doi:10.1038/nature12779

Supplementary Table 1: Gene presence and absence data presented in Figure 2 of the main text.

When considering patterns of gene sharing between eukaryotes and the Archaea, at least three classes of genes are evident. The first comprises cases in which orthologues of the eukaryotic gene are conserved across the Archaea, and sometimes also in Bacteria; these include among their number the core gene set used in analyses upon which the tree in Figure 2 is based. The second class includes eukaryotic genes such as actin[1], tubulin[2] and proteins of the ubiquitin modification system[3]. These genes may form large gene families, for example the tubulin/FtsZ gene family, and may have detectable dispersed paralogues among prokaryotes generally, but for which putative orthologues[2] to eukaryotic versions have so far been detected only among the TACK Archaea. The third class includes genes that have so far been found only in eukaryotes and the TACK Archaea; these include several genes involved in transcription and translation, as detailed below.

Gene	Representative archaeal species	NCBI Accession (GI)	Reference	Basis for orthology inference
Actin (Crenactin)	*Thermofilum pendens*	119719444	[1]	Tree
Tubulin (artubulin)	*Nitrosoarchaeum limnia*	494643832	[2]	Tree
Ubiquitin system (Ub, E1, E2, RING-finger containing Ub ligase)	*Caldiarchaeum subterraneum*	343485671, 343485673, 343485672, 343485674	[3]	Sequence similarity, operon structure
Elongation factor Elf1	*Sulfolobus solfataricus*	13813400	[4,5]	Not detected outside group
RNA polymerase RpoG/Rpb8	*Thermofilum pendens*	119719267	[6]	Not detected outside group
Ribosomal protein S25e	*Thermofilum pendens*	119719924	[4]	Not detected outside group
Ribosomal protein S30e	*Thermofilum pendens*	119719279	[4]	Not detected outside group
Ribosomal protein L13e	*Thermofilum pendens*	119719644	[4]	Not detected outside group
Ribosomal protein L38e	*Aeropyrum pernix*	499163706	[4]	Not detected

				outside group

Supplementary Table 2: An amino acid insertion in the elongation factor 1-alpha genes of eukaryotes and certain Archaea. An insertion[7] is present in the Thaumarchaeota, Aigarchaeota and Crenarchaeota highlighted in Figure 2; this insertion is apparently missing from *Korarchaeum cryptophilum*. The coordinates provided refer to the portion of the insertion that is readily alignable among the compared sequences.

Species	**Elongation factor 1-alpha accession**	**Amino acid coordinates**
Caldiarchaeum subterraneum	315425766	120-127
Cenarchaeum symbiosum	118575602	125-132
Nitrosopumilus maritimus	161528542	121-128
Nitrosoarchaeum limnia	494643908	121-129
Thermofilum pendens	119719557	121-128
Pyrobaculum aerophilum	18313751	131-138
Caldivirga maquilingensis	159042306	130-137
Sulfolobus solfataricus	15897164	120-127
Ignicoccus hospitalis	156937938	122-129
Staphylothermus marinus	126465710	121-128
Hyperthermus butylicus	124028427	121-128
Aeropyrum pernix	14601666	120-127

1 Ettema, T. J., Lindas, A. C. & Bernander, R. An actin-based cytoskeleton in archaea. *Mol Microbiol* **80**, 1052-1061 (2011).
2 Yutin, N. & Koonin, E. V. Archaeal origin of tubulin. *Biol Direct* **7**, 10 (2012).
3 Nunoura, T. *et al.* Insights into the evolution of Archaea and eukaryotic protein modifier systems revealed by the genome of a novel archaeal group. *Nucleic Acids Res* **39**, 3204-3223 (2011).

4 Guy, L. & Ettema, T. J. The archaeal 'TACK' superphylum and the origin of eukaryotes. *Trends Microbiol* **19**, 580-587 (2011).
5 Daniels, J. P., Kelly, S., Wickstead, B. & Gull, K. Identification of a crenarchaeal orthologue of Elf1: implications for chromatin and transcription in Archaea. *Biol Direct* **4**, 24 (2009).
6 Koonin, E. V., Makarova, K. S. & Elkins, J. G. Orthologs of the small RPB8 subunit of the eukaryotic RNA polymerases are conserved in hyperthermophilic Crenarchaeota and "Korarchaeota". *Biol Direct* **2**, 38 (2007).
7 Rivera, M. C. & Lake, J. A. Evidence that eukaryotes and eocyte prokaryotes are immediate relatives. Science **257**, 74-76 (1992).