Supplementary Information

Promiscuous behaviour of archaeal ribosomal proteins: Implications for eukaryotic ribosome evolution

Jean-Paul Armache[1,2*], Andreas M. Anger[1,2*], Viter Márquez[1,2#], Sibylle Franckenberg[1,2#], Thomas Fröhlich[1,3], Elizabeth Villa[4], Otto Berninghausen[1,2], Michael Thomm[5], Georg J. Arnold[1,3], Roland Beckmann[1,2,6] and Daniel N. Wilson[1,2,6]

[1] Gene Center and Department of Biochemistry, Ludwig-Maximilians-Universität, Feodor Lynen Str. 25, 81377, Munich, Germany

[2] Center for integrated Protein Science Munich (CiPSM), Ludwig-Maximilians-Universität München, Germany

[3] Laboratory for Functional Genome Analysis LAFUGA, Gene Center and Department of Biochemistry, Ludwig-Maximilians-Universität, Feodor Lynen Str. 25, 81377, Munich, Germany

[4] Department of Molecular Structural Biology, Max Planck Institute of Biochemistry, Am Klopferspitz 18, 82152 Martinsried, Germany.

[5] Lehrstuhl für Mikrobiologie, Regensburg University, Universitätstraße 32, 93053 Regensburg, Germany

[6] Corresponding authors: wilson@lmb.uni-muenchen.de, beckmann@lmb.uni-muenchen.de

* These authors contributed equally to this work

# These authors contributed equally to this work

Supplementary Table 1. Distribution of ribosomal proteins on the ribosomal small subunit in Eukarya (E), Korarchaeota (KA), Crenarchaeota (CA), Euryarchaeota (EA) and Bacteria (B)

Protein names	**E**		**KA**	**CA**				**EA**								**B**	
	Sce	***Hsa***	***Kcr***	***Ape***	***Sma***	***Eco***	***Pae***	***Pfu***	***Tko***	***Mig***	***Mth***	***Mka***	***Afu***	***Hma***	***Tvo***	***Tth***	***Eco***
RACK1	√	√	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
S1e (S3ae)	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	–	–
S2	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
S3	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
S4	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
S4e	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	–	–
S5	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
S6e	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	–	–
S7	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
S7e	√	√	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
S8	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
S8e	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	–	–
S9	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
S10	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
S10e	√	√	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
S11	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
S12	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
S12e	√	√	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
S13	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
S14	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
S15	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
S17	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
S17e	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	–	–
S19	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
S19e	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	–	–
S21e	√	√	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
S24e	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	–	–
S25e	√	√	–	√	√	√	√	–	–	–	–	–	–	–	–	–	–
S26e	√	√	√	√	√	√	√	–	–	–	–	–	–	–	–	–	–
S27e	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	–	–
S28e	√	√	√	√	√	√	√	√	√	√	√	√	√	√	–	–	–
S30e	√	√	√	√	√	√	√	–	–	–	–	–	–	–	–	–	–
S31e (S27ae)	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	Thx	–
L8eS (L7ae)	–	–	u	u	√	√	√	√	√	√	√	u	u	u	√	–	–

Hsa (Homo sapiens), Kcr (Korarchaeum cryptofilum, Ape (Aeropyrum pernix), Sma (Staphylothermus marinus), Sac (Sulfolobus acidocaldarius), Pae (Pyrobaculum aerophilum), Pfu (Pyrococcus furiosus), Tko (Thermococcus kodakaraensis), Mig (Methanococcus igneus), Mth (Methanobacterium thermautotrophicus), Mka (Methanopyrus kandleri), Afu (Archaeoglobus fulgidus), Tvo (Thermoplasma volcanii), Hma (Haloarcula marismortui), Eco (Escherichia coli)

Thx	Protein known to exist in Tth belonging to bacterial S31e family, but unrelated to eukaryotic S31e
√	Present
–	Absent
u	Distribution unknown

Supplementary Table 2. Distribution of large subunit r-proteins in Eukarya (E), Korarchaeota (KA), Crenarchaeota (CA), Euryarchaeota (EA) and Bacteria (B)

Protein names	E		KA	CA				EA								B	
	Sce	*Hsa*	*Kcr*	*Ape*	*Sma*	*Sac*	*Pae*	*Pfu*	*Tko*	*Mig*	*Mth*	*Mka*	*Afu*	*Hma*	*Tvo*	*Eco*	*Tth*
L1	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
L2	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
L3	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
L4	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
L5	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
L6	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
L6e	√	√	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−
L8e(L7ae)	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	−	−
L11	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
L13	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
L13e	√	√	√	√	√	− *	√	−	−	−	−	−	−	−	−	−	−
L14	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
L14e	√	√	√	√	√	√	√	√	√	√	√	√	−	−	−	−	−
L15	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
L15e	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	−	−
L16	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
L18	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
L18e	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	−	−
L19e	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	−	−
L20e(L18ae)	√	√	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−
L21e	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	−	−
L22	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
L22e	√	√	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−
L23	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
L24	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
L24e	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	−	−
L27e	√	√	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−
L28e	−	√	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−
L29	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
L29e	√	√	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−
L30	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
L30e	√	√	√	√	√	√	√	√	√	√	√	−	√	−	−	−	−
L31e	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	−	−
L32e	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	−	−
L33e(L35ae)	√	√	−	√	√	−	−	√	√	−	−	−	−	−	−	−	−
L34e	√	√	√	√	√	√	√	√	√	√	√	√	−	−	−	−	−
L36e	√	√	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−
L37e	√	√	√	√	√	√	−	√	√	√	√	√	√	√	√	−	−
L38e	√	√	−	√	√	−	√	−	−	−	−	−	−	−	−	−	−
L39e	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	−	−
L40e	√	√	√	√	√	√	√	√	√	√	√	√	√	√	−	−	−
L41e	√	√	−	−	−	−	−	√	√	√	−	−	−	−	−	−	−
L43e(L37ae)	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	−	−
L44e	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	−	−
P0(L10)	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
P1(L12)	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
P2(L12)	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	−	−
L8e2(L7ae)	−	−	u	u	√	√	−	√	√	−	√	u	u	− **	u	−	−
L14e2(L14e)	−	−	u	u	√	√	√	√	√	√	√	u	−	−	−	−	−
LX(L20e)	−	−	−	√	√	√	√	√	√	√	√	√	√	√	−	−	−
S24eL	−	−	u	u	−	−	−	√	√	−	−	u	u	−	u	−	−

* Present in *Sulfolobus* species, but not found *S. acidocaldarius*, ** Unknown whether it is absent, or less stably bound, √ Present, − Absent, u Distribution unknown.

Supplementary Table 3. *T. kodakaraensis* 30S r-proteins identified by LC-MS and 2D-PAGE

Protein Name (Family)	LC/MS	2D	Amino acids	gi[a]	Molecular mass (kDa)	pI
S1e (S3ae)	√		200	57641189	23.02	10.1
S2	√	√	201	57641431	23.01	8.9
S3	√	√	209	57641471	23.4	9.6
S4	√	√	180	57641440	21.2	10.2
S4e	√	√	243	57641464	27.8	9.9
S5	√	√	235	57641456	26.3	9.4
S6e	√	√	125	57641886	13.7	10.1
S7	√	√	215	57641012	24.5	10.0
S8	√	√	130	57641461	14.6	9.5
S8e	√	√	130	57641126	14.5	10.8
S9	√	√	135	57641435	15.3	10.5
S10	√	√	102	57640242	11.7	10.1
S11	√	√	140	57641439	15.1	10.3
S12	√	√	147	57641013	16.4	10.6
S13	√	√	149	57641441	16.9	10.8
S14[b]	√		56	57641462	6.6	10.4
S15	√	√	151	57641186	17.5	10.5
S17	√	√	114	57641467	13.2	9.5
S17e	√	√	67	57642227	8	10.3
S19	√	√	133	57641473	15.4	10.3
S19e	√	√	150	57641211	17.3	9.4
S24e	√	√	98	57641631	11.4	5.5
S27e	√	√	65	57641034	7.1	9.2
S28e	√	√	70	57641245	7.9	11.0
S31e (S27ae)	√		57	57641630	6.7	10.2

[a] gi refers to the GenInfo identifier for retrieval from NCBI

[b] Identified with only a single peptide

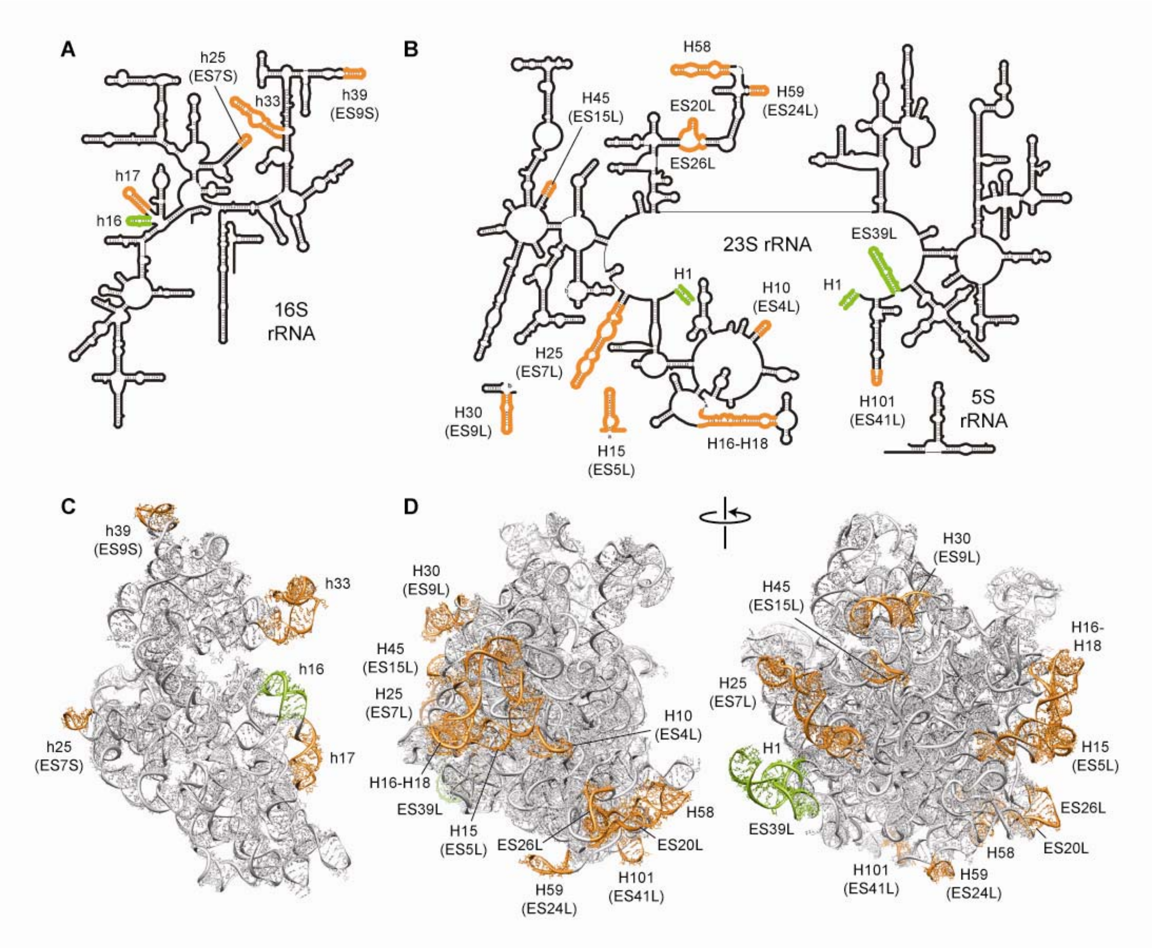
Supplementary Table 4. *T. kodakaraensis* 50S r-proteins identified by LC-MS and 2D-PAGE

Protein Name (Family)	LC/MS	2D	Amino acids	gi[a]	Molecular mass (kDa)	pI
L1	√	√	216	57641352	24.1	9.7
L2	√	√	239	57641474	26.0	10.7
L3	√	√	346	57641477	39	10
L4	√	√	255	57641476	28.7	10.5
L5	√	√	183	57641463	20.9	9.8
L6	√	√	184	57641460	20.8	9.2
L8e (L7ae)	√	√	125	57641246	13.7	5.2
L11	√	√	165	57641353	17.6	5.4
L12			106	57641350	10.8	3.9
L13	√	√	142	57641436	16.3	10.3
L14	√	√	141	57641466	15.2	11.5
L14e	√	√	83	57641448	8.9	9.9
L15	√	√	148	57641454	16.5	10.2
L15e	√	√	194	57641389	22.6	11.2
L16 (L10e)	√	√	182	57641481	21.1	10.3
L18	√	√	201	57641457	22.9	6.5
L18e	√	√	121	57641437	13.8	10.4
L19e	√	√	150	57641458	17.6	10.9
L20e (LX)	√		77	57641257	9.2	9.0
L21e	√	√	98	57640837	11.3	11.5
L22	√	√	156	57641472	17.8	10.7
L23	√	√	86	57641475	9.9	9.7
L24	√	√	121	57641465	14.2	10.0
L24e	√	√	67	57641244	8.1	10.3
L29	√	√	66	57641470	7.9	10.3
L30	√	√	155	57641455	17.8	10.2
L30e[b]	√	√	102	57641015	10.9	8.9
L31e	√	√	90	57641255	10.3	10.6
L32e	√	√	126	57641459	14.7	11.5
L34e	√		90	57641450	10.5	11.7
L33e (L35ae)	√	√	86	57640904	9.5	10.9
L43e (L37ae)	√	√	86	57640550	9.2	11.6
L37e	√		63	57640910	7.5	11.8
L39e	√		51	57641254	6.2	12.6
L40e	√	√	51	57641430	5.8	10.9
L41e			37	57641850	5.0	12.7
L44e	√		94	57641033	11.1	11.1
P0 (L10p)	√		340	57159675	36.8	4.8

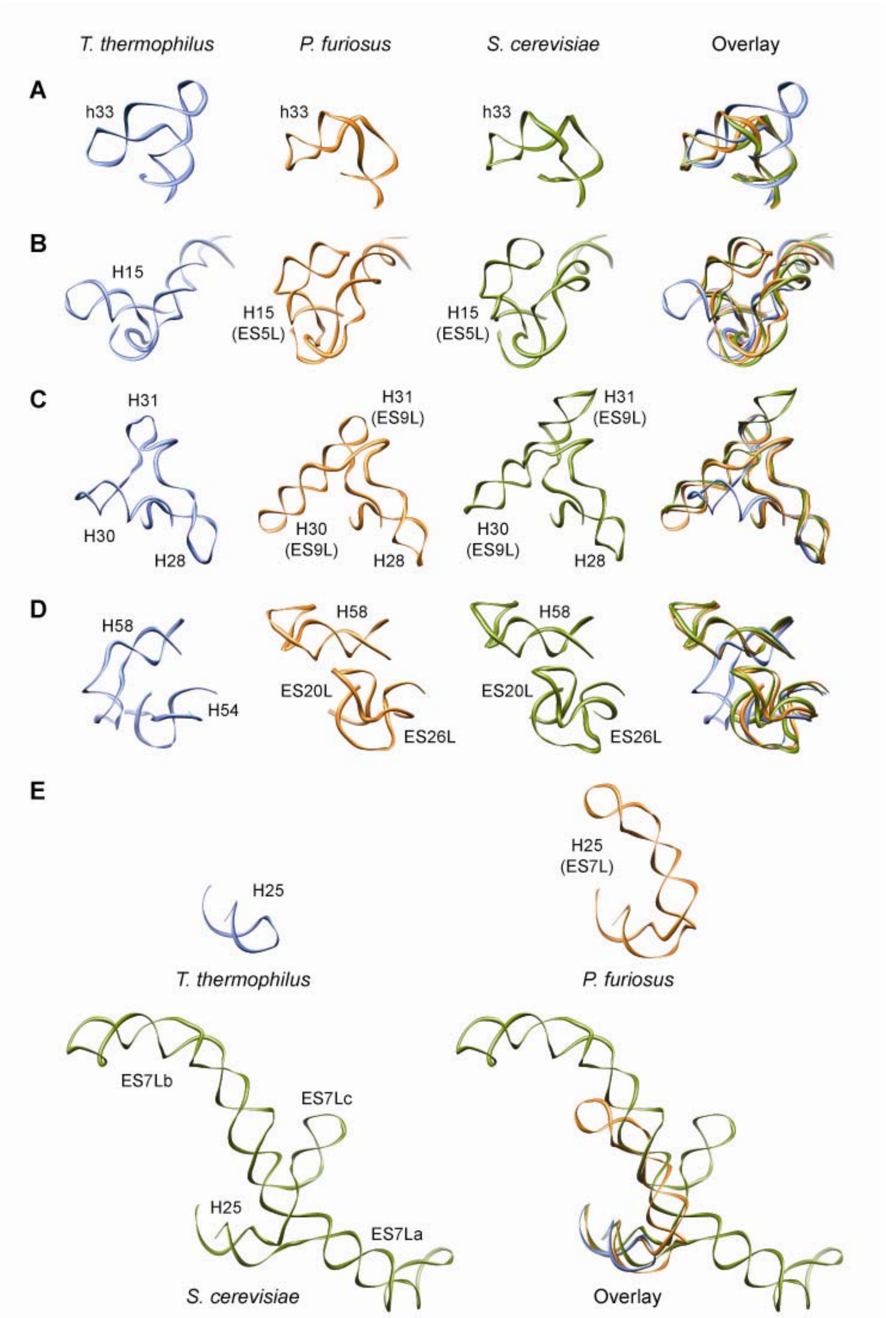
[a] gi refers to the GenInfo identifier for retrieval from NCBI

[b] Identified with only a single peptide

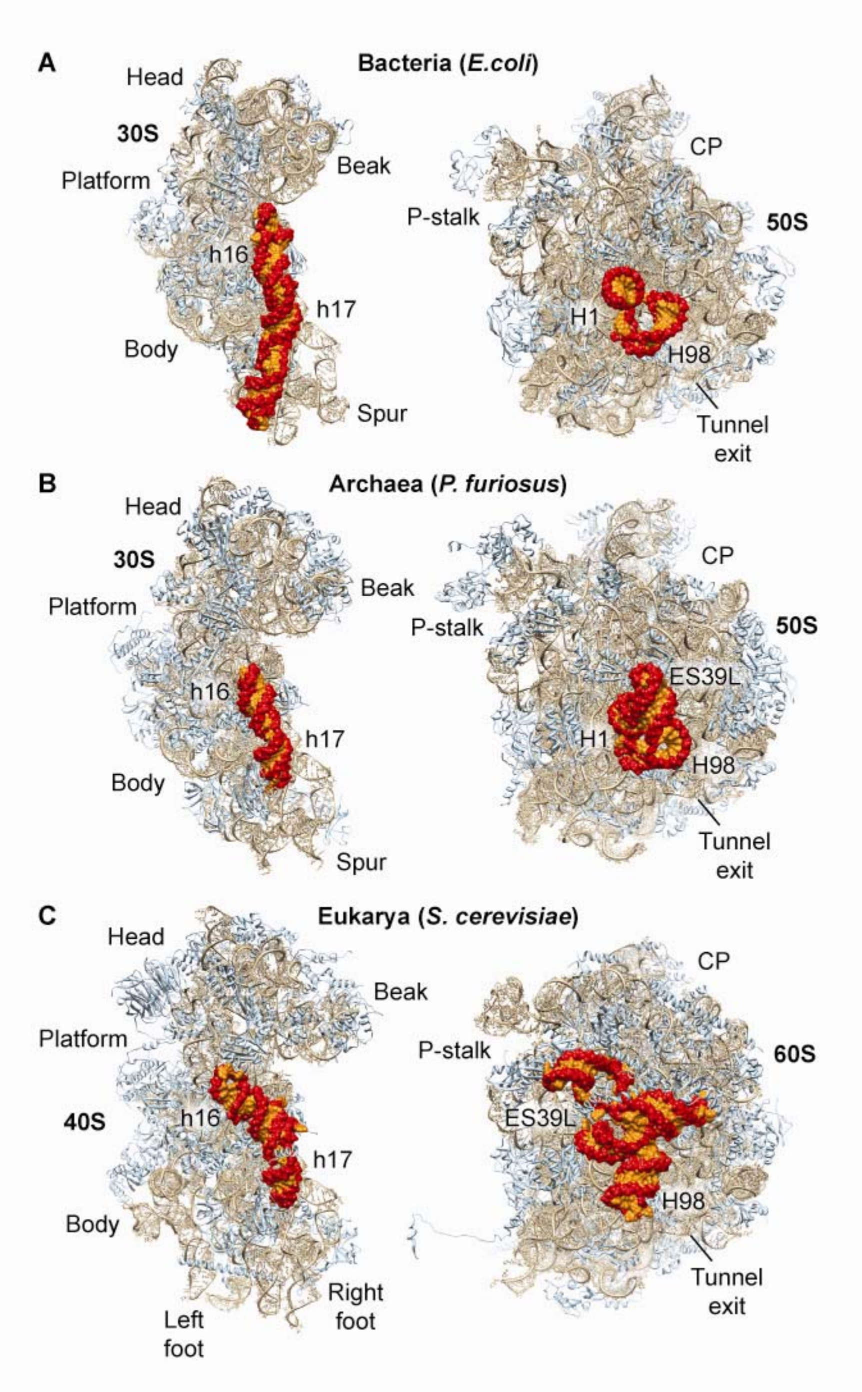
Supplementary Figures



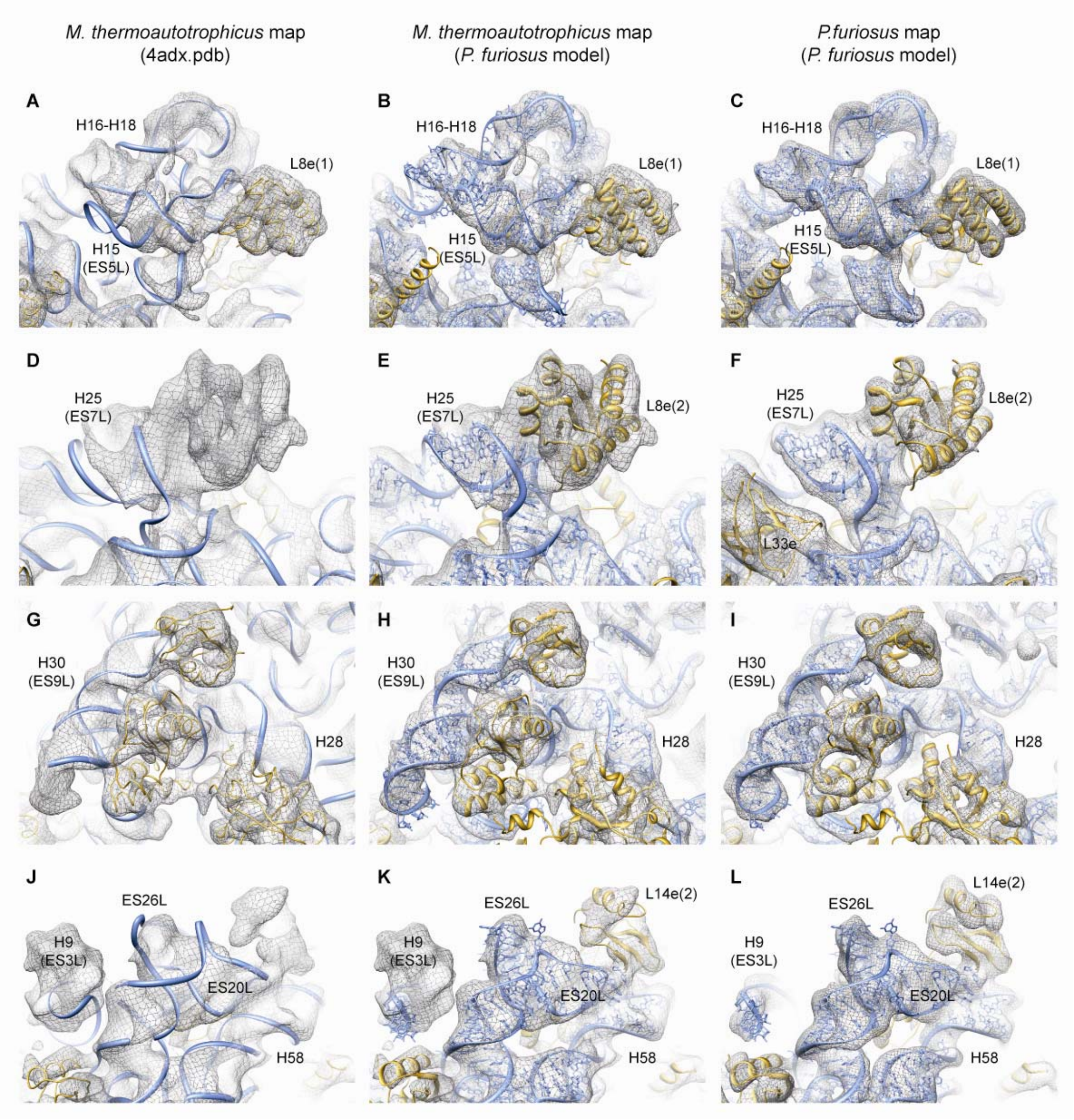
Supplementary Figure 1 Molecular model for rRNA of the *P. furiosus* 70S. (**A-B**) Secondary structure diagrams of the (**A**) 16S and (**B**) 5S and 23S rRNA for *P. furiosus*. Distant parts of the secondary structure drawing are connected by thin lines. (**C-D**) Molecular model of the rRNA for the (**A**) small and (**B**) large subunit of the *P. furiosus* 70S ribosome. Conserved rRNA core is coloured black, eukaryotic-like expansion segments (ES) and variable regions (VR) are coloured orange. ESs and VRs that adopt a unique archaea-specific structure are coloured green.



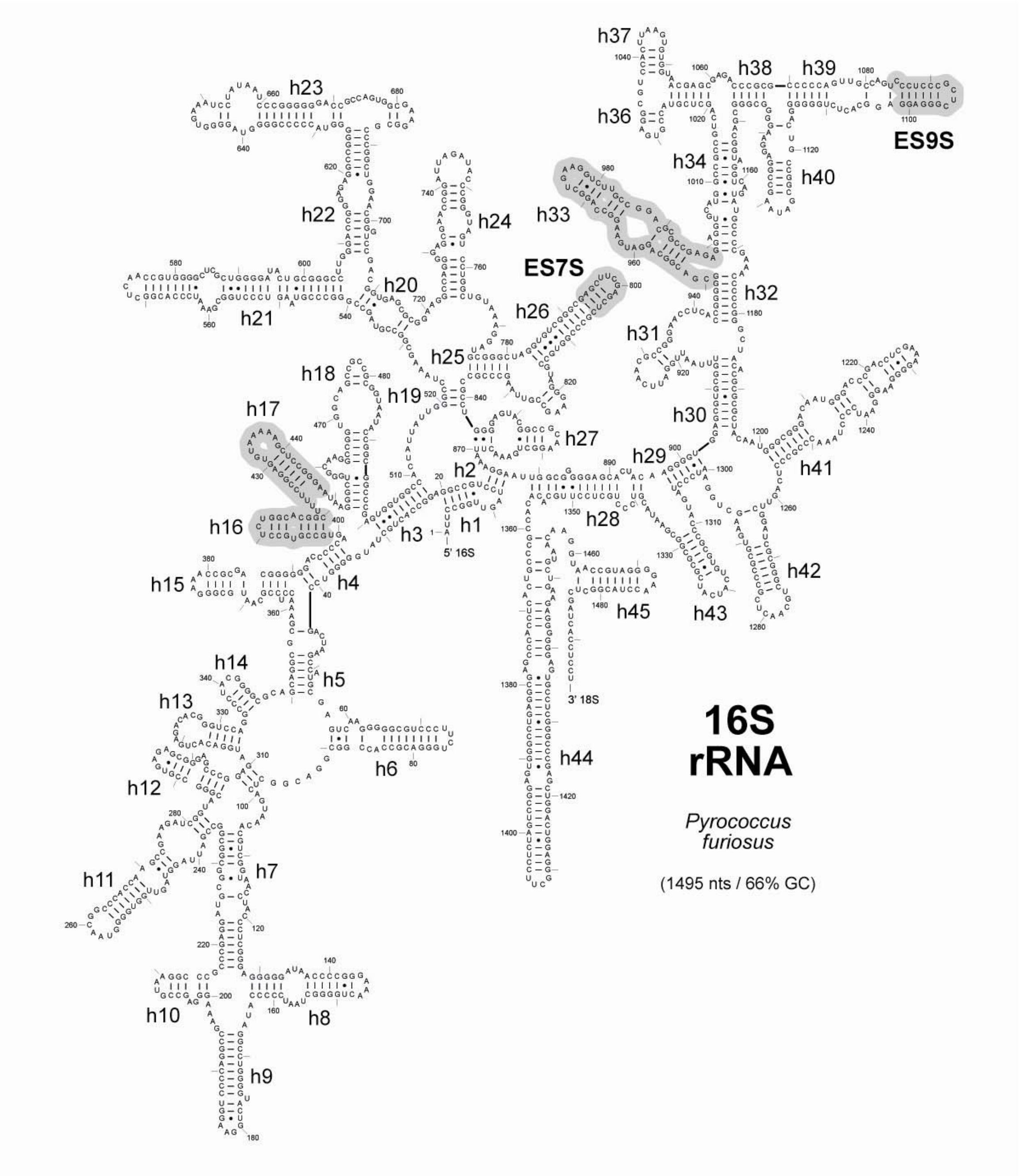
Supplementary Figure 2 Comparison of *P. furiosus* ESs and VRs with bacterial and eukaryotic homologous structures. (**A-E**) Comparison of (A) h33, (B) H15 (ES5L), (C) H28-H31 (ES9L), (D) H54 (ES20L, ES26L), H58 and (E) H25 (ES7L) between the bacteria *T. thermophilus* (blue) (29), the archaea *P. furiosus* (orange), the eukaryote *S. cerevisiae* (green) (7), *S. cerevisiae* ES7L was taken from (32).



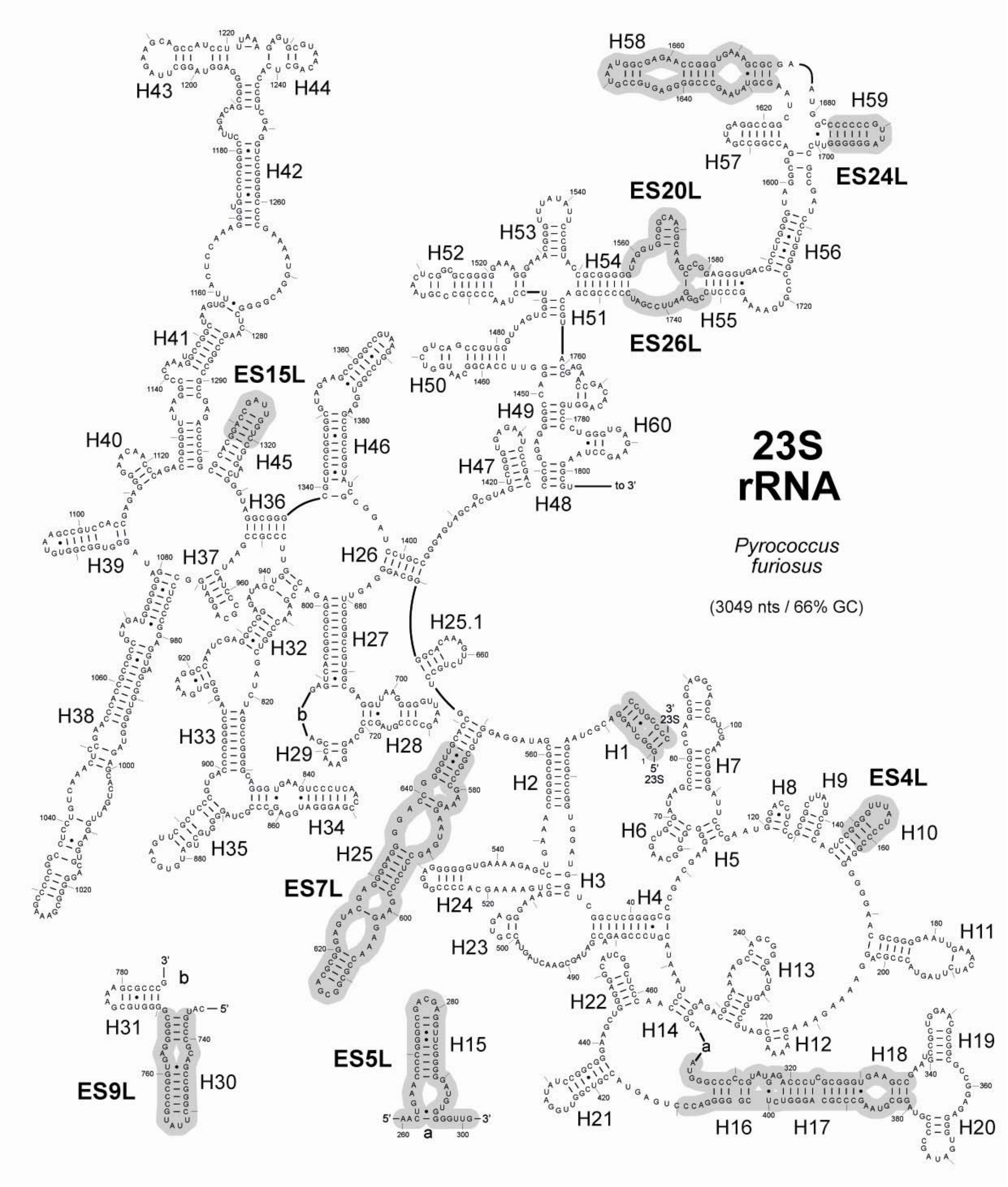
Supplementary Figure 3 Novel rRNA conformations in *P. furiosus* 70S ribosome. (**A-C**) Comparison of h16/17 on the small subunit (left) and H98 (ES39L) on the large subunit (right) between (A) the bacterial (*E. coli*) (30) and (B) archaeal (*P. furiosus*) 70S ribosome, with the (C) eukaryotic (*S. cerevisiae*) 80S ribosome (7). CP marks the central protuberance.



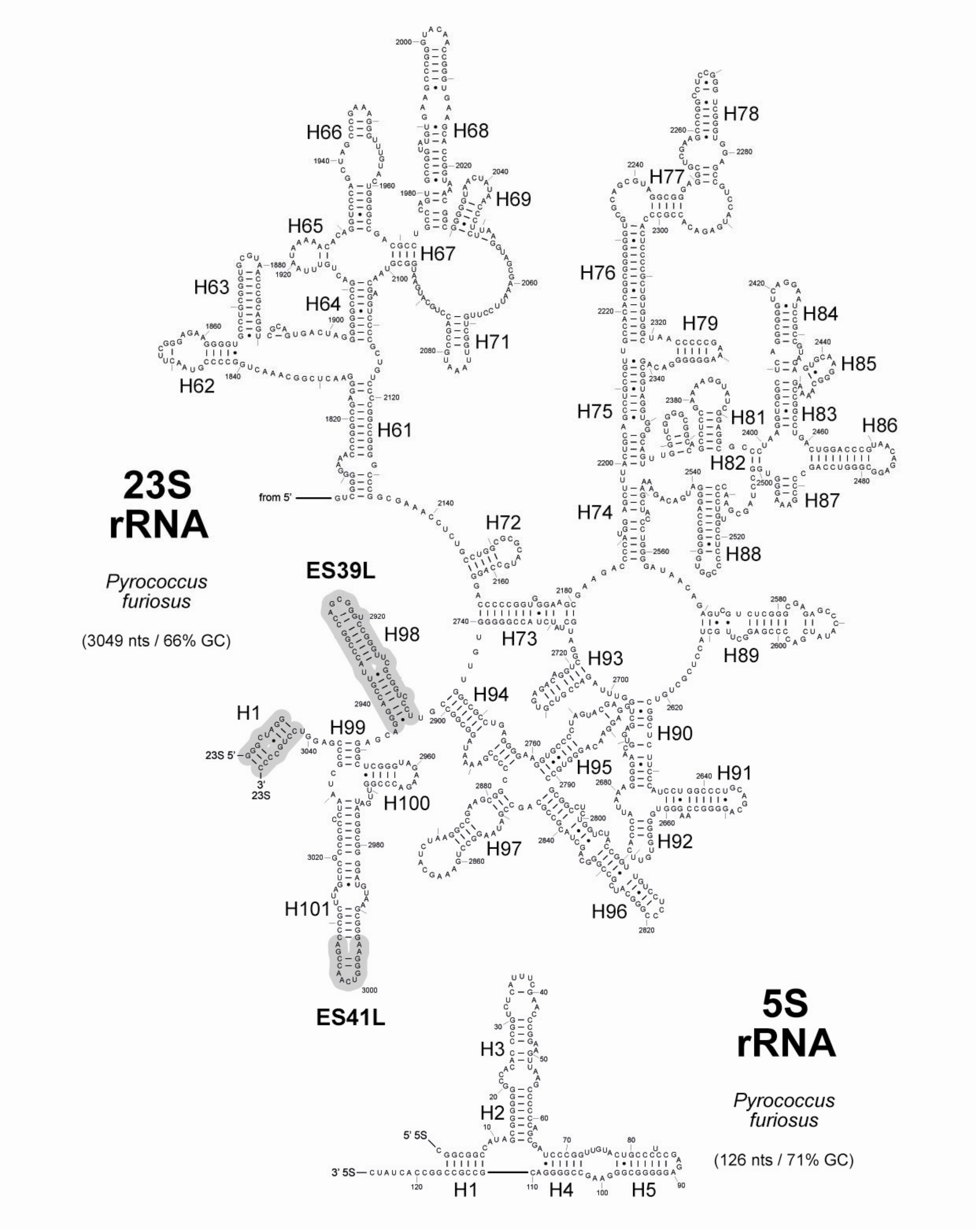
Supplementary Figure 4 Comparison of cryo-EM maps and models for the *M. thermautotrophicus* 50S subunit and the *P. furiosus* 70S ribosome. (**A-L**) Comparison of four distinct regions of the large subunit, namely focusing on (**A-C**) H15 (ES5L), H16-H18 and L8e(1), (**D-F**) H25 (ES7L) and L8e(2), (**G-I**) H28 and ES9L and (**J-L**) ES20L/ES26L and L14e(2). Left panel (A,D,G,J) presents *M. thermautotrophicus* map (EMD-2012) and associated model (PDB4 ADX) (15), middle panel (B,E,H,K) displays *M. thermautotrophicus* map with *P. furiosus* model, and right panel (C,F,I,L) presents the *P. furiosus* map and model. In each case, the cryo-EM density (grey mesh) is shown with rRNA (blue) and proteins (gold).



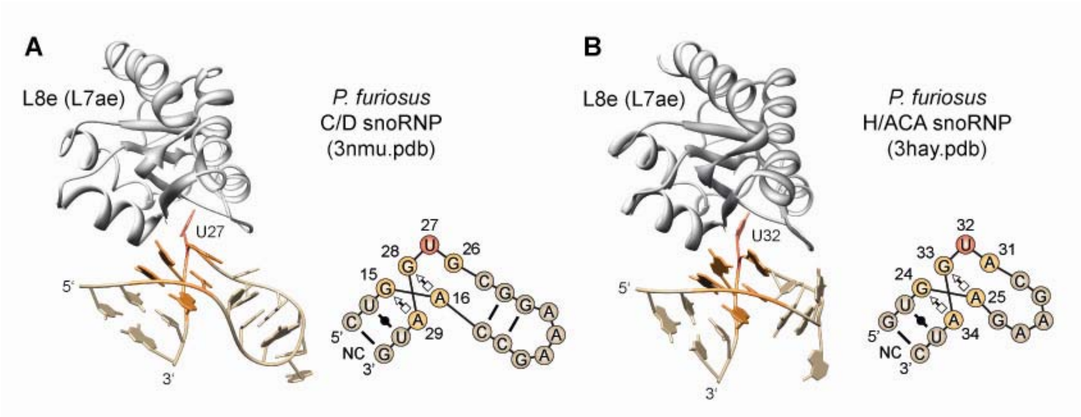
Supplementary Figure 5 Secondary structure diagram of the *P. furiosus* 16S rRNA. The *P. furiosus* 16S rRNA diagram was taken from (51) and adjusted according to the final rRNA model. Grey shaded regions highlight relevant ESs and VRs.



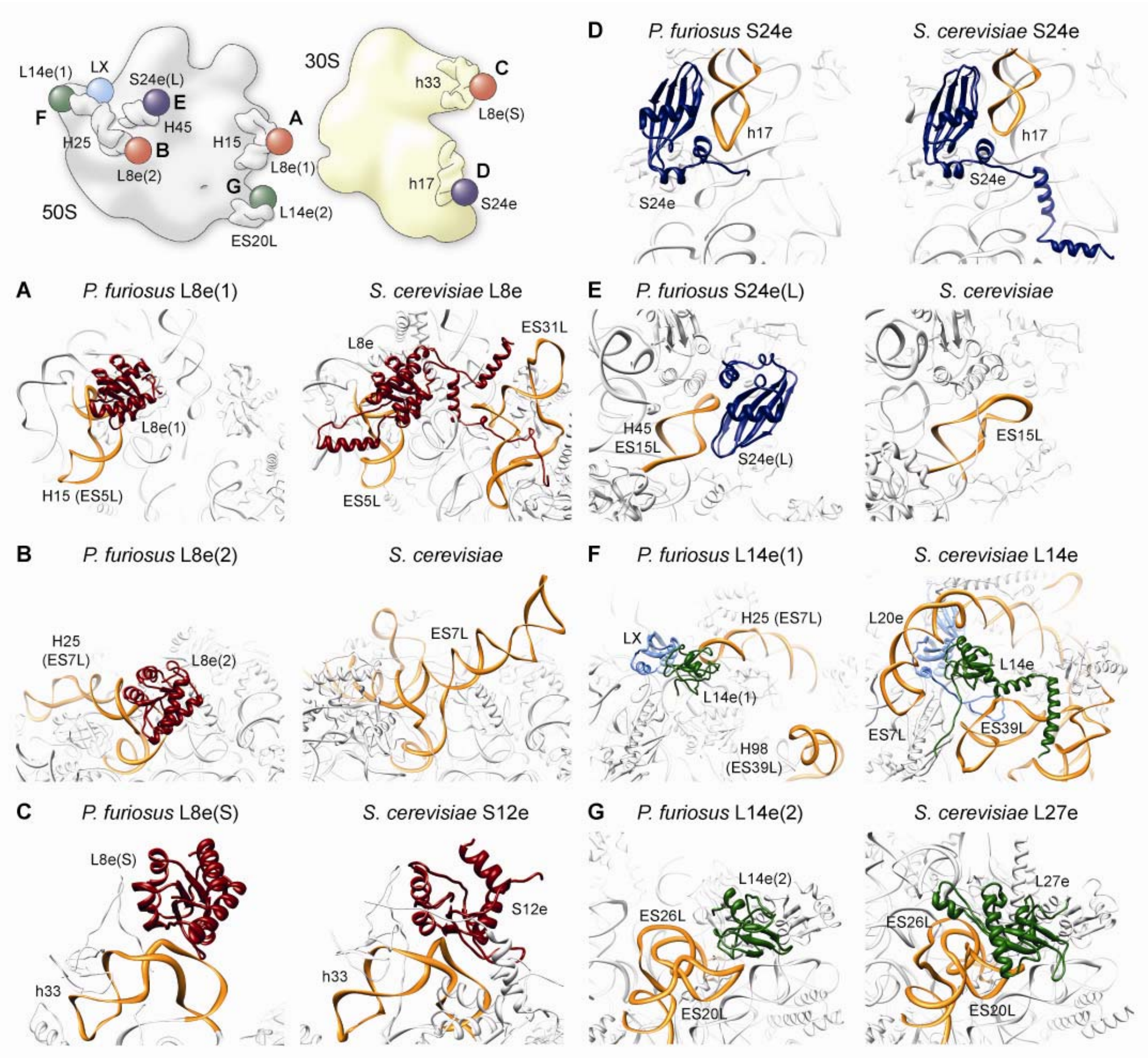
Supplementary Figure 6 Secondary structure diagram of the 5' region of *P. furiosus* 23S rRNA. The *P. furiosus* 23S secondary structure representation is based on the corresponding diagram for *Thermococcus celer* obtained from (51) and was adjusted according to the final rRNA model. Grey shaded regions highlight relevant ESs and VRs.



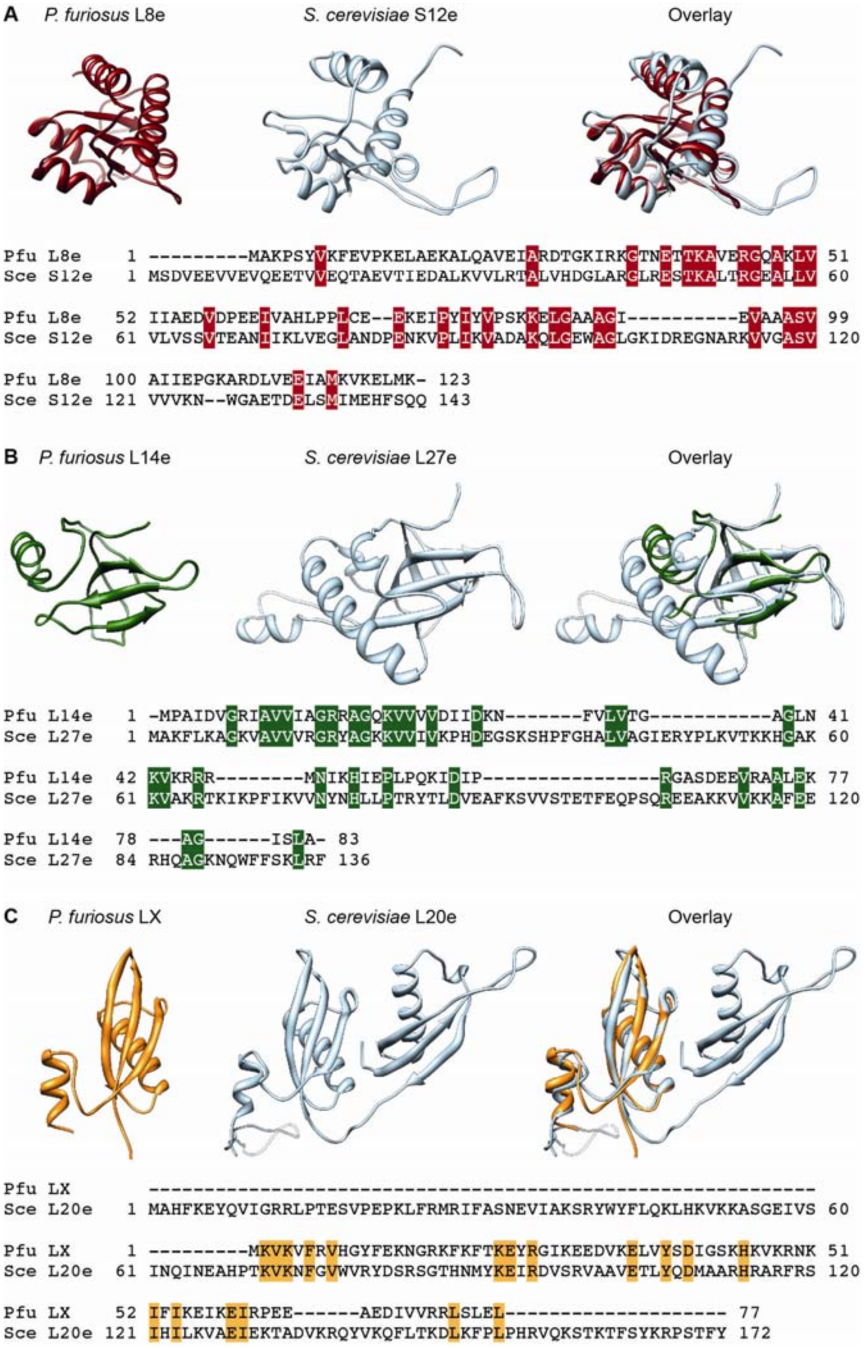
Supplementary Figure 7 Secondary structure diagram of the 3’ region of *P. furiosus* 23S rRNA and complete 5S rRNA. The *P. furiosus* 23S and 5S secondary structure representations are based on the corresponding diagrams for *T. celer* and *Pyrococcus woesei*, respectively obtained from (51) and were adjusted according to the final rRNA model. Grey shaded regions highlight relevant ESs and VRs.



Supplementary Figure 8 *P. furiosus* L8e (L7ae) interaction with kink-turn motifs in C/D and H/ACA snoRNPs. (**A-B**) Interaction of L8e (grey) with bulged out uridine nucleotide (red) within the internal loop of kink-turn motives found in (A) C/D snoRNP (PDB 3NMU) (52) and (B) H/ACA snoRNPs (PDB 2HVY) (42). Insets show secondary structure diagrams of the respective C/D and H/ACA snoRNP kink-turn motives. NC indicates the non-canonical stem.



Supplementary Figure 9 Binding positions of *P. furiosus* L8e, L14e and S24e and respective regions on the eukaryotic ribosome. Comparison of binding positions of (A-C) L8e (red), (D-E) S24e (blue) and (F-G) L14e (green) between *P. furiosus* 70S (left panels) and *S. cerevisiae* 80S (right panels) (7) (Note: ES7L was taken from (32)) ribosomes. Relevant rRNA sections shown as orange ribbon, all other ribosomal components are coloured white. Inset shows a schematic view of the location of L8e (red), L14e (green), LX (light blue) and S24e (blue) on the small (yellow) and large (grey) subunit.

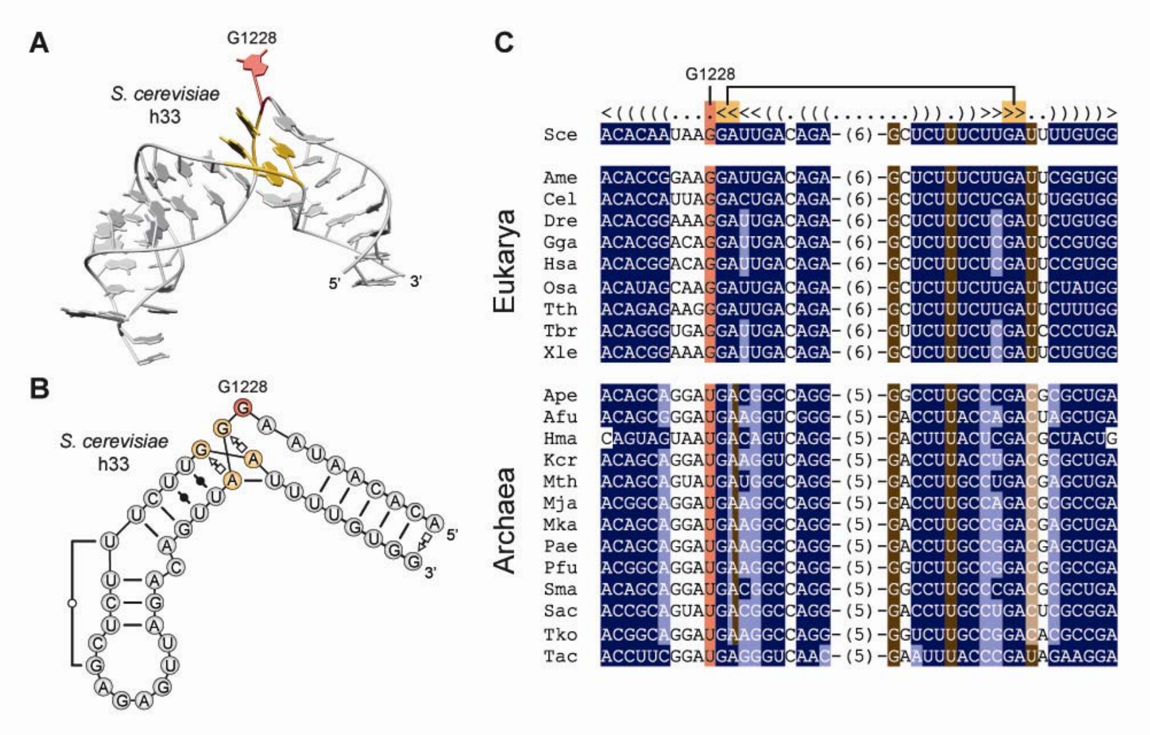


```
Pfu L8e    1 ---------MAKPSYVKFEVPKELAEKALQAVEIARDTGKIRKGTNETTKAVERGQAKLV 51
Sce S12e   1 MSDVEEVVEVQEETVVEQTAEVTIEDALKVVLRTALVHDGLARGLRESTKALTRGEALLV 60

Pfu L8e   52 IIAEDVDPEEIVAHLPPLCE--EKEIPYIYVPSKKELGAAAGI----------EVAAASV 99
Sce S12e  61 VLVSSVTEANIIKLVEGLANDPENKVPLIKVADAKQLGEWAGLGKIDREGNARKVVGASV 120

Pfu L8e  100 AIIEPGKARDLVEEIAMKVKELMK- 123
Sce S12e 121 VVVKN--WGAETDELSMIMEHFSQQ 143
```

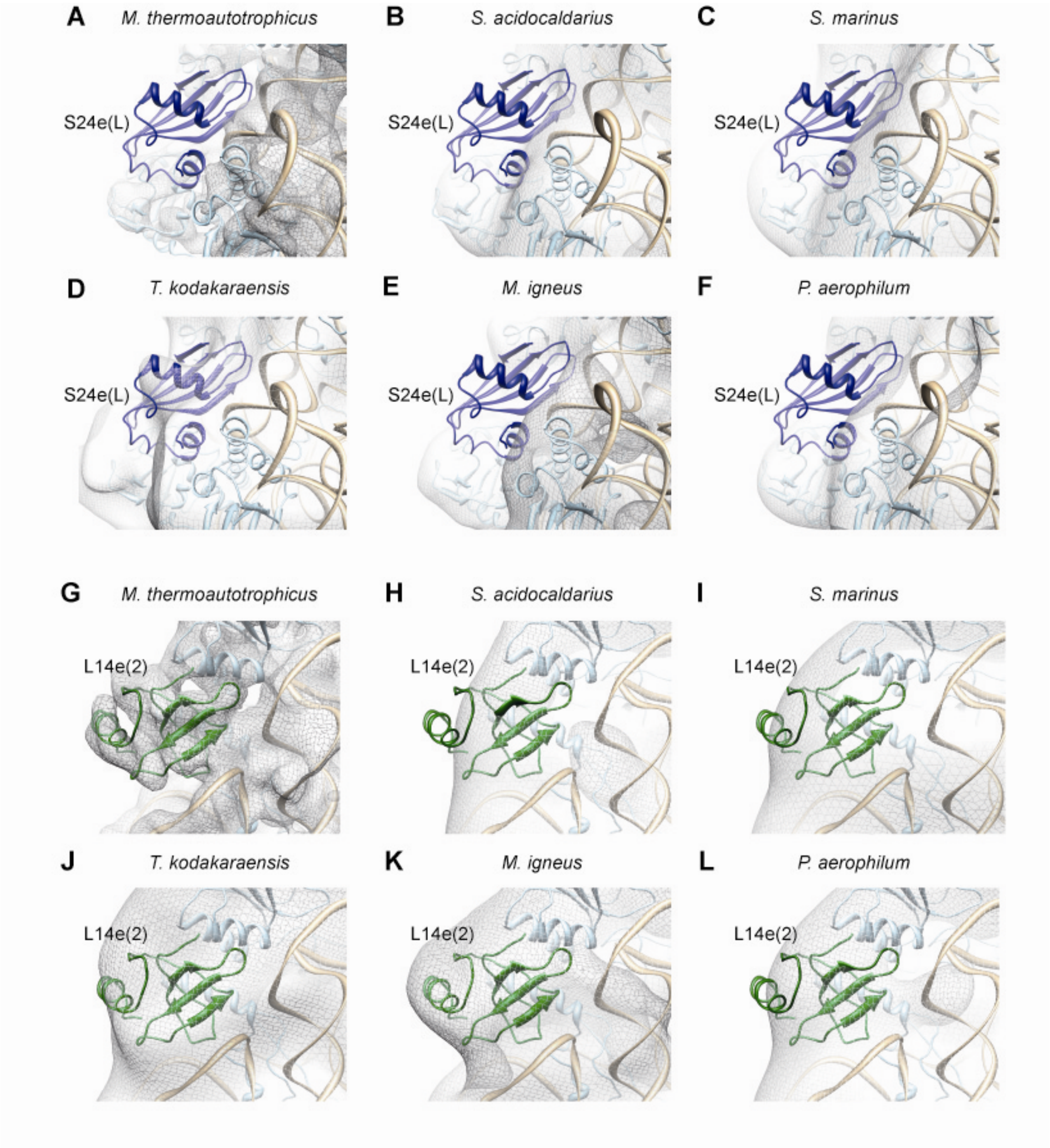
```
Pfu L14e   1 -MPAIDVGRIAVVIAGRRAGQKVVVVDIIDKN-------FVLVTG-----------AGLN 41
Sce L27e   1 MAKFLKAGKVAVVVRGRYAGKKVVIVKPHDEGSKSHPFGHALVAGIERYPLKVTKKHGAK 60

Pfu L14e  42 KVKRRR--------MNIKHIEPLPQKIDIP----------------RGASDEEVRAALEK 77
Sce L27e  61 KVAKRTKIKPFIKVVNYNHLLPTRYTLDVEAFKSVVSTETFEQPSQREEAKKVVKKAFEE 120

Pfu L14e  78 ---AG------ISLA- 83
Sce L27e  84 RHQAGKNQWFFSKLRF 136
```

```
Pfu LX         ------------------------------------------------------------
Sce L20e   1 MAHFKEYQVIGRRLPTESVPEPKLFRMRIFASNEVIAKSRYWYFLQKLHKVKKASGEIVS 60

Pfu LX     1 ---------MKVKVFRVHGYFEKNGRKFKFTKEYRGIKEEDVKELVYSDIGSKHKVKRNK 51
Sce L20e  61 INQINEAHPTKVKNFGVWVRYDSRSGTHNMYKEIRDVSRVAAVETLYQDMAARHRARFRS 120

Pfu LX    52 IFIKEIKEIRPEE------AEDIVVRRLSLEL-------------------- 77
Sce L20e 121 IHILKVAEIEKTADVKRQYVKQFLTKDLKFPLPHRVQKSTKTFSYKRPSTFY 172
```

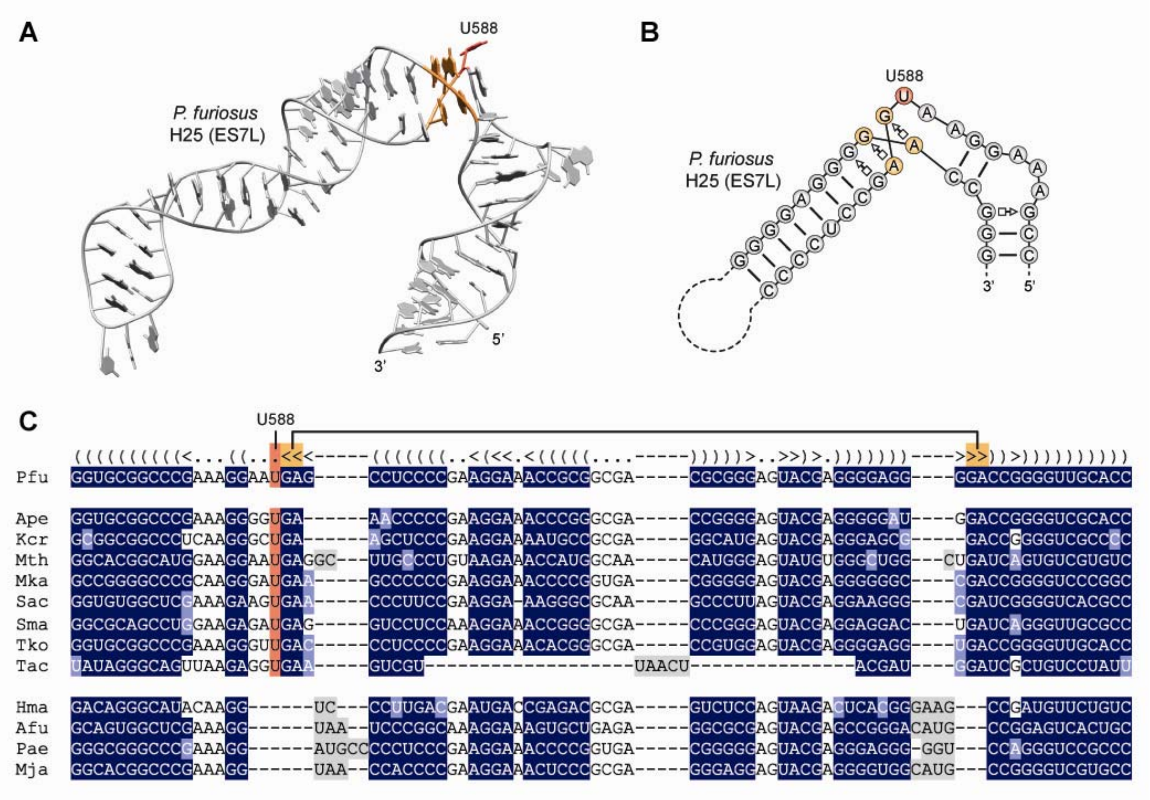
Supplementary Figure 10 Structure and sequence alignments of L8e/L14e/LX with S12e/L27e/L20e. (**A-C**) Superimpositions of the structures of (A) *P. furiosus* L8e with *S. cerevisiae* S12e (7), (B) *P. furiosus* L14e with *S. cerevisiae* L27e (7), and (C) *P. furiosus* LX with *S. cerevisiae* L20e (7). The corresponding sequence alignments shown below were generated using Clustal W (34). Conserved residues are highlighted.



Supplementary Figure 11 Phylogenic distribution of the KT-33 in archaea and eukaryotes. (**A**) Structure of KT-33 with bulged guanine nucleotide (red) from the *S. cerevisiae* ribosome (7). (**B**) Secondary structure diagram of KT-33 from (A). (**C**) Structural alignment of KT-33 showing conservation of guanine in eukaryotes and uridine in archaea (equivalent to position G1228 in *S. cerevisiae* shaded red). The two tandem sheared A/G base pairs characteristic for the KT motif (indicated in orange) are conserved in Archaea and Eukarya. The alignment was generated in S2S (27) using the yeast (Sce) KT-33 structure as template. Colors highlight structural conservation according to the Leontis-Westhof classification (53) implemented in S2S (27), with dark colors indicating isosteric secondary (blue) or tertiary (brown) interactions. Light colors indicate interactions that are geometrically possible, but not necessarily isosteric with the template base-pairs. Nucleotides highlighted in two colors are simultaneously involved in secondary and tertiary interactions. Abbreviations: *Apis mellifera* (Ame), *Caenorhabditis elegans* (Cel), *Danio rerio* (Dre), *Gallus gallus* (Gga), *Homo sapiens* (Has), *Oryza sativa* (Osa), *Saccharomyces cerevisiae* (Sce), *Tetrahymena thermophila* (Tth), *Trypanosoma brucei* (Tbr), *Xenopus laevis* (Xle), *Aeropyrum pernix* (Ape), *Archaeoglobus fulgidus* (Afu), *Haloarcula marismortui* (Hma), *Korarchaeum cryptofilum* (Kcr), *Methanobacterium thermoautotrophicus* (Mth), *Methanococcus jannaschii* (Mja), *Methanopyrus kandleri* (Mka), *Pyrobaculum aerophilum* (Pae), *Pyrococcus furiosus* (Pfu), *Staphylothermus marinus* (Sma), *Sulfolobus acidocaldarius* (Sac), *Thermococcus kodakaraensis* (Tko), *Thermoplasma acidophilum* (Tac).

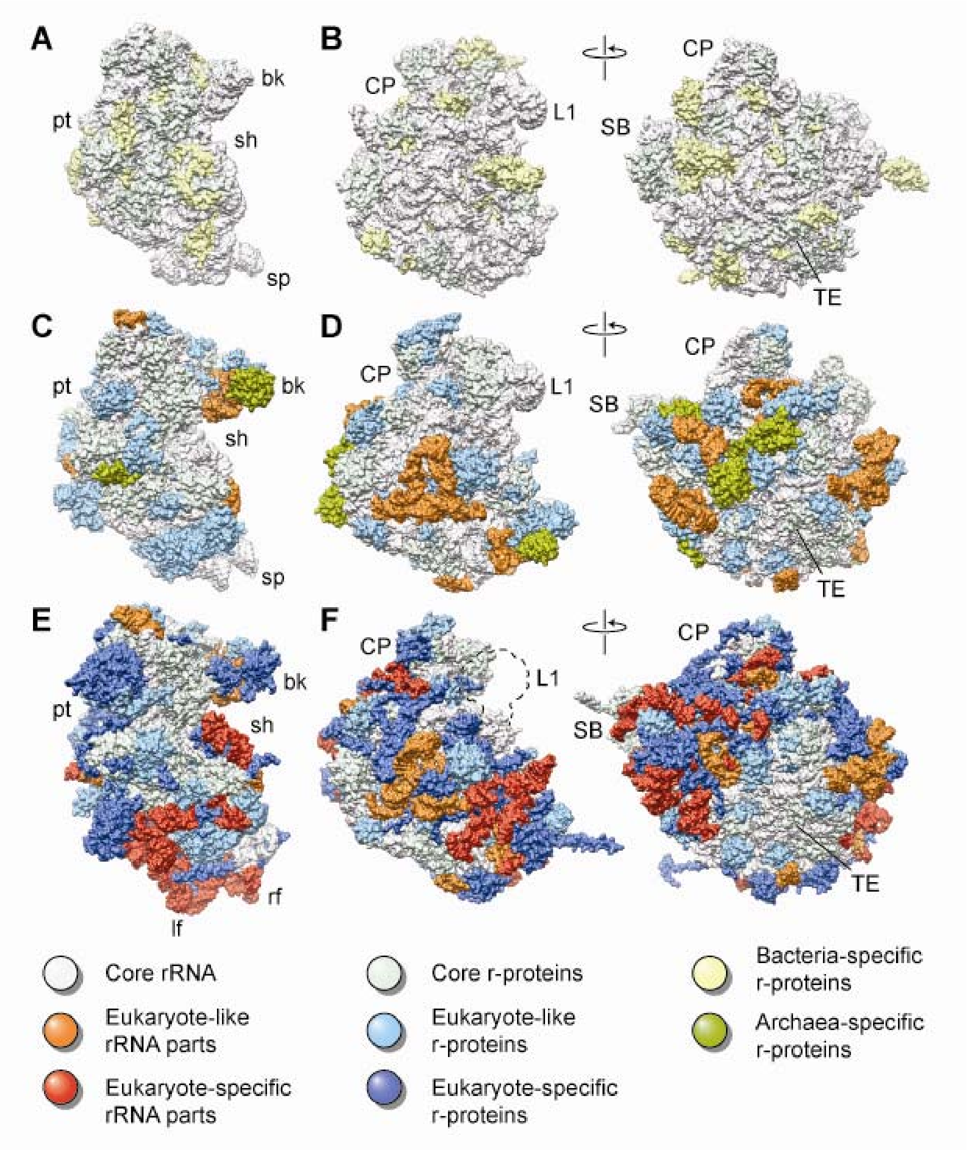


Supplementary Figure 12 Distribution of S24e(L) and L14e(2) binding sites in archaeal ribosomes. (**A-L**) Fit of molecular models of (**A-F**) S24e (blue) and (**G-H**) L14e (green) into the cryo-EM density (mesh) of the (**A,G**) *M. thermautotrophicus* 50S (EMD-2012) (15), (**B,H**) *S. acidocaldarius* 50S (EMD-1797) (21), (**C,I**) *S. marinus* 50S, 50S subunit of the (**D,J**) *T. kodakaraensis* and (**E,K**) *M. igneus* 70S ribosomes, and the (**F,L**) *P. aerophilum* 50S (EMD-1796) (21). Ribosomal RNA is shown in tan with additional r-proteins coloured light blue.

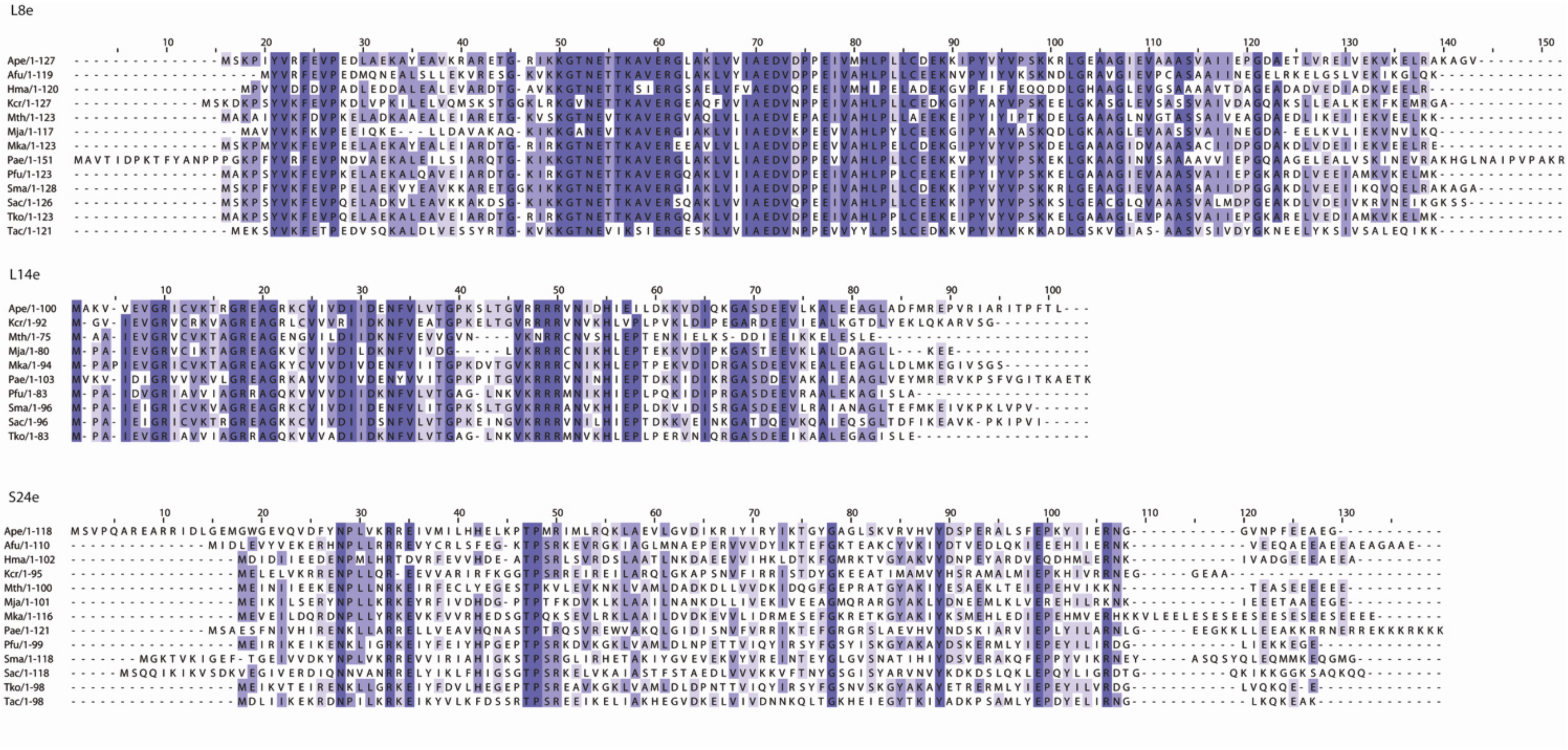


```
    ((((((((((<...((...<<<-----(((((((..<(<<.<((((((....-----))))))>..>>)>.))))))))---->>>))>))))))))))
Pfu GGUGCGGCCCGAAAGGAAUGAG-----CCUCCCCGAAGGAAACCGCGGCGA-----CGCGGGAGUACGAGGGGAGG----GGACCGGGGUUGCACC

Ape GGUGCGGCCCGAAAGGGGUGA------AACCCCCGAAGGAAACCCGGGCGA-----CCGGGGAGUACGAGGGGGAU----GGACCGGGGUCGCACC
Kcr GCGGCGGCCCUCAAGGGCUGA------AGCUCCCGAAGGAAAAUGCCGCGA-----GGCAUGAGUACGAGGGAGCG-----GACCGGGGUCGCCCC
Mth GGCACGGCAUGGAAGGAAUGAGGC---UUGCCCUGUAAGAAACCAUGGCAA-----CAUGGGAGUAUGUGGGCUGG---CUGAUCAGUGUCGUGUC
Mka GCCGGGGCCCGCAAGGGAUGAA-----GCCCCCCGAAGGAAACCCCGGUGA-----CGGGGGAGUACGAGGGGGGC----CGACCGGGGUCCCGGC
Sac GGUGUGGCUCGAAAGAAGUGAA-----CCCUUCCGAAGGA-AAGGGCGCAA-----GCCCUUAGUACGAGGAAGGG----CGAUCGGGGUCACGCC
Sma GGCGCAGCCUGGAAGAGAUGAG-----GUCCUCCAAAGGAAACCGGGGCGA-----CCCGGGAGUACGAGGAGGAC----UGAUCAGGGUUGCGCC
Tko GGUGCGGCCCGAAAGGGUUGAC-----CCUCCCCGAAGGAAACACGGGCGA-----CCGUGGAGUACGAGGGGAGG----UGACCGGGGUUGCACC
Tac UAUAGGGCAGUUAAGAGGUGAA-----GUCGU-----------------UAACU----------------ACGAU----GGAUCGCUGUCCUAUU

Hma GACAGGGCAUACAAGG------UC---CCUUGACGAAUGACCGAGACGCGA-----GUCUCCAGUAAGACUCACGGGAAG---CCGAUGUUCUGUC
Afu GCAGUGGCUCGAAAGG------UAA--UCCCGGCAAAGGAAAGUGCUGAGA-----GGCGCCAGUACGAGCCGGGACAUG---CCGGAGUCACUGC
Pae GGGCGGCCCGAAAGG------AUGCCCCCUCCCGAAGGAAACCCCGGUGA-----CGGGGGAGUACGAGGGAGGG-GGU---CCAGGGUCCGCCC
Mja GGCACGGCCCGAAAGG------UAA--CCACCCCGAAGGAAACUCCCGCGA-----GGGAGGAGUACGAGGGGUGGCAUG---CCGGGGUCGUGCC
```

Supplementary Figure 13 Phylogenic distribution of the KT-25 in archaea. (**A**) Structure of KT-25 with bulged uridine nucleotide (red) from the *P. furiosus* ribosome. (**B**) Secondary structure diagram of KT-25 from (A). (**C**) Structural alignment of KT-25 showing conservation of uridine (shaded red, equivalent to position U588 in *P. furiosus*) and KT motif characteristic tandem sheared A/G base pairs (orange). The two KT features can be found in many archaea, but are absent in *Haloarcula marismortui* (Hma), *Archaeoglobus fulgidus* (Afu), *Pyrobaculum aerophilum* (Pae) and *Methanococcus jannaschii* (Mja). The alignment was generated in S2S (27) using the *P. furiosus* (Pfu) KT-25 structure as template. Colors highlight structural conservation according to the Leontis-Westhof classification (53) implemented in S2S (27), with dark colors indicating isosteric secondary (blue) interactions. Non-alignable sequences are shaded grey. Abbreviations used as in Supplementary Figure 11.



Supplementary Figure 14 Intermediate complexity of archaeal ribosomes compared to bacteria and eukaryotes. (**A-F**) Comparison of the structures of the (**A-B**) bacterial (29), (**C-D**) archaeal and (**E,F**) eukaryotic ribosomes (7). Core rRNA (white) and r-proteins (pale green) are conserved in bacteria, archaea and eukaryotes. The bacterial (**A**) small and (**B**) large subunit are shown with bacterial-specific r-proteins coloured yellow, whereas for the archaeal (**C**) small and (**D**) large subunit, the archaeal-specific r-proteins are coloured green, the eukaryotic-like r-proteins and rRNA parts are coloured light blue and orange, respectively. In the eukaryotic 80S ribosome, the eukaryotic-specific r-proteins and rRNA parts are coloured blue and red, respectively.



Supplementary Figure 15 Sequence alignments of archaeal r-proteins L8e, L14e and S24e. Increasing conservation is shown with darker shading of the amino acids (<40%, white, >40%, >60% to >80%). Abbreviations used as in Supplementary Figure 11.