACC	TGGTCAGGG	TT			ACCCCAA	CGGGTTATC	r 7556
						-				<u>Y</u>	-
							8	0	0	,	
	· · · · · · · · · · · · · · · · · · ·						-			condemocratic	
E.c.	TGAAAAAATTAGCGGATGAC	TTGTGGCTGG	GG-GTGAAAG	GCCAATC/	AAACCGGGGAG	-TAGCTGG	TTTCCGC	AAAGCTAT	ATGAGTA	GTAATTTAAT	C 7654
	TGIAIAGAIAICCGAAGA			FROMMANO							1051
		10				_	¹⁰				-
E.c.	ATTCATCTCCGGGGGGTAGA	GCACTGTTTC	GGCAAGGGG	GTCATCCC	GACTT-AC	CAACCCGA	TGCAA	ACTGCGA	ATA		- 920
mt	AA-CATCTTAGCAGGTACA	GAACTGTGAT	CTCGGAT	AATATCAA	TIGATTTAT	CAACATCG	GGGGGGTTGT	AGTGACT	TACTGGA	GAGTATTTGCA	C 7751
		1-					-				
			1	1	1	111	12	12'			_
									· · ·		
E.c.	CCGGAGAATGT	TATCACGGGA	GACA-CACG	GCGGGTCC	TAACGTCCGT	CGTGAAGA	GGGAAACA	CCCAGAC	CGCCAGCT	AAGGTCCCAAA	G 1011
E.c. mt	TCGGAATGCCAAAG-ATG	TATCACGGGA	GACA-CACG	GCGGGTGC TAGTGCGA	TAACGTCCG1 TAACGTICGT/	TGTCAAGA	GGGAAACAA GGGAAACAA	CCCAGAC	CGCCAGCT CAAGAGTT	AAGGTCCCAAA	G 1011 A 7849
E.c. mt	TCGGAATGGCAAAG-ATG	TATCACGGGA	GACA-CACG	GCGGGTGC TAGTGCGA	TAACGTCCG1	TGTCAAGA	GGGAAACAA GGGAAACAA	CCCAGAC CCCAGAA	CGCCAGCT	AAGGTCCCAAA	G 1011 A 7849
E.c. mt	CCGGAGAATGI TCGGAATGGCAAAG-ATG/	TATCACGGGA	GACA-CACG	GCGGGTGC TAGTGCGA	TAACGTCCGI TAAGGT[FGT/	TGTCAAGA	GGGAAACAA GGGAAACAC	CCCAGAC CCCAGAA	COCCAGCT	AAGGTCCCAAA	G 1011 A 7849
E.c. mt	TCGGAATGGCAAAG-ATG	13	AATCOGACA	GCGGGTGC	TAACGTCCGT TAAGGTIGT/ 14	TGTGAAGA	14 '		CGCCAGCT CAAGAGTT	AAGGTTCCCAAA	G 1011 A 7849
E.c. mt E.c. mt	TCGGAATGCCAAAG-ATG	13 ACGATGTGGG	AATCOGACA	ACAGCCAG	TAACGTCCGT TAAGGTIGT/ 14 GATGTTGGCT GAAATAGGC1	TGTCAAAGA	14 '	CCCAGAC CCCAGAAA TAAAGAA	AGCGTAAC	AAGGTCCCAAA AAGGTTCCAAA AGGTCACTGGT AGAGCACTGAT	G 1011 A 7849 C 1109 T 7948
E.c. mt E.c. mt	TCGGAATGGCAAAG-ATG	13 LACGATGTGGG	AATCOGACA	GCGGGTGC TAGTCCGA ACAGCCAG ACAGCCAG	TAACGTCCGI TAACGTIGT/ 14 GATGTTCGCI GAAATAGGCI	TGTCAAGA	14 '	CCCAGAC CCCAGAAA TAAAGAA TGAAGAA	AGCGTAAC	AAGGTCCCAAA AAGGTTCCAAA AGCTCACTGGT AGAGCACTGAT	G 1011 A 7849 C 1109 T 7948
E.c. mt E.c. mt	TCGGAATGGCAAGATG	13 ACGATGTGGGG	AATCOGACA	ACAGCCAG ACAGCCAG	TAACGTCCGT TAACGTICGT 14 GATGTTCGCCI GAAATAGGCI	TGTCAAGA TGTCAAAA TGAGAAGQA TJAGAAGGG	14 '	TAAAGAA	AGCGTAAC	AAGGTCCCAAA AAGGTTCCAAA AGCTCACTGGT AGAGCACTGAT	G 1011 A 7849 C 1109 T 7948
E.c. mt E.c. mt	TCGGAATGGCAAAG-ATG	13 IACGATGTTGAAT	AAGGGCCAG	ACACCCAG ACACCCAG ACAACCAG	TAACGTCCGT TAACGTICGT 14 GATGTTGGC1 GAAATAGGC1	TGTCAAGA	14 '	CCCAGAC CCCAGAAA TTAAAGAA TGAAGAAC	AGCGTAAT	AAGGTCCCAAA AAGGTTCCAAA AGGCTCACTGGT AGAGCACTGAT	G 1011 A 7849 C 1109 T 7948
E.c. mt E.c. mt	СССССАААТСЯ ТОССАААТСОСАААССААТСЯ ТОЛТОСТТАЛОТОСССААТСЯ ТТАТТОСТТАЛОТОСССААТСЯ ТАТТОСТТАЛОТОСССААТСЯ САСТОСОСССАТОЛОСИИ САСТОСОСССАТОЛОСИИ	TATCACGGGA	AAGGGCCAG	ACAGCCAG ACAGCCAG ACAATCAG	TAACGAAGCTC	TAGAAGA	IL I I I I I I I I I I I I I I I I I I	TTAAAGAA TTAAAGAA TTGAAGAA	AGCGTAAT	AGGTCACTGGT AGGTCACTGGT AGAGCACTGAT TATGGGT	G 1011 A 7849 C 1109 T 7948 T 1181
E.c. mt E.c. mt E.c. mt	TCGGAATGGCAAGAATG TCGGAATGGCAAGAATG TCATCGTTAAGTCGGC-AA TTATTGTTCAGTGGAATT	TATCACGGGA -ATGTTGAAT 13 IACGATGTGGG IAGGATGTGGG IAGGAAGATGTA CGGAAGATGTA CGAAAATTTA	AAGGGCCAG	ACAGCCAG ACAGCCAG ACAATCAG 13' AACCATGC	TAACGTCCGT TAACGT[TGT/ 14 GATGTTCGCC GAAATAGGCT ACCGAAGCCT ACCGAAGCCT	TOTCAAGA	IL I	TTAAAGAA TTAAAGAA TTGAAGAA TGAAGAC	AGCGTAAT CTCGTAAC	AGGTCCCAAA AAGGTTCCAAA AGGCTCCCAAA AGGCCACTGGT AGAGCACTGAT TATGCGT TTAATATATATATATATATATATATATATATATATAT	G 1011 A 7849 C 1109 T 7948 T 1181 T 181
E.c. mt E.c. mt	COGANEATTO TCGGAATGGCAAAG-ATG TCATGGTTAAGTGGG-AA TTATTGTTGAGTGGAG-AA TTATTGTTGAGTGGAATTA	TATCACGGGA ATGTTGAAT 13 13 14ACGATGTGGGG 14GGATGTTA 	GACA-CACG JAATCGCACA 	ACACCCAG ACATCAG 13' AACCATCA AACAATAT	TANEGTCCGT TAACGTICGT/ 14 GATGTTCGCCI GAAATAGGCI ACCGAAGCTC ACCGAAGCTC	TTAGAAGQA TTAGAAGQA TTAGAAGQA TTAGAAGQA TTAGAAGQA 15 CCCCCACCCC TTGTACACA	ILI GCCATCATT GCCATCATT GCCATCATT ACGQT ACGQT ACGQT		AGCGTAAT CTCGTAAT CTCGTAAC	AGGTCCCAAA AGGTTCCAAA AGGCTCACTGGT AGAGCACTGAT 	G 1011 A 7849 C 1109 T 7948 T 1181 T 181
E.c. mt E.c. mt E.c. mt	COGGAATGCAAAGAATG TOGGAATGCCAAAGAATG TTATTGTTATGTGAGTGAAATT TATTGTTGAGTGAAATT GAGTCGGCCTGCGG AATTTTAGTAATAAACA 81	TATCACGGGA -ATCTTGAAT 13 IACGATGTGGGG IAGGATGTTGGG IAGGATGTTTA	GACA-CACG JAACCGCCAG TTATTATIAAT-	CCCCCAC ACACCCAC ACATCAG 13' AACCATCA AACAATAT 33	TAACCTCCCT TAACCTTCGT 14 GATGTTGGCT GAAATAGGCT ACCCAAGCTC ACCCAAGCTC	TAGAAGA	IGGAAACAA 14 ' GCCATCATT GCCATTTTT GCCATTTTT ACGQT	TAAAGAA TIGAAGAA TIGAAGAA TIGAAGAA	AGCGTAAT CTCGTAAT CTCGTAAC	AGGTCCCAAA AGGTTCCAAA AGGCTCACTGGT AGAGCACTGAT 15' 	G 1011 A 7849 C 1109 T 7948 T 1181 T 1181
E.c. mt E.c. mt E.c. mt	ССССАЛАТТО ТССАЛАСТАЛССАЛАС-ЛАТ ТСАЛСССТАЛАССАЛАС-ЛАТ ТАТТССТСАСССС	TATCACGGGA ATGTTGAAT 13 IACCGATGTGGG IAGGATGTTTA CGGAAGATGTTA CGGAAGATGTA CGAAGATGTA	GACA-CACG MATCOCACA 	ACAGCCAG ACAGCCAG ACAATCAG 13' AACCATCA AACAATAT 33 AGGTGTGC	TANEGTECET TANEGTICT 14 GATGTTEGEC GAAATAGGCT ACCEAAGCTC ACCEAAACCT TGTGA	TAGAAGGA TAGAAGGA TAGAAGGA 15 CCCCCACCC TOTACACCA	CCGAAACAA CCGAAACAA 14 ' GCCATCATT GCCATCATT ACCGT TATTATATAA	TAAAGAA TGAAGAA TGAAGAC	AGCGTAAT CTCGTAAC CTCGTAAC CTCGTAAC	AAGGTTCCAAA AAGGTTCCAAA AAGGTTCCAAA AAGGTCCAAC AGAGCACTCGT 15' 	G 1011 A 7849 C 1109 T 7948 T 1181 T 8048 G 1239
E.c. mt E.c. mt E.c. mt	CATTOGATAGOCAACAATO	TATCACGGA ATCTGAAT 	GACA-CACG MATCGACA AACCOCCAG TTATTAAT- AACCOCCAG AACCOCCAG AACCOCCAG AACCOCCAG	ACAGECEAG ACAGECEAG ACAATCAG 13' AACCATGE AACAATATA 33 AGGTGTGE TATTATTA	TALEGTCCET TALEGTCET 14 GATGTTGGCT GAAATAGGCT ACCEAAGCTC ACCEAAGCTC ACCEAAGCTC ACCEAAGCTC ACCEAAGCTC ACCEAAGCTC	TAGAAGGA TAGAAGGA TAGAAGGA TAGAAGGG TAGAAGGG TGTAGACG TGTAGACAG	TATAAATA	СССАБАС СССАБАА ТАЛАБАА ГАЛТАТАТА ГАЛТТАТА	CGCCAGCT CAAGAGTT 	AAGGTTCCAAA AAGGTTCCAAA AAGGTTCCAAA AAGGTCCAC AGAGCACTGAT 15' 	G 10111 A 7849 C 1109 T 7948 T 1181 T 1181 T 1181 G 12398 S 148
E.c. mt E.c. mt E.c. mt	CONTROLOGIA	13 13 13 13 13 13 13 13 13 13 13 13 13 1	GACA-CACG AATCOGACA 	ACACCCAG ACACCCAG ACATCAG 13' AACCATCAG AACAATAT 33 AGGTGTGC	TAACGTCCGT TAACGTCGTC 14 GATGTTGGCT GAAATAGGCT ACCGAAGCTC AC	TAGAAGGA TAGAAGGA TAGAAGGA TAGAAGGG TAGAAGGG 15 SCCCCCCCC TGTACACCA TGTACACACA TATAATAAG	ILCOMAACAA ILL I ILL I GCCATCATT GCCATCATT GCCATCATT GCCATCATT GCCATCATATA GCCATCATA GCCATCATA GCCATCATA GCCATCATA GCCATCATA GCCATCATATA GCCATCATATA GCCATCATATA GCCATCATATATATATATATATATATATATATATATATAT	БССДАБАА СССДАБАА ГГАЛАБАА ГГАЛТАГА ГГАЛТТТА	AGCGTAAT TOTATATAA TATATAAA TATATAAA	AAGGTCCCAAA AAGGTCCCAAA AAGGTCCCAAA AAGGTCCCAAA 15' 	G 1011 A 7849 C 1109 T 7948 T 1181 T 1181 T 1181 G 1239 A 8148
E.c. mt E.c. mt E.c. mt	COGANEATTO TCCGANTGCCANG-ANG TCCATCGCTAACTOCC-AN TCATCGTTAACTOCC-AN TTATTGTTCATCATCAACA AATTTTAGTAATAACA 81 CT-TCCCTTCCCCTACCCCG	13 13 13 13 13 13 13 13 13 13	AAGGGCCAG AATCOGACAA AAGGGCCAG ATTATIAAT-	ACACCCAG ACCACCCAG ACCACCCAG ACCATCAG 13' AACCATCAG AACAATAT 33 AGCATCAC 33 AGCATCAC AACAATAT	TAAGGTCGT TAAGGTCGT 14 GATGTTGGGC GAAATAGGCT ACCGAAGCTC ACGAAGCTC ACCGAAGCTC ACCGAAGCTC ACCGAAGCTC ACCGAAGCTC ACCGAAGCTC ACCGAAGCTC ACCGAAGCTC ACCGAAGCTC ACCGAAGCTC ACCGAAGCTC ACCGAAGCTC ACCGAAGCTC ACCGAAGCTC ACCGAAGCTC ACCGAAGCTC ACCGAAGCTC ACCGAAGCTC ACCGA	ГОЛГАЛАСА ТОГОЛАЛА ТАСАЛОС ТАСАЛОС ТОГОЛАСАС ТОГОЛАСАСА ТОГОЛАСАСА ТОГОЛАСАСА ТОГОЛАСАСА ТОГОЛАСАСА ТОГОЛАСАСА ТОГОЛАСАСА ТОГОЛАСАСА ТОГОЛАСАСА ТОГОЛАЛАСА ТОГОЛАСА ТОГОЛАЛАСА ТОГОЛАЛАСА ТОГОЛАЛАСА ТОГОЛАЛАСА ТОГОЛ	БОДАЛАСА <u>ВО</u> ДАЛАСА <u>14'</u> <u>GCCATCATT</u> <u>GCCATCATT</u> <u>GCCATCATT</u> <u>GCCATCAT</u> <u>GCCATCAT</u> <u>GCCATCAT</u> <u>GCCATCAT</u> <u>GCCATCAT</u>	ГССДАСАС СССДАСАДА ТГАААСАА ГСААСАС ГСААТТАТА ГГАТТТТА	AGCGTAAT CTCGTAAC TTCGTAAC TATATTAA	AAGGTTCCAAA AAGGTTCCAAA AAGGTTCCAAA AAGGTCCAA TCCAA TCCAA TCCAA AAATTTATCAAT 4,	G 1011 A 7849 C 1109 T 7948 T 1181 T 8048 G 1239 A 8148
E.c. mt E.c. mt E.c. mt E.c. mt	COGATOR COGATO	ТАТСАСБСАНА АТСТТБААТ 13 13 13 13 13 13 13 13 13 13	GACA-EACCO AATCIGOACA AATCIGOACA AACGOGOCCAG ITTATIAAT- AACGOGOCTA AACGOGOCTA AACTATIATA AACGOGATAATA AACTATIATA	CCCCCTC TAGTOCCA ACACCCAG ACATCAG 13' AACCATTCAG 33 ACCATTCAG 33 ACCATTATTA 1 CCCCCCCAG	14 GATGTTGGC 14 GATGTTGGC ACCGAAGCTC ACCGAAGCTC TDTCA	СССТСАЛАСА <u>ITAGAAGGE</u> <u>ITAGAAGGE</u> <u>IS</u> <u>SCCCCCACCG</u> <u>IS</u> <u>SCCCCCACCG</u> <u>IS</u> <u>SCCCCCACCG</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u> <u>IS</u>	50GAAACAA 66GAAACAA 14 14 66CCATCATT 66CATCATT 66CATCATT 70 70 70 70 70 70 70 70 70 70 70 70 70	TAAAGAA TAAAGAA TGAAGAA TGAATTATA TAATTATA	AGCGTAAT CTCGTAAC TTATATAA TATATAA 4 TTCCAACGT	AAGGTCCCAAA AAGGTCCCAAA AAGGTCCCAA AAGGTCCA 15' 	G 1011 A 7849 C 11099 T 7948 T 1181 T 1181 C 12398 G 12397 G 1337
E.c. mt E.c. mt E.c. mt E.c. mt	ССССАБАТСТ ТСССАТАСТАЛСТСССАЛАТТС ТССАТСССТАЛСТАСССАЛАСТИ ТАТТССТСАССССС	ТАТСАСССА - АТСТТСААТ - АТСТТСААТ - 3 - 3 - 3 - 3 - 4 - 4 - 4 - 4 - 4 - 4 - 4 - 4	GACA-EACCO AATCORACA AATCORACA AAGOGOCCAG TTATTATATAT AAGOGOCCTA AAGOGOCTA AAGOGOCTA AAGOGOATATATATAT AGOGATATATATATATATA COATATATATATA	ACAGCCAG ACAGCCAG ACATCAG 13' AACCATCAG AACAATAT 33 AGCTATCAG AACAATAT 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	талестросст талестрост талестроти аластрост алагодска ассалатовст ассалатовст ассалатовст ассалатовст ассалатовст тота тота тота тота тота тота тота то	ССТСАЛАСА ТГАСАЛССА ТГАСАЛССА ТГАСАЛССА ТГАСАЛССА ТОТАСАЛССА ТОТАСАЛССА ТОТАСАЛССА СССССАСАЛ- СССССАСАЛ- СССССАСАЛ- СССССАЛАСА СССССАЛАСА СССССАЛАСА СССССАЛАСА СССССАЛАСА СССССАЛАСА СССССАЛАСА СССССАЛАСА СССССАЛАСА СССССАЛАСА СССССАЛАСА СССССАЛАСА СССССАЛАСА СССССАЛАСА СССССАЛАСА СССАЛАСА СССА	БОДАЛАСА <u>вод</u> аласа <u>14</u> <u>GCCATCATT</u> <u>GCCATCATT</u> <u>GCCATCATT</u> <u>GCCATCATT</u> <u>GCCATCATT</u> <u>GCCATCAT</u> <u>GCCATCAT</u>	TAAAGAA TAAAGAA TAAAGAA TAAATAA TAATTATA TAATTATA TAATTATA	CEGCAGET CAAGAGTT AGCGTAAT TOTOGTAAC TATATTAA TATATTAA 4 TCCAACGT TCCAACGT TTAAATT	AAGGTCCCAAA AAGGTCCCAAA AAGGTCATCCAA AAGGTCAAC AAGGTCAAC AAGTCAACAAC 33' TCCCCGCC	G 1011 A 7849 C 1109 T 7948 T 1181 T 1181 T 1181 G 12399 A 8148 G 13370 A 8240
E.c. mt E.c. mt E.c. mt E.c. mt	ССССАЛАТТ ТССАТССТАЛСТСССАЛАТТ ТССАТССТТАЛСТСССА-ЛА ТСАТССТТАЛСТСССС-ЛА ТТАТСССТТАЛСТСССС-ЛА САСТСССССТСССС АЛАТТТТАСТАЛАТАЛАСАА 81 СТ-ТСССТАССССАЛССССА ТАТСАСАДСТСССАЛСССА ТАТСАСАДСТСССАЛСССА ТАТСАСАДСТСССАЛСССА ТАТСАСАДСТСССАЛСССА	13 13 13 13 13 13 13 13 13 13	GACA-FACG MATCORACA ALGOGOCCAG TTATTMAT- INCOGOCOCTA INCOCTA I	GCGGGGTGC TAGTGCGAG ACAGCCAG 13' AACCATCAG 33 AGCGTGTGCC TATTATA 1	ТАЛЕСТССКІ ТАЛССТРОСКІ 14 САТОТГОССІ АСССАЛАТАССТ ТОТСА	Согдаласа <u>IT</u> адаласа <u>IT</u> адаса <u>IT</u> адаса	50GAAACAA ессаласаа 14 ' 60CATCATT 60CATCATT 60CATCATT 60CATCATT 74 ' 60CATCATT 74 ' 74 '	TTAAAGAA TTAAAGAA TTAAAGAA TTAATTATA TTATTTTA	CCCCAGCT CAAGAGT CTCGTAAC CTCGTAAC CTCGTAAC CTCGTAAC CTCGTAAC CTCGTAAC CTCGTAAC CTCGTAAC CTCGTAAC CTCGTAAC CTCGTAAC CTCGTAAC CTCGTAAC CTCGTAAC CTCGTAAC CTCGTAAC	AAGGTCCCAAA AAGGTCCCAAA AAGGTCCCAAC AAGGTCAAC 15' 	G 1011 A 7849 C 1109 T 7948 T 1181 T 1181 T 1181 G 12398 A 8148 G A 8240
E.c. mt E.c. mt E.c. mt E.c. mt	ССССАВАТСЯ ТОССАНТСКАЛАСТАТСЯ ТОССАНТСКАЛАСТАТСЯ ТОТАТТОТТАКТОСССАЛАСТАТСЯ ТАТТОТТАКТОКСАЛОССАЛАСТА САСТСОССССТОССС АЛАТТТТАСТАТАТАЛАСА 81 СТ-ТФССТАБОССАЛОССАЛОССА СТАТТАСТАКАТОТОССАЛОССА ТАТТСКАЛОТОССАЛОССАЛОССАЛОССАЛОССАЛОССАЛОССАЛО	тателевана 		GCGGGTGC TAGTGCGA ACACCCAG ACATCAG 13' AACCATCAG 13' AACCATCAG 33 AGGTGTGC 1 CCGGTGCAA 4 CCGGGTGCAA	тысеттосей такеептоси 14 актотосесс далатладест актотосесс актотосесс актотосе актосе актос актос актос актос актос актос актос актос акто	COTGALACA ITAGALOGA TAGALOGA TAGALOGA TAGALOGA 15 COCCACCO CATALATALO 2 COCCOCACO CATALATALO 2 COCCOCAL COCTALO 5 5	БОСАЛАСАЛ ессаласам 14: СССАТСАТ СССАТСАТ СССАТСТТ АСССТ ТАТАТАТАТАТ ТАТТАТАТАТАТ З БАССДА-сс СТТАТАТС	TAAAGAA TTAAAGAA TTAAAGAA TTAAAGAA TTAATTATA TTATTATA TTATTATA	CECCAGET CAAGAGTAAT CTCGTAAT CTCGTAAC FATAATAAT 4 TCCAACGT TTAAATAF	AAGGTCCCAAM AAGGTCCCAAM AAGGTCAAT AAGGTCAAT AAGGTCAAT AAGGTCAAT 15 15 15 15 15 15 15 15 15 15 15 15 15	G 1011 A 7849 C 1109 T 7948 T 1181 T 8048 G 1239 A 8148 G 1337 A 8240
E.c. mt E.c. mt E.c. mt E.c. mt	ССССААТСТ ССССААТС ССССААТС СССАТСОСТТАЛСТССС ТССАТСОСТТАЛСТССС ТАТТССТТСАСТСАССА САСТСОССС СССССТСССС АЛТТТТЛСТТСАСТСССАТСС ССССАСТСССССС В' СТ-ТСССТСССССССССССС ТАТСАСАТСТСССАТССССС ТАТСАСАТСТСССАТССССАТССС ТАТСАСАТСТСССАСТССССАТССС ССССАСТССССССССАТССС ССССАСТСССССССАТССССАТССС ССССАСТССАССАТССССАТССС ССССАСТССАССАТСССАТС	13 13 13 13 13 13 13 13 13 13		GCGGGTGC TAGTGCGA ACAGCCAG ACCATCCA 13' AACCATTGC AACCATTGC AACCATTGC ACCATTGCA 1 CCGGTGCAA GCGCGGCAA	ТАЛЕСТРОССТ ТАЛЕСТРОСТ 14 сатотносст салатност салатност салатносст салатност салатнос салатнос салатност сал	Согдаласа Тадаласа Тадаласа Тадаласа Тадаласа 15 Сосодаласа Сосодала 2 Сосодала Сосодала Сосодала Сосодала Сосодала Сосодала Сосодаласа Сосода Сосодаласа Сосодаласа С	БОСАЛАСАЛ 14 - 14 - GCCATCATT GCCATCATT GCCATTATT АСССДТ		CCCCAGCT AGCGTAAT CCCGTAAC TATATTAA TATATTAA CCCAACGT TCCAACGT TCCAACGT TCCAACGT TCCAACGT TCCAACGT TCCAACGT TCCAACGT		G 1011 T 7849 C 11099 C 11099 T 7948 T 11818 T 11818 T 11818 C 12398 8 148 G 8240 C 13370 C 12398 C 12398 C 13370 C 12398 C 12398
E.c. mt E.c. mt E.c. mt E.c. mt	ССССАЛАТТ ССССАЛАТСКА СССАТССТАЛСТСКА ССАТССТАЛСТСКА САСТСССТСАЛССАЛАТТ САСТСССССТСССС АЛАТТТТАСТАЛАТАЛАЛАСА 81 СТ-ТСССТАССССАЛССС ССССАЛССАЛСССАЛССС ТАТСАСАРТССАССССАЛССС СССТСАЛССАЛСССАЛСССА СССТСАЛССАЛССССАЛСССА СССТСАЛССАЛССССАЛАССА СССТСАЛССАЛССАЛАССАЛССА СССТСАЛССАЛССАЛАССАЛССА СССТСАЛССАЛССАЛАССАЛССА СССТСАЛССАЛССАЛАССАЛССА СССТСАЛССАЛССАЛАССАЛССА СССТСАЛССАЛССАЛАССАЛССА СССТСАЛССАЛССАЛАССАЛССА СССТСАЛССАЛССАЛАССАЛССА СССТСАЛССАЛССАЛАССАЛССА СССТСАЛССАЛССАЛАССАЛССА СССТСАЛССАЛССАЛАССАЛССА СССТСАЛССАЛССАЛАССА СССТСАЛССАЛССАЛАССАЛССА СССТСАЛССАЛССАЛАССАЛССА СССТСАЛССАЛСАЛСАЛСАЛССА СССТСАЛССАЛСАЛСАЛСА СССТСАЛССАЛСАЛСАЛСАЛСА СССТСАЛССАЛСАЛСАЛСАЛСАЛСА СССТСАЛССАЛСАЛСАЛСАЛСАЛСА СССТСАЛССАЛСАЛСАЛСАЛСАЛСАЛСАЛСАЛСАЛСА СССТСАЛССАЛСАЛСАЛСАЛСАЛСАЛСАЛСАЛСАЛСАЛСАЛСАЛ	13 13 13 13 13 13 13 13 13 13	GACA-FACG MATCORACA ALGOGOCCAG ITATTAATAA ALGOGOCCAG ITATTAATAA ALGOGOCTAG ALGOGOTAG ACGATAATAA ALGOCATAATAA ALGOCATAATAA	GCGGGGGAACAATAT 13' AACAATCAAGA 13' AACAATCAAGA AACAATAT 33 AGGTGTGC 1 CGGGGGAAA GAAT CGATGGGAA	ТАЛЕСТРОССТ ТАЛССТРОСТ 14 САТОТИТОССТ САТОТИТОССТ САТОТИССТ СТТСТАВСТ СТТСТАВСТ 1 1 АСССОЛЛОСТА СТТСТАВСТ 1 1 АСССОЛЛОСТ 1 1 АСССОЛЛОСТ 1 1 АСССОЛЛОСТ 1 1 АСССОЛЛОСТ 1 1 АСССОЛЛОСТ 1 1 АСССОЛЛОСТ 1 1 АСССОЛЛОСТ 1 1 АСССОЛЛОСТ 1 1 АСССОЛЛОСТ 1 1 4 АСССОЛЛОСТ 1 4 АСССОЛПОСТ 1 4 АСССОЛТОСТ 1 АСССОЛТОСТОСТ 1 А	Сотанаса <u>IT</u> ладалода <u>IT</u> ладалода <u>IT</u> ладалода <u>IT</u> ладалода <u>IT</u> ладалода <u>I</u>	БОДАЛАСА БОДАЛАСА 14. GCCATCAT GCCATCAT GCCATCAT GCCATCAT ACCOL TATAAATAT 3 БАССАА-GC CTTAINATAT 3 БАССАА-GC CTTAINATAT GTA-ACTCC		CACGAACT ACCGTAAT FATATTAA FATATTAA FCCAACCT CACCCT CACCT CACCTT CACCCT CACCTT CACCCT CACCTT CACTTT CACCTT CACTTT CACCTT CACCTT CACTTT CACCTT CACCTT CACTTT CACCTTT CACTTT CACCTT CACTTT CACCTTT CACCTTT CACTTT CACTTT CACTTT		G 1011 7849 C 11099 C 10994 T 11818 10948 11818 18148 12398 13370 18240 3370 18240 3370 18240 3370 18240 3370 18240 3370 18240 1937 1948 1948 1948 1948 1948 1948 1948 1948
E.c. mt E.c. mt E.c. mt E.c. mt		тателевание 	GACA-EACG MATCHAR AACGACAC AAGGOCCAG TTTATTATTATT AACGGOGGCTA AACGGOGGCTA AACGGACAC MATATTATT CACACACAC ACCACACACACACACACACAC	CGGGGGGGA CGGGGGGGA CGATGGGA CGGGGGGGA CGGGGGGA CGGGGGGA CGGGGGGA CGGGGGGA CGGGGGGA CGATGGGA CGATGGGA CGATGGGA	ТАЛЕСТРСССТ ТАЛЕСТРСССТ 14 14 14 14 14 14 14 14 14 14 14 14 14	Согдаласа Трасалося	БОДАЛАСАЛ БОДАЛАСА БОДАЛАСА 14-1 GCC 4. ТОСКАТ GCC 4. ТОСКАТ GCC 4. ТОСКАТ ССС 4. ТОСКАТА ТАТТАТАТАТА ТАТТАТАТАТА З ССТТАЦА. ТОСКАТАТА ССТТАЦА. ТОСКАТАТАТА ССТТАЦА. ТОСКАТАТАТА ССТТАЦА. ТОСКАТАТАТАТА ССТТАЦА. ТОСКАТАТАТАТАТАТАТАТАТАТАТАТАТАТАТАТАТАТАТ		COCCAGOT AGCGTAAT TCGTAAC TCGTAAC TATATTAA TATATTAA TCCAACGT TTAATTA TCCAACGT TTAATTTAA TCCAACGTAA TCCAACGTAA	AAGGTCCCAAA AAGGTCCCAAA AAGGTCACTCCAA AAGGTCACTCCAA AAGGTCACTCCAA AAGGTCCCAA IS' IS' IS' IS' IS' IS' IS' IS' IS' IS'	G 1011 G 1109 G 1109 T 1181 T 1181 T 1181 G A 8148 G A 8148
E.c. st E.c. st E.c. st E.c. st E.c. st E.c. st	ССССАЛАЛТСЯ ТССАЛСОСТТАЛСТССС-ЛАТСЯ ТСАТСОСТТАЛСТССС-ЛАТСЯ ТАТТССТСАСССС-ЛАТСЯ САСТСОСССТСССС АЛАТТТТАСТАСАТСАСССС 8 СТ-ТСССТССССССТСССС АЛАТТТТАСТАСАТССССАЛССС- СПАТАТЕТСССАСССССАЛССС ТАТСАСАДОТССАЛССССТАЛССС АЛТАТСАЛСАЛССАСССССТАЛССС АЛТАТСАЛАЛСТАСАЛАССТАЛ	ТАТСАСССА 13 13 13 13 13 13 13 13 14 13 14 14 15 15 15 15 15 15 15 15 15 15		ACCACTOC ACCACCAC ACCATCAC ACCATCAC ACCATCAC AACCATCAC AACCATCAC AACCATCAC AACCATCAC AACCATCAC AACCATCAC AACCATCAC AACCATCAC AACCATCAC AACCATCAC COCOCTOCACTAA COCOCTOCACTAA	ТАЛЕСТРОССТ ТАЛЕСТРОСТ 14 СТОТИТОВСТ САЛАТАВСТРОСТ АСССАЛАССТ ТОТОА	Сотдаласа 11 11 11 12 13 13 13 13 13 15 13 15 15 15 15 15 15 15 15 15 15	БОДАЛАСАД БОДАЛАСАД				G 1011 A 7849 C 11099 T 17948 T 11811 T 1814 G 12399 G 12397 G 12377 G 123777 G 123777 G 123777 G 1237777 G 123777777777777777777777777777777777777
E.c. mt E.c. mt E.c. mt E.c. mt	ССССАЛАТТ ТССАТССТАЛСТСАСС-ЛА ТСАТССТАЛСТСАСС-ЛА ТАТТСТТАЛСТСАСС-ЛА САСТСОСССТСССС АЛАТТТТАСТАЛТАТАЛАСАА 81 СТ-ТСОСТАСССССТСССС АЛАТТТТАСТАЛАТАЛАСАА ТАТСАСАДОТССАЛСССА ССТАЛАТСТСАССССАЛСССА ССТСАЛСТСАЛСССАЛАССАЛССС ССТСАЛСТСАЛСССАЛАССАЛССС ССТСАЛСТСАЛСССАЛАССАЛССС ССТСАЛСТСАЛСССАЛАССАЛССС ССТСАЛСТСАЛСССАЛАССАЛССС ССТСАЛСТСАЛССССАЛАССАЛССА ССТСАЛСТСАЛСССАЛААССАЛССА ССТСАЛСТСАЛССССАЛАССАЛССА ССТСАЛСТСАЛССССАЛАССАЛАССАЛ ССТСАЛСТСАЛСССССТАЛССАЛАССАЛАССАЛАССАЛА	13 13 13 13 13 13 13 13 13 13	GACA-EACG AAGGCCCAG ITATTAAT- AAGGCCCAG ITATTAAT- AAGGCCAGCAA AAGGCCAGCAA AAGGCCACAG AAGCCAAAAGG ACCAAAAAAAA	CGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG	ТАЛССТРСССТ ТАЛССТРССТ 14 актогтосст далогтоссс далагладосс алагога актогтоссс аласта актогтоссс тгсталадосс аласта тгсталадосс аласта тгсталадосс аласта тгсталадосс аласта тгсталадосс аласта актогтоссс аласта актогтоссс аласта актогтоссс аласта актогтоссс аласта актогтоссс аласта актогтоссс аласта актогтоссс аласта актогтоссс аласта актогтоссс аласта актогтоссс аласта актогтоссс аласта актогтоссс аласта актогтоссс аласта актогтоссс аласта актогтоссс аласта актогтоссс аласта актогтоссс аласта актогтоссс аласта актостоссс аласта актогтоссс аласта актогтоссс аласта актогтоссс аласта актогтоссс аласта актогтоссс аласта актогтоссс аласта актосс аласта актогтоссс аласта актосс аласта актосс аласта актосс аласто актосс аласто актосс аласта актосс аласто актоссс аласто актоссс аласто актоссса актоссс актос актоссса актоссс актос актоссс актоссс актоссс актосс актос актоссс актосс актоссс актоссс актос актоссс актоссс актоссс актоссс актоссс актоссс актоссс актоссс актоссс актоссс актоссс актоссс актоссс актоссс актосссс актоссс актоссс актосссс актоссс актоссс актоссс актосссс актоссс актосссс актоссс актосссс актос актоссс актоссс актоссс актоссс актоссс актоссс актоссс актоссс актоссс актоссс актоссс актоссс актоссс актоссс актоссс актосс актосс актоссс актоссс актосс актосс актосс актоссс актосс актоссс актосс актосс актосс актосс актосс актоссс актосс актоссс актосс актосс актосс актосс актосс актосс актосс актосс актосс актосс актосс актосс актосс актосс актосс актосс актос актос актосс актос актосс актос актос актосс актос актос актосс актос актос актосс актос актос актосс актос актос актосс актос актосс актос актос актосс актос актос актосс актос актос актосс актос актос актосс актос ак		БОДАЛАСАЛ БОДАЛАСАЛ 14-1 GCCATCAT GCCATCAT GCCATCAT GCCATCAT TATTATATAT ТАТТАТАТАТ ТАТТАТААТАТ З БАССЛА-С4 СТТАТАТСТ СТТАТАТСТ СТТАТС СТТАТСТ СТТАТ		CAAGAGTAAT AGCGTAAT TCGTAAC TCGTAAC TATATTAA ATAATAATAA TCCAACGT TTAATTAA CGAACGT CGAACGT CGAACGT CGAACGTAA		G 1011 A 7849 C 11099 T 7948 T 1181 T 1181 G 12394 G 12399 G 12399 G 13370 C 13350 C 13500 C 1
E.c. st E.c. st E.c. st E.c. st	ССССАААТТО ТССАТСССААССААТТО ТССАТСССТААССААТО ТСАТСССТААССААТО ТАТТССТРАСТССААССААТО ТАТТСОТСАССССС-ААССААТО САСТССССССТСССС ААТТТТ АСТААТАТАА 81 СТ-ТОСТ АССССССААССССАА СТАТОСТАССААСССССААСССА СТАТССАСАТОСССААСССССААССС ААТТТТ АСТСАССССССТААСССААТОСТ ТАТССАСТССАСССССТААСССААТОСТ СОСТСААСТССАСССССТААСССААТОСТ СОСТСААСТССАСССССТААСССААТОСТ СОСТСААСТССАСССССТААСССААТОСТ СОСТСАСТССАСССССТААСССААТОСТ СОСТСАСТССАССССССТААСССААТОСТ СОСТСААСТССАСССССТААСССААТОСТ СОСТСАСТССАСССССТААСССААСССССТААСССААТОСТ СОСТСАСТССАСССССТААСССААТОСТ СОСТСАСТССАСССССТААСССААССССТААСССААССС	TATELEGGA -AGGTEGGA -AGGTEGGA -AGGAGGATGTA 	GACA-EACG MATCGACAC AAGGGCCCAG TTATTATAAT AACGGGGCCAG TTATTATAAT AACGGGACTAA AACGGATCAA AACAGAAAAG CONTATATATAAT ACGGATAATATAATAT ACGGATAATATAAT	GCGGGGTGC ACADCCAG ACADCCAG ACADCCAG ACADCCAG ACADCCAG ACADCCAG ACCATCG AACCATCG COTOCT-ACAD COTOCT-ACAD	ТАДЕСТССЕТ ТАДЕСТССЕТ 14 В СТОТОВССС САЛАТАВСТ САЛАТАВСТ САЛАТАВСТ СОСТАСТАЛ 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	Сотанлен Табалас Т	БОДАЛАСАД БОДАЛАСАД БОДАЛАСАД 14.9 БОДАЛАСАД БОДАЛАСАД САТОАЛСАТТА ТАТТАТАТАТАТА ТАТТАТАТАТАТА З БАССАЛ-САД-СОД СТТАЦАЛАСАД СТТАЦАД СТ		COCCAGOT ANGAOTH COCCGTAAT COCCGTAAT COCCGTAAC COCCGTAC COCCGTACCGT COCCGTACCGTAC COCCGTAC COCCGTACCGTAC COCCGTACCGTAC COCCGTACCGTAC COCCG		G 10111 A 7849 C 11099 T 1181181 T 1181818848 G 12399 A 1337 G 12399 G 12399 G 12399 G 12399 G 1337 G 13377 G 1337 G 13377 G 133777 G 133777 G 13377777777777777777777777777777777777
E.c. mt E.c. mt E.c. mt E.c. mt	ССССААТСКАТС ССССААТСКАТСКА СССССААТСКАТСКАТСКА САСТСОСТСААТСКАТСКАТС САСТСОССССТСССС АЛАТТТТАСТАСАТСКАТСКАТС В' СТ-ТСССАТСССССААСССА ССССААТСКАТСКАТСКАТСКАТСКАТСКАТСКАТСКАТС			ACCACTOC ACCACCAG ACCATCAG ACCATCAG ACCATCAG ACCATCAG AA		Сотанса, <u>IT</u> асалоса, <u>IT</u> асалоса, <u>IT</u> асалоса, <u>IT</u> асалоса, <u>IT</u> асалоса, <u>IT</u> асалоса, <u>IT</u> асалоса, <u>IT</u> асалоса, <u>IT</u> асалоса, <u>ITTACANOCA</u> , <u>ITTACANOCA</u> , <u>ITT</u>	БОДАЛАСАЛ БОДАЛАСАЛ 14. GCCATCATT GCCATCATT GCCATCATT GCCATCATT GCCATCATT GCCATCATT GCCATCAT TATAAATAT 3 БАССДА-CQ CTTAIATAT 3 БАССДА-CQ CTTAIATAT 4 СТТАІАТАТАТАТАТАТАТ СТТАІАТАТАТАТАТАТАТАТАТАТАТАТАТАТАТАТАТА				G 1011 A C 11099 T 1181 T 1181 T 18048 G 1239 G 1239 G 1239 G 1239 G 1239 G 1239 G 1239 G 1239 G 1337 C 1 531 G 1 531 G 1 531 G 1 531 G 1 531 G 1 531 C 1 531
E.c. st E.c. st E.c. st E.c. st		TATELEGGA -AGGTTELEGGA -AGGTTELEGGA -AGGATETTA 		COGOCIOCAL CONTOLICAL CONTOL							G 1011 A C 11099 C 11099 C 11099 T 1181 R 1181 R 1181 G 12399 G 12399 G 14355 G 14355 G 14355 G 14355 G 14355 G 15311 C 153111 C 15311 C 15311 C 15311
E.c. st E.c. E.c. st E.c. st E.c. st E.c. st E.c. st E.c. st E.c. st		ТАТСАСССА - АЛСТТСААТ 13 13 13 13 13 13 14 14 15 16 16 16 17 17 16 16 16 16 16 16 16 16 16 16		COGOTOCT	ТАЛЕСТСССТ ТАЛССТРСССТ 14 сталостроти 14 сталостроти алессаласст салатладог салатла салатла салатла салатла салатла сала		БОДАЛАСАЛ БОДАЛАСАЛ 14. БОДАЛАСАТ 14. СССАТСАТТ СССАТСАТТ СССАТСАТТ СССАТСАТТ СССАТСАТТ СССАТСАТСАТ ССССАТСАТТ ССССАТСАТТ ССССАТСАТСАТСАТСАТСАТСАТСАТСАТСАТСАТСАТС		COCCARCE ANGAOTH TOTAATATTAA TATATTAA TATATTAA TTAATTA TTAAATTA COCCARCE COCARCE TTAAATTA COCARCECE COCARCECE COCARCECE COCARCECE		G 1011 A 7849 C 11099 T 1181 T 1181 T 1181 G 12398 G 1337 G 13
E.c. st E.c. E.c. t E.c. t E.c. t E.c. t E.c. t E.c. t E.c. t E.c. t E.c. t E.c.	ССССАЛАСАНТСЯ ТССАТСОСТТАЛСТСОСС-ЛАТ ТСАТСОСТТАЛСТСОСС-ЛАТ ТТАТССТТАЛСТСОСС-ЛАТ ТТАТССТТАЛСТСОСС-ЛАТ САСТСОСССТСССС АЛАТТТТАСТАЛАТАЛАЛАСА 81 СТ-ТСОСТАСССССТСССС АЛАТТТТАСТАЛСТСОСАЛССС	13 13 13 140047070000 140047070000 140047070000 14004707000 10004707000 10004707000 1000470000 1000470000 1000470000 1000470000 1000470000 10004700 10004700 100040000000000000000000000000000000	GACA-FACE MATCORACA AAGGCCCAG TTATTATAT- AAGGCCCAG TTATTATAT- ACCCGGGCCAA AATATTATTATA ACCCATATATT CCCATATATAT	CGGCGGGAAATCA	ТАЛЕСТРОССТ ТАЛССТРОСТ 14 сталования салагания салагания салагания сосолованост сосоловано сосоловано сососост сосост сосоловано сососост со	Сотанаса <u>IT</u> ладаласа <u>IT</u> ладал	БОДАЛАСАД БОДАЛАСАД ВОДАЛАСАН 14.1 GCCATCATT GCCATCATT GCCATCATT GCCATCATT GCCATCATT GCCATCAT TATAAATAT 3 БАССДА-04 СТТАДАТАТАТАТАТАТАТАТАТАТАТАТАТАТАТАТАТА		COLORGE CARCE		G 1011 A 7849 C 11099 C 11099 T 11811 B 4848 G A 8140 G A 81400 G A 8140 G A 8140 G A 81400 G A 81400 G A 81400 G A 8140
E.c. st E.c. t E.c. st E.c. S		ТАТСАССССА 1. АССТОСАТА 1. АССТОСАТА СССАЛСТИТА СССАЛСТИТА СССАЛСТИТА СССАЛСТИТА СССАЛСТИТА СССАЛСТИТА СССАЛСССАТА СССАЛСССАТА СССАЛСССАТА СССАЛСССАТА СССАЛСССАТА СССАЛСССАТА СССАЛСССАТА СССАЛСССАТА СССАЛСССАТА СССАЛСССАТА СССАЛСССАТА СССАЛСССАТА СССАЛСССАТА СССАЛСССАТА СССАЛСССАТА СССАЛСССАТА СССАЛСССАТА СССАЛСССАТА СССАЛСС	GACA-EACG MATCHACA AAGGOCCAG TTTATTAAT AAGGOGCCAG TTTATTAAT AAGGOGCCAG AAGGOGCAG AAGGOGCAG AAGGOGCAG AAGGOGCAG AAGGOGCAG AAGGOGCAG AAGGOGCAG AAGGOGCAG AAGGOGCAG AAGGOGCAG AAGGOGCAG AAGGOGCAG TTTTAGGAAAA	CCCCCCC ACCACCACCAC ACACCACCAC ACATCAC ACATCAC ACCATCAC ACCATCAC ACCATCAC ACCATCAC ACCATCAC ACCATCAC ACCATCAC ACCATCAC CCCCCCCC	ТАЛЕСТРОССТ ТАЛЕСТРОССТ ТАЛЕСТРОСТИ 14 64ТОТТОССТ 644АТАОССТ 644АТАОССС 644АТАОССС 644АТАОССС 644АТАОССС 7570 7570 7570 7570 7570 7570 7570 7	Сотанаса 11 15 15 15 15 15 15 15 15 15			COCCAACT AGCCGTAAT TOTATTAA TATATTAA TOTAACTAA TOTAACTAA TOTAACTAA TOTAACTAA TOTAACTAA TOTAACTAA TOTAACCCC TAAACTAA TOTAACCCC TAAACCCC TAAACCCC TAAACCCC		G 1011 A 7849 C 11099 C 17948 T 11811 R 11811 R 129 G A 1239 G A 1340 G A 13400 G A 13400 G A 1340 G A 13400 G A 13400 G A 13400 G A 1340



Figure 1. Alignment of the A. <u>nidulans</u> mitochondrial L-rRNA gene sequence (nucleotides 7097 to 9547 of exon I and nucleotides 11226 to 11641 of exon II) (7) with the <u>E. coli</u> 23S rRNA gene sequence (nucleotides 180 to 2807) (1). The boxed regions form base-pairs according to the secondary structure of Noller et al. (6). The six domains of this model are indicated in roman numbers. The thick arrow gives the position of the 1678 b intron (splice point), the thin arrow the position of a 194 b "mini-insert" (14).

No significant homology could be found between the two flanking sequences of the mitochondrial gene (166 bp downstream the proline tRNA gene and 120 bp upstream the threenine tRNA gene) (7) and the two terminal regions of the \underline{E} .

<u>coli</u> sequences corresponding to the 5.8S and 4.5S rRNA sequences of eukaryotic L-rRNAs (6). It remains to be shown by RNA sequencing whether these two regions are deleted from the mature mt L-rRNA or are replaced by rapidly diverging sequences of reduced functional importance.

The position of the 1678 b intron sequence within domain V (splice point) could accurately be localized by the strong homology between the two flanking exon sequences and nucleotides 2430 to 2520 of <u>E. coli</u> 23S rRNA. A comparison with a corresponding sequence around the splice point of yeast (w^{+}) mitochondrial L-rRNA (14) indicates a strong conservation of the splice point in both fungal mitochondria.

The <u>A</u>. <u>midulans</u> mt L-rRNA sequence exhibits several inserts (21 to 53 nucleotides) relative to the bacterial molecule, most of them in domain II. A longer insert of 194 nucleotides is found at the same position in domain V



<u>Figure 2.</u> Secondary structure model of <u>A.</u> <u>nidulans</u> mt L-rRNA, based on the alignment of Fig. 1 and the <u>E. coli</u> 23S rRNA model of Noller et al. (6). The dotted parts of the molecule are found only in the bacterial species, the boxed numbers are nucleotides inserted in the mitochondrial species. Heavy lines indicate those parts of domains IV and V which are conserved in all sequenced L-rRNA species, including mammalian mt L-rRNAs (15-17). (thin arrow of Fig. 1) as the "mini-insert" (66 nucleotides) of mt L-rRNA in ω^+ strains of yeast (14). Possibly this insert represents another intron not detected by electron microscopy of RNA/DNA hybrids (10), since the corresponding yeast mini-insert is absent in ω^- strains, like the large intron (14).

The alignment of Fig. 1 demonstrates a high coincidence of primary and secondary structure homology. The <u>A</u>. <u>nidulans</u> mt L-rRNA molecule deduced from the gene sequence can be folded almost exactly like <u>E</u>. <u>coli</u> 23S rRNA, as shown in Fig. 2.

Only few of the bacterial hairpin structures are deleted or shortened in the mitochondrial species, and one of the bacterial hairpin stems (stem 3 of domain IV) is extended by a mitochondrial insert. The other inserts are of low potential secondary structure and appear to extend hairpin loops or side loops.

The parts of domains IV and V drawn in heavy lines are conserved in all mitochondrial L-rRNAs, and all splice points of L-rRNAs coded by nuclear and mt genes are found in the conserved regions (6).

Fig. 3 shows a compilation of domain V sequences (around the fungal mitochondrial splice point) from nine species, including six mitochondrial sequences. All sequences have the potential to form the stem structures 15 to 19 (see Figs. 1 and 2).

Fig. 4 shows a phylogenetic tree based on the sequence compilation of Fig. 3. The difference matrix indicates that all organelle sequences are more related to the bacterial than to the nuclear sequence. As previously reported for conserved S-rRNA gene regions (9) the eubacterial affinity decreases with the order: chloroplasts (84 % homology to the E. coli sequence), fungal mt

123456789	C T	TA CG GT TT		CA T	TA	-A -A -A	CCC CC	AAG	T GT. T T T T	CT GGGCCC	A A CCCG TA TA TA TA TA	GGGA	. TAA (T	TG GCTGA G A T CGCA CGCA CGC	C C C C	3 A CCG TTT -T TA TA TA	GT GA T 1 TT TT1	G	C GAQ1	AG C	C	T T T T A	GT GT	AAT AAT	CTT GTG C AT TT	TTT	ATT GGCAC C ACG ACG ACG	TT	ATGT		GCTC T A	T G G G G G G G G G	CT GCA T TT	ATCC A	AC C C C C C C C A C	123456789
123456789		CGATTTTT	A	A TO	CALCTCCTC		TOTTCCCCCC	GGI CCC ATG AAG ATI ATI		с 366	TG TAT TC T T T T C C C	AT GGC1 A A TT TT	GTT		CAC CATTI CA GA GA GA GA GA	•	FAG AGT F C A	GGI C T CCCC CCCC	AC	AT GCG/ T T T T T	AGCT T T T	A A AA	GTT		C C C C	GTC AT GA GA G-	GTG A CA -CA	AGAC	G T AGTTC - AT G G G	GGTC T T T	TT.	ACC FATC	A TGC T T A A T	A A CGI TAG T A T T TTA		GA GA AA LAT LAT	123456789

Figure 3. Alignment of nine L-rRNA sequences (3'-terminal part of domains V) from yeast cytosol (nucleotides 2784-2992, ref. 3), (2) <u>E. coli</u> (2418-2627, ref. 1), (3) maize chloroplast (2515-2722, ref. 2), (4) <u>A. nidulans mt</u> (9518-11405, ref. 7), (5) <u>S. cerevisiae</u> mt (-90-118, ref. 14), (6) <u>Paramecium primau-relia</u> mt (1888-2097, ref. 18), (7) human mt (2907-3113, ref. 15), (8) mouse mt (2472-2680, ref. 16) and rat mt (1269-1469, ref. 17). Nucleotides homologous to the <u>E. coli</u> sequence are omitted.



Figure 4. Difference matrix and evolutionary tree derived from L-rRNA gene sequences of Fig. 3. The tree was constructed as previously described (12,9).

(68-70 %) and animal mt (62-64 %). The ratio: bacterial affinity versus nuclear affinity of all organelle sequences is higher for the L-rRNA domain V than for the six conserved S-rRNA domains previously tested (9), indicating a higher functional constraint and a stronger preservation of the ancestral eubacterial L-rRNA domain V in mitochondria.

The compilation also includes a protozoan mt sequence (from <u>Paramecium</u> <u>primaurelia</u>) (18) having a surprisingly high affinity (82 %) to the bacterial sequence (but only 60 % homology to the nuclear one).

Many different tree topologies were tested, but only the topology of Fig. 4 produced increasing difference values for successively constructed nodal points (12). This topology is identical to that of the S-rRNA trees (9): the nuclear sequence diverges first from a common ancestor of organelle and eubacterial sequences. The two groups of fungal and mammalian mt sequences do not exhibit a common root. Instead, the fungal mt protosequence is more related to the bacterial/chloroplastic protosequence than to the mammalian mt protosequence. The protozoan mt sequence is more related to the fungal than to the mammalian mt protosequence.

In contrast to recent claims (19) the eubacterial ancestry of both fungal and animal mitochondria is quite evident from all available sequence data, and any alternative to the endosymbiont hypothesis will be increasingly difficult to reconcile with experimental data.

REFERENCES

- Brosius, J., Dull, T.J. and Noller, H.F. (1980) Proc. Natl. Acad. Sci. USA <u>77</u>, 201-204.
- 2. Edwards, K. and Kössel, H. (1981) Nucl. Acids Res. 9, 2853-2869.
- 3. Georgiev, O.I., Nikolaev, N., Hadjiolov, A.A., Skryabin, K.G., Zakharyev, V.M. and Bayev, A.A. (1981) Nucl. Acids Res. 9, 6953-6958.
- Veldman, G.M., Klootwijk, J., de Regt, V.C.H., Planta, R.J., Branlant, C., Krol, A. and Ebel, J.P. (1981) Nucl. Acids Res. <u>9</u>, 6935-6951.

- Branlant, C., Krol, A., Machatt, M.A., Pouyet, J. and Ebel, J.P. (1981) Nucl. Acids Res. 9, 4303-4323.
- Noller, H.F., Kop, J.A., Wheaton, V., Brosius, J., Gutell, R.R., Kopylov, A.M., Dohme, F., Herr, W., Stahl, D.A., Gupta, R. and Woese, C.R. (1981) Nucl. Acids Res. <u>9</u>, 6167-6185.
- Netzker, R., Köchel, H.G., Basak, N. and Küntzel, H. (1982) Nucl. Acids Res. <u>10</u>, 4783-4794
- 8. Köchel, H.G. and Küntzel, H. (1981) Nucl. Acids Res. 9, 5689-5695.
- 9. Küntzel, H. and Köchel, H.G. (1981) Nature 293, 751-755.
- Lazarus, C.M., Lünsdorf, H., Hahn, U., Stepien, P.P. and Küntzel, H. (1980) Molec. Gen. Genet. <u>177</u>, 389-397.
- 11. Bartnik, E., Biderman, A.W., Hahn, U., Küntzel, H. and Stepien, P.P. (1981) Molec. Gen. Genet. <u>182</u>, 332-335.
- 12. Küntzel, H., Heidrich, M. and Piechulla, B. (1981) Nucl. Acids Res. 9, 1451-1461.
- 13. Köchel, H.G., Lazarus, C.M., Basak, N. and Küntzel, H. (1981) Cell 23, 625-633.
- 14. Dujon, B. (1980) Cell 20, 185-197.
- 15. Eperon, I., Anderson, S. and Nierlich, D. (1980) Nature 286, 460-466.
- 16. Van Etten, R., Walberg, M. and Clayton, D. (1980) Cell 22, 157-170.
- 17. Saccone, C., Cantatore, P., Gadaleta, G., Gallerani, R., Lanave, C., Pepe, G. and Kroon, A.M. (1981) Nucl. Acids Res. 9, 4139-4148.
- 18. Seilhamer, J.J. and Cummings, D.J. (1981) Nucl. Acids Res. 9, 6391-6406.
- 19. Gray, M.W. and Doolittle, F. (1982) Microbiol. Rev. 46, 1-42.