Molecular Phylogenies Based on Ribosomal Protein L11, L1, L10, and L12 Sequences

Daiqing Liao, Patrick P. Dennis

Canadian Institute for Advanced Research, Program in Evolutionary Biology, and Department of Biochemistry and Molecular Biology, University of British Columbia, 2146 Health Sciences Mall, Vancouver, B.C., Canada V6T 1Z3

Received: 22 February 1993 / Revised: 7 June 1993

Summary. Available sequences that correspond to the E. coli ribosomal proteins L11, L1, L10, and L12 from eubacteria, archaebacteria, and eukaryotes have been aligned. The alignments were analyzed qualitatively for shared structural features and for conservation of deletions or insertions. The alignments were further subjected to quantitative phylogenetic analysis, and the amino acid identity between selected pairs of sequences was calculated. In general, eubacteria, archaebacteria, and eukaryotes each form coherent and well-resolved nonoverlapping phylogenetic domains. The degree of diversity of the four proteins between the three groups is not uniform. For L11, the eubacterial and archaebacterial proteins are very similar whereas the eukaryotic L11 is clearly less similar. In contrast, in the case of the L12 proteins and to a lesser extent the L10 proteins, the archaebacterial and eukaryotic proteins are similar whereas the eubacterial proteins are different. The eukaryotic L1 equivalent protein has yet to be identified. If the root of the universal tree is near or within the eubacterial domain, our ribosomal protein-based phylogenies indicate that archaebacteria are monophyletic. The eukaryotic lineage appears to originate either near or within the archaebacterial domain.

Key words: Molecular phylogeny — Universal tree — Ribosomal proteins — Evolution — Archaebacteria

Introduction

Ribosomes are subcellular particles that play a structural and functional role in the template-directed synthesis of protein. Ribosomes were already present in the common primordial ancestor, and their basic structural and functional features have been preserved in all its diverse descendants. As a result, the macromolecular components of the ribosome, especially the small-subunit ribosomal RNA, have been useful chronometers with which to measure evolutionary relationships among extant organisms.

Journal

of Molecular Evolution

© Springer-Verlag New York Inc. 1994

In the E. coli ribosome, a pentameric complex, consisting of four copies of protein L12 and a single copy of protein L10, binds cooperatively along with another protein, L11, to a region in the 23S rRNA between nucleotides 1030 and 1120 (Ryan et al. 1991; Egebjerg et al. 1990; Dijk et al. 1979). This interaction produces a distinct and easily recognizable stalk on the large ribosomal subunit. This structure is essential for the binding of the extrinsic factors EF-Tu and EF-G and participates in conformational rearrangements of the ribosome that are accompanied by the hydrolysis of GTP. (For reviews, see Liljas 1982; Shimmin et al. 1989). Quaternary complexes similar to the E. coli (L12)₄L10-L11-rRNA complex are structurally and functionally conserved in the ribosomes of archaebacteria and eukaryotes (Uchiumi et al. 1987; Beauclerk et al. 1985; Casiano et al. 1990; El-Baradi et al. 1987). A fourth protein, L1, binds to large subunit RNA between nucleotides 2100 and 2200 (Branlant et al. 1981). It functions to stabilize peptidyl tRNA binding to the ribosome P site and participates indirectly in the factor-

Correspondence to: P. Dennis

dependent GTP hydrolysis (Subramanian and Dabbs 1980; Lake and Strycharz 1981; Sander 1983).

In *E. coli*, the genes encoding L11, L1, L10, and L12 form a complex transcription unit that also contains the genes for the two large subunits of RNA polymerase. It was somewhat surprising to find that the clustering of the genes encoding these four ribosomal proteins was conserved not just in eubacteria but also in a range of distantly related archaebacterial species including *Halobacterium cutirubrum* (Shimmin and Dennis 1989), *Haloferax volcanii* (Shimmin and Dennis, unpublished results), *Haloarcula marismortui* (Arndt and Weigel 1990), and *Sulfolobus solfataricus* (Ramirez et al. 1989). In eukaryotes, these genes are not linked (Newton et al. 1990) and the L12 gene has undergone a very ancient duplication that possibly predates the earliest eukaryotic organism.

In this paper, we have aligned and analyzed available L11, L1, L10, and L12 gene and protein sequences from eubacterial, archaebacterial, and eukaryotic organisms. We observed that for each of the gene-protein analyses, there is strong coherence that supports grouping organisms into the three primary domains: Eubacteria, archaebacteria, and eukaryotes. That is, the gene or protein sequences of organisms from within any one of the three domains are more closely related to each other than they are to sequences from the other two domains. The patterns of divergence for the L11, L10, and L12 proteins between eubacteria, archaebacteria, and eukaryotes are surprisingly dissimilar considering their intimate physiological interactions on the ribosome.

Materials and Methods

The molecular sequences (nucleotide sequences and/or amino acid sequences) for ribosomal proteins L11, L1, L10, and L12 were obtained from sequence data banks (EMBL, GenBank, and Swiss-Prot data banks) associated with the GeneWorks package (IntelliGenetics, Inc., Mountain View, CA, USA). Sequences not available from the data banks were obtained from the literature. The abbreviations used as organism identifiers in sequence alignments and phylogenetic trees and the reference for each sequence are listed in Table 1.

Sequence Alignment. The amino acid sequences of ribosomal proteins L11, L1, L10, and L12 from eubacteria, archaebacteria, and eukaryotes were aligned using the alignment algorithm in the GeneWork package. The resulting alignments were visually inspected to minimize the alignment gaps and to maximize amino acid identities. In the cases of ribosomal proteins L10 and L12, the previous evolutionary models were consulted in order to preserve predicted structural features (Shimmin et al. 1989). Our L12 alignments center on the conserved arginine-tryptophan residue at position 88. When required for analysis, nucleotide sequence alignments colinear to the depicted amino acid sequence alignments were used. Consensus of sequence alignments was determined visually by a somewhat flexible majority rule, where conservative amino acid replacements at each alignment position were taken into consideration. For example, at position 279 in the five archaebacterial L10 proteins there are 2 Ds, 1 E, 1 K, and 1 T. Because of the chemical similarity between D and E, D was chosen as the consensus residue even though it does not represent the majority residue at this position.

Phylogenetic Reconstruction. Parsimony analysis of the aligned amino acid sequences using the heuristic and/or branch and bound tree search options and bootstrap analysis were carried out using PAUP (Swofford 1989). When the heuristic tree search option was used, random addition of sequences with 10 replications was used to generate the parsimony tree. For bootstrap analysis of the L12 alignments, random addition of sequences with one replication was used because of limitation in computing capacity. The tree bisection-reconnection (TBR) algorithm was used in the heuristic tree searches (Swofford 1989). The distance matrix method was also employed to construct distance matrix trees using DNADIST, FITCH, KITSCH, and NEIGH-BOR programs in the PHYLIP Package (Felsenstein 1991).

Results and Discussion

Alignment and Phylogeny of L11 Proteins

There are five eubacterial sequences and one chloroplast sequence, which is encoded by the nuclear genome, available for ribosomal protein L11. They align from end to end with only two gaps in the alignment at positions 2-5 and 53 (Fig. 1). The high degree of amino acid sequence identity among these five sequences clearly suggests that the chloroplast sequence is of eubacterial origin.

The three available archaebacterial L11 protein sequences can be easily accommodated to this alignment. The archaebacterial proteins retain 7 of the 8 proline residues that are conserved in the eubacterial alignment at positions 24, 26, 27, 30, 60, 79, and 98; an eighth proline at position 80 has been replaced only in the *S. solfataricus* sequence. The archaebacterial L11 proteins are further characterized by a shorter amino terminus and by a 25–32-amino-acid-long extension at the carboxy terminus when compared to the eubacterial L11 sequences.

The proteins designated "L15" from S. cerevisiae (Pucciarelli et al. 1990) and "L12" from R. rattus (Suzuki et al. 1990) are homologs. They align end-to-end without gaps and are identical at 115 of the 165 positions. Based upon (1) immunological cross-reactivity (Juan-Vidales et al. 1983), (2) a limited degree of amino acid sequence similarity, and (3) a common binding site within mouse 28S rRNA (El-Baradi et al. 1987), these eukaryotic proteins have been implicated as homologs of the L11 protein of E. coli. The eukaryotic L11 sequences can be accommodated in the alignment by the inclusion of only two internal gaps (positions 66 and 77). Of the seven positions where proline is conserved in the archaebacterial and eubacterial proteins, only two (positions 30 and 79) are retained in the eukaryotic proteins.

Table 1 (Organisms and their abbreviations from which the sequences of the ribosomal proteins L11, L1, L10, and L12 are available

Organism	Abbreviation	Protein ^a	Reference		
Eubacteria					
Bacillus stearothermophilus	Bst	L1	Kimura et al. 1985		
-		L12	Garland et al. 1987		
Bacillus subtilis	Bsu	L12	Itoh and Wittman-Liebold 1979		
Desulfovibrio vulgaris	Dvu	L12	Itoh and Otaka 1984		
Escherichia coli	Eco	L11, L1, L10, L12	Post et al. 1979		
Haloanerobium prevalens	Hpr	L12	Matheson et al. 1987		
Halophilic eubacterium NRCC 41227	Heu	L12	Falkenberg et al. 1985		
Micrococcus lysodeikticus	Mlv	L12	Itoh 1981		
Proteus vulgaris	Pvu	L11, L1	Sor and Nomura 1987		
Rhodnseudomonas spheroides	Rsp	L12	Itoh and Higo 1983		
Serratia marscescens	Sma	L11. L1	Sor and Nomura 1987		
Salmonella tynhimurium	Sty	L10 L12	Paton et al. 1990a.b		
Sninacea oleracea (chloroplast)	Sol(c)	L12	Bartsch et al. 1982		
Spinacea oreracea (enteroptasi)	501(0)	L11	Smooker et al. 1991		
Strantommag grisgus	Sor	L 12	Itoh 1982		
Streptomyces griseus	Syi	L12 I 11	Okamoto et al. 1992		
Surreptomyces virginiae	See		Sibold and Subramanian 1000		
Synechocysus sp. PCC 6805	Sec Terra		Lice and Donnis 1002		
I nermotoga maritima	Tina	LII, LI, LI0, LI2	Liao and Dennis 1992		
Lukaryotes	A ==	L 10H(al 10')	Among at al. 1070, 1082		
Artemia salina	Asa	L12II(eL12)	Allions et al. 1979, 1982		
	DĽ		Delete et al. 1001		
Dictyostelium discoideum	Ddi	L10(P0)	Prieto et al. 1991		
Drosophila melanogaster	Dme	L10(P0)	Kelley et al. 1989		
		L12II(rp21C), L12I(rpA1)	Wigboldus 1987; Qian et al. 1987		
Gallus gallus	Gga	L12II(P1)	Ferro and Reinach 1988		
Homo sapiens	Hsa	L10(P0), L12II(P1)	Rich and Steiz 1987		
		L12I(P2)			
Mus musculus	Mmu	L10(P0)	Krowczynska et al. 1989		
Rattus norvegicus	Rno	L10(P0)	Chan et al. 1989		
Rattus rattus	Rra	L12II(P1) L12I(P2)	Wool et al. 1990		
		L11(L12)	Suzuki et al. 1990		
Saccharomyces cerevisiae	Sce	L10(P0), L12IA, L12IB,	Newton et al. 1990		
		L12IIA, L12IIB	Mitsui and Tsurugi 1988; Remacha et al. 1988		
		L11(L15)	Pucciarelli et al. 1990		
Schizosaccharomyces pombe	Spo	L12I(A4), $L12IB(A2)$	Beltrame and Bianchi 1990		
Semiced Senier Conjects period	~F~	$L_{12II}(A_1) L_{12IIB}(A_3)$			
Trypanosoma cruzi	Ter	L12I(P2)	Schiiman et al. 1990		
Tetrahymena thermophila	Tth	L12I(L37)	Hansen et al. 1991		
Archaebacteria		21211(207)			
Halobacterium cutirubrum	Heu	L11, L1, L10, L12	Shimmin et al. 1989		
Halobacterium halobium	Hha	L11 L1 L10 L12	Itoh 1988		
Haloarcula marismortui	Hma	L11 L1 L10 L12	Arndt and Weigel 1990		
Haloferax volcanii	Hvo	L11 L1 L10 L12	Shimmin and Dennis		
			(unpublished data)		
Methanococcus vannielli	Mva	L1, L10, L12	Baier et al. 1990		
Sulfolobus acidocaldarius	Sac	L12	Matheson et al. 1988		
Sulfolobus solfataricus ^b	Sso	L11, L1, L10, L12	Ramirez et al. 1989		
		,,			

^a The protein designations used in this paper are based on the sequence similarity to the *E. coli* L11, L1, L10, and L12. The original nomenclature where appropriate is given in parentheses

^b Our recent unpublished data indicate that the organism used to clone these ribosomal protein genes was actually *S. acidocaldarius* and not *S. solfataricus*. Nonetheless, we have here retained the species designation of Ramirez et al. (1989)

The phylogenetic relationships between the eleven L11 protein sequences were analyzed using PAUP (Fig. 1B). The eubacteria were contained within a well-defined domain. The location and branching order of three species within this domain, *Streptomyces virginiae*, spinach chloroplast, and *T. maritima*, are not rigorously defined. The two eukaryotic L11 sequences form another well-defined branch that originates from the *S. solfataricus* lineage within the archaebacterial group. If the ancestral root of the tree is located near or within the eubacterial domain (below the position of the arrow in Fig. 1B), then the archaebacteria would appear to be mono-

700

Α											
	10		20	30	40	50	60	70	80	90	100
SceL11	MPPKFDPNEVI	KYLYLRA	VGGEVGAS	AALAPKIG	PLGLSPKKVGE	DIAKATKE-	FKGIKVYVQLK	I-QNRO-A	AASV-VPSASSI	VITALKEPH	RDRKKDK
RraL11	MPPKFDPNEI	KVVYLRC	TGGEVGAI	SALAPKIG	PLGLSPKKVGE	DIAKATGD-	WKGLRITVKLI	I-QNRQ-A	QIEV-VPSASAI	IIKALKEPH	RDRKKQK
SsoL11	MPTKT	IKIMV	EGGSAKPG	PPLGPTLS	LGLNVQEVVK	KINDVTAQ-	FKGMSVPVTIE	IDSSTKKY	DIKVGVPTTTSI	LLKAINAQE	PSGDPAH
HcuL11	MAE-T	IEVLV	AGGQADPO	PPLGPELGE	TPVDVQAVVQ	EINDQTEA-	FDGTEVPVTIE	YEDDGS-F:	SIEVGVPPTAAI	VKDEAGFD	GSGEPQE
HmaL11	MAG-T	IEVLV	PGGEANPO	PPLGPELGI	PTPVDVQAVVÇ	EINDQTAA-	FDGTEVPVTVK	YDDDGS-F1	EIEVGVPPTAEL	IKDEAGFET	GSGEPQE
Sol (c) L11	KAKKVI	GVIKLAL	EAGKATPA	PPVGPALGS	KGVNIMAFCK	DYNARTAD-	KPGFVIPVEIT	VFDDKS-F	TFILKTPPASVI	LLKASGAE	GSKDPQM
EcoL11	MAKKVQA	AYVKLQV.	AAGMANPS	PPVGPALGÇ	QGVNIMEFCK	AFNAKTDSI	EKGLPIPVVII	VYADRS-F	TFVTKTPPAAVI	LKKAAGIKS	GSGKPNK
SmaL11	MAKKVQA	AYVKLQV	AAGMANPS	PPVGPALG	QGVNIMEFCK	AFNAKTDSI	EKGLPIPVVIT	VYSDRS-F	TFVTKTPPAAVI	LKKAAGIKS	SGSGKPNK
PvuL11	MAKKVQ2	AYIKLQV	SAGMANPS	PPVGPALG	QGVNIMEFCK	AFNAKTESV	EKGLPIPVVII	VIADRS-F	TFVTKTPPAAVI	LKKAAGVKS	SGSGKPNK
SviL11	MPPK-KKKVTC	GLIKLQI	KAGAANPA	PPVGPALG	QHGVNIMEFCK	AYNAATES-	QRGMVVPVEIT	VYDDRS-F	TFITKTPPAARI	ILKHAGIER	GSGEPHK
TmaL11	MAKKVA2	AQIKLQL	PAGKATPA	PPVGPALG	HGVNIMEFCK	RFNAETAD-	KAGMILPVVII	VYEDKS-F	TFIIKTPPASFI	LKKAAGIEF	GSSEPKR
	110		120	130	140	150	160	170	180		
SceL11	NVKHSGNIOLI	DEIIEIA	ROMRDKSF	GRTLASVT	EILGTAOSVG	CRVDFKNPH	DIIEGINAGEI	EIPEN			
RraL11	NIKHNGNITFI	DEIVNIA	RQMRHRSI	ARELSGTI	EILGTAQSVG	CNVDGRHPH	DIIDDINSGAV	ECPAS			
SsoL11	KIGNLDLI	EQIADIA	IKKKPQLS	AKTLTAAIH	SLLGTARSIG	ITVEGKDPK	DVIKEIDQGKY	NDLLTNYE	QKWNE-AEG		
HcuL11	FVADLSI	EQLKTIA	EQKKPDLI	AYDARNAAF	EVAGTCASLO	VTIEGEDAR	TFNERVDDGDY	DDVLGD	ELAAA		
HmaL11	FVADLSVI	DQVKQIA	EQKHPDLI	SYDLTNAAF	EVVGTCTSLG	VTIEGENPR	EFKERIDAGEY	DDVFAA	E-AQA		
Sol(c)L11	EKVGKITII	DQLRGIA	TEKLPDLN	CTTIESAM	RIIAGTAANMG	IDIDPP	ILVKKKKEVIF	·			
EcoL11	DKVGKISR	AQLQEIA	QTKAADMI	GADIEAMTH	RSIEGTARSMG	LVVED					
SmaL11	DKVGKVTR	AQVREIA	ETKAADMI	GSDVEAMTH	RSIEGTARSMG	LVVED					
PvuL11	EKVGKITSA	AQVREIA	ETKAADLI	GADVEAMM	RSIAGTARSMO	LVVED					
SviL11	TKVAKLTA	AQVKEIA	ELKMPDLN	ANDIDAAVI	LIAGTARSMG	VTVEG					
TmaL11	KIVGKVTRJ	KQIEEIA	KTKMPDLN	ANSLEAAM	(IIEGTAKSMG	IEVVD					

В



Fig. 1. Alignment of the amino acid sequences of the ribosomal protein L11 family and phylogenetic tree based on this alignment. **A** The L11 proteins are from 5 eubacteria, 1 chloroplast, 3 archaebacteria, and 2 eukaryotes. The leader peptide required for import of the chloroplast L11 protein into the organelle is not included in the alignment. The *numbers* indicate the common alignment positions. Abbreviations are listed in Table 1. **B** A parsimony analysis of aligned sequences was carried out using PAUP with the branch and bound search option and the most parsimonious tree is illustrated. There are 130 informative sites; all of these are included in the parsimony analysis and yield a tree with 472 steps and a consistency index of 0.886. The *numbers* indicate percent confirmation of grouping of species to the right of the node by bootstrapping analysis with 2,000 replications. Only values greater than 50% are indicated. Below the position of the *arrow* indicates the portion of the tree that would contain the root if the root were located either within the eubacteria or between the eubacteria and archaebacteria.

phyletic but not holophyletic. However, bootstrap analysis indicates that the positioning of *S. solfataricus* relative to eukaryotes is tenuous. In the DNA parsimony tree (and in all other distance method trees), the archaebacteria are both monophyletic and holophyletic; the bootstrap confidence for this arrangement was 0.82.

Alignment and Phylogeny of L1 Protein Sequences

There are five eubacterial and six archaebacterial L1 equivalent protein sequences available (Fig. 2A). Although the proportion of conserved amino acid residues within the L1 family is relatively high, the alignment is interrupted by gaps at approximately 15 different positions. Many of these gaps, particularly the five gaps located beyond amino acid position 125, clearly differentiate the archaebacterial proteins from the eubacterial proteins. Deletion-insertion events are generally rare and their co-occurrence in multiple sequence alignments is a strong indication of common ancestry.

In E. coli, protein L1 binds to nucleotides 2100–2200 of the E. coli 23S rRNA (Branlant et al. 1981). The sequence and secondary structure of this binding domain within large-subunit rRNA of archaebacteria and eukaryotes are highly conserved and the E. coli protein can protect these sites in vitro from ribonuclease digestion (Zimmerman et al. 1980; Gourse et al. 1981). In E. coli protein L1 is also an autogenous regulator of translation of the mRNA containing the L11, L1, L10, and L12 cistrons. A region within the leader of the mRNA exhibits primary sequence and secondary structural similarity to the authentic L1 binding domain in 23S rRNA. Any deficiency in the production of rRNA results in L1 protein accumulation; the excess protein binds to the structural mimic on the mRNA and prevents translation of the L11 and L1 cistrons (Dean and Nomura 1980; Yates and Nomura 1981; Baughman and Nomura 1983; Thomas and Nomura 1987; Kearney and Nomura 1987). Similar mimics of the L1 rRNA binding site have been identified in the mRNAs of other eubacterial, as well as halophilic and methanogenic, archaebacterial species (Sor and Nomura 1987; Liao and Dennis 1992; Shimmin and Dennis 1989; Baier et al. 1990). Thus, both structural and regulatory features of the L1 family of proteins are conserved within eubacteria and at least some groups of archaebacteria. The eukaryotic homolog to protein L1 has not been identified.

The PAUP analysis of the L1 protein sequences produced two equally parsimonious trees that group eubacteria and archaebacteria in separate and well-resolved domains. The two trees differ only in their placement of *S. solfataricus;* in the first case it branches separately and somewhat closer to eubacteria (solid branch position in Fig. 2B; 53% bootstrap confirmation), and in the second case it branches with *M. vannielli* (dashed branch) and separately from the halophilic L1 sequences. Distance and DNA parsimony methods position *S. solfataricus* and *M. vannielli* together although the grouping is tenuous.

The Sequence Alignments and Phylogeny of L10 Proteins

Between eubacteria and archaebacteria, the L10 proteins are in general less conserved than are the L11 and L1 proteins. However, because of domain conservation within L10 proteins, a reasonable alignment can be achieved with little difficulty. By using L10 sequences from the archaebacterial species *H. cutirubrum* and *S. solfataricus* as "bridges," Shimmin et al. (1989) demonstrated that the eukaryotic "P0" proteins are actually homologs of the bacterial L10 proteins.

The sequence alignment of the L10 protein family from 4 eubacteria, 5 archaebacteria, and 6 eukaryotes is illustrated in Fig. 3A. Amino acid identity among all the L10 proteins is highest within the amino terminal 121 residues. The most conspicuous feature is the presence of several highly conserved basic residues at alignment positions 17 (lys), 51(arg), 68 (lys or arg), 74 (lys or arg), and 121 (lys). There are also many positions in this region which have a high incidence of hydrophobic residues. These features suggest that secondary structures in this domain may be highly similar if not identical and that this domain may be involved in rRNA binding (Gudkov et al. 1980; Pettersson 1979; Mitsui et al. 1989). It is difficult to align with certainty the carboxyl domain of the eubacterial L10 sequences beyond position 121 with the eukaryotic and archaebacterial sequences. Nonetheless, the sequence RNLVYVLNAI of T. maritima L10 near the carboxyl end is highly similar to the archaebacterial sequences around position 240 (e.g., RNL-SV-NAA in H. cutirubrum; Fig. 3C). This sequence was used as a starting point to achieve the depicted alignment between positions 173 and 248.

The archaebacterial and eukaryotic proteins exhibit a carboxy-terminal extension of approximately 80–100 residues that is clearly not present in the eubacterial protein. This extension is characterized in part by a cluster of charged amino acids (approximately position 320–359). In the eukaryotic proteins, this charged region is preceded by an alanine-proline-rich region that is either shortened in, or absent from, the archaebacterial proteins. It has been suggested that these features are a result of a partial duplication of the L12 gene that has been fused to the end of the L10 gene (Shimmin et al. 1989). Within any species of archaebacteria or eukaryote, substantial sequence identity is always apparent between the carboxy termini of the respective L10 and L12 proteins. For example, the identical sequences at the 410 .

A	1.0	~~			5.0	60			~ ~	
	10	20	30	40	50	60	70	80	90	100
HCuL1	MADNDIE-EAVAR	-ALEDAPQR		NFRETVDI	LAVNLRDLD	LNDPSQRVDEGV	VLPSGTGQE	TQIVVFADGE	TAV-RADDVAI	DVLDE
HhaL1	MADNDIE-EAVAR	-ALEDAPQR		NFRETVDI	LAVNLRDLD	LNDPSQRVDEGV	VLPSGTGQE	TQIVVFADGE	TAV-RADDVAI	DVLDE
HmaL1	MADQEIE-NAVSR	-ALEDAPER		NFRETVDI	LAVNLRDLD	LNDPSNRVDESV	VLPAGTGQE	TTIVVFAEGE	TAL-RAEEVAL	DDVLDE
HvoL1	MAD-TIV-DAVSR	-ALDEAPGR		NFRETVDI	LAVNLRDLD	LNDPSKRVDESI	VLPSGTGQI	TQIVVFATGE	TPAEDAAI	DEVLGP
MvaL1	MDSAQIQ-KAVKE	-ARTRKPR-NF	RQSVDLIV	NFTQSVDI	IVNLKELD	TRPENRLKEQI	VLPSGKGKI	TKIAVIAKGD	LAA-QAAEMGI	JTVIRQ
SsoL1	MKKVLAD-KESLI	EALKLALS	STEYNV-KR-	NFTQSVE	LILTFKGID	IKKGDLKLREIV	PLPKQPSKA	KRVLVVPSFE	QLEYAKKASP	IVVITR
BstL1	MPKVDKKYLEALK	-LVDRSKAYPI	AQAIEIVKKT	NVAKFDATVEV	AFRL-GVD	PKKADQQIRGAV	VLPHGTGKV	ARVLVFAKGE	KAK-EAEAAG	ADYVG-
EcoL1	MAKLTKRMRVI-R	EKVDATKQYDII	NEAIALLKEL	ATAKFVESVD	AVNL-GID	ARKSDQNVRGAT	VLPHGTGRS	VRVAVFTQGA	NAE-AAKAAG	AELVG-
SmaL1	MAKLTKRMRV1-R	DKVDATKQYDI	FEAIALLKEL	ATAKFVESVD\	AVNL-GID	ARKSDQNVRGAT	VLPHGTGRS	VRVAVFTQGA	NAE-AAKAAGA	ELVG-
PvuL1	MAKLTKRMRNI-RI	EKVEVTKOYEL	AEAVALLKEL	ATAKFVESVD	AVNL-GID	ARKSDONVRGAT	VLPHGTGRS	VRVAVFAOGA	NAE-AAKEAGA	ELVG-
TmaL1	MPKHSKRYLEA-R	KLVDRTKYYDLI	DEAIELVKKT	ATAKFDETIEI	HIQT-GID	RKPEQHIRGTI	VLPHGTGKE	VKVLVFAKGE	KAK-EALEAG	ADYVG-
	110	120	130	140	150	160	170	180	190	200
HcuL1	DDLSDLADDTDAAL	KDLADETDFFV	AEAPM	MODIVGALGO	LGPRGKMP	PLOPDDDVV	DTVNRMKNI	-VQIRSRDRR	TFHTRVGAED	ISAEDI
HhaL1	DDLSDLADETDAAL	KDLADETDFFV	AEAPM	MODIVGALGOV	LGPRGKMP	PLOPDDDVV	DTVNRMKNI	-VQIRSRDRR	TFHTRVGAED	ISAEDI
HmaL1	DELEELGGDDDAAI	KDLADDTDFF1/	AEKGL	MODIGRYLGT	LGPRGKMPI	EPLOPDDDVV	EVIERMKNT	-VOLRSGERR	TFHTRVGAED	ISAENI
HvoL1	DELEDFGDDTDAA	KDLADETDFFV/	AEAGL	MODIGRYLGT	LGPRGKMP	PLOPADDVV	ETVNRMKNI	-VOLRTRDRR	TFHTRVGEDD	TPDEI
MvaL1	EELEELGKNKKAA	KRIANEHGFFI/	AOADM	MPLVGKSLGP	LGPRGKMP	PLPGNA NLA	PLVARFKKT	-VAINTRDKS	LFOVYIGTEAN	SDEEI
SsoL1	EELOKLOGOKRPV	KKLAIONEWFL	INOES	MALAGRILGPA	LGPRGKFP	PLPNTADIS	EYINRFKRS	-VIVKTKDOP	OVOVFIGTED	KPEDL
		~	~					-		
BstL1	-D-TEYIN	KIQQGWFDI	DVVVATPDM	MGEVGK-LGR	IGPKGLMP	VPKTGTVTFDVA	KAVQEIKAG	KVEYRVDKAG	NIHVPIGKVSH	DMEKL
EcoL1	MEDLAD	QIKKGEMN	DVVIASPDA	MRVVGQ~LGQ\	/LGPRGLMPI	IPKVGTVTPNVA	EAVKNAKAG	QVRYRNDKNG	IIHTTIGKVDE	DADKL
SmaL1	MEDLAE	0IKKGEMNI	FDVVIASPDA	MRVVGO-LGOI	SGPRGLMPI	NPKVGTVTPNVA	EAVKNAKAG	QVRYRNDKNG	IIHTTIGKVD	DADKL
PvuL1	MDDLAA	KVKAGEMDI	DVVIASPDA	MRVVGQ-LGQI	LGPRGLMPI	VPKVGTVTPNVA	EAVKNAKAG	OVRYRNDKNG	IINTTICKVVS	TKHKL
TmaL1	AEDLVE	K-I-EKEGFLDE	FDVAIATPDM	MRIIGR-LGKI	LGPRGLMP	SPKSGTVTQEVA	EAVKEFKKG	RIEVRTDKTG	NIHIPVGKRSI	DNEKL
						-				
	210	220	230	240						
HcuL1	ASNIDVIMRRLHAD	NLEKGPLNVI	OSVYVKTTMG	PAVEVA						
HhaL1	ASNIDVIMRRLHA	NLEKGPLNVI	DSVYVKTTMG	PAVEVA						
HmaL1	ADNIDVILRRLHA	DLEKGPLNI	DTVYVKTTMG	PAMEVA						
HvoL1	ARTSNVIVRRLEA	TLEKGPLNII	OSVYVKTTMG	PSVEVPA						
MvaL1	AANAEAILNVVAK	KYEKGLYHVI	KSAFTKLTMG	AAAPISK						
SsoL1	AENAIAVLNAIEN	KA-KVETNL	RNIYVKTTMG	KAVKVKRA						
BstL1	KENFAAVYEAIIK	AKPAAAKGTYVI	KNVTI TSTMG	PGIKVDPTTV-	-AVAQ					
EcoL1	KENLEALLVALKK.	AKPTQAKGVYII	KKVSISTTMG	AGVAVDQAGLS	SASVN					
SmaL1	KENLEALLVALKK.	AKPSQAKGMYII	KKVSLSTTMG	AGVAIDQSGLS	SAAAN					
PvuL1	KENLEALLVALKK	AKPSAAKGVYII	KKVSLSTTMG	AGVAIDQASLS	SATV-					
TmaL1	KENIIAAIKQIMQ	MKPAGVKGQFII	KKVVLASTMG	PGIKLNLQSL-	-LK-E					
R										



Fig. 2. Alignment of the amino acid sequences of the ribosomal protein L1 family and the phylogenetic tree based on this alignment. A The L1 proteins from five eubacteria and six archaebacteria are aligned. The numbers indicate the common alignment positions. Abbreviations are listed in Table 1. B A parsimony analysis of the L1 sequences was carried out using PAUP with branch and bound tree search options. One of the two shortest trees found with 627 steps is depicted. The other tree differs only in the positioning of S. solfaraticus, which is indicated by a dashed line; the branch length of this alternative lineage is arbitrary. There are 176 informative sites; all of them are included in the parsimony analysis. The consistency index for the two shortest trees is 0.900. The numbers indicate percent confirmation of grouping of species to the right of the node by bootstrapping analysis with 2,000 replications. Only values greater than 50% are indicated. Below the position of the arrow indicates the portion of the tree that would contain the root if the root were located either within the eubacteria or between the eubacteria and archaebacteria.

	10	20	30	40	50	60	70	80	90	100
DdiL10	MSGAGS-K	RKKLFIEKAT	KLFTTYDKMI	VAEADFVGS	SQLQKIRKSI	RGIGAVL	MGKKTMIRK	VIRDLAD	SKPELDALNTY	LKQNTC
DmeL10	MVRENKAA	WKAQYFIKV\	ELFDEFPKCE	FIVGADNVGS	KQMQNIRTSI	RGLAVVL	MGKNTMMRB	AIRGHLE	NNPQLEKLLPH	IKGNVG
HsaL10	MPREDRAT	WKSNYFLKI	QLLDDYPKCH	TIVGADNVGS	KQMQQIRMSI	RGKAVVL	MGKNTMMRF	AIRGHLE	NNPALEKLEPH	RGNVG
MmuL10	MPREDRAT	WKSNYFLKII	OL LODY PKCE	TVGADNVGS	KOMOOT RMSI	RGKAVVL	MGKNTMNRF	AIRGHLE	NNPALEKLLPH	LINGINUG
SceL10	MGGIRE-K	KAEYFAKLI	REYLEEYKSLI	VVGVDNVSS	QOMHEVRKEI	RGRAVVL	MGKNTMVRE	AIRGFLS	DLPDFEKLLPF	VKGNVG
HcuL10	M-SAEEQRTTEEVI	PEWKRQEVAEL	DLLETYDSVO	GVVNVTGIPS	KQLQDMRRGI	HGQAAVR	MSRNTLLVE	ALEEAGD	GLDTLTEY	VEGEVG
HhaL10	M-SAEEQRTTEEVI	PEWKRQEVAEL	DLLETYDSVO	SVVNVIGI PS		HGQAALR	MSKNTLLVI VSRNTLLFI	ALEEAGD	GLDTLTEI	TTGOVG
Mvall0	MTDAKSEHKT	APWKIEEVDAL	ELLKSANVI	ALIDMMEVPA	VOLOEIRDKI	RDOMTLK	MSRNTLIKE	AVEEVAEE	GNPEFAKLVDY	LDKGAA
SsoL10	M-IGLAVTTTKKI	AKWKVDEVAEL	TEKLKTHKTI	IIANIEGFP	DKLHEIRKKI	RGKADIK	VTKNNLFNI	ALKNAGY	DTKLFESY	LTGPNA
Ecol.10	MALNLOD	KOAIVAEVS	SEVAKGALSAV	VVADSRGVT	DKMTELRKA	REAGVYMR	VVRNTLLR	AVEGT	PFECLKDA	FVGPTL
SecL10	MGRTREN	KATVISDV	ELFQDAQMT	VIIDYQGLTV	AEITDLRNRI	RPLGGTCK	IAKNTLVR	ALAGQ-E	AWSPMEEF	LTGTTA
StyL10	MALNLQD	KQAIVAEVS	SEVAKGALSA	VVADSRGVTV	/DKMTELRKAG	REAGVYMR	VVRNTLLRI	VVEGT	QFECLKDT	FVGPTL
TmaL10	M-LTRQQ	KELIVKEM	SEIFKKTSLII	LFADFLGFT	ADLTELRSRI	REKYGDGARFR	VVRNTLLR	AVENA	EYEGYEEF	LKGPTA
	110	120	130	140	150	160	170	180	190	200
DdiL10	IIFCKDNIAEVKR	VINTQRVGA	PAKAGVFAPNI	DVIIPAGPTO	SMEPTQ-TSFI	QDLKIATKINF	GQIDIVNE	HIIKTGQK	/GASEATLLQKL	NIKPFT
DmeL10	FVFTKGDLAEVRD	KLLESKVRA	PARPGAIAPL	HVIIPAQNTO	SLGPEK-TSFI	QALSIPTKISK	GTIEIIND	PILKPGDK	/GASEATLLNML	NISPFS
HsaL10	FVFTKEDLTEIRD	MLLANKVPA	ARAGAIAPC	EVTVPAQNT	SLGPEK-TSFI	QALGITTKISF	GTIEILSD	QLIKTGDK	GASEATLLNML	NISPFS
MmuL10	FVFTKEDLTEIRD	MLLAN KVPA	ARAGAIAPC	EVTV PAQNIX	ILGPEK-TSFI	OALGITIKISE MALGITTKISE	GTIEILSD	OUTKUCOK	VGASEATLINMI.	NISPES
SceL10	FVFTNEPLTEIKN	VIVSNRVAA	PARAGAVAPE	DIWVRAVNTO	SMEPGK-TSFI	QALGVPTKIAF	GTIEIVSD	KVVDAGNK	VGQSEASLLNLL	NISPFT
HcuL10	LVATNONPEGLYQ	QLENSKTPA	PINAGEVAPNI	DIVVPEGDT(STDPGPFVGE	OTIGANARIQE	GSTOVLDD	VVTEEGE'I'	VSDDVSNVLSEL VSDDVSNVLSEL	GIEPKE
HnaL10	LIGTODNPESLEO	QLENSKTPA	PIGAGEVAPN	DIVIPEGDIN	GVDPGPFVGEI	OSVGADARIOE	GSIQVIDD	STVLDTGEE	VSOELSNVLNEL	GIEPKE
MvaL10	IVVTEMNPFKLFK	TLEESKSPA	PIKGGAIAPC	DIEVKSGST	MPPGPFLSE	KAVGIPAAIDK	GKIGIKED	VVAKEGDV:	ISPKLAVVLSAL	GIKPVT
SsoL10	FIFTDTNPFELQL	FLSKFKLKR	YALPGDKADE	EVVVPAGDT	GIAAGPMLSVI	GKLKIKTKVQI	GKIHILQD	PTVAKPGDE:	IPADIVPILQKL	GIMPVY
Fcol.10	TAVSMEHP-GAAA	RLFKEFAK						ANAKFEVK	AAAFEGELIPAS	OIDRL-
SecL10	ILVLKEDL-GGAI	KAYKKFOK					D	KKTELR	GVLEGKSLTQA	DVEAI-
StyL10	IAYSMEHP-GAAA	RLFKEFAK						ANAKFEVK	AAAFEGELIPAS	QIDRL-
TmaL10	VLYVTEGDPVEAV	KIIYNFYK					D	KADLSRLK	GGFLEGKKFTAE	EVENI-
	210	220	230	240	250	260	270	280	290	300
DdiL10	YGLEPKIIYDAGA	CYSPSISE	EDLINKFKQG	IFNIAAI-S	L-EIGYPTVA	SIPHSVMNAFKN	ILLAISFET	SYTFD	AAEKFKS	AÁA-AA
DmeL10	YGLIVNQVYDSGS	IFSPEILDIKP.	EDLRAKFQQG	VANLAAV-C	L-SVGYPTIA:	SAPHSIANGFKN	ILLAIAATT	SVEFK	EATTIKE	YIK
			THE REAL PROPERTY AND ADDRESS OF THE PROPERTY ADDRESS					100000000000000000000000000000000000000	T 3 11773 777 3	
HSaL10	FGLVIQQVFDNGS	IYNPEVLDITE	ETLHSRFLEG	VRNVASV-C	L-QIGYPTVA	SVPHSIINGYKE	VLALSVET	OYTFP	LAEKVKA	FLA FLA
MmuL10 RnoL10	FGLVIQQVFDNGS FGLIIQQVFDNGS FGLIIQQVFDNGS	IYNPEVLDITE IYNPEVLDITE IYSPEVLDITE	ETLHSRFLEG QALHSRFLEG OALHTRFLEG	VRNVASV-C VRNVASV-C VRNVASV-C	L-QIGYPTVA: L-QIGYPTVA: L-OIGYPTVA:	SVPHSIINGYKF SVPHSIINGYKF SVPHSIINGYKF	VLALSVET VLALSVET	OYTFP SYTFP OYTFP	LAEKVKA LTEKVKA LAEKVKA	FLA FLA FLA
MmuL10 MmuL10 RnoL10 SceL10	FGLVIQQVFDNGS FGLIIQQVFDNGS FGLIIQQVFDNGS FGLTVVQVYDNGQ	IYNPEVLDITE IYNPEVLDITE IYSPEVLDITE VFPSSILDITD	ETLHSRFLEG QALHSRFLEG QALHTRFLEG EELVSHFVSA	VRNVASV-C VRNVASV-C VRNVASV-C VSTIASI-S	L-QIGYPTVA L-QIGYPTVA L-QIGYPTVA L-AIGYPTLP	SVPHSIINGYKE SVPHSIINGYKE SVPHSIINGYKE SVGHTLINNYKI	VLALSVET VLALSVET VLALSVET)LLAVAIAA	OYTFP SYTFP OYTFP SYHYP	LAEKVKA LTEKVKA LAEKVKA EIEDLVD	FLA FLA FLA RIE
Hsall0 MmuL10 RnoL10 SceL10	FGLVIQQVFDNGS FGLIIQQVFDNGS FGLIIQQVFDNGS FGLTVVQVYDNGQ	IYNPEVLDITE IYNPEVLDITE IYSPEVLDITE VFPSSILDITD	ETLHSRFLEG QALHSRFLEG QALHTRFLEG EELVSHFVSA	VRNVASV-C VRNVASV-C VRNVASV-C VSTIASI-S	L-QIGYPTVA: L-QIGYPTVA: L-QIGYPTVA: L-AIGYPTLP:	SVPHSIINGYKE SVPHSIINGYKE SVPHSIINGYKE SVGHTLINNYKI	XVLALSVET XVLALSVET XVLALSVET XVLALSVET XLAVAIAA	YTFP SYTFP YTFP SYHYP	LAEKVKA LTEKVKA EIEDLVC	FLA FLA FLA RIE
Hsalio Mmulio Rnolio Scelio Hculio Hbalio	FGLVIQQVFDNGS FGLIIQQVFDNGS FGLIIQQVFDNGS FGLTVVQVYDNGQ VGLDLRGVFSEGV VGLDLRGVFSEGV	IYNPEVLDITE IYNPEVLDITE IYSPEVLDITE VFPSSILDITD LFTPEELEIDV LFTPEELEIDV	ETLHSRFLEG QALHSRFLEG QALHTRFLEG EELVSHFVSA DEYRADIQSA DEYRADIQSA	VRNVASV-C: VRNVASV-C: VRNVASV-C: VSTIASI-S: AASARNL-S' AASARNL-S'	L-QIGYPTVA: L-QIGYPTVA: L-QIGYPTVA: L-AIGYPTLP: V-NAAYPTER' V-NAAYPTER'	SVPHSIINGYKF SVPHSIINGYKF SVPHSIINGYKF SVGHTLINNYKF FAPDLIAKGRGF FAPDLIAKGRGF	XVLALSVET XVLALSVET XVLALSVET DLLAVAIAA ZAKSLGLQA ZAKSLGLQA	DYTFP SYTFP SYHYP SVESPDLAD	LAEKVKA LTEKVKA EIEDLVC DLVSKADAQVRA DLVSKADAQVRA	FLA FLA FIE RIE LAAQID LAAOID
Hsalio Mmulio Rnolio Scelio Hculio Hhalio Hmalio	FGLVIQQVFDNGS FGLIIQQVFDNGS FGLIIQQVFDNGS FGLTVVQVYDNGQ VGLDLRGVFSEGV VGLDLRGVFSEGV VGLDLRAVFADGV	IYNPEVLDITE IYNPEVLDITE IYSPEVLDITE VFPSSILDITD LFTPEELEIDV LFTPEELEIDV LFTPEELEIDV	ETLHSRFLEG QALHSRFLEG QALHTRFLEG EELVSHFVSA DEYRADIQSA DEYRADIQSA DEYRSDIQAA	VRNVASV-C: VRNVASV-C: VRNVASV-C: VSTIASI-S: AASARNL-S' AASARNL-S' AGRAFNL-S'	L-QIGYPTVA: L-QIGYPTVA: L-QIGYPTVA: L-AIGYPTLP: V-NAAYPTER V-NAAYPTER V-NAAYPTER	SVPHSIINGYKF SVPHSIINGYKF SVGHTLINNYKI FAPDLIAKGRGF FAPDLIAKGRGF FAPTMLQSDRGN	XVLALSVET XVLALSVET XVLALSVET XVLALSVET XLAVAIAA EAKSLGLQA XAKSLGLQA XAKSLALQA	DYTFP SYTFP DYTFP SYHYP SVESPDLAD SVESPDLAD	LAEKVKA LTEKVKA EIEDLVC DLVSKADAQVRA DLVSKADAQVRA DLVSKADAQVRA	FLA FLA FIE LAAQID LAAQID LASQID
Hsalio MmuL10 RnoL10 SceL10 HcuL10 Hhal10 Hmal10 MvaL10	FGLVIQQVFDNGS FGLIIQQVFDNGS FGLIIQQVFDNGS FGLTVVQVYDNGQ VGLDLRGVFSEGV VGLDLRAVFADGV VGLNVLGVYEEGV	IYNPEVLDITE IYNPEVLDITE IYSPEVLDITE VFPSSILDITD LFTPEELEIDV LFTPEELEIDV LFEPEELEDI IYTSDVLRIDE	ETLHSRFLEG QALHSRFLEG QALHTRFLEG EELVSHFVSA DEYRADIQSA DEYRADIQSA DEYRSDIQAA EEFLGKLQKA	VRNVASV-C: VRNVASV-C: VRNVASV-C: VSTIASI-S: AASARNL-S' AASARNL-S' AGRAFNL-S' YTNAFNL-S'	L-QIGYPTVA: L-QIGYPTVA: L-QIGYPTVA: L-AIGYPTLP: V-NAAYPTER V-NAAYPTER V-NAAYPTER V-NADYPTAT	SVPHSIINGYKF SVPHSIINGYKF SVPHSIINGYKF SVGHTLINNYKI FAPDLIAKGRGE FAPDLIAKGRGE FAPTMLQSDRGE FIETIVQKAFNI	XVLALSVET XVLALSVET XVLALSVET DLLAVAIAA ZAKSLGLQA XAKSLGLQA XAKSLALQA DAKAVSVES	DYTFP SYTFP SYHYP SVESPDLAD SVESPDLAD AIEDPEVVP	LAEKVKA LTEKVKA EIEDLVC DLVSKADAQVRA DLVSKADAQVRA DLVSKADAQVRA AILGKAHAQMIA	FLA FLA RIE LAAQIE LAAQIE LASQIE VA-KLA
Hsalio Mmullo Rnollo Scello Hcullo Hhallo Hmallo Ssollo	FGLVIQQVFDNGS FGLIIQQVFDNGS FGLIIQQVFDNGS FGLTVVQVVDNGQ VGLDLRGVFSEGV VGLDLRGVFSEGV VGLDLRAVFADGV VGLNVLGVYEGGV VKLNIKIAYDNGV	IYNPEVLDITE IYNPEVLDITE IYSPEVLDITE VFPSSILDITD LFTPEELEIDV LFTPEELEIDV LFEPEELELDI IYTSDVLRIDE IIPGDKLSINL	ETLHSRFLEG QALHSRFLEG QALHTRFLEG EELVSHFVSA DEYRADIQSA DEYRADIQSA DEYRSDIQAA EEFLGKLQKA DDYTNEIRKA	VRNVASV-C: VRNVASV-C: VSTIASI-S: AASARNL-S' AGRAFNL-S' AGRAFNL-S' YTNAFNL-S' HINAFAV-A'	L-QIGYPTVA L-QIGYPTVA L-QIGYPTVA L-AIGYPTLP V-NAAYPTER V-NAAYPTER V-NADYPTAT V-NAVIPTSA T-EIAYPEPK	SVPHSIINGYKF SVPHSIINGYKF SVGHTLINNYKF FAPDLIAKGRGF FAPTMLQSDRGN FIETIVQKAFNI /LEFTATKAN	AVLALSVETI AVLALSVETI AVLALSVETI AVLAVAIAA: EAKSLGLQA: AKSLGLQA: AKSLALQA: DAKAVSVES: ARNALALAS:	YTFP SYTFP SYHYP SVESPDLADD SVESPDLADD AIEDPEVVPI AFITEKTAD SIGYITQET.	LAEKVKA LTEKVKA EIEDLVU DLVSKADAQVRA DLVSKADAQVRA DLVSKADAQVRA AILGKAHAQMIA AQAVFTKAVMKA	FLA FLA RIE LAAQID LAAQID LASQID VA-KLA YAAVAS
Hsalio Mmullo Rnollo Scello Hcullo Hhallo Hmallo Mvallo Ssollo Ecollo	FGLVIQQVFDNGS FGLIIQQVFDNGS FGLIIQQVFDNGS FGLTVVQVVDNGQ VGLDLRGVFSEGV VGLDLRGVFSEGV VGLDLRAVFADGV VGLNVLGVFEGV VKLNIKIAYDNGV ATLPTYEE-A	IYNPEVLDITE IYNPEVLDITE IYSPEVLDITE VFPSSILDITD LFTPEELEIDV LFTPEELEIDV LFEPEELEDI IYTSDVLRIDE IIPGDKLSINL IAR-LMATMKE	ETLHSRFLEG QALHSRFLEG QALHTRFLEG EELVSHFVSA DEYRADIQSA DEYRADIQSA DEYRSDIQAA EEFLGKLQKA DDYTNEIRKA ASAGKLVRTL	VRNVASV-C: VRNVASV-C: VRNVASV-C: VSTIASI-S: AASARNL-S' AGRAFNL-S' YINAFNL-S' HINAFAV-A' A-AVRD-A-	L-QIGYPTVA: L-QIGYPTVA: L-AIGYPTVA: L-AIGYPTLP: V-NAAYPTER' V-NAAYPTER' V-NAVIPTSA' T-EIAYPEPK'	SVPHSIINGYKF SVPHSIINGYKF SVGHTLINNYKI FAPDLIAKGRGF FAPDLIAKGRGF FAPTMLQSDRGA FIETIVQKAFNI /LEFTATKAM	WLALSVET WLALSVET WLALSVET DLLAVAIAA ZAKSLGLQA WAKSLGLQA WAKSLGLQA WAKSLGLQA WAKSLGLQA WAKAVSVES WRNALALAS	YTFP YTFP SYHYP SVESPDLAD SVESPDLAD SVESPDLAD AIEDPEVVP AFITEKTAD SIGYITQET.	LAEKVKA LTEKVKA EIEDLVC DLVSKADAQVRA DLVSKADAQVRA AILGKAHAQWIA AILGKAHAQWIA AQAVFTKAVMKA	FLA FLA RIE LAAQID LAAQID LASQID VA-KLA YAAVAS
HSALIO MmuL10 RnoL10 SceL10 HcuL10 HhaL10 HmaL10 MvaL10 SsoL10 SecL10	FGLVIQQVFDNGS FGLIIQQVFDNGS FGLIIQQVFDNGS FGLTVVQVYDNGQ VGLDLRGVFSEGV VGLDLRAVFADGV VGLDLRAVFADGV VGLNVLGVFEGV VKLNIKIAYDNGV ATLPTYEE-A GDLPSKEQ-L	IYNPEVLDITE IYNPEVLDITE IYSPEVLDITE VFPSSILDITD LFTPEELEIDV LFTPEELEIDV LFTPEELEIDI IYTSDVLRIDE IIPGDKLSINL IAR-LMATMKE MGQ-IAGGIN-	ETLHSRILEG QALHSRILEG ZALHTRFLEG EELVSHFVSA DEYRADIQSA DEYRADIQSA DEYRSDIQAA EEFLGKLQKA DDYTNEIRKA ASAGKLVRTL ALATKIAIGI	VRNVASV-C: VRNVASV-C: VRNVASV-C: VSTIASI-S: AASARNL-S' AASARNL-S' HINAFAV-A' A-AVRD-A- KEVPASVAR(L-QIGYPTVA: L-QIGYPTVA: L-QIGYPTVA: L-AIGYPTLP: V-NAAYPTER' V-NAAYPTER' V-NAVIPTSA' T-EIAYPEPK' GLQHV	SVPHSIINGYKF SVPHSIINGYKF SVGHTLINNYKI FAPDLIAKGRGF FAPDLIAKGRGF FAPTMLQSDRCH FIETIVQKAFNI /LEFTATKAM	WLALSVET WLALSVET WLALSVET WLALSVET DLLAVAIAA EAKSLGLQA MAKSLALQA MAKSLALQA MANALALAS	YTFP YTFP SYHYP SVESPDLAD SVESPDLAD SVESPDLAD SVESPDLAD SVESPDLAD SVESPDLAD SUGSPTATA SUGSTTATA SIGYITQET	LAEKVKA LTEKVKA ELEDLVC DLVSKADAQVRA DLVSKADAQVRA AILGKAHAQMIA AQAVFTKAVMKA	FLA FLA FLA RIE LAAQID LAAQID LASQID VA-KLA YAAVAS
HSAL10 MmuL10 RnoL10 Scel10 HcuL10 HhaL10 MwaL10 SsoL10 Sccl10 SecL10 StyL10 TmaL10	FGLVIQQVFDNGS FGLIIQQVFDNGS FGLIIQQVFDNGS FGLTVVQVYDNGQ VGLDLRGVFSEGV VGLDLRAVFADGV VGLNVLGVYEEGV VKLNIKIAYDNGV ATLPTYEE-A GDLPSKEQ-L ATLPTYEE-A AKLPSKEA-L	IYNPEVLDITE IYNPEVLDITE VFPSSILDITD VFPSSILDITD LFTPEELEIDV LFTPEELEIDV LFTPEELEIDI IYTSDVLRIDE IIPGDKLSINL IAR-LMATMKE MGQ-IAGGIN- IAR-LMATMKE YAM-LWGRVK-	ETLHSRILEG QALHSRILEG ZALHTRFILEG EELVSHFVSA DEYRADIQSA DEYRADIQSA DEYRSDIQAA EEFLGKLQKA DDYTNEIRKA ASAGKLVRTL ALATKIALGI ASAGKLVRTL	VRNVASV-C: VRNVASV-C: VSTIASI-S: AASARNL-S' AGRAFNL-S' YTNAFNL-S' YINAFNL-S' HINAFAV-A' A-AVRD-A- KEVPASVAR A-AVRD-A- KEVPASVAR A-AVRD-A-	L-QIGYPTVA: L-QIGYPTVA: L-QIGYPTVA: L-AIGYPTLP: V-NAAYPTER' V-NAAYPTER' V-NAVIPTSA' T-EIAYPEPK' GLQHV VLNAI	SVPHSIINGYKF SVPHSIINGYKF SVGHTLINNYKI FAPDLIAKGRGF FAPDLIAKGRGF FIETIVQKAFNI /LEFTATKAN	WLALSVET WLALSVET WLALSVET WLALSVET EAKSLGLQA EAKSLGLQA MAKSLALQA MAKAVSVES MRNALALAS	YTFP YTFP SYHYP SVESPDLAD SVESPDLAD AIEDPEVVPI AFITEKTAD SIGYITQET,	LAEKVKA LTEKVKA EIEDLVC DLVSKADAQVRA DLVSKADAQVRA AILGKAHAQWIA AILGKAHAQWIA AQAVFTKAVMKA	FLA FLA RIE LAAQID LAAQID LASQID VA-KLA YAAVAS
Hsallo MmuL10 RnoL10 Scel10 Hcul10 Hhal10 MvaL10 Ssol10 Scol10 Scol10 Scol10 Styl10 Tmal10	FGLVIQQVFDNGS FGLIIQQVFDNGS FGLIIQQVFDNGS FGLTVVQVYDNGQ VGLDLRGVFSEGV VGLDLRAVFADGV VGLNVLGVYEGQV VKLNIKIAYDNGV ATLPTYEE-A GDLPSKEQ-L ATLPTYEE-A AKLPSKEE-L	IYNPEVLDITE IYNPEVLDITE VFPSSILDITD LFTPEELEIDV LFTPEELEIDV LFTPEELEIDV IIYTSDVLRIDE IIPGDKLSINL IAR-LMATMKE MGQ-IAGGIN IAR-LMATMKE YAM-LVGRVK-	ETLHSRILEG QALHSRILEG ZALHTRILEG EELVSHFVSA DEYRADIQSA DEYRADIQSA DEYRSDIQAA EEFLGKLQKA DDYTNEIRKA ASAGKLVRTL ALATKIALGI ASAGKLVRTL APITGLVFAL	VRNVASV-C: VRNVASV-C: VSTIASI-S: AASARNL-S' AGRAFNL-S' AGRAFNL-S' HINAFAV-A' HINAFAV-A' A-AVRD-A- KEVPASVAR' A-AVRD-A- SGILRNLVY'	L-QIGYPTVA: L-QIGYPTVA: L-AIGYPTVA: L-AIGYPTLP: V-NAAYPTER' V-NAAYPTER' V-NAVIPTSA' T-EIAYPEPK' SLQHV VLNAI	SVPHSIINGYKF SVPHSIINGYKF SVGHTLINNYKI FAPDLIAKGRGF FAPDLIAKGRGF FIETIVQKAFNI /LEFTATKAN	AVLALSVETT AVLALSVETT AVLALSVETT AVLALSVETT EAKSLGLQA AKSLGLQA AKSLGLQA AKSLGLQA AKAVSVES ARNALALAS	YTFP SYTFP SYTFP SYHYP SVESPDLAD SVESPDLAD ALEDPEVVPI AFITEKTAD BIGYITQET	LAEKVKA LTEKVKA EIEDLVU DLVSKADAQVRA DLVSKADAQVRA AILGKAHAQMIA AQAVFTKAVMKA	FLA FLA RIE LAAQID LAAQID LASQID VA-KLA YAAVAS
Hsallo MmuLlo RnoLlO Scello Hcullo Hhallo Hmallo Mvallo Ssollo Ecollo Secllo Styllo Tmallo	FGLVIQQVFDNGS FGLIIQQVFDNGS FGLIIQQVFDNGS FGLTVVQVYDNGQ VGLDLRGVFSEGV VGLDLRAVFADGV VGLNVLGVYEEGV VKLNIKIAYDNGV ATLPTYEE-A GDLPSKEQ-L ATLPTYEE-A AKLPSKEE-L 310	IYNPEVLDITE IYNPEVLDITE VFPSSILDITD VFPSSILDITD LFTPEELEIDV LFTPEELEIDV IFTPEELEIDI IYTSDVLRIDE IIPGDKLSINL IAR-LMATMKE MGQ-IAGGIN- IAR-LMATMKE YAM-LVGRVK- 320	224 224 224 224 224 224 224 225 2224 2244 2224 2224 2224 2244 2224 2224 2224 2224 2224 2224 2224 2224 2224	VRNVASV-C: VRNVASV-C: VRNVASV-C: VSTIASI-S: AASARNL-S' AGRAFNL-S' YTNAFNL-S' HINAFAV-A' A-AVRD-A- KEVPASVAR A-AVRD-A- SGILRNLVY' 340	L-QIGYPTVA: L-QIGYPTVA: L-QIGYPTVA: L-AIGYPTLP: V-NAAYPTER' V-NAAYPTER' V-NAVIPTSA' T-EIAYPEPK' GLQHV VLNAI 350	SVPHSIINGYKE SVPHSIINGYKE SVPHSIINGYKE SVGHTLINNYKI FAPDLIAKGRGE FAPTMLQSDRGA FIETIVQKAFNI /LEFTATKAN	WLALSVET WLALSVET WLALSVET WLAVAIAA EAKSLGLQA AKSLGLQA NAKSLGLQA NAKSLALQA IRNALALAS	9YTFP SYTFP SYTFP SYHYP SVESPDLAD SVESPDLAD SVESPDLAD SVESPDLAD SVESPDLAD SU	LAEKVKA LAEKVKA LAEKVKA EIEDLVI DLVSKADAQVRA DLVSKADAQVRA AILGKAHAQMIA AQAVFTKAVMKA	FLA FLA RIE LAAQID LAAQID LASQID VA-KLA YAAVAS
Hsalio MmuLlo RnoLlo Scello Hcullo Hhalio Hmallo MvaLlo SsoLlo Secllo StyLio Tmallo DdiLlo	FGLVIQQVFDNGS FGLIQQVFDNGS FGLIQQVFDNGS FGLTVVQVYDNGQ VGLDLRGVFSEGV VGLDLRGVFSEGV VGLDLRAVFADGV VGLNVLGVYEGV VGLNVLGVYEGV VKLNIKIAYDNGV ATLPTYEE-A GDLPSKEQ-L ATLPTYEE-A AKLPSKEE-L 310 -PVRAAP DCCUPALA	IYNPEVLDITE IYNPEVLDITE VFPSSILDITD VFPSSILDITD LFTPEELEIDV LFTPEELEIDV LFTPEELEIDV IYTSDVLRIDE IIPGDKLSINL IAR-LMATMKE MGQ-IAGGIN IAR-LMATMKE YAM-LVGRVK 320 -SAAAPRAAA-	ETLHSRILEG QALHSRILEG QALHTRFLEG EELVSHFVSA DEYRADIQSA DEYRADIQSA DEYRSDIQAA EEFLGKLQKA DDYTNEIRKA ASAGKLVRTL ASAGKLVRTL ASAGKLVRTL 330 KKVVVE	VRNVASV-C: VRNVASV-C: VRNVASV-C: VSTIASI-S: AASARNL-S' AGRAFNL-S' AGRAFNL-S' HINAFAV-A' A-AVRD-A- KEVPASVARA A-AVRD-A- SGILRNLVY 340 EKK	L-QIGYPTVA: L-QIGYPTVA: L-QIGYPTVA: L-AIGYPTLP: V-NAAYPTER' V-NAVIPTSA' T-EIAYPEPK' SLQHV SLQHV 350 -EESD	SVPHSIINGYKE SVPHSIINGYKE SVPHSIINGYKE SVGHTLINNYKI FAPDLIAKGRGE FIETIVQKAFNI /LEFTATKAN 	VLALSVET VLALSVET VLALSVET VLAVAIAA SAKSLGLQA JAKSLALQA JAKAVSVES IRNALALAS	9YTFP SYTFP	LAEKVKA LAEKVKA EIEDLVU DLVSKADAQVRA DLVSKADAQVRA AILGKAHAQMIA AQAVFTKAVMKA	FLA FLA RIE LAAQID LAAQID LASQID VA-KLA YAAVAS
Hsall0 MmuL10 RnoL10 Scel10 Hcul10 Hhal10 Hmal10 MvaL10 SsoL10 Secl10 Secl10 Styl10 Tmal10 DdiL10 DdiL10 Dmel10 Hsal10	FGLVIQQVFDNGS FGLIQQVFDNGS FGLIQQVFDNGS FGLTVVQVYDNGQ VGLDLRGVFSEGV VGLDLRGVFSEGV VGLDLRAVFADGV VGLNVLGVYEEGV VKLNIKIAYDNGV ATLPTYEE-A GDLPSKEQ-L ATLPTYEE-A AKLPSKEQ-L 310 -PVRAAP DPSKFAAA DPSKFAAA	IYNPEVLDITE IYNPEVLDITE VFPSSILDITD LFTPEELEIDV LFTPEELEIDV LFTPEELEIDV IYTSDVLRIDE IIPGDKLSINL IAR-LMATMKE MGQ-IAGGIN IAR-LMATMKE YAM-LVGRVK- 320 -SAAAPAAGAA	ETLHSRILEG QALHSRILEG QALHTRFLEG EELVSHFVSA DEYRADIQSA DEYRADIQSA DEYRSDIQAA EEFLGKLQKA DDYTNEIRKA ASAGKLVRTL ASAGKLVRTL ASAGKLVRTL 330 KKVVVE TEKKEEA	VRNVASV-C: VRNVASV-C: VSTIASI-S: AASARNL-S' AASARNL-S' AGRAFNL-S' YINAFNL-S' HINAFAV-A' A-AVRD-A- KEVPASVARA A-AVRD-A- SGILRNLVY 340 EKK KKPESE:	L-QIGYPTWA: L-QIGYPTWA: L-AIGYPTWA: L-AIGYPTPR' V-NAAYPTER' V-NAAYPTER' V-NAVIPTSA' T-EIAYPEPK' GLQHV SLQHV SLQHV SLQHV SESD SEESD	SVPHSIINGYKF SVPHSIINGYKF SVGHTLINNYKI FAPDLIAKGRGF FAPTHLQSDRGA FIETIVQKAFNI /LEFTATKAN 360 DDMGMG- DDMGFG-	VULALSVETT VULALSVETT VULALSVETT VULAVAIAA: SAKSLGLQA: VAKSLGLQA: VAKAVSVESJ	9YTFP SYTFP SYTFP SVESPDLAD SVESPDLAD ALEDPEVVP AFITEKTAD SIGYITQET	LAEKVKA LTEKVKA EIEDLVU DLVSKADAQVRA DLVSKADAQVRA AILGKAHAQWIA AQAVFTKAVMKA	FLA FLA RIE LAAQID LAAQID LASQID VA-KLA YAAVAS
Hsalio MmuLio RnoLio Scelio Hculio Hhalio Hhalio Mvalio Ssolio Seclio Seclio Stylio DdiLio DdiLio Dmelio Hsalio Mmulio	FGLVIQQVFDNGS FGLIIQQVFDNGS FGLIIQQVFDNGS FGLTVVQVYDNGQ VGLDLRGVFSEGV VGLDLRGVFSEGV VGLDLRAVFADGV VGLNVLGVYEEGV VKLNIKIAYDNGV ATLPTYEE-A GDLPSKEQ-L ATLPTYEE-A AKLPSKE-L 310 -PVRAAP DPSKFAAAA DPSAFAAAAPAAA DPSAFAAAAPAAA	IYNPEVLDITE IYNPEVLDITE VFPSSILDITD LFTPEELEIDV LFTPEELEIDV IYTSDVLRIDE IIPGDKLSINL IAR-LMATMKE MGQ-IAGGIN IAR-LMATMKE YAM-LVGRVK- 320 -SAAAPAAAA ATTAAPAAAAA	ETLHSRILEG QALHSRIEG QALHTRFLEG EELVSHFVSA DEYRADIQSA DEYRADIQSA DEYRSDIQAA EEFLGKLQKA DDYTNEIRKA ASAGKLVRTL ASAGKLVRTL ASAGKLVRTL 330 KKVVVE TEKKEEA PA-KAEA	VRNVASV-C: VRNVASV-C: VSTIASI-S: AASARNL-S' AASARNL-S' YINAFNL-S' HINAFAU-S' HINAFAU-S' HINAFAU-S' SGILRNLVY' 340 EKKEE: KEE:	L-QIGYPTVA: L-QIGYPTVA: L-AIGYPTPA: V-NAAYPTER' V-NAAYPTER' V-NAVPTER' V-NAVIPTSA' T-EIAYPEPK' GLQHV GLQHV SLQHV SEESD SEESD SEESD	SVPHSIINGYKF SVPHSIINGYKF SVGHTLINNYKI FAPDLIAKGRGF FAPTLQSDRCH FIETIVQKAFNI /LEFTATKAN 360 DMGFG- EDMGFG- EDMGFG-	VLALSVET VLALSVET VLALSVET VLAVAIAA: SAKSLGLQA: ZAKSLGLQA: JAKSLALQA DAKAVSVES, MRNALALAS 	9YTFP SYTFP SYTFP SVESPDLAD NEDPEVVP AFITEKTAD SIGYITQET.	LAEKVKA LTEKVKA EIEDLVU DLVSKADAQVRA DLVSKADAQVRA AILGKAHAQWIA AQAVFTKAVMKA	FLA FLA RIE LAAQID LAAQID LASQID VA-KLA YAAVAS
Hsalio MmuLio RnoLio Scelio Hculio Hhalio Hhalio Mvalio Ssolio Seclio Seclio Stylio DdiLio Dmelio Hsalio Mmulio RnoLio	FGLVIQQVFDNGS FGLIIQQVFDNGS FGLIIQQVFDNGS FGLTVVQVYDNGQ VGLDLRGVFSEGV VGLDLRGVFSEGV VGLDLRAVFADGV VGLNVLGVYEGV VKLNIKIAYDNGV ATLPTYEE-A GDLPSKEQ-L ATLPTYEE-A AKLPSKEE-L 310 -PVRAAP DPSKFAAAA DPSAFVAAAPVAA DPSAFAAAAPLAA DPSAFAAAAPLAA	IYNPEVLDITE IYNPEVLDITE VFPSSILDITD LFTPEELEIDV LFTPEELEIDV LFEPEELEDI IYTSDVLRIDE IIPGDKLSINL IAR-LMATMKE YAM-LVGRVK- 320 -SAAAPRAAA- SASAAPAAGAA ATTAAPAAAAA	ETLHSRILEG QALHSRILEG QALHTRFLEG EELVSHFVSA DEYRADIQSA DEYRADIQSA DEYRSDIQAA EEFLGKLQKA DDYTNEIRKA ASAGKLVRTL ASAGKLVRTL ASAGKLVRTL 330 KKVVVE TEKKEEA PA-KAEA PA-KVEA	VRNVASV-C: VRNVASV-C: VSTIASI-S: AASARNL-S' AASARNL-S' YINAFNL-S' YINAFNL-S' HINAFAV-A' A-AVRD-A- KEVPASVARA A-AVRD-A- SGILRNLVY' 340 EKK EKEE KEE	L-QIGYPTVA: L-QIGYPTVA: L-AIGYPTUA: L-AIGYPTUA: V-NAAYPTER: V-NAAYPTER: V-NAVPTER: V-NAVIPTSA: T-EIAYPEPK: SLQHV	SVPHSIINGYKF SVPHSIINGYKF SVGHTLINNYKI FAPDLIAKGRGF FAPTULQSDRCH FIETIVQKAFNI /LEFTATKAN 360 DDMGFG- EDMGFG- EDMGFG-	VLALSVET VLALSVET VLALSVET VLAVAIAA: SAKSLGLQA: ZAKSLGLQA: JAKSLALQA JAKAVSVES, IRNALALAS 	YTFP SYTFP SYTFP SVESPDLAD SVESPDLAD AIEDPEVVP AFITEKTAD SIGYITQET	LAEKVKA LTEKVKA EIEDLVU DLVSKADAQVRA DLVSKADAQVRA AILGKAHAQWIA AQAVFTKAVMKA	FLA FLA RIE LAAQID LAAQID LASQID VA-KLA YAAVAS
Hsalio MmuLio RnoLio Scelio Hculio Hhalio Hhalio MvaLio Ssolio Seclio Stylio Tmalio DdiLio DdiLio Dmelio Hsalio RnoLio Scelio	FGLVIQQVFDNGS FGLIIQQVFDNGS FGLIIQQVFDNGS FGLTVVQVYDNGQ VGLDLRGVFSEGV VGLDLRGVFSEGV VGLDVLGVYEGV VKLNIKIAYDNGV ATLPTYEE-A GDLPSKEQ-L ATLPTYEE-A AKLPSKEQ-L ATLPTYEE-A AKLPSKEQ-L DSKFAAAA DPSKFAAAA DPSAFAAAAPAAA DPSAFAAAAPAAA DPSAFAAAAPLAA	IYNPEVLDITE IYNPEVLDITE VFPSSILDITD LFTPEELEIDV LFTPEELEIDV IYTSDVLRIDE IIPGDKLSINL IAR-LMATMKE MGQ-IAGGIN IAR-LMATMKE YAM-LVGRVK- 320 -SAAAPRAAA- SASAAPAAGAA ATTAAPAAAAA ATTAAPAAAAA	ETLHSRILEG QALHSRILEG QALHTRFLEG EELVSHFVSA DEYRADIQSA DEYRSDIQAA EEFLGKLQKA DDYTNEIRKA ASAGKLVRTL ASAGKLVRTL ASAGKLVRTL 330 KKVVVE TEKKEEA PA-KAEA PA-KVEA PA-KVEA ASGDAAPA	VRNVASV-C: VRNVASV-C: VSTIASI-S: AASARNL-S' AASARNL-S' AASARNL-S' YINAFNL-S' HINAFAV-A' A-AVRD-A- KEVPASVARA A-AVRD-A- SGILRNLVY 340 EKKEE: KEE: KEE: EEAAEE:	L-QIGYPTVA: L-QIGYPTVA: L-AIGYPTPA: V-NAAYPTER' V-NAAYPTER' V-NAVPTER' V-NAVIPTSA' T-EIAYPEPK' GLQHV GLQHV SEESD SEESD SEESD SEESD SEESD	SVPHSIINGYKF SVPHSIINGYKF SVGHTLINNYKI FAPDLIAKGRGF FAPTLQSDRCH FIETIVQKAFNI /LEFTATKAN 360 DDMGFG- EDMGFG- EDMGFG- EDMGFG-	VLALSVET VLALSVET VLALSVET VLAVAIAA: SAKSLGLQA: ZAKSLGLQA: JAKSLALQA DAKAVSVES MRNALALAS 	YTFP SYTFP YTFP SVESPDLAD SVESPDLAD AIEDPEVVP AFITEKTAD SIGYITQET	LAEKVKA LTEKVKA EIEDLVU DLVSKADAQVRA DLVSKADAQVRA AILGKAHAQWIA AQAVFTKAVMKA	FLA FLA RIE LAAQID LAQID LAQID LAQID VA-KLA YAAVAS
Hsalio MmuLio RnoLio Scelio Hculio Hhalio Hhalio MvaLio Ssolio Seclio Stylio Tmalio Ddillo Ddillo Ddillo RnoLio RnoLio Hsalio Heulio Haulio Hali	FGLVIQQVFDNGS FGLIIQQVFDNGS FGLIIQQVFDNGS FGLTVVQVYDNGQ VGLDLRGVFSEGV VGLDLRGVFSEGV VGLDLRAVFADGV VGLNVLGVYEGV VKLNIKIAYDNGV ATLPTYEE-A GDLPSKEQ-L ATLPTYEE-A GDLPSKEQ-L ATLPTYEE-A AKLPSKEQ-L 310 -PVRAAP DPSKFAAAA DPSAFAAAAPAAA DPSAFAAAAPAAA DPSAFAAAAPAAA DPSAFAAAAPAAA DPSAFAAAAPAAA	IYNPEVLDITE IYNPEVLDITE VFPSSILDITD LFTPEELEIDV LFTPEELEIDV IYTSOVLRIDE IIPGDKLSINL IAR-LMATMKE MGQ-IAGGIN- IAR-LMATMKE YAM-LVGRVK- 320 -SAAAPRAAAA ATTAAPAAAAA ATTAAPAAAAA ATTAAPAAAAA ATTAAPAAAAA	ETLHSRILEG QALHSRILEG QALHTRFLEG EELVSHFVSA DEYRADIQSA DEYRSDIQAA EEPLGKLQKA DDYTNEIRKA ASAGKLVRTL ASAGKLVRTL ASAGKLVRTL 330 KKVVVE TEKKEEA PA-KVEA PA-KVEA PA-KVEA TTADEOS-DE	VRNVASV-C: VRNVASV-C: VSTIASI-S: AASARNL-S' AASARNL-S' AASARNL-S' YINAFNL-S' HINAFAV-A' A-AVRD-A- KEVPASVARA A-AVRD-A- SGILRNLVY 340 EKKEE: KEE: KEE: EEAAAEE TOASE-ADD	L-QIGYPTVA: L-QIGYPTVA: L-AIGYPTVA: L-AIGYPTPR: V-NAAYPTER' V-NAAYPTER' V-NAVIPTSA' T-EIAYPEPK' GLQHV SLQHV SEESD SEESD SEESD SEESD SEESD SEESD	SVPHSIINGYKF SVPHSIINGYKF SVGHTLINNYKI FAPDLIAKGRGF FAPTLQSDRCH FIETIVQKAFNI /LEFTATKAN 360 DDMGFG- EDMGFG- EDMGFG- EDMGFG- DDMGFG- DDMGFG- DMGFG- 	VLALSVET VLALSVET VLALSVET VLAVAIAA: SAKSLGLQA: ZAKSLGLQA: JAKSLALQA DAKAVSVES MRNALALAS 	YTFP SYTFP YTFP SVESPDLAD SVESPDLAD AIEDPEVVP AFITEKTAD SIGYITQET	LAEKVKA LTEKVKA EIEDLVU DLVSKADAQVRA DLVSKADAQVRA AILGKAHAQMIA AQAVFTKAVMKA	FLA FLA RIE LAAQID LAQID LAQID LAQID VA-KLA YAAVAS
Hsalio MmuLio RnoLio Scelio Hculio Hhalio Hhalio MvaLio Ssolio Seclio Stylio DdiLio DdiLio DdiLio DdiLio RnoLio Scelio Hsalio Husiio Hculio	FGLVIQQVFDNGS FGLIIQQVFDNGS FGLIIQQVFDNGS FGLTVVQVYDNGQ VGLDLRGVFSEGV VGLDLRGVFSEGV VGLDLRGVFSEGV VKLNIKIAYDNGV ATLPTYEE-A GDLPSKEQ-L ATLPTYEE-A AKLPSKEQ-L ATLPTYEE-A AKLPSKEQ-L 310 -PVRAAP DPSKFAAAA-PAA DPSAFAAAAPAAA DPSAFAAAAPAAA DPSAFAAAAPAAA NPEKYAAA DEDALPEELQDVD DEDALPEELQDVD	IYNPEVLDITE IYNPEVLDITE VFPSSILDITD LFTPEELEIDV LFTPEELEIDV IYTSOVLRIDE IIYTSOVLRIDE IIPGDKLSINL IAR-LMATMKE MGQ-IAGGIN- IAR-LMATMKE YAM-LVGRVK- 320 -SAAAPRAAA ATTAAPAAAAA ATTAAPAAAAA ATTAAPAAAAA ATTAAPAAAAA ATTAAPAAAAA ATTAAPAAAAA ATTAAPAAAAA	ETLHSRILEG QALHSRILEG QALHSRILEG EELVSHFVSA DEYRADIQSA DEYRADIQSA DEYRSDIQAA EEFLGKLQKA DDYTNEIRKA ASAGKLVRTL ASAGKLVRTL ASAGKLVRTL 330 KKVVVE TEKKEEA PA-KAEA PA-KAEA PA-KAEA TTADEQS-DE TTADEQS-DE	VRNVASV-C: VRNVASV-C: VSTIASI-S: AASARNL-S' AASARNL-S' AASARNL-S' YINAFNL-S' HINAFAV-A' A-AVRD-A- KEVPASVARA A-AVRD-A- SGILRNLVY 340 EKKEE KEE KEE KEE EEAAAEE TQASE-ADD	L-QIGYPTVA: L-QIGYPTVA: L-AIGYPTVA: L-AIGYPTPR: V-NAAYPTER: V-NAVPTER' V-NAVPTER' V-NAVIPTSA' T-EIAYPEPK' GLQHV GLQHV SLQHV SEESD SEESD SEESD SEESD SEESD SEESD SEESD	SVPHSIINGYKF SVPHSIINGYKF SVPHSIINGYKF SVGHTLINNYKI FAPDLIAKGRGF FAPTULQSDRCH FIETIVQKAFNI /LEFTATKAN 360 DDMGFG- EDMGFG- EDMGFG- EDMGFG- DMGFG- DMGFG- DMGFG- DMGFG- 	VLALSVET VLALSVET VLALSVET VLASVET VLAVAIAA: SAKSLGLQA: ZAKSLGLQA: ZAKSLGLQA: ZAKSLGLQA: ZAKSLGLQA: ZAKSLGLQA: XXAKSLGLQA: XXAKSLGLQA: XXAKSLGLQA: XXAKSLGLQA: XXAKSLGA: XXXAKSLGA: XXXAKSLGA: XXXAKSLGA: XXXAKSLGA: XXXAKSLGA	YTFP SYTFP YTFP SVESPDLAD ALEDPEVVP AFITEKTAD SIGYITQET	LAEKVKA LTEKVKA EIEDLVU DLVSKADAQVRA DLVSKADAQVRA AILGKAHAQMIA AQAVFTKAVMKA	FLA FLA RIE LAAQID LAAQID LASQID VA-KLA YAAVAS
Hsalio MmuLio RnoLio Scelio Hculio Hhalio MvaLio Ssolio Ecolio Seclio Seclio Stylio DdiLio DmeLio Hsalio Mmulio RnoLio Scelio Hculio Hhalio Hhalio Halio	FGLVIQQVFDNGS FGLIQQVFDNGS FGLIIQQVFDNGS FGLIIQQVFDNGS FGLTVVQVYDNGQ VGLDLRGVFSEGV VGLDLRGVFSEGV VGLDLRGVFSEGV VKLNIKIAYDNGV ATLPTYEE-A GDLPSKEQ-L ATLPTYEE-A AKLPSKEQ-L ATLPTYEE-A AKLPSKEE-L 310 -PVRAAP DPSAFVAAAPLAA DPSAFAAAA DPSAFVAAAPVAA DPSAFAAAAPLAA NPEKYAAA DEDALPEELQDVD DEDALPEELQVD DEDALPEELQVD	IYNPEVLDITE IYNPEVLDITE VFPSSILDITD LFTPEELEIDV LFTPEELEIDV LFEPEELEDI IYTSDVLRIDE IIPGDKLSINL IAR-LMATMKE MGQ-IAGGIN- IAR-LMATMKE YAM-LVGRVK- 320 -SAAAPRAAA- SASAAPAAGGA ATTAAPAAAAA ATTAAPAAAAA ATTAAPAAAAA ATTAAPAAAAA ATTAAPAAAAA ATTAAPAAAAA ATTAAPAAAAA ATTAAPAAAAA ATTAAPAAAAA ATTAAPAAAAA ATTAAPAAAAA	ETLHSRILEG QALHSRILEG QALHSRILEG EELVSHFVSA DEYRADIQSA DEYRADIQSA DEYRADIQSA DEYRSDIQAA EEFLGKLQKA DDYTNEIRKA ASAGKLVRTL ASAGKLVRTL ASAGKLVRTL 330 KKVVVE TEKKEEA PA-KAEA PA-KAEA PA-KAEA TTADEQS-DE TTADEQS-DE TTADEQS-DE	VRNVASV-C: VRNVASV-C: VSTIASI-S: AASARNL-S' AASARNL-S' AASARNL-S' YINAFNL-S' HINAFAV-A' A-AVRD-A- KEVPASVARA A-AVRD-A- SGILRNLVY' 340 EKKEE KEE KEE EAAAEE TQASE-ADD ADADDAAEE	L-QIGYPTVA: L-QIGYPTVA: L-AIGYPTVA: L-AIGYPTPR: V-NAAYPTER: V-NAAYPTER: V-NAVPTER: J-DAYPTAT V-NAVIPTSA: T-EIAYPEPK: J-DAYPTAT V-NAVIPTSA: J-DAYPTAT V-NAVIPTSA: J-DAYPTAT SEESD SEESD SEESD SEESD SEESD SEESD SEESD SEESD SEESD	SVPHSIINGYKF SVPHSIINGYKF SVPHSIINGYKF SVGHTLINNYKI FAPDLIAKGRGF FAPDLIAKGRGF FIETIVQKAFNI /LEFTATKAN 360 DDMGFG- DDMGFG- DDMGFG- DDMGFG- DDMGFG- DDMGFG- DDMGFG- DDMGFG- DDMGFG- DDMGFG- 	VLALSVET VLALSVET VLALSVET VLASVET VLAVAIAA: SAKSLGLQA: EAKSLGLQA: MRNALALAS MRNALALAS CLFD- LFD- LFD- LFD- LFD- LFD- LFD- LFD-	YTFP SYTFP SYTFP SVESPDLAD ATEDPEVVP AFITEKTAD SIGYITQET	LAEKVKA LTEKVKA EIEDLVU DLVSKADAQVRA DLVSKADAQVRA AILGKAHAQMIA AQAVFTKAVMKA	FLA FLA RIE LAAQID LAAQID LASQID VA-KLA YAAVAS
Hsalio MmuLio RnoLio Scelio Hculio Hhalio Mvalio Ssolio Seciio Stylio Ddilio Ddilio Ddilio Mmulio Rsalio Hculio Hculio Hazio Mazio Mvalio	FGLVIQQVFDNGS FGLIIQQVFDNGS FGLIIQQVFDNGS FGLIIQQVFDNGS FGLTVVQVYDNGQ VGLDLRGVFSEGV VGLDLRGVFSEGV VGLDLRGVFSEGV VGLNVLGVYEEGV VKLNIKIAYDNGV ATLPTYEE-A AKLPSKEQ-L ATLPSKEQ-L ATLPSKEE-L 310 -PVRAAP DPSAFVAAAPVAA DPSAFAAAA DPSAFVAAAPVAA DPSAFAAAAPAAA NPEKYAAA DEDALPEELQDVD DEBALPEELQOVD DEBALPEELQGVE GDEALDDLKEQI	IYNPEVLDITE IYNPEVLDITE VFPSSILDITD VFPSSILDITD LFTPEELEIDV LFTPEELEIDV LFEPEELEDI IYTSDVLRIDE IIPGDKLSINL IAR-LMATMKE YAM-LVGRVK- 320 -SAAAPRAAA- SASAAPAAGGA ATTAAPAAAAA ATTAAPAAAAA ATTAAPAAAAA ATTAAPAAAAA ATTAAPAAAAA ATTAAPAAAAA ATTAAPAAAAAA	ETLHSRILEG QALHSRILEG QALHSRILEG DELVSHFVSA' DEYRADIQSA DEYRADIQSA DEYRADIQSA DEYRSDIQAA EEFLGKLQKA DDYTNEIRKA' ASAGKLVRTL ASAGKLVRTL ASAGKLVRTL 330 KKVVVE TEKKEEA PA-KVEA ASGDAAPA TTADEQS-DE TTADEQS-DE TTADEQS-DE TTADEQS-DE	VRNVASV-C: VRNVASV-C: VRNVASV-C: VSTIASI-S: AASARNL-S' AASARNL-S' AASARNL-S' YINAFNL-S' HINAFAV-A' A-AVRD-A- KEVPASVARA A-AVRD-A- SGILRNLVY' 340 EKKEE KEE KEE EAAAEE TQASE-ADD TQASE-ADD ADADDAAEE EKK	L-QIGYPTVA: L-QIGYPTVA: L-AIGYPTVA: L-AIGYPTPR: V-NAAYPTER: V-NAAYPTER: V-NAYPTER: V-NAVPTRT V-NAVPTRT V-NAVIPTSA: T-EIAYPEPK' 	SVPHSIINGYKF SVPHSIINGYKF SVGHTLINNYKI FAPDLIAKGRGF FAPDLIAKGRGF TAPTMLQSDRGN TIETIVQKAFNI /LEFTATKAN 360 DDMGFG- EDMGFG- EDMGFG- 	VLALSVET VLALSVET VLALSVET VLALSVET VLAVAIAA: CAKSIGLQA CAKSIGLQA AAKSIGLQA AAKAVSVES (RNALALAS) 	YTFP SYTFP SYHYP SVESPDLAD SVESPDLAD AFITEKTAD SIGYITQET.	LAEKVKA LTEKVKA EIEDLVC DLVSKADAQVRA DLVSKADAQVRA AILGKAHAQWIA AQAVFTKAVMKA	FLA FLA RIE LAAQID LAAQID LASQID VA-KLA YAAVAS
Hsalio MmuLio RnoLio Scelio Hculio Hhalio Mvalio Ssolio Seciio Ddilio Ddilio Ddilio Ddilio Mmulio Rsalio Hculio Haalio Hculio Malio Scelio Scelio Scelio Stylio	FGLVIQQVFDNGS FGLIIQQVFDNGS FGLIIQQVFDNGS FGLIIQQVFDNGS FGLTVVQVYDNGQ VGLDLRGVFSEGV VGLDLRGVFSEGV VGLDLRGVFSEGV VGLNVLGVYEGV VKLNIKIAYDNGV ATLPTYEE-A GDLPSKEQ-L ATLPTYEE-A AKLPSKEE-L 310 -PVRAAP DPSKFAAAA DPSAFVAAAFVAA DPSAFAAAAPLAA NPEKYAAA DEDALPEELQDVD DEEALPEELQOVD DEEALPEELQOVD DEEALPEELQGVE GDEALDDLKEQI ISGKVDLGVQI	IYNPEVLDITE IYNPEVLDITE VFPSSILDITD VFPSSILDITD LFTPEELEIDV LFTPEELEIDV IFPPELEDU IIYTSDVLRIDE IIPGDKLSINL IAR-LMATMKE YAM-LVGRVK- 320 -SAAAPRAAA- SASAAPAAGGA ATTAAPAAAAA ATTAAPAAAAA ATTAAPAAAAA ATTAAPAAAAA ATTAAPAAAAA ATTAAPAAAAA ATTAAPAAAAA ATTAAPAAAAA ATTAAPAAAAAAA	ETLHSRILEG QALHSRILEG QALHSRILEG EELVSHFVSA' DEYRADIQSA DEYRADIQSA DEYRADIQSA DEYRSDIQAA EEFLGKLQKA DDYTNEIRKA' ASAGKLVRTL ASAGKLVRTL ASAGKLVRTL 330 KKVVVE TEKKEEA PA-KVEA PA-KAEA PA-KVEA TTADEQS-DE TTADEQS-DE TTADEQS-DE TTADEQS-DE P-KAETKE EKKEEKKEE	VRNVASV-C: VRNVASV-C: VRNVASV-C: VSTIASI-S: AASARNL-S' AASARNL-S' AASARNL-S' YINAFNL-S' HINAFAV-A' A-AVRD-A- KEVPASVARA A-AVRD-A- SGILRNLVY' 340 EKKEE KEE KEE EAAAEE TQASE-ADD ADADDAAEE -EKK EKKGP	L-QIGYPTVA: L-QIGYPTVA: L-AIGYPTVA: L-AIGYPTPR: V-NAAYPTER: V-NAAYPTER: V-NAYPTER: V-NAVPTRT V-NAVPTRT V-NAVIPTSA T-EIAYPEPK' JUNAI SEESD	SVPHSIINGYKF SVPHSIINGYKF SVGHTLINNYKI FAPDLIAKGRGF FAPDLIAKGRGF TAPTMLQSDRGN TIETIVQKAFNI /LEFTATKAN 360 DDMGFG- DDMGFG- 	VLALSVET VLALSVET VLALSVET VLALSVET VLASVET AKSIGLQA ZAKSIGLQA ZAKSIGLQA JAKAVSVES (RNALALAS) VLFD- LFD- LFD- LFD- LFD- LFD- LFD- LFD-	YTFP SYTFP SYHYP SVESPDLAD SVESPDLAD AFITEKTAD SIGYITQET	LAEKVKA LTEKVKA EIEDLVI DLVSKADAQVRA DLVSKADAQVRA AILGKAHAQWIA AQAVFTKAVMKA	FLA FLA RIE LAAQID LAAQID LASQID VA-KLA YAAVAS
Hsalio MmuLio RnoLio Scelio Hculio Hhalio Mvalio Ssolio Ecolio Stylio Ddilio Dmelio Hsalio Mmulio Rociio Hculio Haalio Hculio Haalio Ecolio Scelio Scelio Stylio Ddilio Scelio Scelio Scelio Scelio Scelio Scelio Scelio Scelio Hculio Haalio Hculio Haalio Hculio Haalio Scelio Scelio Scelio Scelio Scelio Scelio Hculio Haalio Hculio Haalio Hculio Haalio Hculio Scelio Scelio Scelio Scelio Scelio Scelio Scelio Hculio Haalio Hculio Haalio Hculio Haalio Hculio Hculio Haalio Hculio Scelio Scelio Scelio Scelio Hculio Hc	FGLVIQQVFDNGS FGLIIQQVFDNGS FGLIIQQVFDNGS FGLTVVQVYDNGQ VGLDLRGVFSEGV VGLDLRGVFSEGV VGLDLRGVFSEGV VGLNVLGVYEGV VKLNIKIAYDNGV ATLPTYEE-A GDLPSKEQ-L ATLPTYEE-A AKLPSKEE-L 310 -PVRAAP DPSAFVAAAPVAA DPSAFAAAA DPSAFVAAAPVAA DPSAFAAAAPLAA NPEKYAAA DEDALPEELQDVD DEDALPEELQDVD DEDALPEELQVD DEDALPEELQVG GDEALDDDLKEQI ISGKVDLGVQI	IYNPEVLDITE IYNPEVLDITE VFPSSILDITD VFPSSILDITD LFTPEELEIDV LFTPEELEIDV LFEPEELEDI IYTSDVLRIDE IIPGDKLSINL IAR-LMATMKE YAM-LVGRVK- 320 -SAAAPRAAA- SASAAPAAGGA ATTAAPAAAAA ATTAAPAAAAA ATTAAPAAAAA ATTAAPAAAAA ATTAAPAAAAA ATTAAPAAAAA ATTAAPAAAAA ATTAAPAAAAA ATTAAPAAAAA	ETLHSRILEG QALHSRILEG QALHSRILEG EELVSHFVSA DEYRADIQSA DEYRADIQSA DEYRADIQSA DEYRSDIQAA EEFLGKLQKA DDYTNEIRKA ASAGKLVRTL ASAGKLVRTL ASAGKLVRTL ASAGKLVRTL 330 KKVVVE TEKKEEA PA-KAEA PA-KAEA PA-KAEA PA-KAEA PA-KAEA PA-KAEA PA-KAEA PA-KAEA PA-KAEA ETTADEQS-DE TTADEQS-DE TTADEQS-DE 	VRNVASV-C: VRNVASV-C: VRNVASV-C: VSTIASI-S: AASARNL-S' AASARNL-S' AASARNL-S' YINAFNL-S' HINAFAV-A' A-AVRD-A- KEVPASVARA A-AVRD-A- SGILRNLVY' 340 EKKEE KEE EAAAEE TQASE-ADD ADADDAAEE EKK EKKGP	L-QIGYPTVA: L-QIGYPTVA: L-AIGYPTUA: L-AIGYPTUA: V-NAAYPTER' V-NAAYPTER' V-NAVPTER' V-NAVPTER' I-EIAYPEPK' SLQHV SLQHV SLQHV SEESD	SVPHSIINGYKF SVPHSIINGYKF SVPHSIINGYKF SVGHTLINNYKI FAPDLIAKGRGF FAPDLIAKGRGF FIETIVQKAFNI /LEFTATKAN 360 DDMGFG- DDMGFG- DDMGFG- DDMGFG- DDGNAGAEGLGF DDCNAGAEGLGF 	VLALSVET VLALSVET VLALSVET VLASVET VLASVET VLASVES EAKSIGLQA EAKSIGLA EAKSIG EAKSIG EAKSIGLA EAKSIGLA EAKSIGLA EAKSIG EAKSIG EAKSIG EAKSIGLA EAKSIG EAKSIG EAKSIG EAKSIG EAKSIG EAKSIGLA EAKSIGLA EAKSIGLA EAKSIGLA EAKSIGLA EAKSIG EAKSIG EAKSIGLA EAKSIGLA EAKSIGLA EAKSIGLA EAKSIG EAKSIG EAKSIGLA EAKSIGLA EAKSIGLA EAKSIGLA EAKSIGLA EAKSIG EAKSIGLA EAKSIGLA EAKSIGLA EAKSIGLA EAKSIGLA EAKSIGLA EAKSIGLA	YTFP SYTFP SYTFP SVESPDLAD ATEDPEVVP AFITEKTAD SIGYITQET	LAEKVKA LTEKVKA EIEDLVU DLVSKADAQVRA DLVSKADAQVRA AILGKAHAQMIA AQAVFTKAVMKA	FLA FLA RIE LAAQID LAAQID LASQID VA-KLA YAAVAS
Hsalio MmuLio RnoLio Scelio Hculio Hhalio MvaLio Ssolio Ecolio Setuio DdiLio DmeLio Hsalio MmuLio RnoLio Scelio Hculio Hhalio Hculio Scelio Sc	FGLVIQQVFDNGS FGLIIQQVFDNGS FGLIIQQVFDNGS FGLTVVQVYDNGQ VGLDLRGVFSEGV VGLDLRGVFSEGV VGLDLRGVFSEGV VGLNVLGVYEGV VKLNIKIAYDNGV ATLPTYEE-A GDLPSKEQ-L ATLPSKEC-L ATLPSKEE-L ATLPSKEE-L 310 -PVRAAP DPSAFVAAAPVAA DPSAFAAAA DPSAFVAAAPVAA DPSAFAAAAPLAA NPEKYAAA DEDALPEELQDVD DEEALPEELQOVD DEEALPEELQGVE GDEALDDLKEQI ISGKVDLGVQI	IYNPEVLDITE IYNPEVLDITE VFPSSILDITD VFPSSILDITD LFTPEELEIDV LFTPEELEIDV LFEPEELEDI IYTSDVLRIDE IIPGDKLSINL IAR-LMATMKE YAM-LVGRVK- 320 -SAAAPRAAA- SASAAPAAGGA ATTAAPAAAAA ATTAAPAAAAA ATTAAPAAAAA ATTAAPAAAAA ATTAAPAAAAA ATTAAPAAAAA ATTAAPAAAAA ATTAAPAAAAA ATTAAPAAAAAA ATTAAPAAAAAA ATTAAPAAAAAA ATTAAPAAAAAA	ETLHSRILEG QALHSRILEG QALHSRILEG EELVSHFVSA' DEYRADIQSA DEYRADIQSA DEYRADIQSA DEYRSDIQAA EEFLGKLQKA DDYTNEIRKA' ASAGKLVRTL ASAGKLVRTL ASAGKLVRTL ASAGKLVRTL 330 KKVVVE TEKKEEA PA-KVEA PA-KVEA PA-KVEA TTADEQS-DE TTADEQS-DE TTADEQS-DE TTADEQS-DE CONDTASEDD P-KAETKE EKKEEKKEE	VRNVASV-C: VRNVASV-C: VRNVASV-C: VSTIASI-S: AASARNL-S' AASARNL-S' AASARNL-S' YINAFNL-S' HINAFAV-A' A-AVRD-A- KEVPASVARA A-AVRD-A- SGILRNLVY' 340 EKK	L-QIGYPTVA: L-QIGYPTVA: L-AIGYPTVA: L-AIGYPTPR: V-NAAYPTER: V-NAAYPTER: V-NAYPTER: V-NAVPTRT V-NAVPTRT V-NAVIPTSA T-EIAYPEPK' JUNAI SEESD DDSDDDDDDD	SVPHSIINGYKF SVPHSIINGYKF SVGHTLINNYKI FAPDLIAKGRGF FAPDLIAKGRGF TAPTMLQSDRGN TIETIVQKAFNI /LEFTATKAN 360 DDMGFG- 	VULALSVETT VULALSVETT VULALSVETT VULALSVETT VULAVAIAA: SAKSIGLQA: JAKAUSVES; MRNALALAS: 	YTFP SYTFP SYHYP SVESPDLAD SVESPDLAD AFITEKTAD SIGYITQET	LAEKVKA LTEKVKA EIEDLVI DLVSKADAQVRA DLVSKADAQVRA AILGKAHAQWIA AQAVFTKAVMKA	FLA FLA RIE LAAQID LAAQID LASQID VA-KLA YAAVAS
Hsalio MmuLio RnoLio Scelio Hculio Hhalio Hhalio Mvalio Ssolio Ecolio Seclio DdiLio Dmelio Hsalio Mmulio RnoLio Scelio Hculio Halio Halio Ecolio Scelio Scelio Scelio Scelio Scelio Scelio Scelio Hali	FGLVIQQVFDNGS FGLIIQQVFDNGS FGLIIQQVFDNGS FGLTVVQVYDNGQ VGLDLRGVFSEGV VGLDLRGVFSEGV VGLDLRGVFSEGV VGLNVLGVYEEGV VKLNIKIAYDNGV ATLPTYEE-A GDLPSKEQ-L ATLPSKEQ-L ATLPSKEQ-L ATLPSKEZ-L 310 -PVRAAP DPSAFVAAAPVAA DPSAFAAAA DPSAFVAAAPVAA DPSAFAAAAPAAAAPAAA DPSAFAAAAPAAAAPAAAAAAAAAAAAAAAAAAAAAAAAA	IYNPEVLDITE IYNPEVLDITE VFPSSILDITD VFPSSILDITD LFTPEELEIDV LFTPEELEIDV IFPPELEIDV IIYTSDVLRIDE IIPGDKLSINL IAR-LMATMKE MGQ-IAGGIN- IAR-LMATMKE YAM-LVGRVK- 320 -SAAAPRAAA- SASAAPAAGGA ATTAAPAAAAA ATTAAPAAAAA ATTAAPAAAAA ATTAAPAAAAA ATTAAPAAAAA ATTAAPAAAAA ATTAAPAAAAA ATTAAPAAAAA ATTAAPAAAAA ATTAAPAAAAAA ATTAAPAAAAAA ATTAAPAAAAAA ATTAAPAAAAAA	ETLHSRILEG QALHSRILEG QALHSRIFLEG EELVSHFVSA DEYRADIQSA DEYRADIQSA DEYRADIQSA DEYRSDIQAA EEFLGKLQKA DDYTNEIRKA ASAGKLVRTL ASAGKLVRTL ASAGKLVRTL 330 KKVVVE TEKKEEA PA-KAEA PA-KVEA PA-KVEA PA-KVEA PA-KVEA PA-KVEA PA-KAEA	VRNVASV-C: VRNVASV-C: VRNVASV-C: VSTIASI-S: AASARNL-S' AGRAFNL-S' HINAFAV-A' A-ASARNL-S' HINAFAV-A' A-AVRD-A- KEVPASVAR A-AVRD-A- SGILRNLVY' 340 EKK KPESE: KEE: KEE: KEE: EEAAAEE: TQASE-ADD. ADADDAAEE: EKKGP	L-QIGYPTVA: L-QIGYPTVA: L-QIGYPTVA: L-QIGYPTA: V-NAYPTER' V-NAYPTER' V-NAYPTER' V-NAYPTER' J- GLQHV	SVPHSIINGYKF SVPHSIINGYKF SVGHTLINNYKI FAPDLIAKGRGF FAPDLIAKGRGF TAPTMLQSDRGM TIETIVQKAFNI /LEFTATKAM 360 	VULALSVETT VULALSVETT VULALSVETT VULALSVETT VULAVAIAA: CAKSIGLQA: CAKSIGLQA: VAKSIGLQA:	YTFP SYTFP SYTFP SYHYP SVESPDLAD SVESPDLAD AIEDPEVVP AFITEKTAD BIGYITQET	LAEKVKA LTEKVKA EIEDLVU DLVSKADAQVRA DLVSKADAQVRA AILGKAHAQMIA AQAVFTKAVMKA	FLA FLA FLA RIE LAAQIE LAAQIE LAAQIE VA-KLA YAAVAS

Fig. 3. Alignment of amino acid sequences of the ribosomal protein L10 family and the phylogenetic trees based on the L10 alignment. A The L10 ribosomal proteins from 4 eubacteria, 5 archaebacteria, and 6 eukaryotes are aligned. In the eukaryotes, these proteins were previously designated "P0." Abbreviations are as in Table 1. **B** The consensus of the alignment was generated manually by majority rule. When majority is not evident at an alignment position, chemically similar amino acid residues were considered to determine the consensus. *Question marks* (?) indicate that there is no simple consensus at such positions. **C** Alignment of the L10 sequences of *T. maritima* and *H. cutirubrum* at positions 239–248. **D** A parsimony analysis of the aligned sequences was carried out using PAUP with the branch and bound tree search option and two of the six equally shortest trees are illustrated (*tree 1* and *tree 2*). Four other trees simply rearrange the three mammalian species (Hsa, Rno, and Mmu). There are 289 informative sites; all were included in the parsimony analysis. The consistency index for the shortest trees is 0.849. The *numbers* indicate percent confirmation of grouping of species to the right of the node by bootstrapping analysis with 1,000 replications. Only values greater than 50% are indicated. Below the position of the *arrow* indicates the portion of the tree that would contain the root if the root were located either within the eubacteria or between the eubacteria and archaebacteria.

B CONSENSUS

	10	20	30	40	50	60	70	80	90	100
Eukary	M?GPREDRAT	-WKSNYFLKI	IQLLDDYPKC	FIVGADNVGS	KQMQQIRMSL	RGKAVV	LMGKNTMMRF	AIRGHLEN	NPALEKLLPH	IRGNVG
Archae	MISAE?ERTTEEIP	EWK?EEVAEL	VELLETYDSV	GVVNI?GIPS	KQLQDMRR?L	HGQA?L	RMSRNTLL?F	ALEEAGDETG	NPGLD?L?EY	lEGEVG
Eubact	MALTRQD	KQAIVAEV	SEVfKGALSA	VVADSRGVTV	AKMTELRKRL	REKYGAGVYM	RVVRNTLLRF	AVEGT-E	?FECLEEF	LVGPTA
	110	120	130	140	150	160	170	180	190	200
Eukary	FVFTKEDLTEIRDM	LLANKVPA	AARAGAIAPC	EVTVPAQNTG	LGPEK - TSFF	OALGITTKIS	RGTIEILSDV	OLIKTGDKVG	ASEATLLNML	NISPFS
Archae	LV?TDDNPF?LFQQ	LENSKLKTPA	?INAGEVAPN	DIVVPEGDTG	IDPGPFVGEL	- OTVGANARIC	EGSIOVLDDS	VV?EEGE?VS	DDLSNVLSEL	GIEPKE
Eubact	ILYSMEHPPGAAAK	LFKEFAK					DK	ANAKFELKGG	ALEGKLITAS	OVERI-
										-
	210	220	230	240	250	260	270	280	290	300
Eukary	FGLIIQQVFDNGSI	YSPEVLDITE	EDLHSRFLEG	VRNVASV-CL	-QIGYPTVAS	VPHSIINGYK	RVLALSVETE	YTFP	LAEKVKA	FAA-LA
Archae	VGLDLRGVF?EGVL	FTPEELEIDV	DEYR?DIQ?A	A??AFNL-SV	-NAAYPT?RT	APTLI?K?RG	EAKSL?LQA?	IESPDLADDL	VSKADAQVRA	LAAQID
Eubact	ATLPSYEE-LI	AR-LMGTMKE	ASATKLVRTL	A?AVRDLAYV	LNAI					
	310	320	330	340	350	360				
Eukary	DPSAFAAAAPLAAA	ГТААРААААА	PAKKVEA?	EKK??EES	EESD	EDMGFG	-LFD-			
Archae	DEEALPEELQDVDA	??A????EAD	???DEQSKDE	TQA?E?ADDA	DD?DDDDDDD	DGNAGAEGLG	?MFGG			
Eubact				KE	KEAE					
~										
C										

240 HcuL10 RNL-SV-NAA TmaL10 RNLVYVLNAI

D



Fig. 3. Continued.

carboxy terminus of the L10 and L12 proteins from *S. solfataricus* are "QAAEKKEEEKKG-PSEEEIGGLSSLFG and from human are "KEESEESD (D/E)DMGFGLFD.

The carboxy terminal four to six amino acid residues for the four eubacterial L10 protein contain a high proportion of charged acidic or basic residues. This region is possibly the functional analog to the region of high charge density within archaebacterial and eukaryotic L10 proteins. In the depicted alignment these residues are somewhat arbitrarily placed at positions 343–348.

The analysis of the L10 protein sequences by PAUP

yields six equally parsimonious tree configurations. These six trees divide into the two types designated tree 1 and tree 2 in Fig. 3D. The L10 proteins from human, rat, and mouse are identical except for a few conservative amino acid replacements and a single deletion in the rat protein at position 324. The three subtypes within the type 1 and type 2 trees result from the rearrangement of these closely related mammalian L10 sequences.

The type 1 and type 2 trees differ from each other in two respects: The first is the branching order within the eubacterial domain and the second is the positioning of S. solfataricus. In the type 1 tree, Synechocystis is the deepest branch within the eubacteria and the eukaryotes branch from the S. solfataricus lineage within the archaebacterial group. In the type 2 tree, Synechocystis and T. maritima group together within the eubacteria and the eukaryotes branch from the methanogen/ halophile lineage within the archaebacterial group. Neither of these two positions for the origin of the eucaryotic domain is supported by bootstrapping. And again, if the root of the tree is within the eubacterial domain (below the position of the arrow in Fig. 3D) the archaebacteria appear monophyletic but not holophyletic.

Some regions of the L10 protein alignment are less certain than others. When positions 249–369, representing the region of uncertain alignment, were excluded from parsimony analysis, the shortest trees found exhibited a topology identical to the two types of tree illustrated in Fig. 3D. When only alignment positions 1 to 121 were used for parsimony analysis, the branch pattern within the eukaryotic lineage was not well defined, and branching within the archaebacterial group was reorganized: Halophiles were closer to eukaryotes, *M. vannielli* was closer to eubacteria, and *S. solfataricus* was between the two (data not shown).

The Sequence Alignments and Phylogeny of L12 Proteins

In spite of the major structural discontinuity that occurs between eubacterial L12 sequences and archaebacterial-eukaryotic L12 sequences, biochemical and genetic evidence strongly suggests that all L12 proteins are homologous. First, the organization of the genes encoding ribosomal proteins L11, L1, L10, and L12 is maintained in organisms as divergent as eubacteria and archaebacteria; the L12 gene is always located at the end of the L11, L1, L10, L12 tetragenic cluster. Second, ribosomes from all organisms contain multiple copies of the L12 protein. As a group, these L12 proteins are very acidic, alanine- and proline-rich, and similar in size, ranging between about 110 and 120 amino acids in length. Four copies of the L12 protein along with a single copy of L10 form a distinct stalk on the large ribosome subunit that functions in factor-dependent GTP hydrolysis and mediates structural rearrangements of the ribosome during the protein synthesis cycle. Furthermore, *E. coli* L12 can form an active hybrid with yeast core ribosomes from which the acidic proteins have been removed (Sanchez-Madrid et al. 1981).

In eukaryotic organisms, there are two distinct L12 proteins that have been described. These have been designated type I and type II (or "P2" and "P1," respectively; Amons et al. 1979, 1982; Rich and Steitz 1987; Shimmin et al. 1989; Newton et al. 1990). In the yeast lineage that includes *S. cerevisiae* and *S. pombe*, each of the two genes has been reduplicated to give types IA, IB, IIA, and IIB (Newton et al. 1990; Beltrame and Bianchi 1990).

The alignment of 12 eubacteria and 1 chloroplast, 7 archaebacterial and 9 type I and 10 type II eukaryotic proteins of the L12 family is illustrated in Fig. 4A. All but one of the eukaryotic type II proteins contain a conserved tryptophan at position 88; this aligns to a conserved arginine in the type I, the archaebacterial, and the eubacterial L12 proteins. It is interesting, and perhaps significant, that the extension at the amino terminus of type II proteins shows some sequence similarity to the amino terminus of the eubacterial L12 proteins (alignment positions 1–18). Another salient feature of all L12 proteins, especially the archaebacterial and eukaryotic proteins, is the highly charged carboxyl terminus. The alignment reflects this feature. The two large alignment gaps near the C-terminus within the eubacterial L12 sequences are located within the loops connecting β sheet [B] and α helix [C], and α helix [C] and β sheet [C], respectively (according to the crystal structure of the C-terminal domain of E. coli L12 protein; Leijonmarck et al. 1980). Consequently deletions (or insertions) in these regions could be accommodated without dramatically altering the overall protein structure.

In eukaryotic and archaebacterial species, the L12 carboxy terminal sequences are preceded by an alanine-proline-rich region and exhibit substantial similarity to the carboxy terminus of protein L10. (See above.) Eubacterial L12 proteins have a similar alanine-proline-rich region, but it is located more proximally to the amino-terminus in the protein at positions 39-60. In all the proteins, these alanine-proline-rich regions are believed to be highly flexible and to serve as "hinges" between two distinct domains (Leijonmarck et al. 1980; Leijonmarck and Liljas 1987; Shimmin et al. 1989). The relocation of this hinge to a more amino-terminal position in eubacterial L12 proteins cannot be easily explained. Recent biochemical studies on the S. solfataricus L12 protein have concluded that the aminoand carboxyl-terminal domains of the protein are func-

	10	20	30	40	50	60	70	80	90	100
AsaL12II	MAS-K-DELAC	VYAAI	LILL-DDDVD	ITTEKVN				-TILRAAGVSV	SPYWPGLFTKA	LEGL-
DmeL12II	MSTK-AELAS	VYASI	ILV-DDDVA	VTGEKIN				-TILKAANVEV	EPYWPGLFAKA!	LEAI-
GgaL12II	MASVS-E-LAC	IYSAI	LILH-DDEVT	VTEDKIN				-ALIKAAGVNV	EPFWPGLFAKA	LANI -
HsaL12II	MASVS-E-LAC	IYSAI	ILH-DDEVT	VTEDKIN			 -	-ALIKAAGVNV	EPFWPGLFAKA	LANV-
RraL12II	MASVS-E-LAC	IYSAD	LILH-DDEVT	VTEDKIN				-ALIKAAGVNV	EPFWPGLFAKA	LANV-
SceL12IIA	MS-T-E-SAL	SVAAJ	TLA-DEETE	TSSEKL				-TT. TNAANVDD	ENTWADT FAVA	I DCO-
SCAL12TTB	Manager	CFARE	TIA-DAGLE	TRODUCI				- THINKING LO	MARA DIVIN	LDGQ-
Pool 12TTA	M	OVON	TLA-DECIE	TRODELL					DIVINITE NAME	LEGA-
apout 121 TR	M	SISAL	JILA-DEGIE	I I SDKDG				-SLIKAANVDV	SPIWATIPAKA	LEGA-
aborisiin	MSAS-E-LAT	SISAL	JILA-DEGIS	TTSDKLL				-SLIKAANVDV	SPIWATIFAKA	LEGK-
TCHEIZII	WSLL-R-IEKAA	KGASYSAI	PPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP	TTAANIA				-ALFKTAKLNG	AETTFKTFEDFI	LKTN-
AsaL12I		MRYVAAYI	LAALSGNAD	PSTADIE				-KILSSVGIEC	1PSQLQKVMNEI	LKGK -
DmeL12I		MRYVAAYI	LAVLGGKDS	PANSDLE				-KILSSVGVEV	DAERLTKVIKE	LAGK-
HsaL12I		MRYVASYL	LAALGGNSS	PSAKDIK				-KILDSVGIEA	ODDRLNKVISE	LNGK -
RraL12I		MRYVASYI	LAALGGNSN	PSAKDIK				-KILDSVGIEA	DERLNKVISE	LNGK -
Scel12IA		MKYLAAYI	LLVQGGNAA	PSAADIK				-AVVESVGAEV	DEARINELLSS	LEGK-
SceL121B		MKYLAAYI	LLN-AAGNT	PDATKIK				-AILESVGIEI	EDEKVSSVLSA	LEGK-
Spol12IA		MKYLAAYI	LLTVGGKQS	PSASDIE				-SVLSTVGIEA	SAERVESLISE	LNGK -
SpoL12IB		MKYLAAYI	LLTVGGKDS	PSASDIE				-SVLSTVGIEA	SERIETLINE	LNGK -
TerL121		MKYLAAYA	LVGLSGG-T	PSKSAVE				-AVLKAAGVPV	PSRVDALFAE	FAGK-
HcuL12		MEYVYAAL	ILN-EADEE	LTEDNIT				-GVLEAAGVDVI	SESRAKALVAAI	LEDV-
HhaL12		MEYVYAAL	LILN-EADEE	LTEDNIT				-GVLEAAGVDV	SESRAKALVAA	LEDV-
HvoL12		MEYVYAAI	ILN-ESDEE	VNEENIT				-AVLEAAGVDV	SESRVKALVAA	LEDV~
HmaL12		MEYVYAAD	TIN-BADEE	INFONT				-DVLDAAGVDV	RESEVENTIVAN	LEDV-
MyaL12		MEYTYAAT	LLN-SANKE	VTEEAVK				- AVI.VAGGTEM	TARVKAT.VAA	LEGV
Sact 12		MEVIVA CI	TTH-ANKE	TOPPNIK				-NUL OF ACTUM	SEVEL PRUSSAS	TEEV.
Paol 13		MEVIVACI	ITU-ANKKE	LOEDNIK				-NVLSAAGIIV	SEVELENA VOAN	
380012		MAITINSL	JUDA-AAKKO	ISCENIK				-11015/0601101	JEVRUNAVAAAI	PPEAS
Bat (12	M					TA SA DIV		DA - CAOK TRUT		
BSUDI2 Bent 12	M INTERTION	VENNI VESSINI	SUVE-AISSS	POULTAAR VV	VAGGAAAGA	DAAAEN	BEDVILLA	DR-GROATEVI		DADAK
DSUL12	MARA LAIEBILASV	A MATYLEUNE	UVN-ALSES	FOV TRABLYA	VAGGAAAGGAA	EE	- 8106164	MAS-GSQK1KVI	VVR-51TGLG	DREAK
DVUL12	maSITKEQVVEFI	ANTIVLELSE	STIN-SUBER	I GV SAAAPAM	MAVAAGPA]	AAPASEEK	BrDVILK	AA~GANKIGVI	VVN-ALTGLG	LREAK
SCOL12	MSITKDQIIEAV	AMSVMDVVI	SLIS-AMEEK	r GV SAAAPVA	VAAGEV	KAAEEK	SFOVILK	AA-GANKVAVII	AVR-GATGLGI	JK EAK
HeuL12	MALTQEDIINAV	AEMSVMEVAE	SLVS-AMEEK	FGVSAAAAVV	AGPGGG	A - BEABEQ	refdlvlt	SA-GEKKVNVII	(VVR-EITGLGI	LKEAK
HprL12	MNKEEIMSAI	BEMSVLELSE	ELVB-DLEEK	FGVSAAAPVA	VAGGAA-GAGA	VAEBK	SEFDVFLA	DI-GGKKIKVI	KAVR-ELTGLGI	LKEAK
MlyL12	MNKEQILEAI	KAMIVLELNI	DLVK-AIEEE	FGVTAAAPVV	A-GGAAAAA	EEK	refovla	SA-GAEKIKVII	(VVR-EITGLG	LKEAK
RspL12	MADLNKLAEDI	VGLTLLEAQE	ilkt-ilkdk	YGIEPAAGGA	VMMAGPAAGAA	PAEEEKT	refdvgli	DAAGANKINVII	(EVR-AITGLG)	LKEAK
SecL12	MSAAT-DOILEOL	KSLSLLEASE	ELVK-QIEEA	FGVSAAAPVG	GMVMAAAAAA P	EAAEEKT	TEFDVILE	EVPADKKIAEL	(VVR-TITGLG)	LKEAK
SgrL12	MAKLSQDDLLAQF	EEMTLIELSE	FVK-AFEEK	FDVTAAAAVA	VAGPAAGGAPA		FDVILT	GA-GEKKIQVII	(VVR-ELTSLG)	LKEAK
Sol(c)L12	MAVEAPEKIEQLGTQL	SGLTLEEARV	LVD-WLQDK	LGVSAASFAP	AAAVAAPGAPAI	AAPAVEEKT	FEFDVSIC	EVPSNARISVI	AVR-ALTSLG	LKEAK
StyL12	MSITKDOILEAV	SAMSVMDVVE	LIS-AMEEK	FGVSAAAAVA	VAAGPA	EAAEEK	PEFDVILK	AA-GANKVAVII	KAVR-GATGLG	LKEAK
TmaL12	MTIDEILEAL	EKLTVSELAB	LVK-KLEDK	FGVTAAAPVA	VAAAPVAGAAA	AAOEEK	FEFDVVLK	SF-GONKIOVI	VVR-EITGLG	LKEAK
	110	120	130	140	150	160	170			
Acal.12TT	DL-KSMITNVGS	CUGAAPAAGO	T	EA-PAA	-KERKKERKKE	SEEEDRI	MGPGLET			
Deci 1711	NV-KDLTTN	CUCA & PACCA	ADAAAAAD		-BG-KYFFYKKI	FESDORDO	MORGIEL			
Cast 1371	DI-CELICNVCA	OCOLDA AN A C	A COL & DACO	CANDA	-PERCEPTIVE		MOROLEL			
Noni 1011		CODE DE LOS	DICORDING		- PERVICE VV P		MOROL PL			
Dwellor	NI-GSLICHVGA	COPAPAROAR	PAGGPAPAI.	AAAPA	PERFORMENT PERFORMANCE		MOROL PL			
REALIZIT	NI-GSLICNVGA	GGENERAGAA	PAGGPAPSA	AAAPA	- ABAR VBARADI	10852362				
SCELIZITA	NL-KULLVNF-S	AGAAAPAGVA	GGVAGG-EA	GEAEA	-BREESBARDE	sool	MORGLEL			
Scel1211B	DL-KEILSGFHN	AGPVAGAGAA	SGAAAAGGD	AAA	-BEEKBEEAAE		DMGFGLFL			
Spol1211A	DL-KELLLNIGS	GAGAAPVAGG	AAAPAAA-D	GEAPA	-EEKEEAKBEEI	SSDEI	MGFGLFL			
SpoL1211B	DL-KELLLNIGS	AAAAPAAGGA	GAPAAAAGG	BAAA	- EEQKEEAKEEI	SESDEI	DMGFGLFE)		
TthL12II	PI-TNYIGAIGG	Sapaaassaf	PA	KK	EEPKKEEPKKEI	epkeeetdmi	MG-DLFC			
AsaL12I	DL-EALIAEGQTKLAS	MPTGGAPAAA	AGGAATA-P.	AAE	AKEAKKEEKKEI	ISEEEDEI	MGFGLFD	1		
DmeL12I	SI-DDLIKEGREKLSS	MPVGGGGAVA	AADAAPAAA	AGGD	KKEAKKEEKKEI	SESEDDI	MGFALFE			
Hsall2I	NI~EDVIAQGIGKLAS	VPAGGAVAVS	iaapgsaapa	Agsapaaa	EEKKDEKKEI	SEESDDI	MGFGLFI)		
RraL12I	NI-EDVIAQGVGKLAS	VPAGGAVAVS	aapgsaapa.	AGSAPAAA	EBKKDEKKE	SEESDDI	MGFGLFE	1		
SceL12IA	GS-LEEI IAEGQKKFA	TVPTGGAS	saaagaaga	AAGGDAAA	BEEKEEEKEI	5\$DDI	DMGFGLFE	1		
SceL12IB	SV-DELITEGNEKLAA	VPAAGPASA-	GG-AAA	AGGDAAA	EEEKEEEAAI	BSDDI	MGFGLFI	1		
Spol121A	NI-EELIAAGNEKLST	VPSAGAVATE	PAAGGAAGAE	ATSAA	EEAKEEEAAI	SESDEI	DMGFGLFI	•		
SpoL12IB	DI-DELIAAGNEKLAT	VPTGGAASAA	PAAAAGGAA	PAA	EEAAKEEAKI	SEEE SDEI	DMGFGLFI	•		
TcrL121	DF-DTVCTEGKSKLVG	GVTRPNAATA	SAPTAAAAA	SSGAAAPAAA	A-EEE	EDDI	DMCFGLFE)		
HcuL12	DI-EEAVEEAAA	APAAAPAASG	SDDEAAADD	GDDDEEA-DA	DEAAEAEDAGDI	DDEEPSGE	GLG-DLFG	1		
HhaL12	DI-BEAVEEAAA	APAAAPAASG	SDDEAAADD	GDDDEEA-DA	DEAAEAEDAGDI	DDEEPSGE	LG-DLFG	;		
HvoL12	DI-BEAIETAAA	APAPAAGGSA	GGEVEAADD	DDEED-A-EB	EAADEGGDDDGI	DDEEADGE	SLG-ALFG			
HmaL12	DI-EEAVDQAAA	APVPASGGAA	APAEGDADE	ADEADEEAEE	EAADDGGDDDDI	DEDDEASGE	SLG-ELFG	1		
MvaL12	DI-AEAIAKAAI	APVAAAAPVA	аллара		-EVKKEE-KKEI	угт-алалас	GLG-ALFN	r		
SacL12	NI-DEILKTATA	MPVAAVAAPA	GQQTQQAA-	EK	KEBKKEEEKKGI	SE-EEIGG	LS-SLFG	ł		
SsoL12	NI-DEILKTATA	MPVAAVAAPA	GOOTQOAA-	EK	KBEKKEEEKKGI	SE-EEIGG	LS-SLFC	1		
BstL12	DLVDNTPKPIKE	GIA			-KEEAEBIKAA	EEA	GAKVEIK-			
BsuL12	SLVDNTPKPKE	GIA			-KEBAEELKAKI	.EEV	ASVEVK-			
DvuL12	DKVDGAPSTI.KE	AVS			-KEEAEEAKKO	VEA	AEVEVK-			
Ecol12	DIVESAPAA	WS			-KDDAFALKKAI	.EE ar	AEVEVE-			
Heulij2	AAVOGAPAT				-KEDGDELETEL	EEA	ASVELY-			
Wort.10	(3000) a DOM				-KEDAPENVPV					
MUL12	ENTRY ADVA	GVS			- KUEFELAN		TA GUDUN			
niyuiz Demi 12	DIVE ACC				-NUDADBINAN	V(365-57 E V E -			
ASDU12	DUVE-AGGAVKE	nva			- ARURBAMKKK	A				
Sech12	BUVESTPKAIKE	N.1.C			-KUDAEAIKKQ	155A(JUKAAVK-			
sgrui2	DEVDGTPKPVLE	5.VA			-ABAAEKAAESI	A	MSVEVK-			
SOI(C)L12	ELIEGLPKKLKE	GVS			~KDDAEDAKKQI	EDA	JAKVSIV-			
StyL12	DLVESAPAALKE	GVS			-KDDAEALKKS	.EEA(GAEVEVK-			
TmaL12	DLVEKAGSPDAV-IKS	gvs			-KEEABEIKKK	JEEA	BAEVELK -			
CONSENSUS	10	20	30	40	50	60	70	80	90	100
EukaryII	MASVS-EELACVV	KGASYSAL	ILA-DDEVE	ITSDKIN				-TLTKAAGVNV	SPFWPGLFAKA	LEG?-
EukaryI		MKYLAAYL	LL?LGGN?S	PSASDIE				-KILSSVGIEA	DERL?KVISE	LNGK -
Archae		MEYVYAAI	ILN-EADEE	ITEENIT				-?VLEAAGVDV	EESRVKALVAA	reda-
Eubact	MAVESSLTKEQIIEAI	KEMIVLELNE	ELVK-ALEEK	FGVSAAAPVA	VAGGAAAGAAA	AAEAAEEK"	refovila	?APGANKIKVI	KVVR-EITGLG	LKEAK
	110	120	130	140	150	160	170			
EukaryII	DL-KELI?NIGG	GGGAAAAGAA	AAAAAAAG?.	АААРААКК	EEEKKEEAKKEI	SEEEESDDI	MGFGLFE	•		
EukaryI	DI-EELIAEGNEKLAS	VPTGGAAAVA	аардааааа	AGGA?AAAAE	AKEEKKEEKKEI	ESEESDDI	MGFGLFE	•		
Archae	DI-EEAVETAAA	APVAA?AASA	GDDEQAADD	GDEDEEAAEE	DEAKEEEDKKD	DDEEASGE	GLG-DLFG	1		
Eubact	DLVDGAPKADAV-LKE	GVS			-KEDAEEIKKKI		GASVEVK-			

Fig. 4. Alignment of the amino acid sequences of the ribosomal protein L12 family. **A** The L12 equivalent proteins from 13 eubacteria, 7 archaebacteria, and 19 eukaryotic L12 equivalent proteins were aligned. The eukaryotic proteins divide into two types designated as type I and type II. Abbreviations are as in Table 1. **B** The consensus

Α

В

of the alignment which was generated manually by a flexible majority rule. When majority was not evident at an alignment position, chemically similar amino acid residues were considered to determine the consensus. *Question marks* (?) indicate that there was no simple consensus at such positions. tionally equivalent to the corresponding amino- and carboxyl-terminal domains of the *E. coli* L12 protein (Köpke et al. 1992); this result supports a colinear alignment. To simplify visualization and comparison, a consensus of the eukaryotic type I and II, the archaebacterial, and the eubacterial L12 proteins are aligned in Fig. 4B.

It should be stressed here that in any alignment (and in particular this L12 alignment) the assumption of common ancestry of each amino acid at a given alignment position is less than certain. That is, alignments simply reflect a guess, hopefully a best guess, of common ancestry at every position.

The phylogenetic relationships among the L12 family of protein sequences were determined using parsimony (Fig. 5) and distance matrix methods (not shown). Because of the uncertainty in generating a reliable alignment between eubacterial and archaebacterial-eukaryotic L12 sequences, we first determined the phylogenies of eubacteria, archaebacteria, and eukaryotes separately, and then for comparison we determined the "universal" phylogeny. In general, the branch patterns within the eukaryotic, archaebacterial, and eubacterial groups were essentially identical in the "universal" tree and the three individual trees. The universal parsimony tree (shown in Fig. 5) and a Fitch-Margoliash distance tree (not shown) both indicated that the eubacterial sequences form a single coherent group that is confirmed by bootstrap analysis. However, the branching order within this group is not substantiated by bootstrap analysis.

The archaebacterial L12 sequences also appear to form a coherent group that is both mono- and holophyletic. By bootstrap resampling, the confirmation of this grouping was 57% for the protein alignment and 58% for the corresponding nucleic acid alignment analyzed by PAUP (data not shown). In contrast, the eukaryotic L12 sequences clearly resolve into two groups corresponding to the type I and type II proteins. This distinct division implies that the duplication of the L12 gene occurred very early in the eukaryotic lineage.

Phylogenetic Considerations

The alignment and phylogenetic analysis presented above using L11, L1, L10, and L12 protein sequences generally support the concept that organisms divide into three distinct and well-defined groups: Eubacteria, archaebacteria, and eukaryotes. The ribosomal protein sequences from member species within a group are in most cases more similar to each other based on amino acid identity than to the sequences from species outside the group. Furthermore, numerous deletions, insertions, or structural rearrangements in these ribosomal protein sequences confirm this three-part delineation and demarcation.

If the root in these ribosomal protein-based trees is near or within the eubacterial domain, then it is clear that the archaebacteria appear monophyletic, originating from a common ancestor that is distinct from eubacteria. The origin of the eukaryotes is more problematic. They appear to originate as a distinct branch either outside of the archaebacterial group as suggested by the L12 protein phylogeny or alternatively from within the archaebacterial group as suggested by the L11 and L10 protein phylogenies.

Although ribosomal proteins at first glance might be considered good candidates for phylogenetic analysis, in reality they are less than perfect for a number of reasons. First, they are relatively small proteins, and second, their divergence and structural rearrangements often make alignments difficult and ambiguous. Because of these limitations, the origin of the eukaryotic lineage either from within or outside of the archaebacterial group cannot be statistically substantiated.

Phylogenetic analysis of rRNA sequences and translational elongation factors Tu and G sequences suggests that the hyperthermophilic eubacterium T. maritima is a representative of a deep branching lineage within the eubacterial group (Achenbach-Richter et al. 1987; Bachleitner et al. 1989; Tiboni et al. 1991). Representatives of deep branching lineages within the archaebacteria are also hyperthermophilic. This has led to the suggestion that the ancestor of eubacteria and archaebacteria (i.e., the common ancestor represented as the root of the universal tree) was hyperthermophilic (Achenbach-Richter et al. 1987; Burggraf et al. 1992; Stetter 1993). This would place the position of the root either deep within the eubacterial or archaebacterial groups or somewhere between the two groups. Previous analyses of translational elongation factors and subunits of ATPase have placed the root somewhere between eubacteria and archaebacteria (Iwabe et al. 1989; Gogarten et al. 1989).

In contrast to the phylogenetic analysis based on rRNA and the elongation factors Tu and G (Achenbach-Richter et al. 1987; Bachleitner et al. 1989; Tiboni et al. 1991), our analysis using L11, L1, L10, and L12 ribosomal protein sequences is less definitive with respect to the placement of *T. maritima* within the eubacteria. The resolution of our trees is limited by the relatively small size of these proteins and in some cases by the limited number of sequences available for analysis. The tree for the L12 protein, containing 13 eubacterial sequences, is virtually devoid of resolution that is confirmable by bootstrap analysis. In the L11 tree, the mesophile *S. virginiae* appears to branch more deeply than *T. maritima*. These observations seem to suggest that different molecules, although they are all compo-



Fig. 5. Phylogenetic tree inferred from the alignment of L12 amino acid sequences. A parsimony analysis of aligned sequences was carried out by using PAUP with the heuristic tree search option. Illustrated is the majority rule consensus of the 14 equally shortest trees. There are 147 informative sites in the alignment; all of them were used for parsimony analysis. The consistency index is 0.598. When the first 18 alignment positions, and the flexible hinge regions (position 43 to 74 for eubacteria and 119-146 for archaebacteria and eukaryotes) were excluded from analysis, 20 shortest trees were found; the majority rule consensus of these trees has essentially the same topology as the tree shown here. The numbers refer to the percent confirmation of grouping of the species to the right of the node by bootstrap analysis with 100 replications. Only values greater than 50% are indicated. Below the position of the arrow indicates the portion of the tree that would contain the root if the root were located either within the eubacteria or between the eubacteria and archaebacteria.

nents of the protein synthesis apparatus, can diverge to some extent independently and give rise to incongruent phylogenies. The "true" organismal phylogeny will hopefully become apparent from a consensus of molecular phylogenies.

Rivera and Lake (1992) have suggested that the eukaryotic lineage arose as a branch from the sulfur-metabolizing thermophilic lineage (i.e., the "eocytes" or euryarchaeota) within the archaebacterial group. Other analyses indicate that the eukaryotic lineage originates outside of the archaebacterial domain. Our data neither confirm nor refute either of these two positionings. However, our analysis clearly highlights the major discontinuity that separates archaebacterial and eukaryotic ribosomal protein sequences. The sequence (amino acid identity) and structure (deletions, insertions, and rearrangements) of a ribosomal protein from organisms within a group (i.e., eubacteria, archaebacteria, or eukaryotes) are clearly more similar to each other than to the sequence and structure of the protein from organisms outside the group.

Acknowledgments. We thank Joe Felsenstein for providing the PHYLIP package, Jeffrey Bromberger for assistance in using PHYLIP, David Swofford for help in using PAUP, Allen Delaney for data bank searches, Lawrence Shimmin for discussion of sequence alignments, and James Lake for his interest and comments. This work was supported by grants from the Medical Research Council of Canada (to P.P.D.). P.P.D. is a Fellow of the Canadian Institute for Advanced Research.

References

- Achenbach-Richter L, Gupta R, Stetter KO, Woese CR (1987) Were the original eubacteria thermophiles? System Appl Microbiol 9: 34–39
- Amons R, Pluijms W, Kriek T, Möller W (1982) The primary struc-

ture of proteins eL12'/eL12'-P from the large subunit of Artemia salina ribosomes. FEBS Lett 45:143-147

- Amons R, Pluijms W, Mlöller W (1979) The primary structure of ribosomal eL12/eL12-P from Artemia salina 80S ribosomes. FEBS Lett 104:85–89
- Arndlt E, Weigel C (1990) Nucleotide sequence of the genes encoding the L11, L1, L10, and L12 equivalent ribosomal proteins from the archaebacterium *Halobacterium marismortui*. Nucleic Acids Res 18:1285
- Bacheitner M, Ludwig W, Stetter KO, Schleifer KH (1989) Nucleotide sequence of the gene coding for the elongation factor Tu from extremely thermophilic eubacterium *Thermontoga* str operon. FEBS Lett 57:115–120
- Baier G, Piendl W, Redl B, Stoeffler G (1990) Structure, organization and evolution of the L1 equivalent ribosomal protein gene of the archaebacterium *Methanococcus vanielli*. Nucleic Acids Res 18:719–724
- Bartsch M, Kimura M, Subramanian AR (1982) Purification, primary structure, and homology relationships of a chloroplast ribosomal protein. Proc Natl Acad Sci USA 79:6871–6875
- Baughman G, Nomura M (1983) Localization of the target site for translational regulation of the L11 operon and direct evidence for translational coupling in *Escherichia coli*. Cell 34:979–988
- Beauclerk AAD, Hummel H, Holmes DJ, Bock A, Cundliffe E (1985) Studies of the GTPase domain of archaebacterial ribosomes. Eur J Biochem 151:242–255
- Beltrame M, Bianchi ME (1990) A gene family for acidic ribosomal proteins in *Schizosaccharomyces pombe:* two essential and two non-essential genes. Mol Cell Biol 10:2341–2348
- Branlant C, Krol A, Machatt A, Ebel JP (1981) The secondary structure of the protein L1 binding region of ribosomal 23S RNA: homologies with putative secondary structures of the L11 mRNA and a region of mitochondrial 16S rRNA. Nucleic Acids Res 9:293– 307
- Burggraf S, Olsen GJ, Stetter KO, Woese CR (1992) A phylogenetic analysis of Aquifex pyrophilus. System Appl Microbiol 15: 352–356
- Casiano C, Matheson AT, Traut RR (1990) Occurrence in the archaebacterium *Sulfolobus solfataricus* of a ribosomal protein complex corresponding to *Escherichia coli* (L7/L12)₄ · L10 and eukaryotic (P1)₂/(P2)₂ · PO. J Biol Chem 265:18757–18761
- Chan Y-L, Paz V, Wool IG (1989) The primary sequence of rat acidic ribosomal phosphoprotein P0. EMBL accession number X15096
- Dean D, Nomura M (1980) Feedback regulation of ribosomal protein gene expression in *Escherichia coli*. Proc Natl Acad Sci USA 77: 3590–3594
- Dijk J, Garret RA, Muller R (1979) Studies on the binding of the ribosomal protein complex L7/L12-L10 and protein L11 to the end one third of the 23S RNA: a functional center of the 50S subunit. Nucleic Acids Res 6:2717–2729
- Egebjerg J, Douthwaite SR, Liljas A, Garrett RA (1990) Characterization of the binding sites of protein L11 and the L10 · (L12)₄ pentameric complex in the GTPase domain of 23S ribosomal RNA from *Escherichia coli*. J Mol Biol 213:275–288
- El-Baradi TAL, de Regt CHF, Einerhand SWC, Teixido J, Planta RJ, Ballesta JPG, Raue HA (1987) Ribosomal proteins EL11 from Escherichia coli and L15 from Saccharomyces cerevisiae bind to the same site in both yeast 26S and mouse 28S rRNA. J Mol Biol 195: 909–917
- Falkenberg P, Yaguchi M, Roy C, Zurker M, Matheson AT (1985) The primary structure of the ribosomal A-protein (L12) from the moderate halophile NRCC 41227. Biochem Cell Biol 64:675–680
- Felsenstein J (1991) PHYLIP (Phylogeny Inference Package, version 3.4). University of Washington

Ferro JA, Reinach FC (1988) The complete sequence of chicken-mus-

cle cDNA encoding the acidic ribosomal protein P1. Eur J Biochem 177:513-516

- Garland WG, Louie KA, Matheson AT, Liljas A (1987) The complete amino acid sequence of the ribosomal 'A' protein (L12) form *Bacillus stearothermophilus*. FEBS Lett 220:43–46
- Gogarten JP, Kibak H, Dittrich P, Taiz L, Bowman EJ, Bowman BJ, Manolson MF, Poole RJ, Date T, Oshima T, Konoshi J, Denda K, Yoshida M (1989) Evolution of the vacular H⁺-ATPase: implications for the origin of eukaryotes. Proc Natl Acad Sci USA 86: 6661–6665
- Gourse RL, Thurlow DL, Gerbi SA, Zimmerman RA (1981) Specific binding of a prokaryotic ribosomal protein to a eukaryotic ribosomal RNA: implications for evolution and autoregulation. Proc Natl Acad Sci USA 78:2722–2726
- Gudkov AT, Tamanova LG, Gongadze GH, Bushuev VN (1980) Role of different regions of ribosomal proteins L7 and L10 in their complex formation and in the interaction with the ribosomal 50S subunit. FEBS Lett 109:34–38
- Hansen TS, Andreasen PH, Dreisig H, Højrup P, Nielsen H, Engberg J, Kristiansen K (1991) *Tetrahymena thermophila* acidic ribosomal protein L37 contains an archaebacterial type C-terminus. Gene 105:143–150
- Huber R, Wilharm T, Huber D, Trincone A, Burggaf S, König H, Rachel R, Rockinger I, Fricke H, Stetter KO (1992) Aquifex pyrophilus gen. sp. nov., represents a novel group of marine hyperthermophilic hydrogen-oxidizingbacteria. System Appl Microb 15:340–351
- Itoh T (1981) Primary structure of an acidic ribosomal protein from Micrococcus lysodeikticus. FEBS Lett 127:67-70
- Itoh T, Higo KI (1983) Complete amino acid sequence of an L7/L12type ribosomal protein from *Rhodopseudomonas spheroides*. Biochim Biophys Acta 744:105–109
- Itoh T (1988) Complete nucleotide sequence of the ribosomal "A" protein operon from the archaebacterium *Halobacterium halobium*. Eur J Biochem 176:297–303
- Itoh T, Wittmann-Liebold B (1978) The primary structure of Bacillus subtilis acidic ribosomal protein B-L9 and its comparison with Escherichia coli proteins L7/L12. FEBS Lett 96:392–394
- Itoh T, Sugiyama M, Higo KI (1982) The primary structure of an acidic ribosomal protein from *Streptomyces griseus*. Biochim Biophys Acta 701:164–172
- Itoh T, Otaka E (1984) Complete amino-acid sequence of an L7/L12type ribosomal protein from *Desulfovibrio vulgaris* Miyazaki. Biochim Biophys Acta 789:229–233
- Iwabe N, Kuma K, Hasegawa M, Osawa S, Miyata T (1989) Evolutionary relationship of archaebacteria, eubacteria and eukaryotes inferred from phylogenetic trees of duplicated genes. Proc Natl Acad Sci USA 86:9355–9359
- Juan-Vidales F, Sánchez-Madrid F, Saenz-Robles MT, Ballesta JPG (1983) Purification and characterization of two ribosomal proteins of *Saccharomyces cerevisiae*, homologies with proteins from eukaryotic and with bacterial protein ECL11. Euro J Biochem 136: 275–281
- Kearney KR, Nomura M (1987) Secondary structure of the autoregulatory mRNA binding site of ribosomal protein L1. Mol Gen Genet 210:609-618
- Kelley MR, Venugopal S, Harlem J, Deutsch WA (1989) Antibody to a human DNA repair protein allows for cloning a *Drosophila* cDNA that encodes an apurinic endonuclease. Mol Cell Biol 9: 965–973
- Kimura M, Kimura J, Ashman K (1985) The complete primary structure of ribosomal proteins L1, L14, L15, L23, L24 and L29 from *Bacillus stearothermophilus*. Eur J Biochem 150:491–497
- Köpke AKE, Leggatt PA, Matheson AT (1992) Structure function re-

lationships in the ribosomal stalk proteins of archaebacteria. J Biol Chem 167:1382–1390

- Krowczynska AM, Coutts M, Makrides S, Brawerman G (1989) The mouse homologue of the human acidic ribosomal protein P0: a highly conserved polypeptide that is under translational control. Nucleic Acids Res 17:6408–6408
- Lake JA, Strycharz WA (1981) Ribosomal proteins L1, L17, L27 localized at single sites on the large subunit by immune electron microscopy. J Mol Biol 153:979–992
- Leijonmarck M, Eriksson S, Liljas A (1980) Crystal structure of a ribosomal component at 2.6 Å resolution. Nature 286:824–826
- Leijonmarck M, Liljas A (1987) Structure of the C-terminal domain of the ribosomal protein L7/L12 from *Escherichia coli* at 1.7 Å. J Mol Biol 195:555–580
- Liao D, Dennis PP (1992) The organization and expression of essential transcription-translation component genes in the extremely thermophilic eubacterium *Thermotoga maritima*. J Biol Chem 267: 22787–22797
- Liljas A (1982) Structural studies of ribosomes. Prog Biophys Mol Biol 40:161–228
- Matheson AT, Louie KA, Tak BD, Zuker M (1987) The primary structure of ribosomal A-protein (L12) from the halophilic eubacterium *Haloanaerobiumpraevalens*. Biochimie 69:1013–1020
- Matheson AT, Louie AK, Bock A (1988) The complete amino acid sequence of the ribosomal A protein (L12) from the archaebacterium Sulfolobus acidocaldarius. FEBS Lett 231:331–335
- Mitsui K, Nakogawa T, Tsurugi K (1989) The gene and the primary structure of acidic ribosomal protein A0 from yeast Saccharomyces cerevisiae which show partial homology to bacterial ribosomal protein L10. J Biochem 160:223–227
- Mitsui K, Tsurugi K (1988) cDNA and deduced amino acid sequence of acidic ribosomal protein A0 from Saccharomyces cerevisiae. Nucleic Acids Res 16:3573
- Newton CH, Shimmin LC, Yee J, Dennis PP (1990) A family of genes encode the multiple forms of the *Saccharomyces cerevisiae* ribosomal proteins equivalent to the *Escherichia coli* L12 protein and a single form to L10-equivalent ribosomal protein. J Bacteriol 172: 579–588
- Okamoto S, Nihira T, Kataoka H, Suzuki A, Yamada Y (1992) Purification and molecular cloning of butyrolactone autoregulator receptor from *Streptomyces virginiae*. J Biol Chem 267:1093–1098
- Paton EB, Woodmaska MI, Kroupskaya IV, Zhyvoloup AN, Matsuka GK (1990a) Evidence for the ability of L10 ribosomal proteins of Salmonella typhimurium and Klebsiella pneumoniae to regulate rplJL gene expression in Escherichia coli. FEBS Lett 2675:129– 132
- Paton EB, Zolotukhiu SB, Woodmaska MI, Kroupskaya IV, Zhyvoloup AN (1990b) The nucleotide sequence of gene rplJ encoding ribosomal protein L10 of Salmonella typhimurium. Nucleic Acids Res 18:2824
- Pettersson I (1979) Studies on the RNA and protein binding sites of the *E. coli* ribosome protein L10. Nucleic Acids Res 6:2637–2646
- Post LE, Strycharz GD, Nomura M, Lewis H, Dennis PP (1979) Nucleotide sequence of the ribosomal protein gene cluster adjacent to the gene for RNA polymerase subunit in *Escherichia coli*. Proc Natl Acad Sci USA 76:1697–1701
- Prieto J, Candel E, Coloma A (1991) Nucleotide sequence of cDNA encoding ribosomal protein P0 in *Dictyostelium discoideum*. Nucleic Acids Res 19:1342
- Pucciarelli MG, Remacha M, Vilella MD, Ballesta JPG (1990) The 26S rRNA binding ribosomal protein equivalent to bacterial protein L11 is encoded by unspliced duplicated genes in Saccharomyces cerevisiae. Nucleic Acids Res 18:4409–4416

- Qian S, Zhang J-Y, Kay MA, Jacobs-Lorena M (1987) Structural analysis of the *Drosophila* rpA1 gene, a member of the eukaryotic "A" type ribosomal protein family. Nucleic Acids Res 15: 987–1003
- Ramirez C, Shimmin LC, Newton CH, Matheson AT, Dennis PP (1989) Structure and evolution of the L11, L1, L10 and L12 equivalent ribosomal proteins in eubacteria, archaebacteria and eucaryotes. Can J Microbiol 35:234–244
- Remacha M, Saenz-Robler AT, Vilella MD, Ballesta JPG (1988) Independent genes coding for three acidic proteins of the large ribosomal subunit from *Saccharomyces cerevisiae*. J Biol Chem 263:9094–9101
- Rich BE, Steitz JA (1987) Human acidic ribosomal phosphoproteins P0, P1, and P2: analysis of cDNA clones, in vitro synthesis, and assembly. Mol Cell Biol 7:4065–4074
- Rivera MC, Lake JA (1992) Evidence the eucaryotes and eocyte prokaryotes are immediate relatives. Science 257:74-76
- Ryan PC, Lu M, Draper DE (1991) Recognition of highly conserved GTPase center of 23S ribosomal RNA by ribosomal protein L11 and antibiotic thiotrepton. J Mol Biol 221:1257–1268
- Sanchez-Madrid F, Vidales FJ, Ballesta JPG (1981) Functional role of acidic ribosomal proteins. Interchangeability of proteins from bacterial and eukaryotic cells. Biochemistry 20:3263–3266
- Sander G (1983) Ribosomal protein L1 from *Escherichia coli*. Its role in the binding of tRNA to the ribosome and in elongation factor G-dependent hydrolysis. J Biol Chem 258:10098–10102
- Schijman AG, Dusetti NJ, Vazquez MP, Lafton S, Levy-Yeyati P, Levin MJ (1990) Nucleotide cDNA and complete deduced amino acid sequence of a *Trypanosoma cruzi* ribosomal P protein (P-JL5). Nucleic Acids Res 18:3399
- Shimmin LC, Dennis PP (1989) Characterization of the L11, L1, L10 and L12 equivalent ribosomal protein gene cluster of the halophilic archaebacterium *Halobacterium cutirubrum*. EMBO J 8:1225–1235
- Shimmin LC, Ramirez C, Matheson AT, Dennis PP (1989) Sequence alignment and evolutionary comparison of the L10 equivalent and L12 equivalent ribosomal protein from archaebacteria, eubacteria and eukaryotes. J Mol Evol 29:448–462
- Sibold C, Subramanian AR (1990) Cloning and characterization of the genes for ribosomal proteins L10 and L12 from *Synechocystis* Sp. PCC 6803: comparison of gene clustering pattern and protein sequence homology between cyanobacteria and chloroplasts. Biochim Biophys Acta 1050:61–68
- Smooker PM, Schmidt J, Subramanian AR (1991) The nuclear:organelle distribution of chloroplast ribosomal protein genes. Features of a cDNA clone encoding precursor of L11. Biochimie 73: 845–851
- Sor F, Nomura M (1987) Cloning and DNA sequence determination of the L11 ribosomal protein of Serratia marcenscens and proteus vulgaris. Translational feedback regulation of Escherichia coli L11 operon by heterologous L1 proteins. Mol Gen Genet 210:52–59
- Stetter KO (1993) Life at the upper temperature border. In: Than Thanh Van JT, Mounolou JC and McKay C (eds) Colloque interdisciplinaire du comite national de la recherche scientifique. Frontiers of Life, Le Bloris proceedings, Editions Frontièrs, Gifsur-Yvette, pp 195–219
- Subramanian AR, Dabbs ER (1980) Functional studies on ribosomes lacking protein L1 from mutant E. coli. Eur J Biochem 112:425– 430
- Suzuki K, Olvera J, Wool IG (1990) The primary structure of rat ribosomal protein L12. Biochem Biophys Res Commun 172:35–41
- Swofford DL (1989) PAUP Version 3.0 Illinois National History Survey, Champaign, Illinois

- Tiboni O, Cantoni R, Creti R, Cammarano P, Sanangelantoni AM (1991) Phylogenetic depth of *Thermotoga maritima* inferred from analysis of the fus gene: amino acid sequence of elongation factor G and organization of the *Thermotoga* str operon. J Mol Evol 33:142–151
- Thomas MS, Nomura M (1987) Translational regulation of the L11 ribosomal protein operon of *Escherichia coli:* mutations that define the target site for repression by L1. Nucleic Acids Res 15:3085–3096
- Uchiumi T, Wahba AJ, Traut RR (1987) Topography and stoichoimetry of acidic proteins in large ribosomal subunits from Artemia salina as determined by crosslinking. Proc Natl Acad Sci USA 84:5580-5584
- Wigboldus JO (1987) cDNA and deduced amino acid sequence of

Drosophila rp21C, another "A" type ribosomal protein. Nucleic Acids Res 15:10064

- Wool IG, Endo Y, Chan Y-L, Glück A (1990) Structure, function and evolution of mammalian ribosomes. In: Hill WE, Dahlberg A, Garrett RA, Moore PB, Schlessinger D and Warner JR (eds) The ribosome: Structure, function and evolution. American Society for Microbiology, Washington, DC, pp 203–214
- Yates JL, Nomura M (1981) Feedback regulation of ribosomal protein synthesis in *E. coli*: localization of the mRNA target sites for repressor action of ribosomal protein L11. Cell 24:243–249
- Zimmerman RA, Thurlow DL, Finn RS, March TL, Ferrett LK (1980) Conservation of specific protein RNA interactions in ribosomal evolution. In: Osasa S, Ozeki H, Uchida H and Yura T (eds) Genetics and Evolution of RNA polymerase, tRNA and ribosomes. University of Tokyo Press, Tokyo, pp 569–584