doi:10.1038/nature14447

Supplementary Discussions

1. Redundancy of Lokiarchaeum composite genome

Single-marker phylogenies revealed that approximately six highly similar strains from the same DSAG clade were present in LCGC14AMP and LCGC14. Such a high level of micro-diversity prevented us to assemble a complete, non-redundant Lokiarchaeum genome. However, micro-diversity could significantly be reduced by using very stringent filters, based on read coverage (see Suppl. Methods). The final Lokiarchaeum composite genome has an estimated genome redundancy of 1.4 fold. Out of 162 assumed single copy genes, 71 are duplicated and three are triplicated. In the case of triplicated genes, the third copy is always a partial match and shows 100% identity to part of one of the other copies. Two of them are located at contig edges. The median identity between the duplicated single copy genes is 94.94 % and shows the high similarity of the lokiarchaeal strains present in the final bin.

2. Phylogenetic assessment of the Lokiarchaeota-Eukarya affiliation

Phylogenetic inference of deep evolutionary relations are often very hard to resolve as these are potentially influenced by a number of artifacts[1]. This is especially relevant within the framework of the current work, which aims to place eukaryotes in the Tree of Life, which contains several long branches as a result of rapidly evolving lineages and/or poor taxonomic sampling. We have tried to address and minimise the effects of potential artefacts on our phylogenetic inferences in a number of ways. First, we have used methods that have been designed to model sequence heterogeneity[2] by using a Bayesian framework that employs site-heterogeneous mixture models of amino acid replacement[3], which have improved model fit that have been shown to be able to suppress long-branch attraction artifacts (e.g. see references 4,5). Our Bayesian analyses of the concatenated protein datasets of 36 carefully selected single copy marker genes unequivocally supported the phylogenetic affiliation between Lokiarchaeota and eukaryotes with the highest possible support (PP = 1.00; Fig. 2b and Suppl. Fig. S8). While the Lokiarchaeota+eukaryotes support in the ML analyses of the same dataset is still regarded significant (BS = 80), we suspected that this analysis was affected by phylogenetic artefacts that were embedded in the concatenated protein dataset. To investigate this further, a number of ML analyses were performed to assess the robustness of the phylogenetic affiliation between Lokiarchaeota and eukaryotes. Since phylogenetic analyses based solely on ribosomal proteins have been shown to perform poorly in placing deep-rooting lineages due to compositional bias[6,7], an analysis was performed in which the ribosomal protein (RP) markers (21 markers) were analysed separately from the non-ribosomal markers (15 markers). A ML phylogeny solely based on the RP dataset was found to weakly support eukaryotes grouping with Korarachaeota (‘KE’; BS=35; Suppl. Fig. S10), whereas a phylogeny based on non-RP dataset supported, with higher confidence, eukaryotes grouping within Lokiarchaeota (‘LE’; BS=71; Suppl. Fig. S11). These results indicate that the phylogenomic affiliation between Korarchaeota and eukaryotes is probably induced by compositional bias that is embedded in the ribosomal protein sequences. Next, we assessed the effect of amino acid composition biases as follows: sites in the reference alignment were ranked by the amount of their contribution to the overall amino acid bias, and partial alignments were built, adding increasing amounts of either the most or the least diverse sites[8,9]. Maximum likelihood phylogenies were then run and the BS for three critical bipartitions in the tree were assessed: eukaryotes grouping with either Lokiarchaeota (‘LE’), with Korarchaeota (‘KE’), or with all members of the TACK superphylum, including Lokiarchaeota (‘TACKLE’) (Suppl. Fig. S12). This analysis revealed strong support for eukaryotes grouping with TACK (‘TACKLE’) as well as for the Lokiarchaeota-Eukarya affiliation (‘LE’): With the least diverse sites, the overall support for TACKLE reaches 80% whenever 50% of the sites are included. Except when taking 50-60% of the sites, the support for LE was better than for KE, increasing to 80% with all the sites included. With the most diverse sites, the support for TACKLE was >80% and for LE >60%, whenever taking 20% or more of the sites (Suppl. Fig. S12).

We also investigated the effects of the presence or absence of certain taxonomic groups (‘taxonomic sampling’) on the phylogenetic robustness by omitting clades or groups of clades, and inferring ML phylogenies of the remaining datasets (Suppl. Table S5 and Suppl. Fig. S13). These analyses revealed that the Lokiarchaeota-Eukarya affiliation is generally robust with respect to taxonomic sampling: When Bacteria, Thaumarchaeota, Aigarchaeota, Korarchaeota, Lokiarchaeum or SAGs are removed, the support for Lokiarchaeota branching with eukaryotes was >70. When removing DPANN Archaea, the support drops (BS=64). When removing Crenarchaeota, eukaryotes group with Korarchaeota, albeit with poor support (BS=43). Finally, the removal of Korarchaeota resulted in a highly supported cluster comprised of Lokiarchaeaum, Loki2, Loki3 and eukaryotes (BS=100; PP=1; Suppl. Figure S14). Further, removing Lokiarchaeum or Loki2/Loki3 yields different results: in the first case, eukaryotes branch with the remaining of the Lokiarchaeota (BS=86), but in the second case, eukaryotes group with Korarchaeota (BS=80). The latter indicates that reducing the Lokiarchaeota clade to a single lineage induces a long-branch attraction-related artefact that pulls Korarchaeum towards the eukaryotes.

Finally, to compare how well different trees explained the aligned sequence data, Approximately Unbiased (AU) tests[10] were performed on the concatenated alignment of 36 marker proteins. Two alternative topologies were tested against the alignment: (i) the tree resulting from the ML analysis of the 36 concatenated markers present in Loki2/3, showing Lokiarchaeota grouping with Eukaryotes (LE) and (ii) the one resulting from the concatenation of the ribosomal protein genes only, showing Korarchaeota grouping with Lokiarchaeota (KE). The LE tree was shown to explain the data significantly better than the KE tree ($p < 10^{-6}$). Hence, the AU tests supported the Lokiarchaeota-eukaryotes affiliation. We also examined single-marker gene ML trees, but topologies were often inconclusive, with low support values at critical nodes, a phenomenon that has been suggested to be due to the small amount of information contained in single-gene trees[11]. Nevertheless, AU tests were also performed on single-gene alignments: neither tree explained the data significantly better than the other in 30 out of 36 alignments. Out of the six alignments for which one tree explained the data significantly better ($p<0.01$), five (arCOG00412, phenylanalyl-tRNA synthetase beta subunit; arCOG01559, translation elongation factor G; arCOG01762, RNA polymerase B; arCOG04169, SecY; arCOG04256, RNA polymerase A'') supported the Lokiarchaeota-eukaryotes affiliation, and only one (arCOG04113, ribosomal protein L10AE/L16) supported the Korarchaeota-eukaryotes affiliation.

In summary, the phylogenetic analyses we have performed present strong evidence that Lokiarchaeota forms a monophyletic clade with eukaryotes. The sister relation between Korarchaeota and eukaryotes that is observed in a minority of our phylogenetic analyses most likely represents an artefact caused by the fact that Korarchaeota are only represented by a single taxon, represented by a long branch in all trees, with an unusual amino acid bias. The correct phylogenetic position of such taxa is notoriously difficult to resolve.

3. Bacteria-affiliated genes encoded by the Lokiarchaeum genome

The Lokiarchaeum genome was found to contain a relatively high fraction of genes that display highest similarity to genes of bacterial origin (29% of all genes). Although this high fraction seems atypical, similar observations have been made for other archaeal lineages (Fig. 2a), such as the Miscellaneous Crenarchaeal Group (MCG)[12] and Marine Group II archaea[13]. These findings seem in line with recent observations of extensive inter-domain gene exchange between Bacteria and Archaea[14] and are also supportive of the hypothesis that the origin of major archaeal clades coincides with profound gene acquisitions from Bacteria[15]. Importantly, the absence of bacterial marker genes, such as bacterial ribosomal proteins and genes involved in genetic information processing, from the Lokiarchaeum genome bin excludes the possibility that significant part of these genes originate from contaminating bacterial contigs. Rather, while the archaea-related proteins in Lokiarchaeum are assigned to COGs involved in informational processing processes including translation, transcription, replication and posttranslational modification, the lokiarchaeal proteins with highest similarity to bacterial proteins were predominantly assigned to categories related to various aspects of metabolism, signal transduction, defense mechanisms, inorganic ion and lipid transport etc. (Suppl. Figure S15). A sole exception to this pattern involves a lokiarchaeal NAD-dependent DNA ligase (LigA) encoding gene (Suppl. Table S9). This gene, however, seems to have been acquired horizontally by Lokiarchaeum as it is flanked by several genes most closely related to archaeal protein homologs. Furthermore, LigA is also encoded by the *Methanomassilicoccus alvus* genome and was suggested to have been acquired by horizontal gene transfer from Bacteria[16]. Altogether, the high fraction of genes with highest similarity to bacterial genes thus seems to be a genuine feature of the Lokiarchaeum genome.

4. The informational processing machineries of Lokiarchaeum.

Despite of its extensive genetic repertoire concerning putative cytoskeleton, vesicular trafficking and cell division related proteins, the Lokiarchaeum composite genome revealed informational processing machineries most similar to Archaea (Suppl.-Table S9). Yet, the investigation of the ribosomal protein complement of Lokiarchaeum revealed interesting translational features. For instance, except for RPL41e, Lokiarchaeum not only contains all RPs previously found in just a subset of members of the TACK superphylum[17] and shared with eukaryotes (Figure 5; Suppl. Table S7 and S8), but also encodes a putative distant homolog of the eukaryotic specific RPL22, (Suppl. Figure S24). Notably, in Lokiarchaeum, three of these RPs (RPL34E, RPS14 and RPS26E) as well as the Translation initiation factor 1 (eIF-1/SUI1) (arCOG04223) appear to start with the alternative initiation codon AUU, a feature which has not been found in Archaea so far and only used by few Bacteria, including *E. coli,* to initiate transcription of *infC* (encoding IF3) and *pcnB* (encoding *E. coli* poly(A) polymerase)[18]. Two RPs (RPL13 and RPS26) were not part of the final Lokiarchaeum bin, but were present on several LCGC14 contigs that were binned as "Lokiarchaeum" that were not included in the final reassembly due to slightly lower coverage than required by our stringent coverage criteria. Suppl. Table S7 lists the locus tags of these RPs present on contigs binned as Lokiarchaeum and longer than 5kb. We argue that these ribosomal proteins are of lokiarchaeal origin for the following reasons: RPL13 and RPS26 are absent from bacterial genomes and solely present in some members of the TACK superphylum. The only TACK members present in our sample are closely related to the Marine Group I (Thaumarchaeota), which constitute only a minor fraction of the non-amplified LCGC14 metagenome. The RPs encoded by contigs binned as Lokiarchaeum are only distantly related to Thaumarchaeota but close homologs of these are present on several LCGC14 contigs, which is in accordance with the expected micro-diversity of Lokiarchaeota in this sample.

The Lokiarchaeum composite genome encodes all common archaeal RNA polymerase (RPB) subunits (see Table S9, for details and redundancy), yet lacks a homolog of the small RPB8 subunit of the eukaryotic RNA polymerase (RpoG), which so far has only been found in Cren- and Korarchaota (Fig. 5)[19]. Interestingly, the RPB subunits of Lokiarchaeum do not show a clear affiliation with any particular archaeal phylum (top blast hits include both Eury-, Cren- and Thaumarchaeota). In addition, while RPB subunit L does not retrieve significant matches to other sequences in nr, RPB subunit N is most similar to eukaryotic homologs. Altogether, these findings further confirm the suggestion that Lokiarchaeota comprise a distinct archaeal linage of high taxonomic rank.

With its genes encoding DNA polymerase I of the B1-family (PolB1), a putative DNA polymerase of the inactivated B2-family (PolB2) as well as the small and large subunits of DNA polymerase D (DP2), Lokiarchaeum has a complement of replicative DNA polymerases similar to those of Thaum- and Aigarchaeota (see a recent review by Makarova *et al*. for details[20]). While PolB1 is absent from all euryarchaeal genomes, Crenarchaeota lack homologs of the two subunits of DP2. Instead, both Cren- and Euryarchaeota encode Polymerases I of the B3 family (PolB3), which is missing from Thaum- and Aigarchaeal genomes[20] and could neither be identified in Lokiarchaeum.

In contrast, the lokiarchaeal pattern of topoisomerase genes is most similar to Cren- and Euryarchaeaota, rather than to Thaum- and Aigarchaeota. For instance, Lokiarchaeum encodes both the type I topoismerase IA/III (TopoIA) as well as the type II toposimerase IIB (TopoIIB), which introduce transient single- and double-stranded breaks into DNA, respectively[21]. Interestingly, subunit B of TopoIIB was most similar to eukaryotic homologs, while TopoIIB subunit A (encoded by the flanking gene) shares highest similarity with its archaeal counterparts. In addition, we could identify a lokiarchaeal homolog of subunit B of type II topoisomerase IIA (TopoIIA) (Table S9), which is present in some members of the Euryarchaeaota. However, the A subunit of topoisomerase IIA was absent from the final bin (contig break next to TopoIIA, subunit B) but seems to be present in the metagenome. Curiously, Lokiarchaeum did not seem to encode a homolog of type I topoisomerase IB (TopoIB), which is present in some TACK lineages such as Thaum- and Aigarchaeota, as well as in MCGs, and was suggested to share a common ancestry with the eukaryotic TopoIB [22-24]. It remains to be elucidated whether TopoIB has been present in the last common ancestor of all TACK members and was later lost in the cren- kor- and lokiarchaeal lineages (Fig. 5), whether it was acquired by some members of TACK later on, or whether its history involves a more complex evolutionary scenario.

5. A dynamic actin cytoskeleton and a primordial vesicular trafficking machinery in Lokiarchaeum

An in-depth analysis of the Lokiarchaeum composite genome revealed the presence of eukaryotic-like actins as well as small proteins with gelsolin-like domains (Interpro domains IPR007122, IPR007123 and IPR029006; Suppl. Table S10). In eukaryotes, the latter domains are part of actin-binding, capping and modulating proteins such as villin, severin and adseverin, which often contain several gelsolin-domains. These proteins have been proposed to have evolved from small (120-130 amino acid) gelsolin-domain proteins[25], and have so far not been detected in prokaryotic genomes. While most of the lokiarchaeal gelsolin domain proteins retrieve no close homologs in public databases, Lokiarch_14340 and Lokiarch_49620 for instance display significant homology to eukaryotic gelsolins/adseverins (E-value to top Blast hit: 1 e-101, 24% identity) and Lokiarch_42060 was most similar to actin-binding proteins of *Entamoeba* (E-value: 6e-59; 27% identity) after the second Psiblast iteration using an inclusion threshold of 0.001). The presence of these gelsolin-domain proteins in an archaeon opens up the possibility that the archaeal ancestor of eukaryotes already contained a machinery to actively remodel its actin cytoskeleton. Despite of the presence of genes suggestive of a dynamic actin cytoskeleton, we could not identify homologs of eukaryotic or archaeal tubulins in the Lokiarchaeum composite genome, apart from a homolog of the canonical FtsZ proteins (Fig. 5, Suppl. Table S9).

In addition to the cytoskeletal elements mentioned above, Lokiarchaeum contains subunits of all components of the eukaryotic endosomal sorting complexes ESCRT-I, ESCRT-II and ESCRT-III (Extended Data Table 1; Suppl. Table S6). Thus far, only distant homologs of SNF7 (a subunit of the eukaryotic ESCRT-III complex) have been identified in several TACK lineages, while genes encoding ESCRT-related proteins are notably absent from bacterial genomes. While the homology of the lokiarchaeal ESCRT-II with eukaryotic EAP30-domain-containing proteins (Vps36/22) and Vps25 could easily be established (Fig. 4a; Suppl. Figs. S21-S22), sequence homology of the putative lokiarchaeal ESCRT-I subunit Vps28 to putative eukaryotic counterparts was low. Yet, we could successfully model the C-terminal domain of the Lokiarchaeum protein to Vps28 (>90% confidence for 40% of the residues; homology confidence 95.7%; also see Suppl. Fig S23). In addition to the lokiarchaeal ESCRT-I protein Vps28, we identified a gene encoding a low complexity coiled coil protein (Lokiarch_16740) that showed low (insignificant) similarity to a Steadiness domain (IPR017916) at its C-terminus (Suppl. Table S6 + S10). Interestingly, this gene is located in close proximity to a gene encoding a putative Vps20/32/60-type ESCRT-III subunit (Lokiarch_16760) in the Lokiarchaeum genome. Although the insignificant homology prevents us from drawing firm conclusions, the product of this gene could represent a functional homolog of the Steadiness box of eukaryotic ESCRT-I subunit Vps23, which is implicated in the assembly of the ESCRT-I complex in eukaryotes [26].

6. References

1 Philippe, H., Zhou, Y., Brinkmann, H., Rodrigue, N. & Delsuc, F. Heterotachy and long-branch attraction in phylogenetics. *BMC evolutionary biology* **5**, doi:10.1186/1471-2148-5-50 (2005).

2 Foster, P. G. Modeling compositional heterogeneity. *Systematic biology* **53**, 485-495 (2004).

3 Lartillot, N., Rodrigue, N., Stubbs, D. & Richer, J. PhyloBayes MPI: phylogenetic reconstruction with infinite mixtures of profiles in a parallel environment. *Systematic biology* **62**, 611-615, doi:10.1093/sysbio/syt022 (2013).

4 Cox, C. J., Foster, P. G., Hirt, R. P., Harris, S. R. & Embley, T. M. The archaebacterial origin of eukaryotes. *Proc Natl Acad Sci U S A* **105**, 20356-20361, doi:10.1073/pnas.0810647105 (2008).

5 Lartillot, N., Brinkmann, H. & Philippe, H. Suppression of long-branch attraction artefacts in the animal phylogeny using a site-heterogeneous model. *BMC evolutionary biology* **7 Suppl 1**, S4, doi:10.1186/1471-2148-7-S1-S4 (2007).

6 Brochier, C., Gribaldo, S., Zivanovic, Y., Confalonieri, F. & Forterre, P. Nanoarchaea: representatives of a novel archaeal phylum or a fast-evolving

euryarchaeal lineage related to Thermococcales? *Genome biology* **6**, doi:10.1186/gb-2005-6-5-r42 (2005).

7 Guy, L., Spang, A., Saw, J. H. & Ettema, T. J. 'Geoarchaeote NAG1' is a deeply rooting lineage of the archaeal order Thermoproteales rather than a new phylum. *The ISME journal* **8**, 1353-1357, doi:10.1038/ismej.2014.6 (2014).

8 Guy, L., Saw, J. H. & Ettema, T. J. The Archaeal Legacy of Eukaryotes: A Phylogenomic Perspective. *Cold Spring Harbor perspectives in biology*, doi:10.1101/cshperspect.a016022 (2014).

9 Viklund, J., Ettema, T. J. & Andersson, S. G. Independent genome reduction and phylogenetic reclassification of the oceanic SAR11 clade. *Molecular biology and evolution* **29**, 599-615, doi:10.1093/molbev/msr203 (2012).

10 Shimodaira, H. An approximately unbiased test of phylogenetic tree selection. *Systematic biology* **51**, 492-508, doi:10.1080/10635150290069913 (2002).

11 Thiergart, T., Landan, G. & Martin, W. F. Concatenated alignments and the case of the disappearing tree. *BMC evolutionary biology* **14**, 2624, doi:10.1186/s12862-014-0266-0 (2014).

12 Lloyd, K. G. *et al.* Predominant archaea in marine sediments degrade detrital proteins. *Nature* **496**, 215-218, doi:10.1038/nature12033 (2013).

13 Iverson, V. *et al.* Untangling genomes from metagenomes: revealing an uncultured class of marine Euryarchaeota. *Science* **335**, 587-590, doi:10.1126/science.1212665 (2012).

14 Deschamps, P., Zivanovic, Y., Moreira, D., Rodriguez-Valera, F. & Lopez-Garcia, P. Pangenome evidence for extensive interdomain horizontal transfer affecting lineage core and shell genes in uncultured planktonic

thaumarchaeota and euryarchaeota. *Genome biology and evolution* **6**, 1549-1563, doi:10.1093/gbe/evu127 (2014).

15 Nelson-Sathi, S. *et al.* Origins of major archaeal clades correspond to gene acquisitions from bacteria. *Nature* **517**, 77-80, doi:10.1038/nature13805 (2015).

16 Borrel, G. *et al.* Unique characteristics of the pyrrolysine system in the 7th order of methanogens: implications for the evolution of a genetic code expansion cassette. *Archaea* **2014**, 374146, doi:10.1155/2014/374146 (2014).

17 Yutin, N., Puigbo, P., Koonin, E. V. & Wolf, Y. I. Phylogenomics of prokaryotic ribosomal proteins. *PloS one* **7**, e36972, doi:10.1371/journal.pone.0036972 (2012).

18 Laursen, B. S., Sorensen, H. P., Mortensen, K. K. & Sperling-Petersen, H. U. Initiation of protein synthesis in bacteria. *Microbiology and molecular biology reviews : MMBR* **69**, 101-123, doi:10.1128/MMBR.69.1.101-123.2005 (2005).

19 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". *Biology direct* **2**, 38, doi:10.1186/1745-6150-2-38 (2007).

20 Makarova, K. S., Krupovic, M. & Koonin, E. V. Evolution of replicative DNA polymerases in archaea and their contributions to the eukaryotic replication machinery. *Frontiers in microbiology* **5**, 354, doi:10.3389/fmicb.2014.00354 (2014).

21 Forterre, P. & Gadelle, D. Phylogenomics of DNA topoisomerases: their origin and putative roles in the emergence of modern organisms. *Nucleic acids research* **37**, 679-692, doi:10.1093/nar/gkp032 (2009).

22 Brochier-Armanet, C., Gribaldo, S. & Forterre, P. A DNA topoisomerase IB in Thaumarchaeota testifies for the presence of this enzyme in the last common ancestor of Archaea and Eucarya. *Biology direct* **3**, 54, doi:10.1186/1745-6150-3-54 (2008).

23 Meng, J. *et al.* Genetic and functional properties of uncultivated MCG archaea assessed by metagenome and gene expression analyses. *The ISME journal* **8**, 650-659, doi:10.1038/ismej.2013.174 (2014).

24 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 research* **39**, 3204-3223, doi:10.1093/nar/gkq1228 (2011).

25 Way, M. & Weeds, A. Nucleotide sequence of pig plasma gelsolin. Comparison of protein sequence with human gelsolin and other actin-severing proteins shows strong homologies and evidence for large internal repeats. *Journal of molecular biology* **203**, 1127-1133 (1988).

26 Kostelansky, M. S. *et al.* Structural and functional organization of the ESCRT-I trafficking complex. *Cell* **125**, 113-126, doi:10.1016/j.cell.2006.01.049 (2006).

Suppl. Table S1: General characteristics of the LCGC14 and LCGC14AMP metagenomic datasets.

	LCGC14AMP	LCGC14
Raw data (Gb)	56.6	8.6
Reads after preprocessing ($\cdot 10^6$)	188.2	53
Contigs > 1 kb	289,831	70,510
Total assembly size (contigs > 1kb) (Mb)	724.5	227.4
N50 (bp)	2895	4690
CDS	1,070,063	315,060
SSU rRNA genes	6	25

Suppl. Table S2: Description of arCOG markers used in concatenated phylogenies.

arCOG	Lokiarchaeum	Loki2/3	RP/non-RP	COG	Description
arCOG00109	x			COG02890	Methylase of polypeptide chain release factors
arCOG00405	x			COG00423	Glycyl-tRNA synthetase (class II)
arCOG00412	x	x		COG00072	Phenylalanyl-tRNA synthetase beta subunit
arCOG00415	x	x		COG00468	RecA/RadA recombinase
arCOG00779	x		x	COG00200	Ribosomal protein L15
arCOG00785	x	x	x	COG00255	Ribosomal protein L29
arCOG00987	x	x		COG00130	Pseudouridine synthase
arCOG01001	x			COG00024	Methionine aminopeptidase
arCOG01179	x			COG00361	Translation initiation factor 1 (IF-1)
arCOG01183	x	x		COG00533	Subunit of KEOPS complex, contains a domain with ASKHA fold and RIO-type kinase (AP-endonuclease activity)
arCOG01227	x	x		COG00552	Signal recognition particle GTPase
arCOG01228	x	x		COG00541	Signal recognition particle GTPase
arCOG01358	x			COG00621	2-methylthioadenine synthetase
arCOG01559	x	x		COG00480	Translation elongation factor G, EF-G (GTPase)
arCOG01560	x	x		COG00532	Translation initiation factor 2 (IF-2; GTPase)
arCOG01704	x			COG00452	Phosphopantothenoylcysteine synthetase/decarboxylase
arCOG01722	x	x	x	COG00099	Ribosomal protein S13
arCOG01758	x	x	x	COG00051	Ribosomal protein S10
arCOG01762	x	x		COG00085	DNA-directed RNA polymerase subunit B
arCOG04050	x			COG00258	5'-3' exonuclease (including N-terminal domain of PolI)
arCOG04064	x	x		COG00750	Predicted membrane-associated Zn-dependent protease
arCOG04067	x		x	COG00090	Ribosomal protein L2
arCOG04070	x		x	COG00087	Ribosomal protein L3
arCOG04071	x		x	COG00088	Ribosomal protein L4
arCOG04072	x		x	COG00089	Ribosomal protein L23
arCOG04086	x		x	COG01841	Ribosomal protein L30
arCOG04087	x		x	COG00098	Ribosomal protein S5
arCOG04088	x		x	COG00256	Ribosomal protein L18
arCOG04090	x	x	x	COG00097	Ribosomal protein L6P
arCOG04091	x	x	x	COG00096	Ribosomal protein S8
arCOG04092	x	x	x	COG00094	Ribosomal protein L5
arCOG04094	x	x	x	COG00198	Ribosomal protein L24
arCOG04095	x	x	x	COG00093	Ribosomal protein L14
arCOG04096	x	x	x	COG00186	Ribosomal protein S17
arCOG04097	x	x	x	COG00092	Ribosomal protein S3
arCOG04098	x	x	x	COG00091	Ribosomal protein L22
arCOG04099	x	x	x	COG00185	Ribosomal protein S19
arCOG04113	x	x	x	COG00197	Ribosomal protein L10AE/L16
arCOG04121	x	x		COG00164	Ribonuclease HII

arCOG04131	x			COG00030	Dimethyladenosine transferase (rRNA methylation)
arCOG04169	x	x		COG00201	Preprotein translocase subunit SecY
arCOG04184	x			COG00127	Xanthosine triphosphate pyrophosphatase
arCOG04185	x		x	COG00184	Ribosomal protein S15P
arCOG04239	x	x	x	COG00522	Ribosomal protein S4 or related protein
arCOG04240	x	x	x	COG00100	Ribosomal protein S11
arCOG04241	x	x		COG00202	DNA-directed RNA polymerase subunit D
arCOG04242	x	x	x	COG00102	Ribosomal protein L13
arCOG04243	x	x	x	COG00103	Ribosomal protein S9
arCOG04245	x	x	x	COG00052	Ribosomal protein S2
arCOG04254	x	x	x	COG00049	Ribosomal protein S7
arCOG04255	x	x	x	COG00048	Ribosomal protein S12
arCOG04256	x	x		COG00086	DNA-directed RNA polymerase subunit A double prime
arCOG04257	x	x		COG00086	DNA-directed RNA polymerase subunit A prime
arCOG04277	x			COG00231	Translation elongation factor P (EF-P)/translation initiation factor 5A (eIF-5A)
arCOG04288	x		x	COG00244	Ribosomal protein L10
arCOG04289	x	x	x	COG00081	Ribosomal protein L1
arCOG04302	x			COG00008	Glutamyl- or glutaminyl-tRNA synthetase
arCOG04372	x		x	COG00080	Ribosomal protein L11

Suppl. Table S3: Genomes used in arCOG attribution and in phylogenies.

Short name	Full name	NCBI Taxid	Accession	Group	SAG	Phylogenies
Novel archaea (44 in total, 23 selected for phylogenies):						
Ca_Micrarchaeum_acidiphilum_A	Candidatus Micrarchaeum acidiphilum ARMAN-2	425595	ACVJ01000001-8.1	arman	no	yes
Ca_Parvarchaeum_acidiphilum_A	Candidatus Parvarchaeum acidiphilum ARMAN-4	662760	ADCE01000001-45.1	arman	no	no
Ca_Parvarchaeum_acidophilus_A	Candidatus Parvarchaeum acidophilus ARMAN-5	662762	ADHF01000001-73.1	arman	no	yes
Ca_Nitrososphaera_gargensis_G	Candidatus Nitrososphaera gargensis Ga9.2	1237085	NC_018719.1	thaum	no	yes
Uncultured_Marine_Group_II_Eu	Uncultured Marine Group-II Euryarchaeote	274854	CM001443.1	eury	no	yes
Fervidicoccus_fontis_Kam940	Fervidicoccus fontis Kam940	1163730	NC_017461.1	cren	no	yes
Ca_Nanosalina_sp_J07AB43	Candidatus Nanosalina sp J07AB43	889948	AEIY01000001-210.1	eury	no	yes
Ca_Nanosalinarum_sp_J07AB56	Candidatus Nanosalinarum sp J07AB56	889962	AEIX01000001-259.1	eury	no	yes
Geo_NAG1	Geoarchaeote NAG1	1448933	JGI:2504756013*	cren	no	yes
MCG_SCGC_AB539E09	Thaumarchaeota archaeon SCGC AB-539-E09	1198115	NZ_ALXK00000000.1	MCG	yes	yes
MBGD_SCGC_AB539N05	Thermoplasmatales archaeon SCGC AB-539-N05	1198116	NZ_ALXL00000000.1	eury	yes	yes**
MBGD_SCGC_AB539C06	Thermoplasmatales archaeon SCGC AB-539-C06	1242690	NZ_AOSH00000000.1	eury	yes	no**
MBGD_SCGC_AB540F20	Thermoplasmatales archaeon SCGC AB-540-F20	1242866	NZ_AOSI00000000.1	eury	yes	no**
Nanoarchaeote_Nst1	Nanoarchaeote Nst1	1294122	NZ_APJZ00000000.1	nano	yes	yes
Methanomassiliicoccus_lu_B10	Methanomassiliicoccus luminyensis B10	1175296	NZ_CAJE00000000.1	eury	no	yes
Aenigma_AAA011_F07	Aenigmarchaeota archaeon SCGC AAA011-F07	1046938	AQYU00000000.1	aenigma	yes	no***
Aenigma_AAA011_O16	Candidatus Aenigmarchaeum subterraneum SCGC AAA011-O1	743730	ASMQ00000000.1	aenigma	yes	yes***
Aenigma_0000106_F11	Aenigmarchaeota archaeon JGI 0000106-F11	1130284	ASMO00000000.1	aenigma	yes	no***
Aiga_AAA471_F17	Aigarchaeota archaeon SCGC AAA471-F17	1052829	ASPD00000000.1	aiga	yes	yes
Aiga_AAA471_G05	Aigarchaeota archaeon SCGC AAA471-G05	1052830	ASLV00000000.1	aiga	yes	yes
Aiga_0000106_J15	Aigarchaeota archaeon JGI 0000106-J15	1130285	ASPF00000000.1	aiga	yes	yes
Diapher_AAA011_K09	Diapherotrites archaeon SCGC AAA011-K09	1104575	ASMR00000000.1	diapher	yes	no
Diapher_AAA011_E11	Candidatus Iainarchaeum andersonii SCGC AAA011-E11	1104576	AQRS00000000.1	diapher	yes	yes
Diapher_AAA011_N19	Diapherotrites archaeon SCGC AAA011-N19	743731	ASMS00000000.1	diapher	yes	no
Eury_AAA252_I15	Euryarchaeota archaeon SCGC AAA252-I15	913322	AQSC00000000.1	eury	yes	yes
Geo_AAA471_B05	Crenarchaeota archaeon SCGC AAA471-B05	1052801	AQYM00000000.1	cren	yes	yes
Geo_AAA471_B23	Crenarchaeota archaeon SCGC AAA471-B23	1052802	ASPO00000000.1	cren	yes	no
Geo_AAA471_C03	Crenarchaeota archaeon SCGC AAA471-C03	1052803	AQTC00000000.1	cren	yes	no
Geo_AAA471_L13	Crenarchaeota archaeon SCGC AAA471-L13	1052804	AQSU00000000.1	cren	yes	no
Geo_AAA471_L14	Crenarchaeota archaeon SCGC AAA471-L14	1052805	ASMI00000000.1	cren	yes	no
Geo_AAA471_O08	Crenarchaeota archaeon SCGC AAA471-O08	1052806	ASMJ00000000.1	cren	yes	no
Nano_AAA011_D5	Nanoarchaeota archaeon SCGC AAA011-D5	743729	ASMP00000000.1	nano	yes	yes
Nano_AAA011_G17	Nanoarchaeota archaeon SCGC AAA011-G17	1104577	AQRL00000000.1	nano	yes	yes
Thaum_AAA007_O23	Thaumarchaeota archaeon SCGC AAA007-O23	913333	ARWO00000000.1	thaum	yes	yes
Thaum_AB_179_E04	Thaumarchaeota archaeon SCGC AB-179-E04	1104571	ASPK00000000.1	thaum	yes	yes
Ca_Haloredivivus_sp_G17	Candidatus Haloredivivus sp. G17	1072681	AGNT00000000.1	eury	yes	no
SCGC_AAA261C22	Crenarchaeote SCGC AAA261-C22	1130299	AQVF00000000.1	cren	yes	no
SCGC_AAA261F05	Crenarchaeote SCGC AAA261-F05	1130300	AQVG00000000.1	cren	yes	no
SCGC_AAA261G18	Crenarchaeote SCGC AAA261-G18	1130301	AQVH00000000.1	cren	yes	no
SCGC_AAA261L14	Crenarchaeote SCGC AAA261-L14	1130302	AQVI00000000.1	cren	yes	no
SCGC_AAA261N13	Crenarchaeote SCGC AAA261-N13	1130304	AQVJ00000000.1	cren	yes	no
SCGC_AAA261E04	Euryarchaeote SCGC AAA261-E04	1130317	AQVK00000000.1	eury	yes	no
SCGC_AAA261G15	Euryarchaeote SCGC AAA261-G15	1130318	AQVL00000000.1	eury	yes	no
SCGC_AAA261L22	Crenarchaeote SCGC AAA261-L22	1130303	AQVP00000000.1	cren	yes	no
SCGC_AAA261N23	Crenarchaeote SCGC AAA261-N23	1130305	AQVQ00000000.1	cren	yes	no
Selected from Yutin et al (2012) (58 archaea, only the ones selected for phylogenies are shown):						
Methanohalobium_evestigatum_Z	Methanohalobium evestigatum Z-7303	644295	49857	eury	no	yes
Methanosarcina_acetivorans_C2	Methanosarcina acetivorans C2A	188937	57879	eury	no	yes
Methanosaeta_thermophila_PT	Methanosaeta thermophila PT	349307	58469	eury	no	yes
Methanocella_paludicola_SANAE	Methanocella paludicola SANAE	304371	42887	eury	no	yes
Methanosphaerula_palustris_E1	Methanosphaerula palustris E1-9c	521011	59193	eury	no	yes
Methanospirillum_hungatei_JF	Methanospirillum hungatei JF-1	323259	58181	eury	no	yes
Methanoplanus_petrolearius_DS	Methanoplanus petrolearius DSM 11571	679926	52695	eury	no	yes
Methanoculleus_marisnigri_JR1	Methanoculleus marisnigri JR1	368407	58561	eury	no	yes
Methanocorpusculum_labreanum	Methanocorpusculum labreanum Z	410358	58785	eury	no	yes
Haloarcula_marismortui_ATCC_4	Haloarcula marismortui ATCC 43049	272569	57719	eury	no	yes
Halalkalicoccus_jeotgali_B3	Halalkalicoccus jeotgali B3	795797	50305	eury	no	yes
Haloferax_volcanii_DS2	Haloferax volcanii DS2	309800	46845	eury	no	yes
Halobacterium_NRC_1	Halobacterium sp. NRC-1	64091	57769	eury	no	yes
Archaeoglobus_fulgidus_DSM_43	Archaeoglobus fulgidus DSM 4304	224325	57717	eury	no	yes
Ferroglobus_placidus_DSM_1064	Ferroglobus placidus DSM 10642	589924	40863	eury	no	yes
Thermoplasma_acidophilum_DSM	Thermoplasma acidophilum DSM 1728	273075	61573	eury	no	yes
Ferroplasma_acidarmanus_fer1	Ferroplasma acidarmanus fer1	333146	54095	eury	no	yes
Picrophilus_torridus_DSM_9790	Picrophilus torridus DSM 9790	263820	58041	eury	no	yes
Aciduliprofundum_boonei_T469	Aciduliprofundum boonei T469	439481	43333	eury	no	yes
Methanococcus_maripaludis_C6	Methanococcus maripaludis C6	444158	58947	eury	no	yes
Methanotorris_igneus_Kol_5	Methanotorris igneus Kol 5	880724	67321	eury	no	yes
Methanocaldococcus_jannaschii	Methanocaldococcus jannaschii DSM 2661	243232	57713	eury	no	yes
Methanobacterium_AL_21	Methanobacterium AL 21	868132	63623	eury	no	yes
Methanosphaera_stadtmanae_DSM	Methanosphaera stadtmanae DSM 3091	339860	58407	eury	no	yes
Methanothermobacter_thermauto	Methanothermobacter thermautotrophicus Delta H	187420	57877	eury	no	yes
Methanothermus_fervidus_DSM_2	Methanothermus fervidus DSM 2088	523846	60167	eury	no	yes
Methanopyrus_kandleri_AV19	Methanopyrus kandleri AV19	190192	57883	eury	no	yes
Pyrococcus_furiosus_DSM_3638	Pyrococcus furiosus DSM 3638	186497	57873	eury	no	yes
Thermococcus_kodakarensis_KOD	Thermococcus kodakarensis KOD1	69014	58225	eury	no	yes
Nanoarchaeum_equitans_Kin4_M	Nanoarchaeum equitans Kin4-M	228908	58009	nano	no	yes
Sulfolobus_islandicus_M_16_4	Sulfolobus islandicus M.16.4	426118	58841	cren	no	yes
Sulfolobus_solfataricus_P2	Sulfolobus solfataricus P2	273057	57721	cren	no	yes

Metallosphaera_cuprina_Ar_4	Metallosphaera cuprina Ar-4	1006006	66329	cren	no	yes
Metallosphaera_sedula_DSM_534	Metallosphaera sedula DSM 5348	399549	58717	cren	no	yes
Acidianus_hospitalis_W1	Acidianus hospitalis W1	933801	66875	cren	no	yes
Sulfolobus_acidocaldarius_DSM	Sulfolobus acidocaldarius DSM 639	330779	58379	cren	no	yes
Sulfolobus_tokodaii_7	Sulfolobus tokodaii 7	273063	57807	cren	no	yes
Desulfurococcus_kamchatkensis	Desulfurococcus kamchatkensis 1221n	490899	59133	cren	no	yes
Thermosphaera_aggregans_DSM_1	Thermosphaera aggregans DSM 11486	633148	48993	cren	no	yes
Staphylothermus_marinus_F1	Staphylothermus marinus F1	399550	58719	cren	no	yes
Acidilobus_saccharovorans_345	Acidilobus saccharovorans 345-15	666510	51395	cren	no	yes
Aeropyrum_pernix_K1	Aeropyrum pernix K1	272557	57757	cren	no	yes
Ignicoccus_hospitalis_KIN4_I	Ignicoccus hospitalis KIN4/I	453591	58365	cren	no	yes
Hyperthermus_butylicus_DSM_54	Hyperthermus butylicus DSM 5456	415426	57755	cren	no	yes
Pyrolobus_fumarii_1A	Pyrolobus fumarii 1A	694429	73415	cren	no	yes
Ignisphaera_aggregans_DSM_172	Ignisphaera aggregans DSM 17230	583356	51875	cren	no	yes
Pyrobaculum_aerophilum_IM2	Pyrobaculum aerophilum IM2	178306	57727	cren	no	yes
Pyrobaculum_calidifontis_JCM	Pyrobaculum calidifontis JCM 11548	410359	58787	cren	no	yes
Thermoproteus_uzoniensis_768	Thermoproteus uzoniensis 768-20	999630	65089	cren	no	yes
Vulcanisaeta_distributa_DSM_1	Vulcanisaeta distributa DSM 14429	572478	52827	cren	no	yes
Caldivirga_maquilingensis_IC	Caldivirga maquilingensis IC-167	397948	58711	cren	no	yes
Thermofilum_pendens_Hrk_5	Thermofilum pendens Hrk 5	368408	58563	cren	no	yes
Nitrosoarchaeum_koreensis_MY1	Candidatus Nitrosoarchaeum koreensis MY1	1001994	67913	thaum	no	yes
Nitrosoarchaeum_limnia_SFB1	Candidatus Nitrosoarchaeum limnia SFB1	886738	66129	thaum	no	yes
Nitrosopumilus_maritimus_SCM1	Nitrosopumilus maritimus SCM1	436308	58903	thaum	no	yes
Cenarchaeum_symbiosum_A	Cenarchaeum symbiosum A	414004	61411	thaum	no	yes
Caldiarchaeum_subterraneum	Candidatus Caldiarchaeum subterraneum	311458	49157	aiga	no	yes
Ca_Korarchaeum_cryptofilum_OP	Candidatus Korarchaeum cryptofilum OPF8	374847	58601	kor	no	yes
Bacteria (10)						
Bacillus_subtilis_168	Bacillus subtilis 168	224308	57675	bact		yes
Bacteroides_thetaiotaomicron	Bacteroides thetaiotaomicron VPI 5482	226186	62913	bact		yes
Borrelia_burgdorferi_B31	Borrelia burgdorferi B31	224326	57581	bact		yes
Campylobacter_jejuni_NCTC_111	Campylobacter jejuni NCTC 11168 ATCC 700819	192222	57587	bact		yes
Chlamydia_trachomatis_D_UW_3	Chlamydia trachomatis D UW 3 CX	272561	57637	bact		yes
Escherichia_coli_K_12_substr	Escherichia coli K 12 substr. MG1655	511145	57779	bact		yes
Rhodopirellula_baltica_SH_1	Rhodopirellula baltica SH 1	243090	61589	bact		yes
Rickettsia_prowazekii_Madrid	Rickettsia prowazekii Madrid E	272947	61565	bact		yes
Synechocystis_PCC_6803	Synechocystis PCC 6803	1148	57659	bact		yes
Thermotoga_maritima_MSB8	Thermotoga maritima MSB8	243274	57723	bact		yes
Eukarya (10)						
Arabidopsis_thaliana	Arabidopsis thaliana	3702		euk		yes
Dictyostelium_discoideum	Dictyostelium discoideum	44689		euk		yes
Entamoeba_histolytica_HM_1_IM	Entamoeba histolytica HM-1:IMSS	294381		euk		yes
Homo_sapiens	Homo sapiens	9606		euk		yes
Leishmania_infantum	Leishmania infantum	5671		euk		yes
Plasmodium_falciparum	Plasmodium falciparum	5833		euk		yes
Saccharomyces_cerevisiae	Saccharomyces cerevisiae	4932		euk		yes
Tetrahymena_thermophila	Tetrahymena thermophila	5911		euk		yes
Thalassiosira_pseudonana_CCMP	Thalassiosira pseudonana CCMP 1335	35128		euk		yes
Trichomonas_vaginalis	Trichomonas vaginalis	5722		euk		yes

*: Sequences for NAG1 are not deposited at NCBI. Here, the accession corresponds to JGI's taxon ID.

/*: These two groups of three genomes were pooled: paralog from the one marked with 'yes' was used whenever possible, complemented with those marked with 'no' elsewhere

Suppl. Table S4: General characteristics of the Lokiarchaeum assembly.

	Lokiarchaeum
Raw data (Mb)	423.7
Reads after preprocessing	1,532,941
Contigs, length ≥1 kb, coverage ≥ 20X	504
Assembly (filtered contigs) (Mb)	5.143
N50 (bp)	15403
N90 (bp)	5061
Completeness, weighted	0.919
Redundancy, weighted	1.438
GC content (%)	31.1
CDS	5381
SSU/LSU rRNA genes	1/1
tRNA genes (tRNAscan-SE/SPLITSX)	22/26

Suppl. Table S5: Results of the taxon removal phylogenetic analyses.

Removed taxon	Sister clade to Eukarya	Bootstrap Support	TACKLE [1] topology	Panel in Suppl. Fig. S13
Lokiarchaeum	Loki2/3	86	K,(LE,(TA,C))	a
Loki2/Loki3	K	80	L,(KE,(TA,C)	b
Lokiarchaeum/Loki2/Loki3	K	99	KE,(TA,C))	c
Bacteria	L	75	LE,(K,(TA,C))	d
Eukarya	-	-	L,(TA,KC)	e
Bacteria/Eukarya	-	-	L,(TA,KC)	f
DPANN	L	64	LE,(K,(TA,C))	g
Euryarchaeota	L	98	K,(LE,(TA,C))	h
Thaumarchaeota	L	87	LE,((K,TA),C)	i
Crenarchaeota	K	43	L,(KE,TA)	j
Aigarchaeota	L	71	LE,(T,KC)	k
Korarchaeota	L	100	LE,(TA,C)	l
SAGs	L	92	LE,(TA,KC)	m

[1] TACKLE: **T**haumarchaeota, **A**igarchaeota, **C**renarchaeota, **K**orarchaeota, **L**okiarchaeota, **E**ukarya

Suppl. Table S6: Description and annotation of genes discussed in the manuscript and shown in Table 1 including blast hits and InterPro domains

putative ribosomal protein								
Product	**Locus tag**	**Contig**	**Contig length**	**ORF number**	**Protein length (AA)**	**Best Blast Hit**	**IPR domain**	**Note**
putative homolog of eukaryotic ribosomal protein L22e*	Lokiarch_30160	contig150	12534	10	102	**gb|EPR78232.1| 60S ribosomal protein L22 [Spraguea lophii 42_110] - Expect = 0.21**	-	Previously not found in archaea.
putative ESCRT-complexes and additional genes serving as candidates for associated functions								
Product	**Locus tag**	**Contig**	**Contig length**	**ORF number**	**Protein length (AA)**	**Best Blast Hit (database: nr; evalue cutoff: 0.02)**	**IPR-domain**	**Note**
Vps4-like ATPase*	Lokiarch_37470	contig119	14547	11	405	**ref|XP_004249979.1| PREDICTED: vacuolar protein sorting-associated protein 4-like [Solanum lycopersicum] - Expect = 2e-105**	IPR003959 ATPase, AAA-type, core; IPR027417 P-loop containing nucleoside triphosphate hydrolase; IPR003593 AAA+ ATPase domain; IPR007330 MIT-domain;	(distant) homologs present in several archeal lineages, including Eurarchaeota
Vps2/24/46-like protein (ESCRT-III)*	Lokiarch_37480	contig119	14547	12	209	**ref|XP_001958212.1| GF10809 [Drosophila ananassae] - Expect = 4e-08**	IPR005024 Snf7	More distant homologs are also present in several other members of the TACK superphylum. NOTE: CdvA and ar-tubulins appear to be absent from the genome of Lokiarchaeum. ESCRT-related proteins are however absent from bacterial genomes.
Vps20/32/60-like protein (ESCRT-III)*	Lokiarch_16760	contig17	33755	12	218	**ref|XP_656010.1| SNF7 family protein [Entamoeba histolytica HM-1:IMSS] - Expect = 0.027**	IPR005024 Snf7	
EAP30 domain protein (ESCRT-II) (Vps22/36-like)*	Lokiarch_37450	contig119	14547	9	237	**dbj|GAA48847.1| vacuolar-sorting protein SNF8 [Clonorchis sinensis] - Expect = 4e-11**	IPR007286 EAP30	Previously not found in Archaea (and Bacteria).
Vps25-like protein (ESCRT-II)*	Lokiarch_37460	contig119	14547	10	297	**gb|ESW97316.1| hypothetical protein HPODL_01416 [Ogataea parapolymorpha DL-1] - Expect = 1e-10**	IPR014041 ESCRT-II complex, Vps25 subunit, N-terminal winged helix; IPR008570 ESCRT-II complex, vps25 subunit; IPR011991 Winged helix-turn-helix DNA-binding domain	
hypothetical protein with Vps28-like domain*	Lokiarch_10170	contig164	11945	10	218	-	IPR007143 Vacuolar protein sorting-associated, Vps28 (Expect: 2.4E-6), Homology modelling with Phyre2 showed, that the C-terminal domain of this protein models with 95.7% confidence to Vps28.	Vps28 is part of ESCRT-I in eukaryotes and has so far not been found in archaea, however proteins with putative Vps28 domains (IPR007143) might be present in mycoplasma; see http://www.ebi.ac.uk/interpro/entry/IPR007143/taxonomy
hypothetical protein with vacuolar fusion domain MON1	Lokiarch_21780	contig95	16096	6	130	-	IPR004353 Vacuolar fusion protein MON1	Proteins with potential Mon1 domain (Suppl. Table-S10). In *S. cerevisiae*, Mon1 and Ccz1 associate with particular small GTPase and mediate vesicle trafficking to the vacuole (Wang et al., 2002).
hypothetical protein with vacuolar fusion domain MON1	Lokiarch_01670	contig189	10697	11	137	-	IPR004353 Vacuolar fusion protein MON1	
hypothetical protein with vacuolar fusion domain MON1	Lokiarch_15160	contig446	3154	3	128	-	IPR004353 Vacuolar fusion protein MON1	
hypothetical coil protein with FAM92-domain	Lokiarch_46220	contig311	5775	3	216	**ref|XP_002736146.1| PREDICTED: protein FAM92A1-like isoform X1 [Saccoglossus kowalevskii] - E-value = 2e-06**	IPR009602 FAM92 protein; IPR004148 BAR domain	Lokiarch_46220 and Lokiarch_08900 share 97% sequence identity; these proteins are low complexity, coil coil proteins, however, using the superfamily 1.75 database search tool, these proteins show significant hits to proteins belonging to the BAR/IMD domain-like superfamily (SSF103657) (E-value: 3.14e-12 and 5.36e-14,
hypothetical coil protein with FAM92-domain	Lokiarch_08900	contig219	9080	7	216	**ref|XP_002736146.1| PREDICTED: protein FAM92A1-like isoform X1 [Saccoglossus kowalevskii] - Expect = 6e-07**	IPR009602 FAM92 protein; Superfamily: BAR/IMD domain-like superfamily protein	
hypothetical protein with putative steadiness-like domain	Lokiarch_16740	contig17	33755	10	649	-	insignificant hit to IPR017916: Steadiness box (at CTD)	In eukaryotes, a steadiness box domain is part of the ESCRT-I protein Vps23. The CTD of this protein shows also structural similarity to the CTD of ESCRT-I component Vps28 (Superfamily and Phyre predictions), which in eukaryotes interacts with ESCRT-II. However, due to low sequenence conservation and the low complexity sequence of this coiled coil protein, its function in Lokiarchaeum remains elusive.The only link to ESCRT-vesicle trafficking stems from the fact, that this proteins is located in proximity to Lokiarch_16760; one of the SNF7 homologs of Lokiarchaeum
hypothetical protein with longin-like domain	Lokiarch_01890	contig137	13309	3	169	-	IPR011012 Longin-like domain	Proteins with potential longin domains (Suppl.-Table-S10, which are ~150AA in length and part (form N-terminal domain); so far not found in other prokaryotes but present in marine sediment metagenome; longin-domain containing proteins are involved in subcellular trafficking machinery in eukaryotes
hypothetical protein with longin domain	Lokiarch_13110	contig166	11690	10	132	-	IPR010908 Longin domain; IPR011012 Longin-like domain	
hypothetical protein with longin-like domain	Lokiarch_03280	contig220	9042	8	252	-	IPR011012 Longin-like domain	
hypothetical protein with longin-like domain	Lokiarch_22790	contig41	24085	20	130	-	IPR011012 Longin-like domain	
hypothetical protein with longin-like domain	Lokiarch_04850	contig49	23044	15	167	-	IPR011012 Longin-like domain	
putative cytosceleton/cell division related proteins								
Product	**Locus tag**	**Contig**	**Contig length**	**ORF number**	**Protein length (AA)**	**Best Blast Hit (database: nr; evalue cutoff: 0.01)**	**IPR-domain**	**Note**
actin-/Arp1-/Arp2-/Arp3-related proteins*	Lokiarch_44920	contig230	8738	6	381	**ref|XP_001862810.1| actin [Culex quinquefasciatus] - Expect = 2e-177**	arCOG05583; IPR004000, Actin-related protein; IPR020902 Actin/actin-like conserved site	Some Cren- and Aigarchaeota encode crenactins.
	Lokiarch_36250	contig402	3772	1	381	**ref|XP_001862810.1| actin [Culex quinquefasciatus] - Expect = 2e-177**		
	Lokiarch_10650	contig208	9739	4	202	**ref|XP_005662446.1| PREDICTED: beta-centractin [Sus scrofa] - Expect = 2e-21**		
actin-related proteins*	Lokiarch_09100	contig48	23111	18	373	**ref|XP_002174311.1| actin [Schizosaccharomyces japonicus yFS275] - Expect = 3e-88**	IPR004000; Actin-related protein	
	Lokiarch_41030	contig136	13348	6	388	**ref|XP_007012602.1| Actin-11 [Theobroma cacao] - Expect = 8e-40**		
hypothetical proteins with gelsolin-like domain	Lokiarch_14520	contig111	14949	13	233	-	IPR029006 ADF-H/Gelsolin-like domain	gelsolin-domain proteins have reviously not been found in Archaea; in eukaryotes, gelsolin-domains are part of actin-binding, capping and modulating proteins (e.g. villin, severin ,adseverin etc.) often containing several gelsolin-domains. These eukaryotic proteins have been suggested to have evolved from small (120-130AA) gelsolin-domain proteins (Way and Weeds1988). However the function of these proteins in Lokiarchaeum remain to be elucidated. Interestingly, after the second psiblast iteration, Lokiarch_14340 and Lokiarch_49620 showed significant matches to eukaryotic gelsolins/adseverins (E-value to top blast hit: 1e-101, 24% identity) and Lokiarch_42060 was most similar to actin-binding proteins of Entamoeba (E-value: 6e-59; 27% identity) (psi-blast treshold: 0.001).
	Lokiarch_20110	contig129	13873	4	359	-	IPR029006 ADF-H/Gelsolin-like domain	
	Lokiarch_19250	contig114	14820	13	292	-	IPR029006 ADF-H/Gelsolin-like domain	
	Lokiarch_14340	contig144	12812	8	335	**ref|XP_004350112.1| severin [Dictyostelium fasciculatum] - Expect = 2e-15**	IPR007123 Gelsolin-like domain; IPR029006 ADF-H/Gelsolin-like domain; IPR007122 Villin/Gelsolin	
	Lokiarch_49620	contig162	11997	7	335	**ref|XP_004350112.1| severin [Dictyostelium fasciculatum] - Expect = 2e-14**	IPR007122 Villin/Gelsolin; IPR029006 ADF-H/Gelsolin-like domain; IPR007123 Gelsolin-like domain	
	Lokiarch_16460	contig267	7443	10	140	ref|XP_002166077.2| PREDICTED: advillin-like [Hydra vulgaris] - Expect = 0.078	IPR007123 Gelsolin-like domain; IPR029006 ADF-H/Gelsolin-like domain	
	Lokiarch_42060	contig214	9381	14	199	**ref|XP_004255074.1| actin-binding protein, putative, partial [Entamoeba invadens] - Expect = 5e-04**	IPR007122 Villin/Gelsolin; IPR007123 Gelsolin-like domain; IPR029006 ADF-H/Gelsolin-like domain	
	Lokiarch_06290	contig187	10803	7	199	ref|XP_005100380.1| PREDICTED: gelsolin-like protein 2-like isoform X1 [Aplysia californica] - Expect = 0.013	IPR029006 ADF-H/Gelsolin-like domain; IPR007123 Gelsolin-like domain	
	Lokiarch_35550	contig43	23852	15	317	-	IPR003034 SAP domain; IPR029006 ADF-H/Gelsolin-like domain	
	Lokiarch_33820	contig445	3187	4	328	-	IPR029006 ADF-H/Gelsolin-like domain	
	Lokiarch_32420	contig5	44173	41	180	ref|NP_001011294.1| capping protein (actin filament), gelsolin-like [Xenopus (Silurana) tropicalis]; Expect = 0.040	IPR029006 ADF-H/Gelsolin-like domain	
	Lokiarch_14920	contig75	18348	14	235	-	IPR029006 ADF-H/Gelsolin-like domain	
putative ubiquitin modifier system and candidates for associated functions (see also Suppl. Table-S10)								
Product	**Locus tag**	**Contig**	**Contig length**	**ORF number**	**Protein length (AA)**	**Best Blast Hit (cutoff: 0.02)**	**IPR domain**	**Note**
hypothetical protein with ubiquitin-like domain	Lokiarch_29280	contig202	9990	2	133	ref|XP_003340419.2| PREDICTED: ubiquitin D [Monodelphis domestica] - Expect = 0.073	IPR029071 Ubiquitin-related domain; IPR000626 Ubiquitin-like	Ubiquitin modifier system was recently identified in Aigarchaeota; while a canonical E3 ubiquitin ligase is not present in Caldiarchaeum and Lokiarchaeoum, RING domain-containing proteins (IPR001841 and IPR013083) present in these organisms serve as candidates for E3 ubiquitin ligases.
hypothetical protein with ubiquitin-like domain	Lokiarch_29290	contig202	9990	3	266	-	IPR029071 Ubiquitin-related domain	
hypothetical protein with ubiquitin-like domain	Lokiarch_29310	contig202	9990	5	113	-	IPR029071 Ubiquitin-related domain; IPR000626 Ubiquitin-like	
hypothetical protein with ubiquitin-like domain	Lokiarch_37670	contig241	8470	6	88	pdb|4GOC|A Chain A, Crystal Structure Of The Get5 Ubiquitin-like Domain - Expect = 0.077	IPR000626 Ubiquitin-like; IPR029071 Ubiquitin-related domain	
putative membrane-bound protein with ubiquitin-like domain	Lokiarch_28320	contig45	23655	17	1687	-	IPR029071 Ubiquitin-related domain; IPR000626 Ubiquitin-like	
putative E1-like ubiquitin activating protein	Lokiarch_15900	contig10	37281	8	527	ref|WP_020265555.1| hypothetical protein [Aigarchaeota archaeon SCGC AAA471-G05] - Expect = 1e-41	IPR023280 Ubiquitin-like 1 activating enzyme, catalytic cysteine domain; IPR019572 Ubiquitin-activating enzyme	
putative E1-like ubiquitin activating protein	Lokiarch_29320	contig202	9990	6	411	ref|WP_020265555.1| hypothetical protein [Aigarchaeota archaeon SCGC AAA471-G05] - Expect = 2e-53	IPR016040 NAD(P)-binding domain; IPR009036 Molybdenum cofactor biosynthesis, MoeB; IPR000594 UBA/THIF-type NAD/FAD binding fold; IPR000127 Ubiquitin-activating enzyme repeat; IPR019572 Ubiquitin-activating enzyme	
hypothetical protein with ubiquitin-conjugating domain	Lokiarch_16100	contig10	37281	28	493	-	IPR016135 Ubiquitin-conjugating enzyme/RWD-like; IPR006575 RWD domain	
putative E2-like ubiquitin conjugating protein	Lokiarch_10330	contig52	22764	2	170	**gb|AHY75163.1| Ubc1p [Saccharomyces cerevisiae YJM993] - Expect = 2e-20**	IPR016135 Ubiquitin-conjugating enzyme/RWD-like; IPR000608 Ubiquitin-conjugating enzyme, E2	
putative E2-like ubiquitin conjugating protein	Lokiarch_41800	contig184	11095	5	141	**ref|XP_002055124.1| GJ18969 [Drosophila virilis] - Expect = 2e-13**	IPR000608 Ubiquitin-conjugating enzyme, E2; IPR016135 Ubiquitin-conjugating enzyme/RWD-like;	
putative E2-like ubiquitin conjugating protein	Lokiarch_29330	contig202	9990	7	175	**ref|XP_790380.1| PREDICTED: ubiquitin-conjugating enzyme E2 R1-like [Strongylocentrotus purpuratus] - Expect = 1e-16**	IPR000608 Ubiquitin-conjugating enzyme, E2; IPR016135 Ubiquitin-conjugating enzyme/RWD-like	
hypothetical protein with JAB1/MPN/MOV34 metalloenzyme domain	Lokiarch_29340	contig202	9990	8	278	**ref|XP_001385344.2| multicatalytic endopeptidase [Scheffersomyces stipitis CBS 6054] - Expect = 6e-15**	IPR000555 JAB1/MPN/MOV34 metalloenzyme domain	Rpn11-like metalloproteases of Lokiarchaeum might be involved in degradation of ubiquitin-conjugated proteins as part of the 26S proteasome.
hypothetical protein with JAB1/MPN/MOV34 metalloenzyme domain	Lokiarch_43590	contig167	11665	8	524	**ref|XP_003688639.1| hypothetical protein TPHA_0P00470 [Tetrapisispora phaffii CBS 4417] - Expect = 5e-08**	IPR011990 Tetratricopeptide-like helical; IPR000555 JAB1/MPN/MOV34 metalloenzyme domain	
hypothetical protein with JAB1/MPN/MOV34 metalloenzyme domain	Lokiarch_26830	contig3	57269	39	373	**emb|CDS10782.1| hypothetical protein LRAMOSA11268 [Absidia idahoensis var. thermophila] - Expect = 3e-09**	IPR000555 JAB1/MPN/MOV34 metalloenzyme domain	
hypothetical protein with JAB1/MPN/MOV34 metalloenzyme domain	Lokiarch_08140	contig46	23214	11	525	**ref|XP_001645604.1| hypothetical protein Kpol_1033p52 [Vanderwaltozyma polyspora DSM 70294] - Expect = 4e-08**	IPR000555 JAB1/MPN/MOV34 metalloenzyme domain	
hypothetical protein with RING-domain	Lokiarch_34010	contig6	43995	19	175	**emb|CCC10989.1| unnamed protein product [Sordaria macrospora k-hell] - Expect = 3e-04**	IPR013083 Zinc finger, RING/FYVE/PHD-type	Candidate for E3 ubiquitin ligase, top hit after second psiblast iteration: ref|XP_010274322.1| E3 ubiquitin-protein ligase RING1-like [Nelumbo nucifera] - Expect = 2e-40
hypothetical protein with RING-/ and TPR domain	Lokiarch_38540	contig127	13916	4	192	ref|XP_790543.3| PREDICTED: uncharacterized protein LOC585630 [Strongylocentrotus purpuratus] - Expect = 5e-05	IPR001841 Zinc finger, RING-type; IPR013083 Zinc finger, RING/FYVE/PHD-type; IPR011990 Tetratricopeptide-like helical	no significant hit to E3 ubiquitin ligases after PSI-Blast iteration 2
hypothetical protein with RING-domain	Lokiarch_14290	contig144	12812	3	675	partial hit only: ref|XP_004241916.1| PREDICTED: RING-H2 finger protein ATL7-like [Solanum lycopersicum] - Expect = 1e-06	IPR013083 Zinc finger, RING/FYVE/PHD-type; IPR001841 Zinc finger, RING-type	not enough sequences producing significant alignments to perform a second psiblast iteration (E-value cutoff: 0.005)
hypothetical protein with RING-domain	Lokiarch_49580	contig162	11997	3	678	partial hit only: ref|XP_004241916.1| PREDICTED: RING-H2 finger protein ATL7-like [Solanum lycopersicum] - Expect = 5e-06	IPR013083 Zinc finger, RING/FYVE/PHD-type; IPR001841 Zinc finger, RING-type	
hypothetical protein with RING-domain	Lokiarch_36980	contig44	23740	15	67	-	IPR013083 Zinc finger, RING/FYVE/PHD-type; IPR001841 Zinc finger, RING-type	
hypothetical protein with RING-domain	Lokiarch_35850	contig222	8965	9	178	ref|XP_001866046.1| conserved hypothetical protein [Culex quinquefasciatus] - Expect = 0.003	IPR013083 Zinc finger, RING/FYVE/PHD-type	
hypothetical protein with RING-domain	Lokiarch_30770	contig293	6358	2	150	-	IPR013083 Zinc finger, RING/FYVE/PHD-type; IPR013083 Zinc finger, RING/FYVE/PHD-type; membrane domain	
hypothetical protein with RING-domain	Lokiarch_45630	contig31	28546	15	100	-	IPR011011 Zinc finger, FYVE/PHD-type; IPR013083 Zinc finger, RING/FYVE/PHD-type	
hypothetical protein with RING-domain	Lokiarch_07660	contig59	21334	11	320	ref|XP_004241916.1| PREDICTED: RING-H2 finger protein ATL7-like [Solanum lycopersicum] - Expect = 5e-06	IPR013083 Zinc finger, RING/FYVE/PHD-type	
Small G proteins of the small GTPases families (IPR006689 and IPR001806)*, and interacting proteins								
Product	**Locus tag**	**Contig**	**Contig length**	**ORF number**	**Protein length (AA)**	**Best Blast Hit**	**IPR domain**	**Note**
small GTP-binding domain protein, Rab-/Arf-domain signatures	Lokiarch_06040	contig103	15613	2	177	**ref|XP_004534366.1| PREDICTED: ADP-ribosylation factor-like protein 3-like [Ceratitis capitata] - Expect = 2e-12**	IPR003579 Small GTPase superfamily, Rab type; IPR027417 P-loop containing nucleoside triphosphate hydrolase; IPR027417 P-loop containing nucleoside triphosphate hydrolase; IPR006689 Small GTPase superfamily, ARF/SAR type; IPR001806 Small GTPase superfamily	
small GTP-binding domain protein, Rab-/Rho-domain signatures	Lokiarch_15930	contig10	37281	11	175	**ref|XP_004647300.1| PREDICTED: ras-related protein Rab-22A [Octodon degus] - Expect = 3e-39**	IPR020849 Small GTPase superfamily, Ras type; IPR005225 Small GTP-binding protein domain; IPR003579 Small GTPase superfamily, Rab type; IPR003578 Small GTPase superfamily, Rho type; IPR027417 P-loop containing nucleoside triphosphate hydrolase; IPR001806 Small GTPase superfamily	
small GTP-binding domain protein, Ras-/Rab-/Arf-domain signatures	Lokiarch_15960	contig10	37281	14	602	**ref|NP_001002473.1| ADP-ribosylation factor-like protein 1 [Danio rerio] - Expect = 2e-13**	IPR020849 Small GTPase superfamily, Ras type; IPR027417 P-loop containing nucleoside triphosphate hydrolase; IPR003579 Small GTPase superfamily, Rab type; IPR001806 Small GTPase superfamily; IPR005225 Small GTP-binding protein domain; IPR024156 Small GTPase superfamily, ARF type; IPR006689 Small GTPase superfamily, ARF/SAR type	
small GTP-binding domain protein, Ras-/Rab-/Rho-domain signatures	Lokiarch_45790	contig105	15410	1	179	**emb|CAI44537.1| rab_B25 [Paramecium tetraurelia] - Expect = 2e-38**	IPR001806 Small GTPase superfamily; IPR027417 P-loop containing nucleoside triphosphate hydrolase; IPR020849 Small GTPase superfamily, Ras type; IPR027417 P-loop containing nucleoside triphosphate hydrolase; IPR003579 Small GTPase superfamily, Rab type; IPR005225 Small GTP-binding protein domain; IPR003578 Small GTPase superfamily, Rho type	
small GTP-binding domain protein, Ras-/Rab-/Rho-domain signatures	Lokiarch_45850	contig105	15410	7	244	**gb|ETN85986.1| Ras family protein [Necator americanus] - Expect = 3e-17**	IPR001806 Small GTPase superfamily; IPR003579 Small GTPase superfamily, Rab type; IPR020849 Small GTPase superfamily, Ras type; IPR003578 Small GTPase superfamily, Rho type; IPR027417 P-loop containing nucleoside triphosphate hydrolase	
small GTP-binding domain protein, Arf-/Rab-domain signatures	Lokiarch_05490	contig106	15403	13	353	**gb|KDN51599.1| hypothetical protein RSAG8_00144, partial [Rhizoctonia solani AG-8 WAC10335] - Expect = 4e-16**	IPR005225 Small GTP-binding protein domain; IPR027417 P-loop containing nucleoside triphosphate hydrolase; IPR006689 Small GTPase superfamily, ARF/SAR type; IPR003579 Small GTPase superfamily, Rab type; IPR001806 Small GTPase superfamily; IPR024156 Small GTPase superfamily, ARF type;	
small GTP-binding domain protein, Arf-/Rab-/Rho-domain signatures	Lokiarch_05440	contig106	15403	8	346	**ref|XP_007511648.1| ADP-ribosylation factor-like protein 5A [Bathycoccus prasinos] - Expect = 6e-13**	IPR001806 Small GTPase superfamily; IPR027417 P-loop containing nucleoside triphosphate hydrolase; IPR024156 Small GTPase superfamily, ARF type; IPR006689 Small GTPase superfamily, ARF/SAR type; IPR005225 Small GTP-binding protein domain; IPR003579 Small GTPase superfamily, Rab type; IPR003578 Small GTPase superfamily, Rho type	
small GTP-binding domain protein, Arf-domain signature	Lokiarch_42280	contig107	15355	2	346	**ref|XP_005340765.1| PREDICTED: ADP-ribosylation factor-like protein 13A [Ictidomys tridecemlineatus] - Expect = 0.003**	IPR027417 P-loop containing nucleoside triphosphate hydrolase; IPR006689 Small GTPase superfamily, ARF/SAR type	

small GTP-binding domain protein, Ran-/Ras-/Rab-/Rho-domain signatures	Lokiarch_33450	contig116	14764	3	195	**ref|XP_005703007.1| Rab family, other [Galdieria sulphuraria] - Expect = 2e-35**	IPR001806 Small GTPase superfamily; IPR027417 P-loop containing nucleoside triphosphate hydrolase; IPR002041 Ran GTPase; IPR005225 Small GTP-binding protein domain; IPR003578 Small GTPase superfamily, Rho type; IPR003579 Small GTPase superfamily, Rab type; IPR020849 Small GTPase superfamily, Ras type
small GTP-binding domain protein, Ras-/Rab-/Rho-domain signatures	Lokiarch_49520	contig117	14713	16	191	**emb|CDW57011.1| ras protein Rab 8A [Trichuris trichiura] - Expect = 2e-37**	IPR001806 Small GTPase superfamily; IPR003578 Small GTPase superfamily, Rho type; IPR003579 Small GTPase superfamily, Rab type; IPR027417 P-loop containing nucleoside triphosphate hydrolase; IPR020849 Small GTPase superfamily, Ras type; IPR005225 Small GTP-binding protein domain
small GTP-binding domain protein, Ran-/Ras-/Rab-/Rho-domain signatures	Lokiarch_18970	contig122	14305	3	186	**gb|AAX97487.1| small Rab GTPase RabX10 [Trichomonas vaginalis] - Expect = 1e-37**	IPR001806 Small GTPase superfamily; IPR002041 Ran GTPase; IPR003579 Small GTPase superfamily, Rab type; IPR003578 Small GTPase superfamily, Rho type; IPR020849 Small GTPase superfamily, Ras type; IPR005225 Small GTP-binding protein domain; IPR027417 P-loop containing nucleoside triphosphate hydrolase
small GTP-binding domain protein	Lokiarch_17150	contig126	13957	11	162	**ref|XP_004258039.1| GTP-binding protein YPTM1, putative [Entamoeba invadens IP1] - Expect = 5e-04**	IPR027417 P-loop containing nucleoside triphosphate hydrolase; IPR001806 Small GTPase superfamily
small GTP-binding domain protein, Ran-/Ras-/Rab-/Rho-domain signatures	Lokiarch_38550	contig127	13916	5	529	**ref|XP_008384909.1| PREDICTED: ras-related protein Rab2BV [Malus domestica] - Expect = 1e-31**	IPR027417 P-loop containing nucleoside triphosphate hydrolase; IPR001806 Small GTPase superfamily; IPR020849 Small GTPase superfamily, Ras type; IPR005225 Small GTP-binding protein domain; IPR002041 Ran GTPase; IPR003579 Small GTPase superfamily, Rab type; IPR003578 Small GTPase superfamily, Rho type
small GTP-binding domain protein, Ran-/Ras-/Rab-/Rho-domain signatures	Lokiarch_38560	contig127	13916	6	519	**ref|XP_006304197.1| hypothetical protein CARUB_v10010276mg [Capsella rubella] - Expect = 3e-33**	IPR005225 Small GTP-binding protein domain; IPR003579 Small GTPase superfamily, Rab type; IPR001806 Small GTPase superfamily; IPR027417 P-loop containing nucleoside triphosphate hydrolase; IPR003578 Small GTPase superfamily, Rho type; IPR002041 Ran GTPase; IPR020849 Small GTPase superfamily, Ras type
small GTP-binding domain protein, Ran-/ Ras-/Rab-/Rho-/Arf-domain signatures	Lokiarch_02350	contig13	35460	2	398	**ref|XP_007267677.1| GTP binding protein [Fomitiporia mediterranea MF3/22] - Expect = 2e-36**	IPR001806 Small GTPase superfamily; IPR003579 Small GTPase superfamily, Rab type; IPR027417 P-loop containing nucleoside triphosphate hydrolase; IPR020849 Small GTPase superfamily, Ras type; IPR002041 Ran GTPase; IPR003578 Small GTPase superfamily, Rho type; IPR005225 Small GTP-binding protein domain; IPR024156 Small GTPase superfamily, ARF type
small GTP-binding domain protein; Ras-/Rab-/Rho-domain signatures	Lokiarch_12880	contig134	13501	14	186	**ref|XP_005533719.1| PREDICTED: ras-related protein Rab-13, partial [Pseudopodoces humilis] - Expect = 2e-27**	IPR001806 Small GTPase superfamily; IPR027417 P-loop containing nucleoside triphosphate hydrolase; IPR005225 Small GTP-binding protein domain; IPR020849 Small GTPase superfamily, Ras type; IPR003579 Small GTPase superfamily, Rab type; IPR003578 Small GTPase superfamily, Rho type
small GTP-binding domain protein, Ras-/Rab-domain signatures	Lokiarch_09250	contig140	13254	13	203	**gb|EPR78369.1| Rab GTPase [Spraguea lophii 42_110] - Expect = 1e-21**	IPR020849 Small GTPase superfamily, Ras type; IPR027417 P-loop containing nucleoside triphosphate hydrolase; IPR003579 Small GTPase superfamily, Rab type; IPR005225 Small GTP-binding protein domain; IPR001806 Small GTPase superfamily
small GTP-binding domain protein	Lokiarch_18500	contig145	12777	5	209	gb|KCZ71086.1| putative GTPase [Candidatus Methanoperedens nitroreducens] - Expect = 1e-17	IPR006073 GTP binding domain; IPR027417 P-loop containing nucleoside triphosphate hydrolase; IPR001806 Small GTPase superfamily
small GTP-binding domain protein, Ras-/Rab-/Rho-domain signatures	Lokiarch_18530	contig145	12777	8	150	**ref|XP_003677683.1| hypothetical protein NCAS_0H00220 [Naumovozyma castellii CBS 4309] - Expect = 3e-17**	IPR001806 Small GTPase superfamily; IPR027417 P-loop containing nucleoside triphosphate hydrolase; IPR005225 Small GTP-binding protein domain; IPR020849 Small GTPase superfamily, Ras type; IPR003579 Small GTPase superfamily, Rab type; IPR003578 Small GTPase superfamily, Rho type
small GTP-binding domain protein, Ras-/Rab-/Rho-domain signatures	Lokiarch_15320	contig148	12697	2	167	**ref|XP_002588708.1| RAS oncogene protein [Branchiostoma floridae] - Expect = 7e-22**	IPR027417 P-loop containing nucleoside triphosphate hydrolase; IPR020849 Small GTPase superfamily, Ras type; IPR005225 Small GTP-binding protein domain; IPR001806 Small GTPase superfamily; IPR003578 Small GTPase superfamily, Rho type; IPR003579 Small GTPase superfamily, Rab type
small GTP-binding domain protein	Lokiarch_15340	contig148	12697	4	209	gb|KCZ71086.1| putative GTPase [Candidatus Methanoperedens nitroreducens] - Expect = 1e-17	IPR006073 GTP binding domain; IPR027417 P-loop containing nucleoside triphosphate hydrolase; IPR001806 Small GTPase superfamily
GTP-binding domain protein, Arf-/Ras-/Rab-domain signatures	Lokiarch_19980	contig15	34201	27	336	**ref|XP_008930080.1| PREDICTED: ADP-ribosylation factor-like protein 3 [Manacus vitellinus] - Expect = 2e-23**	IPR006689 Small GTPase superfamily, ARF/SAR type; IPR020849 Small GTPase superfamily, Ras type; IPR027417 P-loop containing nucleoside triphosphate hydrolase; IPR024156 Small GTPase superfamily, ARF type; IPR001806 Small GTPase superfamily; IPR003579 Small GTPase superfamily, Rab type; IPR005225 Small GTP-binding protein domain
small GTP-binding domain protein, Ran-/Ras-/Rab-/Rho-domain signatures	Lokiarch_52350	contig16	34146	22	253	**gb|EFA75388.1| Rab GTPase [Polysphondylium pallidum PN500] - Expect = 9e-37**	IPR001806 Small GTPase superfamily; IPR005225 Small GTP-binding protein domain; IPR027417 P-loop containing nucleoside triphosphate hydrolase; IPR020849 Small GTPase; IPR003578 Small GTPase superfamily, Rho type IPR002041 Ran GTPase; IPR003579 Small GTPase superfamily, Rab type superfamily, Ras type
small GTP-binding domain protein, Ras-/Rab-/Rho-domain signatures	Lokiarch_26330	contig171	11494	1	117	**ref|XP_004235823.1| PREDICTED: ras-related protein RABA6a-like [Solanum lycopersicum] - Expect = 1e-21**	IPR003579 Small GTPase superfamily, Rab type; IPR003578 Small GTPase superfamily, Rho type; IPR001806 Small GTPase superfamily; IPR020849 Small GTPase superfamily, Ras type; IPR027417 P-loop containing nucleoside triphosphate hydrolase
small GTP-binding domain protein; Arf-domain signature	Lokiarch_50330	contig1	71539	3	208	ref|WP_012992071.1| gliding motility protein [Thermocrinis albus] - Expect = 2e-26	IPR027417 P-loop containing nucleoside triphosphate hydrolase; IPR006689 Small GTPase superfamily, ARF/SAR type
small GTP-binding domain protein; Arf-domain signature	Lokiarch_50350	contig1	71539	5	229	ref|WP_011340060.1| cell polarity determinant GTPase MglA [Pelobacter carbinolicus] - Expect = 7e-18	IPR006689 Small GTPase superfamily, ARF/SAR type; IPR027417 P-loop containing nucleoside triphosphate hydrolase
small GTP-binding domain protein, Ran-/Ras-/Rab-/Rho-domain signatures	Lokiarch_12240	contig18	33144	14	184	**ref|XP_002670122.1| rab family small GTPase [Naegleria gruberi] - Expect = 4e-31**	IPR003579 Small GTPase superfamily, Rab type; IPR005225 Small GTP-binding protein domain; IPR001806 Small GTPase superfamily; IPR020849 Small GTPase superfamily, Ras type; IPR027417 P-loop containing nucleoside triphosphate hydrolase; IPR002041 Ran GTPase; IPR003578 Small GTPase superfamily, Rho type
small GTP-binding domain protein, Ras-/Rab-domain signature	Lokiarch_12180	contig18	33144	8	418	**ref|XP_001311109.1| Ras-like GTP-binding protein YPT1 [Trichomonas vaginalis G3] - Expect = 5e-27**	IPR027417 P-loop containing nucleoside triphosphate hydrolase; IPR001806 Small GTPase superfamily; IPR003579 Small GTPase superfamily, Rab type; IPR020849 Small GTPase superfamily, Ras type
small GTP-binding domain protein, Ras-/Rab-/Rho-domain signatures	Lokiarch_01690	contig189	10697	13	172	**ref|XP_002671605.1| rab family small GTPase [Naegleria gruberi] - Expect = 2e-35**	IPR005225 Small GTP-binding protein domain; IPR001806 Small GTPase superfamily; IPR003578 Small GTPase superfamily, Rho type; IPR020849 Small GTPase superfamily, Ras type; IPR027417 P-loop containing nucleoside triphosphate hydrolase; IPR003579 Small GTPase superfamily, Rab type
small GTP-binding domain protein, Ras/Rab/Ran/Rho-domain signatures	Lokiarch_01650	contig189	10697	9	174	**ref|XP_004337611.1| ras-related protein Rab-2A, putative [Acanthamoeba castellanii str. Neff] - Expect = 2e-38**	IPR003579 Small GTPase superfamily, Rab type; IPR027417 P-loop containing nucleoside triphosphate hydrolase; IPR001806 Small GTPase superfamily; IPR003578 Small GTPase superfamily, Rho type; IPR005225 Small GTP-binding protein domain; IPR002041 Ran GTPase; IPR020849 Small GTPase superfamily, Ras type
small GTP-binding domain protein, Ras-/Rab-/Rho-domain signatures	Lokiarch_10880	contig199	10083	5	180	**ref|XP_007900912.1| PREDICTED: ras-related protein Rab-30 [Callorhinchus milii] - Expect = 4e-38**	IPR003578 Small GTPase superfamily, Rho type; IPR027417 P-loop containing nucleoside triphosphate hydrolase; IPR005225 Small GTP-binding protein domain; IPR003579 Small GTPase superfamily, Rab type; IPR020849 Small GTPase superfamily, Ras type; IPR001806 Small GTPase superfamily
small GTP-binding domain protein, Ras-/Rab-/Rho-domain signatures	Lokiarch_37110	contig206	9763	5	325	**gb|AAX97453.1| small Rab GTPase Rab7a [Trichomonas vaginalis] - Expect = 5e-23**	IPR003579 Small GTPase superfamily, Rab type; IPR005225 Small GTP-binding protein domain; IPR001806 Small GTPase superfamily; IPR020849 Small GTPase superfamily, Ras type; IPR027417 P-loop containing nucleoside triphosphate hydrolase; IPR003578 Small GTPase superfamily, Rho type
small GTP-binding domain protein, Rab-/Arf-domain signature	Lokiarch_53730	contig226	8815	8	176	**gb|EJY79831.1| ARL3, ARF-like Ras superfamily GTPase [Oxytricha trifallax] - Expect = 2e-14**	IPR027417 P-loop containing nucleoside triphosphate hydrolase; IPR003579 Small GTPase superfamily, Rab type; IPR001806 Small GTPase superfamily; IPR005225 Small GTP-binding protein domain; IPR006689 Small GTPase superfamily, ARF/SAR type
small Arf-related GTPase	Lokiarch_28560	contig23	31031	20	139	**ref|XP_007248151.1| PREDICTED: ADP-ribosylation factor-like protein 2-like [Astyanax mexicanus] - Expect = 9e-15**	IPR027417 P-loop containing nucleoside triphosphate hydrolase; IPR006689 Small GTPase superfamily, ARF/SAR type
small Arf-related GTPase	Lokiarch_46160	contig247	8256	6	268	**ref|XP_001642590.1| hypothetical protein Kpol_333p2 [Vanderwaltozyma polyspora DSM 70294] - Expect = 0.005**	IPR006689 Small GTPase superfamily, ARF/SAR type; IPR027417 P-loop containing nucleoside triphosphate hydrolase
small Arf-related GTPase	Lokiarch_01140	contig2	58255	31	309	**emb|CBN78716.1| GTR1, Ras superfamily GTPase [Ectocarpus siliculosus] - Expect = 5e´14**	IPR006689 Small GTPase superfamily, ARF/SAR type; IPR027417 P-loop containing nucleoside triphosphate hydrolase; IPR024156 Small GTPase superfamily, ARF type; IPR001806 Small GTPase superfamily
small GTP-binding domain protein	Lokiarch_01230	contig2	58255	40	186	gb|KCZ72488.1| small GTP-binding protein domain [Candidatus Methanoperedens nitroreducens] - Expect = 3e-12	IPR001806 Small GTPase superfamily; IPR027417 P-loop containing nucleoside triphosphate hydrolase
small GTP-binding domain protein, Ras-/Rab-/Rho-domain signatures	Lokiarch_34920	contig26	30581	30	192	**ref|XP_005737420.1| PREDICTED: ras-related protein Rab-33B-like [Pundamilia nyererei] - Expect = 1e-25**	IPR001806 Small GTPase superfamily; IPR027417 P-loop containing nucleoside triphosphate hydrolase; IPR003579 Small GTPase superfamily, Rab type; IPR003578 Small GTPase superfamily, Rho type; IPR020849 Small GTPase superfamily, Ras type; IPR005225 Small GTP-binding protein domain
small GTP-binding domain protein, Ras-/Rab-/Rho-domain signatures	Lokiarch_02670	contig263	7607	7	159	**gb|AAX97487.1| small Rab GTPase RabX10 [Trichomonas vaginalis] - Expect = 1e-27**	IPR003579 Small GTPase superfamily, Rab type; IPR005225 Small GTP-binding protein domain; IPR027417 P-loop containing nucleoside triphosphate hydrolase; IPR001806 Small GTPase superfamily; IPR003578 Small GTPase superfamily, Rho type; IPR020849 Small GTPase superfamily, Ras type
small GTP-binding domain protein, Arf-/Gtr1/RagA G protein domain signature	Lokiarch_51460	contig268	7411	6	403	**emb|CBN80651.1| ADP-ribosylation factor 4 [Dicentrarchus labrax] - Expect = 2e-10**	IPR027417 P-loop containing nucleoside triphosphate hydrolase; IPR024156 Small GTPase superfamily, ARF type; IPR006762 Gtr1/RagA G protein; IPR001806 Small GTPase superfamily
small GTP-binding domain protein, Rab-/Rho-/Arf-domain signatures	Lokiarch_44040	contig274	7051	1	342	**ref|XP_008882834.1| ADP-ribosylation factor family protein 1, putative [Hammondia hammondi] - Expect = 6e-14**	IPR027417 P-loop containing nucleoside triphosphate hydrolase; IPR003578 Small GTPase superfamily, Rho type; IPR023114 Elongated TPR repeat-containing domain IPR003579 Small GTPase superfamily, Rab type; IPR005225 Small GTP-binding protein domain; IPR024156 Small GTPase superfamily, ARF type; IPR006689 Small GTPase superfamily, ARF/SAR type
small GTP-binding domain protein, Ras-/Rab-/Rho-domain signatures	Lokiarch_39540	contig29	29291	7	159	**ref|XP_002903160.1| Rab32/38 family GTPase, putative [Phytophthora infestans T30-4] - Expect = 2e-21**	IPR001806 Small GTPase superfamily; IPR027417 P-loop containing nucleoside triphosphate hydrolase; IPR020849 Small GTPase superfamily, Ras type; IPR003579 Small GTPase superfamily, Rab type; IPR005225 Small GTP-binding protein domain; IPR003578 Small GTPase superfamily, Rho type
small GTP-binding domain protein, Ran-/Ras-/Rab-/Rho-domain signatures	Lokiarch_39550	contig29	29291	8	186	**ref|XP_008559290.1| PREDICTED: ras-related protein Rab-23 [Microplitis demolitor] - Expect = 2e-30**	IPR020849 Small GTPase superfamily, Ras type; IPR005225 Small GTP-binding protein domain; IPR002041 Ran GTPase; IPR027417 P-loop containing nucleoside triphosphate hydrolase; IPR003579 Small GTPase superfamily, Rab type; IPR001806 Small GTPase superfamily; IPR003578 Small GTPase superfamily, Rho type
small GTP-binding domain protein, Ras-/Rab-/Rho-domain signatures	Lokiarch_47920	contig32	28483	17	178	**gb|AAW31982.1| CG2532 [Drosophila melanogaster] - Expect = 7e-41**	IPR001806 Small GTPase superfamily; IPR005225 Small GTP-binding protein domain; IPR027417 P-loop containing nucleoside triphosphate hydrolase; IPR020849 Small GTPase superfamily, Ras type; IPR003579 Small GTPase superfamily, Rab type; IPR003578 Small GTPase superfamily, Rho type
small GTP-binding domain protein, Ras-/Rab-/Rho-domain signatures	Lokiarch_27490	contig33	28201	16	178	**gb|AAW31980.1| CG2532 [Drosophila melanogaster] - Expect = 3e-40**	IPR003579 Small GTPase superfamily, Rab type; IPR027417 P-loop containing nucleoside triphosphate hydrolase; IPR020849 Small GTPase superfamily, Ras type; IPR003578 Small GTPase superfamily, Rho type; IPR005225 Small GTP-binding protein domain; IPR001806 Small GTPase superfamily
small GTP-binding domain protein, Ras-/Rab-/Arf-domain signatures	Lokiarch_10070	contig356	4727	3	344	**gb|KDN51599.1| hypothetical protein RSAG8_00144, partial [Rhizoctonia solani AG-8 WAC10335] - Expect = 8e-17**	IPR001806 Small GTPase superfamily; IPR027417 P-loop containing nucleoside triphosphate hydrolase; IPR024156 Small GTPase superfamily, ARF type; IPR006689 Small GTPase superfamily, ARF/SAR type; IPR003579 Small GTPase superfamily, Rab type; IPR020849 Small GTPase superfamily, Ras type
small GTP-binding domain protein, Ras-/Rab-/Rho-domain signatures	Lokiarch_48070	contig383	4065	4	178	**emb|CAI39347.1| rab_B12 [Paramecium tetraurelia] - Expect = 5e-29**	IPR003579 Small GTPase superfamily, Rab type; IPR003578 Small GTPase superfamily, Rho type; IPR001806 Small GTPase superfamily; IPR027417 P-loop containing nucleoside triphosphate hydrolase; IPR005225 Small GTP-binding protein domain; IPR020849 Small GTPase superfamily, Ras type
small GTP-binding domain protein, Ras-/Rab-/Rho-domain signatures	Lokiarch_21100	contig385	3995	2	199	**ref|XP_006341536.1| PREDICTED: ras-related protein RABA6a-like [Solanum tuberosum] - Expect = 3e-26**	IPR001806 Small GTPase superfamily; IPR027417 P-loop containing nucleoside triphosphate hydrolase; IPR005225 Small GTP-binding protein domain; IPR020849 Small GTPase superfamily, Ras type; IPR003578 Small GTPase superfamily, Rho type; IPR003579 Small GTPase superfamily, Rab type

small GTP-binding domain protein, Ras-/Rab-/Rho-domain signatures	Lokiarch_21460	contig386	3992	2	204	**gb|ETW34887.1| hypothetical protein PFTANZ_04420 [Plasmodium falciparum Tanzania (2000708)] - Expect = 2e-26**	IPR001806 Small GTPase superfamily; IPR027417 P-loop containing nucleoside triphosphate hydrolase; IPR020849 Small GTPase superfamily, Ras type; IPR003579 Small GTPase superfamily, Rab type; IPR003578 Small GTPase superfamily, Rho type; IPR005225 Small GTP-binding protein domain
small GTP-binding domain protein, Ran-/Ras-/Rab-/Rho-domain signatures	Lokiarch_04620	contig403	3729	2	195	**ref|XP_003969638.1| PREDICTED: ras-related protein Rab-8B-like [Takifugu rubripes] - Expect = 2e-34**	IPR003579 Small GTPase superfamily, Rab type; IPR001806 Small GTPase superfamily; IPR027417 P-loop containing nucleoside triphosphate hydrolase; IPR002041 Ran GTPase; IPR005225 Small GTP-binding protein domain; IPR003578 Small GTPase superfamily, Rho type; IPR020849 Small GTPase superfamily, Ras type
small GTP-binding domain protein, Arf-domain signature	Lokiarch_35740	contig404	3724	2	206	ref|WP_013885492.1| mutual gliding protein A [Flexistipes sinusarabici] - Expect = 3e-22	IPR001806 Small GTPase superfamily; IPR027417 P-loop containing nucleoside triphosphate hydrolase; IPR006689 Small GTPase superfamily, ARF/SAR type
small GTP-binding domain protein, Ras-/Rab-/Rho-domain signatures	Lokiarch_33750	contig427	3400	1	235	gb|AIF07601.1| small GTP-binding protein (RAB) [uncultured marine group II/III euryarchaeote KM3_205_C10] - Expect = 2e-31	IPR020849 Small GTPase superfamily, Ras type; IPR027417 P-loop containing nucleoside triphosphate hydrolase; IPR005225 Small GTP-binding protein domain; IPR001806 Small GTPase superfamily; IPR003579 Small GTPase superfamily, Rab type; IPR003578 Small GTPase superfamily, Rho type
small GTP-binding domain protein, Ras-/Rab-/Rho-domain signatures	Lokiarch_33770	contig427	3400	3	308	**emb|CCG82788.1| GTPase Ypt3 [Taphrina deformans PYCC 5710] - Expect = 3e-32**	IPR003579 Small GTPase superfamily, Rab type; IPR003578 Small GTPase superfamily, Rho type; IPR005225 Small GTP-binding protein domain; IPR027417 P-loop containing nucleoside triphosphate hydrolase; IPR001806 Small GTPase superfamily; IPR020849 Small GTPase superfamily, Ras type
small GTP-binding domain protein, Ras-/Rab-/Rho-domain signatures	Lokiarch_35450	contig43	23852	5	159	**ref|XP_006347973.1| PREDICTED: ras-related protein RABD2a-like [Solanum tuberosum] - Expect = 3e-28**	IPR001806 Small GTPase superfamily; IPR003578 Small GTPase superfamily, Rho type; IPR020849 Small GTPase superfamily, Ras type; IPR003579 Small GTPase superfamily, Rab-type; IPR027417 P-loop containing nucleoside triphosphate hydrolase; IPR005225 Small GTP-binding protein domain
small GTP-binding domain protein, Arf-domain signature	Lokiarch_20230	contig442	3214	3	198	ref|WP_013638657.1| mutual gliding protein A [Desulfurobacterium thermolithotrophum] - Expect = 2e-32	IPR006689 Small GTPase superfamily, ARF/SAR type; IPR027417 P-loop containing nucleoside triphosphate hydrolase
small GTP-binding domain protein, Ras-/Rab-domain signatures	Lokiarch_52000	contig480	2660	1	137	**emb|CAP22858.2| Protein CBR-RAB-19 [Caenorhabditis briggsae] - Expect = 5e-25**	IPR001806 Small GTPase superfamily; IPR020849 Small GTPase superfamily, Ras type; IPR027417 P-loop containing nucleoside triphosphate hydrolase; IPR003579 Small GTPase superfamily, Rab type
small GTP-binding domain protein, Arf-domain signature	Lokiarch_09000	contig48	23111	8	522	**ref|XP_455999.1| hypothetical protein [Kluyveromyces lactis NRRL Y-1140] - Expect = 3e-06**	IPR006689 Small GTPase superfamily, ARF/SAR type; IPR027417 P-loop containing nucleoside triphosphate hydrolase
small GTP-binding domain protein, Ras-/Rab-domain signatures	Lokiarch_11830	contig515	2182	1	137	**gb|AAX97487.1| small Rab GTPase RabX10 [Trichomonas vaginalis] - Expect = 1e-23**	IPR020849 Small GTPase superfamily, Ras type; IPR027417 P-loop containing nucleoside triphosphate hydrolase; IPR001806 Small GTPase superfamily; IPR003579 Small GTPase superfamily, Rab type; IPR005225 Small GTP-binding protein domain
small GTP-binding domain protein, Rho-/Ras-/Rab-domain signatures	Lokiarch_32470	contig5	44173	46	320	**ref|XP_001020738.1| Ras family protein [Tetrahymena thermophila] - Expect = 1e-24**	IPR003578 Small GTPase superfamily, Rho type; IPR027417 P-loop containing nucleoside triphosphate hydrolase; IPR001806 Small GTPase superfamily; IPR020849 Small GTPase superfamily, Ras type; IPR003579 Small GTPase superfamily, Rab type; IPR005225 Small GTP-binding protein domain
small GTP-binding domain protein, Ran-/Ras-/Rab-/Rho-domain signatures	Lokiarch_36830	contig572	1582	1	338	**gb|AAX97487.1| small Rab GTPase RabX10 [Trichomonas vaginalis] - Expect = 2e-37**	IPR020849 Small GTPase superfamily, Ras type; IPR027417 P-loop containing nucleoside triphosphate hydrolase; IPR002041 Ran GTPase; IPR003578 Small GTPase superfamily, Rho type; IPR001806 Small GTPase superfamily; IPR003579 Small GTPase superfamily, Rab type; IPR005225 Small GTP-binding protein domain
small GTP-binding domain protein	Lokiarch_09940	contig593	1436	1	117	**ref|XP_003872954.1| putative rab1 small GTP-binding protein [Leishmania mexicana MHOM/GT/2001/U1103] - Expect = 2e-21**	IPR003579 Small GTPase superfamily, Rab type; IPR027417 P-loop containing nucleoside triphosphate hydrolase; IPR001806 Small GTPase superfamily; IPR020849 Small GTPase superfamily, Ras type; IPR003578 Small GTPase superfamily, Rho type
small GTP-binding domain protein, Ras-/Rab-/Rho-domain signatures	Lokiarch_44790	contig620	1309	1	179	**ref|XP_002787019.1| RAB6 protein, putative [Perkinsus marinus ATCC 50983] - Expect = 2e-38**	IPR005225 Small GTP-binding protein domain; IPR003579 Small GTPase superfamily, Rab type; IPR027417 P-loop containing nucleoside triphosphate hydrolase; IPR001806 Small GTPase superfamily; IPR020849 Small GTPase superfamily, Ras type; IPR003578 Small GTPase superfamily, Rho type
small GTP-binding domain protein, Ras-/Rab-/Rho-domain signatures	Lokiarch_53220	contig63	20695	12	170	**emb|CAI44483.1| rab_C72 [Paramecium tetraurelia] - Expect = 2e-25**	IPR001806 Small GTPase superfamily; IPR003579 Small GTPase superfamily, Rab type; IPR003578 Small GTPase superfamily, Rho type; IPR027417 P-loop containing nucleoside triphosphate hydrolase; IPR020849 Small GTPase superfamily, Ras type; IPR005225 Small GTP-binding protein domain
small GTP-binding domain protein, Ran-/Ras-/Rab-/Rho-domain signatures	Lokiarch_53190	contig63	20695	9	588	**ref|XP_007135286.1| hypothetical protein PHAVU_010G116500g [Phaseolus vulgaris] - Expect = 2e-46**	IPR001806 Small GTPase superfamily; IPR005225 Small GTP-binding protein domain; IPR003578 Small GTPase superfamily, Rho type; IPR027417 P-loop containing nucleoside triphosphate hydrolase; IPR002041 Ran GTPase; IPR003579 Small GTPase superfamily, Rab type; IPR020849 Small GTPase superfamily, Ras type
small GTP-binding domain protein, Ran-/Ras-/Rab-/Rho-domain signatures	Lokiarch_36400	contig64	20324	13	252	**emb|CCD81400.1| putative rab-2,4,14 [Schistosoma mansoni] - Expect = 5e-38**	IPR001806 Small GTPase superfamily; IPR005225 Small GTP-binding protein domain; IPR003578 Small GTPase superfamily, Rho type; IPR027417 P-loop containing nucleoside triphosphate hydrolase; IPR002041 Ran GTPase; IPR003579 Small GTPase superfamily, Rab type; IPR020849 Small GTPase superfamily, Ras type
small GTP-binding domain protein, Ras-/Rab-/Rho-domain signatures	Lokiarch_33970	contig6	43995	15	174	**emb|CAI39320.1| rab_C09 [Paramecium tetraurelia] - Expect = 3e-32**	IPR001806 Small GTPase superfamily; IPR027417 P-loop containing nucleoside triphosphate hydrolase; IPR003579 Small GTPase superfamily, Rab type; IPR005225 Small GTP-binding protein domain; IPR020849 Small GTPase superfamily, Ras type; IPR003578 Small GTPase superfamily, Rho type
small GTP-binding domain protein, Ras-/Rab-/Rho-domain signatures	Lokiarch_31830	contig65	20269	11	180	**emb|CDS34289.1| small GTP binding protein [Hymenolepis microstoma] - Expect = 7e-36**	IPR001806 Small GTPase superfamily; IPR027417 P-loop containing nucleoside triphosphate hydrolase; IPR003579 Small GTPase superfamily, Rab type; IPR005225 Small GTP-binding protein domain; IPR020849 Small GTPase superfamily, Ras type; IPR003578 Small GTPase superfamily, Rho type
small GTP-binding domain protein, Ras-/Rab-/Rho-domain signatures	Lokiarch_31850	contig65	20269	13	189	**ref|XP_007229676.1| PREDICTED: ras-related protein Rab-43-like isoform X1 [Astyanax mexicanus] - Expect = 1e-41**	IPR001806 Small GTPase superfamily; IPR027417 P-loop containing nucleoside triphosphate hydrolase; IPR003579 Small GTPase superfamily, Rab type; IPR005225 Small GTP-binding protein domain; IPR020849 Small GTPase superfamily, Ras type; IPR003578 Small GTPase superfamily, Rho type
small GTP-binding domain protein, Ran-/Ras-/Rab-/Rho-domain signatures	Lokiarch_31930	contig65	20269	21	176	**gb|EPB90357.1| rab family, other [Mucor circinelloides f. circinelloides 1006PhL] - Expect = 9e-41**	IPR001806 Small GTPase superfamily; IPR005225 Small GTP-binding protein domain; IPR003578 Small GTPase superfamily, Rho type; IPR027417 P-loop containing nucleoside triphosphate hydrolase; IPR002041 Ran GTPase; IPR003579 Small GTPase superfamily, Rab type; IPR020849 Small GTPase superfamily, Ras type
small GTP-binding domain protein, Ras-/Rab-/Rho-domain signatures	Lokiarch_31960	contig65	20269	24	137	**gb|AAX97453.1| small Rab GTPase Rab7a [Trichomonas vaginalis] - Expect = 1e-24**	IPR001806 Small GTPase superfamily; IPR027417 P-loop containing nucleoside triphosphate hydrolase; IPR003579 Small GTPase superfamily, Rab type; IPR005225 Small GTP-binding protein domain; IPR020849 Small GTPase superfamily, Ras type; IPR003578 Small GTPase superfamily, Rho type
small GTP-binding domain protein, Ran-/Ras-/Rab-/Rho-domain signatures	Lokiarch_31810	contig65	20269	9	181	**ref|XP_005848051.1| hypothetical protein CHLNCDRAFT_48772 [Chlorella variabilis] - Expect = 8e-38**	IPR001806 Small GTPase superfamily; IPR005225 Small GTP-binding protein domain; IPR003578 Small GTPase superfamily, Rho type; IPR027417 P-loop containing nucleoside triphosphate hydrolase; IPR002041 Ran GTPase; IPR003579 Small GTPase superfamily, Rab type; IPR020849 Small GTPase superfamily, Ras type
small GTP-binding domain protein, Ran-/Ras-/Rab-/Rho-domain signatures	Lokiarch_18250	contig72	18733	12	182	**gb|AAX97487.1| small Rab GTPase RabX10 [Trichomonas vaginalis] - Expect = 5e-38**	IPR001806 Small GTPase superfamily; IPR005225 Small GTP-binding protein domain; IPR003578 Small GTPase superfamily, Rho type; IPR027417 P-loop containing nucleoside triphosphate hydrolase; IPR002041 Ran GTPase; IPR003579 Small GTPase superfamily, Rab type; IPR020849 Small GTPase superfamily, Ras type
small GTP-binding domain protein, Arf- and Gtr1/RagA G protein domain signatures	Lokiarch_18170	contig72	18733	4	413	**gb|EAR84414.2| ADP-ribosylation factor(Arf)/Arf-like (Arl) small GTPase family protein [Tetrahymena thermophila SB210] - Expect = 5e-11**	IPR001806 Small GTPase superfamily; IPR027417 P-loop containing nucleoside triphosphate hydrolase; IPR024156 Small GTPase superfamily, ARF type; IPR006762 Gtr1/RagA G protein
small GTP-binding domain protein, Ran-/Ras-/Rab-/Rho-domain signatures	Lokiarch_30440	contig74	18368	12	181	**emb|CAI44537.1| rab_B25 [Paramecium tetraurelia] - Expect = 4e-40**	IPR001806 Small GTPase superfamily; IPR005225 Small GTP-binding protein domain; IPR003578 Small GTPase superfamily, Rho type; IPR027417 P-loop containing nucleoside triphosphate hydrolase; IPR002041 Ran GTPase; IPR003579 Small GTPase superfamily, Rab type; IPR020849 Small GTPase superfamily, Ras type
small GTP-binding domain protein, Ras-/Rab-/Rho-domain signatures	Lokiarch_30390	contig74	18368	7	244	gb|EGB07418.1| hypothetical protein AURANDRAFT_6384 [Aureococcus anophagefferens] - Expect = 1e-17	IPR001806 Small GTPase superfamily; IPR027417 P-loop containing nucleoside triphosphate hydrolase; IPR003579 Small GTPase superfamily, Rab type; IPR005225 Small GTP-binding protein domain; IPR020849 Small GTPase superfamily, Ras type; IPR003578 Small GTPase superfamily, Rho type
small GTP-binding domain protein, Rho/Ras/Rab-domain signatures	Lokiarch_04160	contig77	18214	10	521	**ref|XP_002533328.1| protein with unknown function [Ricinus communis] - Expect = 8e-31**	IPR027417 P-loop containing nucleoside triphosphate hydrolase; IPR003578 Small GTPase superfamily, Rho type; IPR020849 Small GTPase superfamily, Ras type; IPR001806 Small GTPase superfamily; IPR005225 Small GTP-binding protein domain; IPR003579 Small GTPase superfamily, Rab type
small GTP-binding domain protein, Rho/Ras/Rab-domain signatures	Lokiarch_04210	contig77	18214	15	170	**ref|XP_003645458.1| hypothetical protein Ecym_3137 [Eremothecium cymbalariae DBVPG#7215] - Expect = 1e-24**	IPR003578 Small GTPase superfamily, Rho type; IPR001806 Small GTPase superfamily; IPR005225 Small GTP-binding protein domain; IPR003579 Small GTPase superfamily, Rab type; IPR020849 Small GTPase superfamily, Ras type; IPR027417 P-loop containing nucleoside triphosphate hydrolase
small GTP-binding domain protein, Rho/Ras/Rab-domain signatures	Lokiarch_22470	contig80	17811	16	348	**emb|CAI39307.1| rab_C39 [Paramecium tetraurelia] - Expect = 1e-36**	IPR027417 P-loop containing nucleoside triphosphate hydrolase; IPR003578 Small GTPase superfamily, Rho type; IPR020849 Small GTPase superfamily, Ras type; IPR001806 Small GTPase superfamily; IPR005225 Small GTP-binding protein domain; IPR003579 Small GTPase superfamily, Rab type
small GTP-binding domain protein, Rho/Ras/Rab-domain signatures	Lokiarch_45420	contig8	42635	19	175	ref|WP_002698198.1| small GTP-binding protein domain [Microscilla marina] - Expect = 1e-31	IPR027417 P-loop containing nucleoside triphosphate hydrolase; IPR003578 Small GTPase superfamily, Rho type; IPR020849 Small GTPase superfamily, Ras type; IPR001806 Small GTPase superfamily; IPR005225 Small GTP-binding protein domain; IPR003579 Small GTPase superfamily, Rab type
small GTP-binding domain protein, Arf- and Gtr1/RagA G protein domain signatures	Lokiarch_03700	contig87	16838	6	413	**ref|NP_001120217.1| ADP-ribosylation factor-like 3, gene 2 [Xenopus (Silurana) - Expect = 2e-12**	IPR027417 P-loop containing nucleoside triphosphate hydrolase; IPR024156 Small GTPase superfamily, ARF type; IPR006762 Gtr1/RagA G protein; IPR001806 Small GTPase superfamily
small GTP-binding domain protein, Arf-domain signature	Lokiarch_45120	contig88	16776	18	220	ref|WP_029551737.1| gliding motility protein [Thermocrinis sp. GBS] - Expect = 1e-12	IPR001806 Small GTPase superfamily; IPR027417 P-loop containing nucleoside triphosphate hydrolase; IPR006689 Small GTPase superfamily, ARF/SAR type
small GTP-binding domain protein, Rho/Ras/Rab-domain signatures	Lokiarch_44980	contig88	16776	4	174	**ref|XP_005993187.1| PREDICTED: ras-related protein Rab-39B-like [Latimeria chalumnae] - Expect = 1e-32**	IPR027417 P-loop containing nucleoside triphosphate hydrolase; IPR003578 Small GTPase superfamily, Rho type; IPR020849 Small GTPase superfamily, Ras type; IPR001806 Small GTPase superfamily; IPR005225 Small GTP-binding protein domain; IPR003579 Small GTPase superfamily, Rab type
small GTP-binding domain protein, Rho/Ras/Rab-domain signatures	Lokiarch_17900	contig92	16479	15	178	**gb|EJY72016.1| Rab GTPase 2b [Oxytricha trifallax] - Expect = 7e-34**	IPR027417 P-loop containing nucleoside triphosphate hydrolase; IPR003578 Small GTPase superfamily, Rho type; IPR020849 Small GTPase superfamily, Ras type; IPR001806 Small GTPase superfamily; IPR005225 Small GTP-binding protein domain; IPR003579 Small GTPase superfamily, Rab type
small GTP-binding domain protein, Ras/Rab-domain signatures	Lokiarch_04470	contig93	16190	16	307	**ref|XP_002175038.1| GTPase Ypt3 [Schizosaccharomyces japonicus yFS275] - Expect = 4e-18**	IPR027417 P-loop containing nucleoside triphosphate hydrolase; IPR020849 Small GTPase superfamily, Ras type; IPR003579 Small GTPase superfamily, Rab type; IPR001806 Small GTPase superfamily
small GTP-binding domain protein, Ran-/Ras-/Rab-domain signatures	Lokiarch_04480	contig93	16190	17	314	**emb|CBK21525.2| unnamed protein product [Blastocystis hominis] - Expect = 3e-25**	IPR001806 Small GTPase superfamily; IPR005225 Small GTP-binding protein domain; IPR027417 P-loop containing nucleoside triphosphate hydrolase; IPR002041 Ran GTPase; IPR003579 Small GTPase superfamily, Rab type; IPR020849 Small GTPase superfamily, Ras type
small GTP-binding domain protein, ARF-domain signature	Lokiarch_43010	contig94	16135	17	280	**ref|XP_004952022.1| PREDICTED: ADP-ribosylation factor-like [Setaria italica] - Expect = 6e-10**	IPR006689 Small GTPase superfamily, ARF/SAR type; IPR027417 P-loop containing nucleoside triphosphate hydrolase
small GTP-binding domain protein, Gtr1/RagA G protein-domain signature	Lokiarch_42900	contig94	16135	6	502	**ref|XP_004249010.1| PREDICTED: ADP-ribosylation factor-related protein 1-like [Solanum lycopersicum] - Expect = 9e-11**	IPR027417 P-loop containing nucleoside triphosphate hydrolase; IPR006762 Gtr1/RagA G protein; IPR001806 Small GTPase superfamily
small GTP-binding domain protein, Ran-/Ras-/Rab-/Rho-domain signatures	Lokiarch_21790	contig95	16096	7	174	**ref|XP_007432282.1| PREDICTED: ras-related protein Rab-35-like isoform X1 [Python bivittatus] - Expect = 1e-39**	IPR001806 Small GTPase superfamily; IPR005225 Small GTP-binding protein domain; IPR003578 Small GTPase superfamily, Rho type; IPR027417 P-loop containing nucleoside triphosphate hydrolase; IPR002041 Ran GTPase; IPR003579 Small GTPase superfamily, Rab type; IPR020849 Small GTPase superfamily, Ras type
small GTP-binding domain protein, ARF-domain signature	Lokiarch_11040	contig98	15711	9	522	**gb|EFX87182.1| hypothetical protein DAPPUDRAFT_187433 [Daphnia pulex] - Expect = 1e-07**	IPR027417 P-loop containing nucleoside triphosphate hydrolase; IPR006689 Small GTPase superfamily, ARF/SAR type

small GTP-binding domain protein, Rho/Ras/Rab-domain signatures	Lokiarch_51330	contig99	15668	4	308	**emb|CDJ56879.1| RAS small GTpase, putative [Eimeria maxima] - Expect = 4e-31**	IPR027417 P-loop containing nucleoside triphosphate hydrolase; IPR003578 Small GTPase superfamily, Rho type; IPR020849 Small GTPase superfamily, Ras type; IPR001806 Small GTPase superfamily; IPR005225 Small GTP-binding protein domain; IPR003579 Small GTPase superfamily, Rab type	
small GTP-binding domain protein, Rho/Ras/Rab-domain signatures	Lokiarch_51340	contig99	15668	5	308	**ref|XP_001633043.1| predicted protein [Nematostella vectensis] - Expect = 2e-29**	IPR027417 P-loop containing nucleoside triphosphate hydrolase; IPR003578 Small GTPase superfamily, Rho type; IPR020849 Small GTPase superfamily, Ras type; IPR001806 Small GTPase superfamily; IPR005225 Small GTP-binding protein domain; IPR003579 Small GTPase superfamily, Rab type	
small GTP-binding domain protein, Rho/Ras/Rab-domain signatures	Lokiarch_51370	contig99	15668	8	328	ref|WP_028458193.1| GTP-binding protein [Chloroflexus sp. Y-396-1] - Expect = 8e-30	IPR027417 P-loop containing nucleoside triphosphate hydrolase; IPR003578 Small GTPase superfamily, Rho type; IPR020849 Small GTPase superfamily, Ras type; IPR001806 Small GTPase superfamily; IPR005225 Small GTP-binding protein domain; IPR003579 Small GTPase superfamily, Rab type	
small GTP-binding domain protein, Rho/Ras/Rab-domain signatures	Lokiarch_00500	contig255	7954	10	180	**ref|XP_005848051.1| hypothetical protein CHLNCDRAFT_48772 [Chlorella variabilis] - Expect = 4e-38**	IPR001806 Small GTPase superfamily, IPR003579 Small GTPase superfamily, Rab type; IPR005225 Small GTP-binding protein domain, IPR020849 Small GTPase superfamily, Ras type, IPR001806 Small GTPase superfamily, IPR027417 P-loop containing nucleoside IPR003578 Small GTPase superfamily, Rho type	
putative small GTPase family protein	Lokiarch_04170	contig77	18214	11	45	-	IPR001806 Small GTPase superfamily, IPR027417 P-loop containing nucleoside triphosphate hydrolase	
Roadblock/LC7 domain proteins (no ESPs), see also Suppl. Table-S10								
Product	**Locus tag**	**Contig**	**Contig length**	**ORF number**	**Protein length (AA)**	**Best Blast Hits (e-value treshold: 0.02)**	**IPR domain**	**Note**
Roadblock/LC7 domain protein	Lokiarch_02330	contig229	8750	7	153	ref|WP_015790751.1| hypothetical protein [Methanocaldococcus fervens] - Expect 1e-11	IPR004942 Dynein light chain-related, IPR004942 Dynein light chain-related	Roadblock/LC7 domain proteins could serve as GTPase activating enzymes similar to MglB of the bacterium Myxococcus xanthus (Zhang et al., 2010)
Roadblock/LC7 domain protein	Lokiarch_03940	contig143	12833	4	142	ref|WP_008316295.1| Hypothetical protein [Wohlfahrtiimonas chitiniclastica] - Expect = 4e-11	IPR004942 Dynein light chain-related, IPR004942 Dynein light chain-related	
hypothetical protein	Lokiarch_06200	contig639	1183	2	131	dbj|BAO45715.1| conserved hypothetical protein [gamma proteobacterium Hiromi1] - Expect = 0.003	IPR004942 Dynein light chain-related, IPR004942 Dynein light chain-related	
Roadblock/LC7 domain protein	Lokiarch_06410	contig187	10803	19	135	dbj|BAL54378.1| roadblock/LC7 family protein [uncultured Chloroflexi bacterium] - Expect = 1e-32	IPR004942 Dynein light chain-related, IPR004942 Dynein light chain-related	
Roadblock/LC7 domain protein	Lokiarch_06430	contig25	30688	2	117	ref|WP_027845797.1| diacylglyceryl transferase [Mastigocoleus testarum] - Expect = 0.004	IPR004942 Dynein light chain-related, IPR004942 Dynein light chain-related	
Roadblock/LC7 domain protein	Lokiarch_06490	contig25	30688	8	100	**gb|EHJ71444.1| roadblock [Danaus plexippus] - Expect = 9e-09**	IPR004942 Dynein light chain-related, IPR004942 Dynein light chain-related	
Roadblock/LC7 domain protein	Lokiarch_08020	contig352	4793	6	117	ref|WP_013644254.1| hypothetical protein [Methanobacterium lacus] - Expect = 6e-13	IPR004942 Dynein light chain-related, IPR004942 Dynein light chain-related	
Roadblock/LC7 domain protein	Lokiarch_25010	contig415	3579	4	117	ref|WP_014453567.1| hypothetical protein [Caldisericum exile] - Expect = 9e-14	IPR004942 Dynein light chain-related, IPR004942 Dynein light chain-related	
Roadblock/LC7 domain protein	Lokiarch_41940	contig214	9381	2	125	dbj|BAL54378.1| roadblock/LC7 family protein [uncultured Chloroflexi bacterium] - Expect = 4e-33	IPR004942 Dynein light chain-related, IPR004942 Dynein light chain-related	
Roadblock/LC7 domain protein	Lokiarch_43650	contig90	16630	2	123	dbj|BAL54378.1| roadblock/LC7 family protein [uncultured Chloroflexi bacterium] - Expect = 1e-24	IPR004942 Dynein light chain-related, IPR004942 Dynein light chain-related	
hypothetical protein	Lokiarch_44990	contig88	16776	5	249	-	IPR004942 Dynein light chain-related	
Roadblock/LC7 domain protein	Lokiarch_54060	contig157	12235	12	142	ref|WP_008316295.1| Hypothetical protein [Wohlfahrtiimonas chitiniclastica] - Expect = 1e-10	IPR004942 Dynein light chain-related, IPR004942 Dynein light chain-related	
Oligosaccharyl transferase complex proteins								
Product	**Locus tag**	**Contig**	**Contig length**	**ORF number**	**Protein length (AA)**	**Best Blast Hit**	**IPR domain**	**Note**
Ribophorin 1 superfamily protein (dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit 1)	Lokiarch_43710	contig90	16630	8	619	**ref|XP_003382687.1| PREDICTED: dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit 1-like [Amphimedon queenslandica] - Expect = 1e-08**	IPR007676 Ribophorin I	Close homolgs in sediment metagenome, absent from other archaeal genomes.
Putative oligosaccharyltransferase complex subunit	Lokiarch_24040	contig91	16548	6	173	**ref|XP_004988908.1|; hypothetical protein PTSG_09995 [Salpingoeca rosetta] - Expect = 1e-05**	IPR021149 putative oligosaccharyl transferase complex, subunit OST3/OST6	Close homolgs in sediment metagenome, absent from other archaeal genomes.
Putative oligosaccharyltransferase complex subunit	Lokiarch_25040	contig188	10797	2	173	**emb|CDJ03393.1| oligosaccharyltransferase complex subunit [Echinococcus multilocularis] - Expect = 5e-04**	IPR021149 Oligosaccharyl transferase complex, subunit OST3/OST6	Close homolgs in sediment metagenome, absent from other archaeal genomes.
Putative oligosaccharyl transferase STT3 subunit	Lokiarch_28460	contig23	31031	10	1285	**ref|XP_002979240.1 oligosaccharyltransferase subunit STT3 (Selaginella moellendorffii) -Expect = 3e-86**	IPR003674 Oligosaccharyl transferase, STT3 subunit	Homologs also present in archaea.

Explanation: *alignments and/or phylogenetic analyses are presented elsewhere in the manuscript for all proteins marked with asterisk (including all small GTPases); proteins showing highest similarities to eukaryotic homologs are highlighted in bold

Suppl. Table S7: Overview of the ribosomal proteins found in Lokiarchaeaum and the LGCG14 metagenome

Ribosomal proteins in Lokiarchaeum based on arCOG attribution								
GeneID	length (AA)	arCOG	COG	COG description		Funclass	Contig	Contig lenght (bp)
Lokiarch_26970	65	arCOG04057		**Ribosomal protein L38E**	J	Translation; ribosomal structure and biogenesis	contig3	57269
Lokiarch_00050	243	arCOG04289	COG00081	**Ribosomal protein L1**	J	Translation; ribosomal structure and biogenesis	contig347	4890
Lokiarch_00040	296	arCOG04288	COG00244	**Ribosomal protein L10**	J	Translation; ribosomal structure and biogenesis	contig347	4890
Lokiarch_03150	102	arCOG04113	COG00197	**Ribosomal protein L10AE/L16**	J	Translation; ribosomal structure and biogenesis	contig225	8877
Lokiarch_40400	225	arCOG04113	COG00197	**Ribosomal protein L10AE/L16**	J	Translation; ribosomal structure and biogenesis	contig81	17799
Lokiarch_00060	139	arCOG04372	COG00080	**Ribosomal protein L11**	J	Translation; ribosomal structure and biogenesis	contig347	4890
Lokiarch_00030	110	arCOG04287	COG02058	**Ribosomal protein L12E/L44/L45/RPP1/RPP2**	J	Translation; ribosomal structure and biogenesis	contig347	4890
Lokiarch_01040	160	arCOG04242	COG00102	**Ribosomal protein L13**	J	Translation; ribosomal structure and biogenesis	contig2	58255
Lokiarch_44410	142	arCOG04095	COG00093	**Ribosomal protein L14**	J	Translation; ribosomal structure and biogenesis	contig290	6486
Lokiarch_24180	142	arCOG04095	COG00093	**Ribosomal protein L14**	J	Translation; ribosomal structure and biogenesis	contig91	16548
Lokiarch_10190	91	arCOG04167	COG02163	**Ribosomal protein L14E/L6E/L27E**	J	Translation; ribosomal structure and biogenesis	contig164	11945
Lokiarch_44290	56	arCOG00779	COG00200	**Ribosomal protein L15**	J	Translation; ribosomal structure and biogenesis	contig290	6486
Lokiarch_45130	56	arCOG00779	COG00200	**Ribosomal protein L15**	J	Translation; ribosomal structure and biogenesis	contig476	2712
Lokiarch_20790	129	arCOG00779	COG00200	**Ribosomal protein L15**	J	Translation; ribosomal structure and biogenesis	contig595	1430
Lokiarch_50940	195	arCOG04209	COG01632	**Ribosomal protein L15E**	J	Translation; ribosomal structure and biogenesis	contig1	71539
Lokiarch_44320	200	arCOG04088	COG00256	**Ribosomal protein L18**	J	Translation; ribosomal structure and biogenesis	contig290	6486
Lokiarch_45160	200	arCOG04088	COG00256	**Ribosomal protein L18**	J	Translation; ribosomal structure and biogenesis	contig476	2712
Lokiarch_01050	119	arCOG00780	COG01727	**Ribosomal protein L18E**	J	Translation; ribosomal structure and biogenesis	contig2	58255
Lokiarch_44330	152	arCOG04089	COG02147	**Ribosomal protein L19E**	J	Translation; ribosomal structure and biogenesis	contig290	6486
Lokiarch_45170	152	arCOG04089	COG02147	**Ribosomal protein L19E**	J	Translation; ribosomal structure and biogenesis	contig476	2712
Lokiarch_48250	241	arCOG04067	COG00090	**Ribosomal protein L2**	J	Translation; ribosomal structure and biogenesis	contig185	11018
Lokiarch_03370	241	arCOG04067	COG00090	**Ribosomal protein L2**	J	Translation; ribosomal structure and biogenesis	contig58	21634
Lokiarch_13720	76	arCOG04175	COG02157	**Ribosomal protein L20A (L18A)**	J	Translation; ribosomal structure and biogenesis	contig14	34782
Lokiarch_16870	107	arCOG04129	COG02139	**Ribosomal protein L21E**	J	Translation; ribosomal structure and biogenesis	contig17	33755
Lokiarch_15120	160	arCOG04098	COG00091	**Ribosomal protein L22**	J	Translation; ribosomal structure and biogenesis	contig465	2904
Lokiarch_24120	160	arCOG04098	COG00091	**Ribosomal protein L22**	J	Translation; ribosomal structure and biogenesis	contig91	16548
Lokiarch_50520	88	arCOG04072	COG00089	**Ribosomal protein L23**	J	Translation; ribosomal structure and biogenesis	contig1	71539
Lokiarch_44400	143	arCOG04094	COG00198	**Ribosomal protein L24**	J	Translation; ribosomal structure and biogenesis	contig290	6486
Lokiarch_24190	143	arCOG04094	COG00198	**Ribosomal protein L24**	J	Translation; ribosomal structure and biogenesis	contig91	16548
Lokiarch_06260	61	arCOG01950	COG02075	**Ribosomal protein L24E**	J	Translation; ribosomal structure and biogenesis	contig187	10803
Lokiarch_19280	68	arCOG00785	COG00255	**Ribosomal protein L29**	J	Translation; ribosomal structure and biogenesis	contig539	1841
Lokiarch_24150	68	arCOG00785	COG00255	**Ribosomal protein L29**	J	Translation; ribosomal structure and biogenesis	contig91	16548
Lokiarch_50500	340	arCOG04070	COG00087	**Ribosomal protein L3**	J	Translation; ribosomal structure and biogenesis	contig1	71539
Lokiarch_44300	176	arCOG04086	COG01841	**Ribosomal protein L30**	J	Translation; ribosomal structure and biogenesis	contig290	6486
Lokiarch_45140	176	arCOG04086	COG01841	**Ribosomal protein L30**	J	Translation; ribosomal structure and biogenesis	contig476	2712
Lokiarch_36610	124	arCOG01752	COG01911	**Ribosomal protein L30E**	J	Translation; ribosomal structure and biogenesis	contig100	15651
Lokiarch_42110	124	arCOG01752	COG01911	**Ribosomal protein L30E**	J	Translation; ribosomal structure and biogenesis	contig310	5802
Lokiarch_13740	151	arCOG04473	COG02097	**Ribosomal protein L31E**	J	Translation; ribosomal structure and biogenesis	contig14	34782
Lokiarch_44340	162	arCOG00781	COG01717	**Ribosomal protein L32E**	J	Translation; ribosomal structure and biogenesis	contig290	6486
Lokiarch_31720	104	arCOG00781	COG01717	**Ribosomal protein L32E**	J	Translation; ribosomal structure and biogenesis	contig596	1426
Lokiarch_16730	98	arCOG04304	COG02451	**Ribosomal protein L35AE/L33A**	J	Translation; ribosomal structure and biogenesis	contig17	33755
Lokiarch_20520	96	arCOG04208	COG01997	**Ribosomal protein L37AE/L43A**	J	Translation; ribosomal structure and biogenesis	contig7	43485
Lokiarch_23020	74	arCOG04126	COG02126	**Ribosomal protein L37E**	J	Translation; ribosomal structure and biogenesis	contig485	2628
Lokiarch_13750	54	arCOG04177	COG02167	**Ribosomal protein L39E**	J	Translation; ribosomal structure and biogenesis	contig14	34782
Lokiarch_50510	261	arCOG04071	COG00088	**Ribosomal protein L4**	J	Translation; ribosomal structure and biogenesis	contig1	71539
Lokiarch_36520	94	arCOG04109	COG01631	**Ribosomal protein L44E**	J	Translation; ribosomal structure and biogenesis	contig100	15651
Lokiarch_05250	94	arCOG04109	COG01631	**Ribosomal protein L44E**	J	Translation; ribosomal structure and biogenesis	contig329	5214
Lokiarch_44380	198	arCOG04092	COG00094	**Ribosomal protein L5**	J	Translation; ribosomal structure and biogenesis	contig290	6486
Lokiarch_24210	198	arCOG04092	COG00094	**Ribosomal protein L5**	J	Translation; ribosomal structure and biogenesis	contig91	16548
Lokiarch_44350	193	arCOG04090	COG00097	**Ribosomal protein L6P**	J	Translation; ribosomal structure and biogenesis	contig290	6486
Lokiarch_31710	193	arCOG04090	COG00097	**Ribosomal protein L6P**	J	Translation; ribosomal structure and biogenesis	contig596	1426
Lokiarch_06270	90	arCOG01751	COG01358	**Ribosomal protein L7AE**	J	Translation; ribosomal structure and biogenesis	contig187	10803
Lokiarch_42080	104	arCOG01751	COG01358	**Ribosomal protein L7AE**	J	Translation; ribosomal structure and biogenesis	contig214	9381
Lokiarch_37590	95	arCOG01758	COG00051	**Ribosomal protein S10**	J	Translation; ribosomal structure and biogenesis	contig224	8914
Lokiarch_12810	131	arCOG04240	COG00100	**Ribosomal protein S11**	J	Translation; ribosomal structure and biogenesis	contig134	13501
Lokiarch_36580	146	arCOG04255	COG00048	**Ribosomal protein S12**	J	Translation; ribosomal structure and biogenesis	contig100	15651
Lokiarch_42140	146	arCOG04255	COG00048	**Ribosomal protein S12**	J	Translation; ribosomal structure and biogenesis	contig310	5802
Lokiarch_12830	164	arCOG01722	COG00099	**Ribosomal protein S13**	J	Translation; ribosomal structure and biogenesis	contig134	13501
Lokiarch_02960	153	arCOG04185	COG00184	**Ribosomal protein S15P**	J	Translation; ribosomal structure and biogenesis	contig39	25066
Lokiarch_19260	114	arCOG04096	COG00186	**Ribosomal protein S17**	J	Translation; ribosomal structure and biogenesis	contig539	1841
Lokiarch_24170	118	arCOG04096	COG00186	**Ribosomal protein S17**	J	Translation; ribosomal structure and biogenesis	contig91	16548
Lokiarch_50390	65	arCOG01885	COG01383	**Ribosomal protein S17E**	J	Translation; ribosomal structure and biogenesis	contig1	71539
Lokiarch_15110	192	arCOG04099	COG00185	**Ribosomal protein S19**	J	Translation; ribosomal structure and biogenesis	contig465	2904
Lokiarch_24110	192	arCOG04099	COG00185	**Ribosomal protein S19**	J	Translation; ribosomal structure and biogenesis	contig91	16548
Lokiarch_13770	142	arCOG01344	COG02238	**Ribosomal protein S19E (S16A)**	J	Translation; ribosomal structure and biogenesis	contig14	34782
Lokiarch_39970	276	arCOG04245	COG00052	**Ribosomal protein S2**	J	Translation; ribosomal structure and biogenesis	contig50	22943
Lokiarch_51720	273	arCOG04245	COG00052	**Ribosomal protein S2**	J	Translation; ribosomal structure and biogenesis	contig89	16658
Lokiarch_06340	118	arCOG04182	COG02004	**Ribosomal protein S24E**	J	Translation; ribosomal structure and biogenesis	contig187	10803
Lokiarch_42010	118	arCOG04182	COG02004	**Ribosomal protein S24E**	J	Translation; ribosomal structure and biogenesis	contig214	9381
Lokiarch_09680	71	arCOG04327	COG04901	**Ribosomal protein S25**	J	Translation; ribosomal structure and biogenesis	contig20	31562
Lokiarch_38480	100	arCOG04327	COG04901	**Ribosomal protein S25**	J	Translation; ribosomal structure and biogenesis	contig37	25983
Lokiarch_06330	70	arCOG04183	COG01998	**Ribosomal protein S27AE**	J	Translation; ribosomal structure and biogenesis	contig187	10803
Lokiarch_42020	70	arCOG04183	COG01998	**Ribosomal protein S27AE**	J	Translation; ribosomal structure and biogenesis	contig214	9381
Lokiarch_36510	32	arCOG04108	COG02051	**Ribosomal protein S27E**	J	Translation; ribosomal structure and biogenesis	contig100	15651
Lokiarch_25970	62	arCOG04108	COG02051	**Ribosomal protein S27E**	J	Translation; ribosomal structure and biogenesis	contig491	2557
Lokiarch_06280	84	arCOG04314	COG02053	**Ribosomal protein S28E/S33**	J	Translation; ribosomal structure and biogenesis	contig187	10803
Lokiarch_42070	87	arCOG04314	COG02053	**Ribosomal protein S28E/S33**	J	Translation; ribosomal structure and biogenesis	contig214	9381
Lokiarch_19290	267	arCOG04097	COG00092	**Ribosomal protein S3**	J	Translation; ribosomal structure and biogenesis	contig539	1841
Lokiarch_24140	264	arCOG04097	COG00092	**Ribosomal protein S3**	J	Translation; ribosomal structure and biogenesis	contig91	16548
Lokiarch_37880	69	arCOG04293	COG04919	**Ribosomal protein S30**	J	Translation; ribosomal structure and biogenesis	contig198	10131
Lokiarch_10570	69	arCOG04293	COG04919	**Ribosomal protein S30**	J	Translation; ribosomal structure and biogenesis	contig52	22764
Lokiarch_02950	249	arCOG04186	COG01890	**Ribosomal protein S3AE**	J	Translation; ribosomal structure and biogenesis	contig39	25066
Lokiarch_12820	173	arCOG04239	COG00522	**Ribosomal protein S4 or related protein**	J	Translation; ribosomal structure and biogenesis	contig134	13501
Lokiarch_44390	252	arCOG04093	COG01471	**Ribosomal protein S4E**	J	Translation; ribosomal structure and biogenesis	contig290	6486
Lokiarch_24200	252	arCOG04093	COG01471	**Ribosomal protein S4E**	J	Translation; ribosomal structure and biogenesis	contig91	16548
Lokiarch_44310	205	arCOG04087	COG00098	**Ribosomal protein S5**	J	Translation; ribosomal structure and biogenesis	contig290	6486
Lokiarch_45150	205	arCOG04087	COG00098	**Ribosomal protein S5**	J	Translation; ribosomal structure and biogenesis	contig476	2712
Lokiarch_26190	135	arCOG01946	COG02125	**Ribosomal protein S6E (S10)**	J	Translation; ribosomal structure and biogenesis	contig130	13861
Lokiarch_06250	135	arCOG01946	COG02125	**Ribosomal protein S6E (S10)**	J	Translation; ribosomal structure and biogenesis	contig187	10803
Lokiarch_36570	206	arCOG04254	COG00049	**Ribosomal protein S7**	J	Translation; ribosomal structure and biogenesis	contig100	15651
Lokiarch_42150	206	arCOG04254	COG00049	**Ribosomal protein S7**	J	Translation; ribosomal structure and biogenesis	contig310	5802
Lokiarch_44360	128	arCOG04091	COG00096	**Ribosomal protein S8**	J	Translation; ribosomal structure and biogenesis	contig290	6486
Lokiarch_31700	128	arCOG04091	COG00096	**Ribosomal protein S8**	J	Translation; ribosomal structure and biogenesis	contig596	1426
Lokiarch_32730	177	arCOG04154	COG02007	**Ribosomal protein S8E**	J	Translation; ribosomal structure and biogenesis	contig313	5698
Lokiarch_41110	127	arCOG04154	COG02007	**Ribosomal protein S8E**	J	Translation; ribosomal structure and biogenesis	contig459	2996
Lokiarch_01030	137	arCOG04243	COG00103	**Ribosomal protein S9**	J	Translation; ribosomal structure and biogenesis	contig2	58255

Ribosomal proteins in Lokiarchaeum starting with alternative start codons (identified by tblastn)									
GeneID	length (AA)	arCOG	COG	COG description		Funclass	predicted start codon	Contig	Contig length (bp)
Lokiarch_26975	54	arCOG04049	COG01552	**Ribosomal protein L40E**	J	Translation; ribosomal structure and biogenesis	GTG	contig3	57269
Lokiarch_10175	90	arCOG04168	COG02174	**Ribosomal protein L34E**	J	Translation; ribosomal structure and biogenesis	ATT	contig164	11945
Lokiarch_44375	50	arCOG00782	COG00199	**Ribosomal protein S14**	J	Translation; ribosomal structure and biogenesis	ATT	contig290	6486
Lokiarch_24225	50	arCOG00782	COG00199	**Ribosomal protein S14**	J	Translation; ribosomal structure and biogenesis	ATT	contig91	16548

Ribosomal proteins present in LCGC14 metagenome and assigned to Lokiarchaeum, but not part of the bin due to coverage filtering criteria									
GeneID	length (AA)	arCOG	COG	COG description		Funclass	method; e-value; (tblastn: start; stop; frame; predicted start codon)	Phymbl-bin	notes
LCGC14_0502170	198	arCOG01013	COG04352	**Ribosomal protein L13E**	J	Translation; ribosomal structure and biogenesis	psiblast; 5e-07	Lokiarchaeum	Additional L13E homologs present in metagenome, however, the contigs seem to be binned incorrectly to Bacteria rather than
LCGC14_1193320	201	arCOG01013	COG04352	**Ribosomal protein L13E**	J	Translation; ribosomal structure and biogenesis	psiblast; 9e-07	Loki2/3	
LCGC14_0804220	202	arCOG01013	COG04352	**Ribosomal protein L13E**	J	Translation; ribosomal structure and biogenesis	psiblast; 1e-05	Lokiarchaeum	
LCGC14_0916275	89	arCOG04305	COG04830	**Ribosomal protein S26E**	J	Translation; ribosomal structure and biogenesis	tblastn (q: *A. hospitalis*); 9e-15; 5472; 5206; -3; ATT	Lokiarchaeum	Additional S26E homologs present in metagenome, however, on short contigs that seem to be binned incorrectly to Bacteria rather than Lokiarchaeum.
LCGC14_0657145	92	arCOG04305	COG04830	**Ribosomal protein S26E**	J	Translation; ribosomal structure and biogenesis	tblastn (q: *A. hospitalis*); 2e-10; 657; 436; -2; ATT	Lokiarchaeum	

Novel eukarya-specific ribosomal proteins potentially present in Lokiarchaeum									
GeneID	length (AA)	arCOG	COG	COG description		Funclass	notes	Contig	Contig length (bp)
Lokiarch_30160	103	-	KOG3434	**putative homolog of eukaryotic ribosomal protein L22e**	J	Translation; ribosomal structure and biogenesis	not found in archaea so far	contig150	12534

Suppl. Table S8: Distribution of Archaea- /Eukarya-specific ribosomal proteins in different archaeal lineages based on arCOG assignments.

		arCOG04108	arCOG04126	arCOG01752	arCOG04167	arCOG04168	arCOG04327	arCOG04305	arCOG04293	arCOG04175	arCOG04304	arCOG04057	arCOG01013	arCOG06624	KOG3434
Species	**Order and/or Phylum**	**S27ae**	**L37e**	**L30e**	**L14e**	**L34e**	**S25e**	**S26e**	**S30e**	**L18ae/L20e/LXa**	**L35ae**	**L38e**	**L13e**	**L41e**	**L22e**
Thermofilum pendens	*Thermoproteales* (Crenarchaeota)	◆	◆	◆	◆	◆	◆	◆	◆	◆	◆		◆		
Geoarchaeon NAG1	*Thermoproteales* (Crenarchaeota)	◆	◆	◆	◆	◆	◆	◆	◆	◆		◆	◆		
Thermoproteus neutrophilus	*Thermoproteales* (Crenarchaeota)	◆	◆	◆	◆	◆	◆	◆	◆	◆		◆	◆		
Pyrobaculum aerophilum	*Thermoproteales* (Crenarchaeota)	◆	◆	◆	◆	◆	◆	◆	◆	◆		◆	◆		
Pyrobaculum arsenaticum	*Thermoproteales* (Crenarchaeota)	◆	◆	◆	◆	◆	◆	◆	◆	◆		◆	◆		
Pyrobaculum calidifontis	*Thermoproteales* (Crenarchaeota)	◆	◆	◆	◆	◆	◆	◆	◆	◆		◆	◆		
Pyrobaculum islandicum	*Thermoproteales* (Crenarchaeota)	◆	◆	◆	◆	◆	◆	◆	◆	◆		◆	◆		
Caldivirga maquilingensis	*Thermoproteales* (Crenarchaeota)	◆	◆	◆	◆	◆	◆	◆	◆	◆		◆	◆		
Thermosphaera aggregans	*Desulfurococcales* (Crenarchaeota)	◆	◆	◆	◆	◆	◆	◆	◆	◆					
Desulfurococcus kamchatkensis	*Desulfurococcales* (Crenarchaeota)	◆	◆	◆	◆	◆	◆	◆	◆	◆					
Acidilobus saccharovorans	*Desulfurococcales* (Crenarchaeota)	◆	◆	◆	◆		◆	◆	◆	◆		◆			
Aeropyrum pernix	*Desulfurococcales* (Crenarchaeota)	◆	◆	◆	◆	◆	◆	◆	◆	◆	◆	◆	◆		
Hyperthermus butylicus	*Desulfurococcales* (Crenarchaeota)	◆	◆	◆	◆	◆	◆	◆	◆	◆	◆	◆	◆		
Ignicoccus hospitalis	*Desulfurococcales* (Crenarchaeota)	◆	◆	◆	◆	◆	◆	◆	◆	◆	◆	◆	◆		
Staphylothermus hellenicus	*Desulfurococcales* (Crenarchaeota)	◆	◆	◆	◆	◆	◆	◆	◆	◆	◆	◆	◆		
Metallosphaera sedula	*Sulfolobales* (Crenarchaeota)	◆	◆	◆	◆	◆	◆	◆	◆	◆					
Sulfolobus acidocaldarius	*Sulfolobales* (Crenarchaeota)	◆	◆	◆	◆	◆	◆	◆	◆	◆					
Sulfolobus islandicus YN1551	*Sulfolobales* (Crenarchaeota)	◆	◆	◆	◆	◆	◆	◆	◆	◆			◆		
Sulfolobus solfataricus	*Sulfolobales* (Crenarchaeota)	◆	◆	◆	◆	◆	◆	◆	◆	◆			◆		
Sulfolobus tokodaii	*Sulfolobales* (Crenarchaeota)	◆	◆	◆	◆	◆	◆	◆	◆	◆			◆		
Korarchaeum cryptofilum	Korarchaeota	◆	◆	◆	◆	◆	◆	◆	◆				◆		
Nitrososphaera gargensis	Thaumarchaeota	◆	◆	◆			◆	◆	◆				◆		
Caldiarchaeum subterraneum	Thaum/ (Aig)archaeota	◆	◆	◆	◆		◆	◆	◆				◆		
Lokiarchaeum	Lokiarchaeota	◆	◆	◆	◆	◆	◆	◆	◆	◆	◆	◆	◆		◆
Nanoarchaeum equitans	Nanoarchaeota	◆	◆	◆	◆	◆		◆		◆	◆			◆	
Thermococcus gammatolerans	*Thermococcales* (Euryarchaeota)	◆	◆	◆	◆	◆				◆	◆			◆	
Methanobrevibacter smithii	*Methanobacteriales* (Euryarchaeota)	◆	◆	◆	◆	◆				◆					
Methanococcus jannaschii	*Methanococcales* (Euryarchaeota)	◆	◆	◆	◆	◆				◆				◆	
Methanopyrus_kandleri	*Methanopyrales* (Euryarchaeota)	◆	◆	◆	◆	◆		◆		◆	◆				
Aciduliprofundum boonei	(Euryarchaeota)	◆	◆	◆						◆				◆	
Archaeoglobus profundus	*Archaeoglobales* (Euryarchaeota)	◆	◆	◆				◆		◆					
Methanoregula boonei	*Methanomicrobiales* (Euryarchaeota)	◆	◆	◆						◆					
Methanocella paludicola	*Methanocellales* (Euryarchaeota)	◆	◆	◆						◆					
Methanosarcina acetivorans	*Methanosarcinales* (Euryarchaeota)	◆	◆	◆						◆					
Thermoplasma acidophilum	*Thermoplasmatales* (Euryarchaeota)	◆	◆											◆	
Haloarcula marismortui	*Halobacteriales* (Euryarchaeota)	◆	◆							◆					
Eukaryotes		◆	◆	◆	◆	◆	◆	◆	◆	◆	◆	◆	◆	◆	◆

Suppl. Table S9: Annotation of important lokiarchaeal informational processing genes

DNA-directed RNA polymerase												
Gene	**Locus**	**Contig**	**Contig length**	**ORF number**	**length (AA)**	**Top Blast Hit***	**arCOG**	**COG**	**COG description**		**Funclass**	**comment**
DNA-directed RNA polymerase, subunit A''	Lokiarch_36620	contig100	15651	12	404	ref|WP_012185255.1| DNA-directed RNA polymerase subunit A'' [Caldivirga maquilingensis] - Expect = 3e-110 (44%)	arCOG04256	COG0086	DNA-directed RNA polymerase subunit A''	K	Transcription	95% identity to each other
DNA-directed RNA polymerase, subunit A''	Lokiarch_42100	contig310	5802	2	402	ref|WP_012185255.1| DNA-directed RNA polymerase subunit A'' [Caldivirga maquilingensis] - Expect = 3e-110 (44%)	arCOG04256	COG0086	DNA-directed RNA polymerase subunit A''	K	Transcription	
DNA-directed RNA polymerase, subunit A'	Lokiarch_36630	contig100	15651	13	990	ref|WP_023991383.1| DNA-directed RNA polymerase subunit A' [Methanobacterium sp. MB1] - Expect = 0.0 (56%)	arCOG04257	COG0086	DNA-directed RNA polymerase subunit A'	K	Transcription	ca. 92% identity to each other
partial DNA-directed RNA polymerase, subunit A'	Lokiarch_42090	contig310	5802	1	151	ref|WP_013799670.1| DNA-directed RNA polymerase subunit A' [Methanotorris igneus] - Expect = 2e-17 (61%)	arCOG04257	COG0086	DNA-directed RNA polymerase subunit A'	K	Transcription	
partial DNA-directed RNA polymerase, subunit A'	Lokiarch_24950	contig474	2743	1	852	ref|WP_013194249.1| DNA-directed RNA polymerase subunit A' [Methanohalobium evestigatum] - Expect = 0.0 (56%)	arCOG04257	COG0086	DNA-directed RNA polymerase subunit A'	K	Transcription	
DNA-directed RNA polymerase, subunit B	Lokiarch_29070	contig291	6442	3	1122	ref|WP_009990478.1| DNA-directed RNA polymerase subunit B [Sulfolobus solfataricus] - Expect = 0.0; (57%)	arCOG01762	COG00085	DNA-directed RNA polymerase subunit B	K	Transcription	99% identity to each other
DNA-directed RNA polymerase, subunit B	Lokiarch_03330	contig58	21634	4	1127	ref|WP_009990478.1| DNA-directed RNA polymerase subunit B [Sulfolobus solfataricus] - Expect = 0.0; (57%)	arCOG01762	COG00085	DNA-directed RNA polymerase subunit B	K	Transcription	
DNA-directed RNA polymerase, subunit D	Lokiarch_01060	contig2	58255	23	277	ref|WP_013143934.1| DNA-directed RNA polymerase subunit D [Staphylothermus hellenicus] - Expect = 1e-64; (41%)	arCOG04241	COG00202	DNA-directed RNA polymerase subunit D	K		
DNA-directed RNA polymerase, subunit E`	Lokiarch_06370	contig187	10803	15	199	ref|WP_014788599.1| DNA-directed RNA polymerase subunit E' [Thermococcus sp. CL1] - Expect = 3e-58 (47%)	arCOG00675	COG01095	DNA-directed RNA polymerase, subunit E'	K	Transcription	98% identity to each other
DNA-directed RNA polymerase, subunit E`	Lokiarch_41980	contig214	9381	6	199	ref|WP_014788599.1| DNA-directed RNA polymerase subunit E' [Thermococcus sp. CL1] - Expect = 7e-58 (46%)	arCOG00675	COG01095	DNA-directed RNA polymerase, subunit E'	K	Transcription	
DNA-directed RNA polymerase, subunit E``	Lokiarch_41990	contig214	9381	7	62	ref|WP_013194048.1| DNA-directed RNA polymerase subunit E'' [Methanohalobium evestigatum] - Expect = 9e-15 (55%)	arCOG04077	COG02093	Transcription elongation factor Spt4/RpoE2, zinc finger protein	K	Transcription	94% identity to each other
DNA-directed RNA polymerase, subunit E``	Lokiarch_06360	contig187	10803	14	62	ref|WP_013194048.1| DNA-directed RNA polymerase subunit E'' [Methanohalobium evestigatum] - Expect = 1e-14 (55%)	arCOG04077	COG02093	Transcription elongation factor Spt4/RpoE2, zinc finger protein	K	Transcription	
DNA-directed RNA polymerase, subunit F	Lokiarch_16880	contig17	33755	24	121	gb|AIE99443.1| DNA-directed RNA polymerase subunit F (rpoF) [uncultured marine thaumarchaeote KM3_110_B01] - Expect = 3e-09 (37%)	arCOG01016	COG01460	DNA-directed RNA polymerase, subunit F (rpoF)	K	Transcription	
DNA-directed RNA polymerase, subunit H	Lokiarch_29060	contig291	6442	2	85	sp|P31815.1|RPOH_THECE RecName: Full=DNA-directed RNA polymerase subunit H [Thermococcus celer] - Expect = 3e-16 (44%)	arCOG04258	COG02012	DNA-directed RNA polymerase, subunit H, RpoH/RPB5	K	Transcription	95% identity to each other
DNA-directed RNA polymerase, subunit H	Lokiarch_03340	contig58	21634	5	85	sp|P31815.1|RPOH_THECE RecName: Full=DNA-directed RNA polymerase subunit H [Thermococcus celer] - Expect = 3e-16 (44%)	arCOG04258	COG02012	DNA-directed RNA polymerase, subunit H, RpoH/RPB5	K	Transcription	
DNA-directed RNA polymerase, subunit K	Lokiarch_17320	contig636	1193	2	100	ref|WP_011751965.1| DNA-directed RNA polymerase subunit K [Thermofilum pendens] - Expect = 9e-19 (50%)	arCOG01268	COG01758	DNA-directed RNA polymerase subunit K/omega	K	Transcription	
putative DNA-directed RNA polymerase, subunit L	Lokiarch_28530	contig23	31031	17	164	-	arCOG04111	COG01761	DNA-directed RNA polymerase, subunit L	K	Transcription	significant domain hit but low sequence conservation (ca. 30%), start codon might be wrongly predicted
hypothetical protein with partial similarity to transcription elongation factor S/DNA-directed RNA polymerase, subunit M	Lokiarch_01920	contig137	13309	6	61	ref|WP_011033344.1| DNA-directed RNA polymerase subunit M [Methanosarcina mazei] - Expect = 5e-04 (35%)	arCOG00579	COG01594	DNA-directed RNA polymerase, subunit M/Transcription elongation factor TFIIS	K	Transcription	
hypothetical protein with partial similarity to transcription elongation factor S/DNA-directed RNA polymerase, subunit M	Lokiarch_03010	contig39	25066	28	83	ref|WP_011972548.1| DNA-directed RNA polymerase subunit M [Methanococcus vannielii] - Expect = 9e-06 (42%)	arCOG00579	COG01594	DNA-directed RNA polymerase, subunit M/Transcription elongation factor TFIIS	K	Transcription	
hypothetical protein with partial similarity to transcription elongation factor S/DNA-directed RNA polymerase, subunit M	Lokiarch_13080	contig166	11690	7	61	ref|WP_018154590.1| DNA-directed RNA polymerase subunit M [Methanothermococcus thermolithotrophicus] - Expect = 4e-04 (42%)	arCOG00579	COG01594	DNA-directed RNA polymerase, subunit M/Transcription elongation factor TFIIS	K	Transcription	
putative DNA-directed RNA polymerase, subunit M/Transcription elongation factor TFIIS	Lokiarch_25960	contig491	2557	4	114	ref|WP_013128916.1| DNA-directed RNA polymerase subunit M [Thermosphaera aggregans] - Expect = 5e-19 (43%)	arCOG00579	COG01594	DNA-directed RNA polymerase, subunit M/Transcription elongation factor TFIIS	K	Transcription	
putative DNA-directed RNA polymerase, subunit M/Transcription elongation factor TFIIS	Lokiarch_30140	contig150	12534	8	116	gb|EWG07791.1| transcription termination factor Tfs [Sulfolobales archaeon AZ1] - Expect = 8e-16 (42%)	arCOG00579	COG01594	DNA-directed RNA polymerase, subunit M/Transcription elongation factor TFIIS	K	Transcription	
DNA-directed RNA polymerase, subunit N	Lokiarch_26180	contig130	13861	1	66	emb|CDS26753.1| DNA directed RNA polymerases I, II, and III [Hymenolepis microstoma] - Expect = 2e-19 (56%)	arCOG04244	COG01644	DNA-directed RNA polymerase, subunit N (RpoN/RPB10)	K	Transcription	
DNA-directed RNA polymerase, subunit P	Lokiarch_34120	contig6	43995	29	49	gb|AAK96096.1|AF393466_34 Zn-ribbon protein [uncultured crenarchaeote 74A4] - Expect = 4e-06 (53%)	arCOG04341	COG01996	DNA-directed RNA polymerase, subunit RPC12/RpoP (contains C4-type Zn-finger)	K	Transcription	
DNA polymerases												
Product	**Locus**	**Contig**	**Contig length**	**ORF number**	**length (AA)**	**Top Blast Hit***	**arCOG**	**COG**	**COG description**		**Funclass**	**comment**
DNA polymerase I (B1-family)	Lokiarch_39310	contig108	15300	5	981	dbj|BAE95222.1| DNA-directed DNA polymerase B [unclutured Candidatus Nitrosocaldus sp.] - Expect = 0.0 (40%)	arCOG00328	COG00417	DNA polymerase elongation subunit (family B)	L	Replication; recombination and repair	see Makarova et al., 2014 for classification archaeal DNA polymerases
putative DNA polymerase (inactivated B2-family)	Lokiarch_35900	contig22	31076	2	804	ref|WP_015020238.1| DNA polymerase B [Candidatus Nitrososphaera gargensis] - Expect = 2e-126 (33%)	arCOG00329	COG00417	DNA polymerase elongation subunit (family B), inactivated	L	Replication; recombination and repair	"N. gargensis/Leptospirillum-like"
DNA polymerase D (II), large subunit	Lokiarch_35660	contig261	7637	3	1153	ref|WP_004068996.1| DNA polymerase II [Thermococcus litoralis] - Expect = 0.0 (44%)	arCOG04447	COG01933	Archaeal DNA polymerase II, large subunit	L	Replication; recombination and repair	92% identity between each other
DNA polymerase D (II), large subunit	Lokiarch_04310	contig286	6610	3	1153	ref|WP_014405073.1| DNA polymerase II [Methanocella conradii] - Expect = 0.0 (45%)	arCOG04447	COG01933	Archaeal DNA polymerase II, large subunit	L	Replication; recombination and repair	
DNA polymerase D (II), small subunit	Lokiarch_35650	contig261	7637	2	586	ref|WP_019262706.1| DNA polymerase II [Methanobrevibacter smithii] - Expect = 4e-90 (35%)	arCOG04455	COG01311	Archaeal DNA polymerase II, small subunit/DNA polymerase delta, subunit B	L	Replication; recombination and repair	
DNA polymerase D (II), small subunit	Lokiarch_04300	contig286	6610	2	586	ref|WP_019262706.1| DNA polymerase II [Methanobrevibacter smithii] - Expect = 3e-91 (35%)	arCOG04455	COG01311	Archaeal DNA polymerase II, small subunit/DNA polymerase delta, subunit B	L	Replication; recombination and repair	
Replication												
Product	**Locus**	**Contig**	**Contig length**	**ORF number**	**length (AA)**	**Top Blast Hit***	**arCOG**	**COG**	**COG description**		**Funclass**	**comment**
ERCC4-like helicase/ Hef nuclease	Lokiarch_33440	contig116	14764	2	771	ref|WP_010885957.1| DEAD/DEAH box helicase [Pyrococcus horikoshii] - Expect = 0.0 (44%)	arCOG00872	COG01111	ERCC4-like helicase	L	Replication; recombination and repair	
ERCC4-like helicase/ Hef nuclease	Lokiarch_04610	contig403	3729	1	717	ref|WP_010885957.1| DEAD/DEAH box helicase [Pyrococcus horikoshii] - Expect = 0.0 (43%)	arCOG00872	COG01111	ERCC4-like helicase	L	Replication; recombination and repair	92.89 % identity to Lokiarch_33440
putative ERCC4-like helicase/ Hef nuclease	Lokiarch_40920	contig79	18091	8	717	ref|WP_011249972.1| DEAD/DEAH box helicase [Thermococcus kodakarensis] - Expect = 1e-115 (39%)	arCOG00872	COG01111	ERCC4-like helicase	L	Replication; recombination and repair	
ERCC4-like nuclease	Lokiarch_11770	contig302	6079	4	251	DEAD/DEAH box helicase [Pyrococcus furiosus] - Expect = 4e-30 (39%), only ERCC4 domain	arCOG04206	COG01948	ERCC4-type nuclease	L	Replication; recombination and repair	
putative GINS (23)-domain protein	Lokiarch_45390	contig8	42635	16	185	-	arCOG00552	COG01711	DNA replication initiation complex subunit, GINS23 family	L	Replication; recombination and repair	no significant hit in nr, very divergent from both other archaea and eukaryotes
putative DNA replication initiation complex subunit, GINS15 family	Lokiarch_28600	contig23	31031	24	239	ref|WP_013483089.1| hypothetical protein [Cenarchaeum symbiosum] - Expect = 5e-05 (41%, partial overlap only)	arCOG00551	COG01711	DNA replication initiation complex subunit, GINS15 family	L	Replication; recombination and repair	
DNA primase small subunit PriS	Lokiarch_28610	contig23	31031	25	427	gb|AJB42955.1| DNA primase small subunit [Thermofilum carboxyditrophus 1505] - Expect = 1e-71 (36%)	arCOG04110	COG01467	Eukaryotic-type DNA primase, catalytic (small) subunit	L	Replication; recombination and repair	
DNA primase large subunit PriL	Lokiarch_28620	contig23	31031	26	396	ref|WP_014585964.1| DNA primase [Methanosaeta harundinacea] - Expect = 9e-28 (27%)	arCOG03013	COG02219	Eukaryotic-type DNA primase, large subunit	L	Replication; recombination and repair	
PCNA	Lokiarch_25940	contig491	2557	2	266	ref|YP_004424199.1| DNA polymerase sliding clamp [Pyrococcus sp. NA2] - Expect = 6e-34 (33%)	arCOG00488	COG00592	DNA polymerase sliding clamp subunit (PCNA homolog)	L	Replication; recombination and repair	
DNA primase (bacterial type, DnaG)	Lokiarch_13690	contig14	34782	10	394	ref|NP_070724.1| DNA primase [Archaeoglobus fulgidus DSM 4304] - Expect = 3e-97 (47%)	arCOG04281	COG00358	DNA primase (bacterial type)	L	Replication; recombination and repair	
ORC1-type DNA replication protein	Lokiarch_29380	contig640	1175	1	382	ref|WP_012036099.1| cell division control protein Cdc6 [Methanocella arvoryzae] - Expect = 2e-97 (45%)	arCOG00467	COG01474	Cdc6-related protein, AAA superfamily ATPase	L	Replication; recombination and repair	
ORC1-type DNA replication protein	Lokiarch_50410	contig1	71539	11	399	ref|WP_018031295.1| MULTISPECIES: hypothetical protein [unclassified Crenarchaeota (miscellaneous)] - Expect = 8e-76 (37%)	arCOG00467	COG01474	Cdc6-related protein, AAA superfamily ATPase	L	Replication; recombination and repair	
ORC1-type DNA replication protein	Lokiarch_29470	contig227	8809	6	399	ref|WP_020863095.1| archaeal cell division control protein 6 [Candidatus Caldiarchaeum subterraneum] - Expect = 2e-32 (29%)	arCOG00467	COG01474	Cdc6-related protein, AAA superfamily ATPase	L	Replication; recombination and repair	
partial Orc1/cdc6 family related protein	Lokiarch_46930	contig216	9258	1	99	dbj|BAJ48975.1| cell division control protein 6 [Candidatus Caldiarchaeum subterraneum] - Expect = 1e-13 (39%)	arCOG00467	COG01474	Cdc6-related protein, AAA superfamily ATPase	L	Replication; recombination and repair	
putative cell division control protein 6	Lokiarch_51010	contig1	71539	70	399	ref|WP_007693106.1| cell division control protein 6-like protein [Halococcus hamelinensis] - Expect = 2e-04 (22%)	arCOG00467	COG01474	Cdc6-related protein, AAA superfamily ATPase	L	Replication; recombination and repair	
partial cdc6-related protein	Lokiarch_25850	contig488	2617	5	77	ref|WP_013897321.1| cell division control protein Cdc6 [Methanosalsum zhilinae] - Expect = 4e-12 (49%)	arCOG00467	COG01474	Cdc6-related protein, AAA superfamily ATPase	L	Replication; recombination and repair	
partial cdc6-related protein	Lokiarch_20300	contig489	2598	5	77	ref|WP_013897321.1| cell division control protein Cdc6 [Methanosalsum zhilinae] - Expect = 1e-12 (48%)	arCOG00467	COG01474	Cdc6-related protein, AAA superfamily ATPase	L	Replication; recombination and repair	
Minichromosome maintenance protein MCM	Lokiarch_45410	contig8	42635	18	709	gb|AJB42837.1| DNA replication helicase protein MCM [Thermofilum carboxyditrophus 1505] - Expect = 5e-169 (38%)	arCOG00439	COG01241	Predicted ATPase involved in replication control, Cdc46/Mcm family	L	Replication; recombination and repair	
RecJ-like exonuclease	Lokiarch_17790	contig92	16479	4	536	ref|WP_007982244.1| recombinase RecJ [Haladaptatus paucihalophilus] - Expect = 1e-78 (34%)	arCOG00427	COG00608	Single-stranded DNA-specific exonuclease RecJ	L	Replication; recombination and repair	
putative RecJ-like exonuclease	Lokiarch_30120	contig150	12534	6	458	ref|WP_004068618.1| single-stranded DNA endonuclease [Thermococcus litoralis] - Expect = 4e-22 (27%)	arCOG00427	COG00608	Single-stranded DNA-specific exonuclease RecJ	L	Replication; recombination and repair	

RecJ-like exonuclease	Lokiarch_00590	contig246	8326	7	536	ref|WP_007982244.1| recombinase RecJ [Haladaptatus paucihalophilus] - Expect = 1e-79; (34%)	arCOG00427	COG00608	Single-stranded DNA-specific exonuclease RecJ	L	Replication; recombination and repair	90% identity to Lokiarch_17790
Replication factor C, small subunit	Lokiarch_28480	contig23	31031	12	323	ref|WP_015791939.1| ATPase AAA [Methanocaldococcus fervens] - Expect = 2e-112 (51%)	arCOG00469	COG00470	ATPase involved in DNA replication HolB, small subunit	L	Replication; recombination and repair	all three genes seem to encode highly divergent rplC, small subunits suggesting that the duplications are genuine and not due to potential microdiversity in the genome bin; this might indicate that Lokiarchaeaum (like eukaryotes) encodes at least three small subunits of Replication factor C
Replication factor C, small subunit	Lokiarch_12520	contig374	4310	2	322	ref|WP_013897746.1| ATPase AAA [Methanosalsum zhilinae] - Expect = 1e-71 (37%)	arCOG00469	COG00470	ATPase involved in DNA replication HolB, small subunit	L	Replication; recombination and repair	
Replication factor C, small subunit	Lokiarch_12530	contig374	4310	3	348	ref|WP_015791939.1| ATPase AAA [Methanocaldococcus fervens] - Expect = 3e-50 (32%)	arCOG00469	COG00470	ATPase involved in DNA replication HolB, small subunit	L	Replication; recombination and repair	
putative fragment of replication factor C, small subunit	Lokiarch_35580	contig43	23852	18	80	ref|XP_003005043.1| replication factor C subunit 2 [Verticillium alfalfae VaMs.102] - Expect = 4e-22 (55%)	arCOG00469	COG00470	ATPase involved in DNA replication HolB, small subunit	L	Replication; recombination and repair	only NTD (80AA in lenght)
putative replication factor C, large subunit	Lokiarch_12540	contig374	4310	4	313	ref|WP_012123440.1| replication protein C [Ignicoccus hospitalis] - Expect = 1e-42 (33%)	arCOG00470	COG00470	ATPase involved in DNA replication HolB, large subunit	L	Replication; recombination and repair	end of contig, thus CTD (ca. 100AA) is missing
fragment of replication factor C, large subunit	Lokiarch_40320	contig425	3405	4	162	ref|WP_013604029.1| replication protein C [Vulcanisaeta moutnovskia] -Expect = 5e-12 (31%, partial)	arCOG00470	COG00470	ATPase involved in DNA replication HolB, large subunit	L	Replication; recombination and repair	
fragment of replication factor C, large subunit	Lokiarch_44840	contig434	3321	4	162	ref|WP_013604029.1| replication protein C [Vulcanisaeta moutnovskia] - Expect = 2e-12 (32%)	arCOG00470	COG00470	ATPase involved in DNA replication HolB, large subunit	L	Replication; recombination and repair	
Flap endonuclease (Fen-1)	Lokiarch_47590	contig27	29742	17	353	ref|WP_011998725.1|endonuclease [Ignicoccus hospitalis] - Expect = 1e-102 (48%)	arCOG04050	COG00258	5'-3' exonuclease (including N-terminal domain of PolI)	L	Replication; recombination and repair	
NAD-dependent DNA ligase	Lokiarch_01740	contig201	10044	4	636	ref|WP_018525347.1| - NAD-dependent DNA ligase LigA [Spirochaeta alkalica] - Expect = 0.0 (48%)	arCOG04754	COG00272	NAD-dependent DNA ligase	L	Replication; recombination and repair	most similar to bacterial homologs and might have been acquired by HGT (similar to what has been reported for this gene in *Methanomassilicoccus alvus* (Borret et al., 2014)) or could be due to contamination. However, several surrounding genes are most closely related to archaeal protein homologs suggesting that this contig is indeed derived from Lokiarchaeum.
ATP-dependent DNA ligase	Lokiarch_12970	contig264	7590	3	585	gb|EWG07367.1| ATP-dependent DNA ligase [Sulfolobales archaeon AZ1] - Expect = 8e-143 (40%)	arCOG01347	COG01793	ATP-dependent DNA ligase	L	Replication; recombination and repair	
hypothetical protein with OB-fold	Lokiarch_34240	contig6	43995	41	243	WP_012310089.1 OB-fold tRNA/helicase-type nucleic acid binding protein [Candidatus Korarchaeum cryptofilum] - Expect = 7e-05 (27%)	arCOG02261	-	Protein containing OB-fold-like and PCI domains	R	General function prediction only	
hypothetical single-stranded DNA-binding protein with nuclease and OB-fold domains	Lokiarch_14650	contig66	20056	5	358	ref|WP_020863706.1| nuclease [Candidatus Caldiarchaeum subterraneum] - Expect = 3e-24 (47%)	arCOG01510	COG01599	Single-stranded DNA-binding replication protein A (RPA), large (70 kD) subunit or related ssDNA-binding protein	**L**	Replication; recombination and repair	homologous regions (detected with Blast) only comprise nuclease domain
putative replication protein A	Lokiarch_33990	contig6	43995	17	574	ref|WP_011033246.1| replication protein A [Methanosarcina mazei] - Expect = 3e-37 (28%)	arCOG01510	COG01599	Single-stranded DNA-binding replication protein A (RPA), large (70 kD) subunit or related ssDNA-binding protein	**L**	Replication; recombination and repair	
archaeal histone	Lokiarch_38600	contig127	13916	10	63	P95669.3| Archaeal histone HAN1 subunit A [Thermococcus zilligii] - Expect = 1e-05 (43%)	arCOG02144	COG02036	Histones H3 and H4	**L**	Replication; recombination and repair	
archaeal histone	Lokiarch_09210	contig140	13254	9	69	WP_010878990.1| histone [Archaeoglobus fulgidus] - Expect = 2e-13 (54%)	arCOG02144	COG02036	Histones H3 and H4	**L**	Replication; recombination and repair	
archaeal histone	Lokiarch_14270	contig144	12812	1	68	ref|WP_014788258.1| histone [Thermococcus sp. CL1] - Expect = 1e-06 (36%)	arCOG02144	COG02036	Histones H3 and H4	**L**	Replication; recombination and repair	
archaeal histone	Lokiarch_49560	contig162	11997	1	68	ref|WP_020962235.1| hypothetical protein [Thermofilum sp. 1910b] - Expect = 3e-05 (40%)	arCOG02144	COG02036	Histones H3 and H4	**L**	Replication; recombination and repair	
archaeal histone	Lokiarch_40960	contig79	18091	12	75	emb|CAI64119.1| archaeal histone [uncultured archaeon] - Expect = 2e-06 (40%)	arCOG02144	COG02036	Histones H3 and H4	**L**	Replication; recombination and repair	
Topoisomerases												
Product	**Locus**	**Contig**	**Contig length**	**ORF number**	**length (AA)**	**Top Blast Hit***	**arCOG**	**COG**	**COG description**		**Funclass**	**comment**
Topoisomerase IA (type III)	Lokiarch_09080	contig48	23111	16	714	ref|WP_008084369.1| DNA topoisomerase I [Aciduliprofundum boonei] - Expect = 8e-91 (35%)	arCOG01305	COG01637	Topoisomerase DNA binding C4 zinc finger fused to uncharacterized N-terminal domain often associated with RecB-like endonuclease of COG1637	L	Replication; recombination and repair	Topoisomerase IB and reverse gyrase are absent from Lokiarchaeum bin and metagenome (except for thaumarchaeal Topo IB homologs)
putative DNA topoisomerase IIB (VI), subunit B	Lokiarch_45360	contig8	42635	13	1114	gb|EGQ43280.1| DNA topoisomerase VI, B subunit [Candidatus Nanosalina sp. J07AB43] - xpect = 3e-27 (37%)	arCOG01165	COG01389	DNA topoisomerase VI, subunit B	L	Replication; recombination and repair	only very distantly related to known sequences
DNA topoisomerase IIB (VI), subunit A	Lokiarch_45370	contig8	42635	14	395	ref|XP_010937777.1| PREDICTED: DNA topoisomerase 6 subunit A3 [Elaeis guineensis] - Expect = 7e-109 (48%)	arCOG04143	COG01697	DNA topoisomerase VI, subunit A	L	Replication; recombination and repair	most similar to eukaryotic homologs (plant TopoVI, subunit A and spo11) (see Forterre and Gadelle, 2009)
Topoisomerase IIA (DNA gyrase/topoisomerase IV), B subunit	Lokiarch_43500	contig244	8418	6	669	ref|WP_038321290.1| DNA gyrase subunit B [candidate division ZIXI] - Expect = 0.0 (53%)	arCOG04371	COG00187	Type IIA topoisomerase (DNA gyrase/topo II, topoisomerase IV), B subunit	L	Replication; recombination and repair	third blast hit to Methanolobus psychrophilus; subunit A not part of final bin due to stringent coverage binning but candidate Loki-contig present in LCGC14 metagenome(contig1198_16831_5.35806_5)
fragment of Topoisomerase IIA (DNA gyrase/topoisomerase IV), B subunit	Lokiarch_46300	contig421	3444	3	173	ref|WP_021058867.1| type IIA topoisomerase (DNA gyrase/topo II, topoisomerase IV), B subunit, partial [Haloquadratum sp. J07HQX50] - Expect = 2e-64 (59%)	arCOG04371	COG00187	Type IIA topoisomerase (DNA gyrase/topo II, topoisomerase IV), B subunit	L	Replication; recombination and repair	overlapping AA region between Lokiarch_46300 and Lokiarch_43500 is 92% identical
Translation factors												
Product	**Locus**	**Contig**	**Contig length**	**ORF number**	**length (AA)**	**Top Blast Hit***	**arCOG**	**COG**	**COG description**		**Funclass**	**comment**
Translation initiation factor 1	Lokiarch_33975	contig6	43995		105	ref|WP_026068920.1| translation initiation factor Sui1 [Candidatus Methanomassiliicoccus intestinalis] (3e-25; 43%id)	arCOG04223	COG00023	Translation initiation factor 1 (eIF-1/SUI1)	J		potentially alternative start codon: ATT; contig location: 12428-12745
Translation elongation factor aEF-2	Lokiarch_18540	contig145	12777	9	752	ref|WP_018193026.1| MULTISPECIES: translation elongation factor aEF-2 [unclassified Thaumarchaeota (miscellaneous)] - Expect = 4e-176 (40%)	arCOG01559	COG00480	Translation elongation factor G, EF-G (GTPase)	**J**	Translation; ribosomal structure and biogenesis	Lokiarch_28370 and Lokiarch_15310 appear to comprise the NTD and CTD domains of aEF2, respectively and are 94% identical to Lokiarch_18540
partial translation elongation factor aEF-2 (CTD)	Lokiarch_15310	contig148	12697	1	442	ref|WP_011822009.1| translation elongation factor aEF-2 [Hyperthermus butylicus] - Expect = 3e-90 (40%)	arCOG01559	COG00480	Translation elongation factor G, EF-G (GTPase)	**J**	Translation; ribosomal structure and biogenesis	
partial translation elongation factor aEF-2 (NTD)	Lokiarch_28370	contig23	31031	1	287	gb|ADX78263.1| elongation factor-2 [Thermococcus sp. LMO-A2] - Expect = 1e-74 (46%)	arCOG01559	COG00480	Translation elongation factor G, EF-G (GTPase)	**J**	Translation; ribosomal structure and biogenesis	
fragment of translation elongation factor aEF2	Lokiarch_16250	contig413	3648	4	65	ref|WP_013303965.1| translation elongation factor aEF-2 [Ignisphaera aggregans] - Expect = 0.001 (41%)	arCOG01559	COG00480	Translation elongation factor G, EF-G (GTPase)	**J**	Translation; ribosomal structure and biogenesis	
partial translation elongation factor aEF-2 (NTD)	Lokiarch_13430	contig619	1315	1	414	ref|WP_018194176.1| MULTISPECIES: translation elongation factor aEF-2 [candidate division YNPFFA] - Expect = 5e-150 (52%)	arCOG01559	COG00480	Translation elongation factor G, EF-G (GTPase)	**J**	Translation; ribosomal structure and biogenesis	39% identical to overlapping region of Lokiarch_18540
partial translation initiation factor aIF-2	Lokiarch_18570	contig145	12777	12	383	ref|WP_011821627.1| translation initiation factor aIF-2 [Hyperthermus butylicus] - Expect = 8e-111 (67%)	arCOG01560	COG00532	Translation initiation factor 2 (IF-2; GTPase)	**J**	Translation; ribosomal structure and biogenesis	overlapping region of Lokiarch_18570 and Lokiarch_28430 shares 92% identity
Translation initiation factor aIF-2	Lokiarch_28430	contig23	31031	7	602	ref|WP_011998037.1| translation initiation factor aIF-2 [Ignicoccus hospitalis] - Expect = 0.0 (48%)	arCOG01560	COG00532	Translation initiation factor 2 (IF-2; GTPase)	**J**	Translation; ribosomal structure and biogenesis	
Translation elongation factor EF-1alpha	Lokiarch_28570	contig23	31031	21	438	emb|CCC55680.1| translation elongation factor EF-1alpha [uncultured archaeon] - Expect = 0.0 (58%)	arCOG01561	COG05256	Translation elongation factor EF-1alpha (GTPase)	**J**	Translation; ribosomal structure and biogenesis	
Translation initiation factor aIF-2	Lokiarch_30090	contig150	12534	3	532	ref|WP_013099590.1| translation initiation factor IF-2 subunit gamma [Methanocaldococcus infernus] - Expect = 1e-151 (55%)	arCOG01563	COG05257	Translation initiation factor 2, gamma subunit (eIF-2gamma; GTPase)	**J**	Translation; ribosomal structure and biogenesis	
Others												
Product	**Locus**	**Contig**	**Contig lenght**	**ORF number**	**length (AA)**	**Top Blast Hit***	**arCOG**	**COG**	**COG description**		**Funclass**	**comment**
putative cell division GTPase, FtsZ	Lokiarch_34180	contig6	43995	35	371	ref|WP_015790785.1| cell division protein FtsZ [Methanocaldococcus fervens] - Expect = 4e-117 (57%)	arCOG02201	COG00206	Cell division GTPase	D	Cell cycle control; cell division; chromosome partitioning	we could not identify homologs of tubulins and/or (ar-) tubulins in Lokiarchaeum and LCGC14

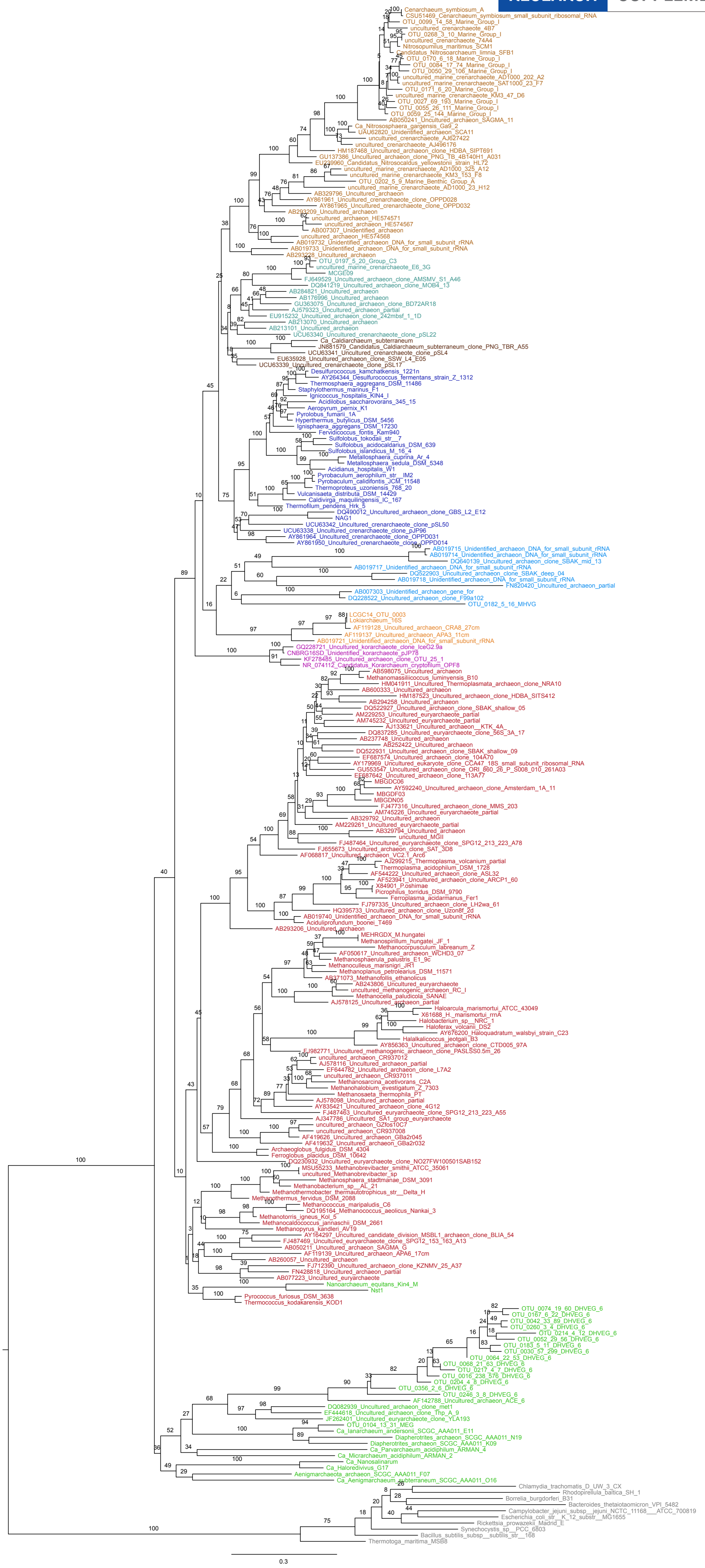
* in case top blast hit was to a single cell genome, the second best blast hit was listed; identity in parenthesis; red: top hit to Eukaryotes, orange: top hit to Bacteria; green/bold: alternative start codon

Suppl. Table S10: InterproScan assignments of selected hypothetical proteins of lokiarchaeal composite genome discussed in the manuscript and/or supplementary discussion

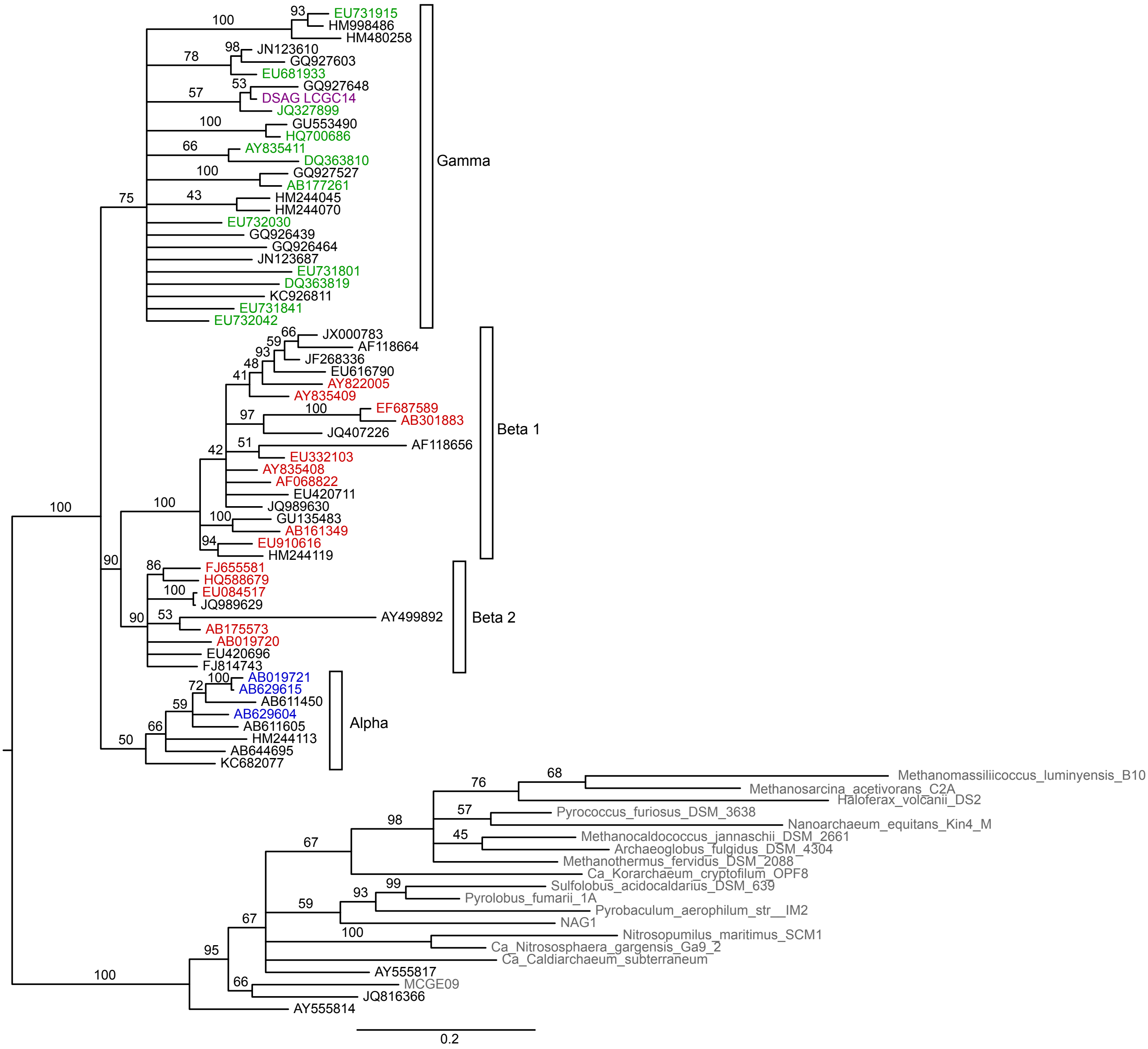
IPR-scan results for roadblock/LC7, gelsolin, longin and RING domain proteins as well as the steadiness containing protein and MON1										
Locus	**Protein Lenght (AA)**	**Analysis**	**Signature Accession**	**Signature Description**	**Start location**	**Stop location**	**E-value**	**InterPro annotations - accession**	**IPR-domain description**	**GO-number**
hypothetical proteins with gelsolin-like domains										
Lokiarch_14520	233	Gene3D	G3DSA:3.40.20.10		31	108	5.1E-5	IPR029006	ADF-H/Gelsolin-like domain	
Lokiarch_14520	233	Coils	Coil		212	233	-			
Lokiarch_14520	233	SUPERFAMILY	SSF55753		36	104	1.99E-6			
Lokiarch_20110	360	Gene3D	G3DSA:3.40.20.10		18	114	2.4E-11	IPR029006	ADF-H/Gelsolin-like domain	
Lokiarch_20110	360	SUPERFAMILY	SSF55753		8	108	3.74E-10			
Lokiarch_19250	293	Gene3D	G3DSA:3.40.20.10		5	113	3.0E-10	IPR029006	ADF-H/Gelsolin-like domain	
Lokiarch_19250	293	SUPERFAMILY	SSF55753		12	114	4.44E-10			
Lokiarch_14340	335	Pfam	PF00626	Gelsolin repeat	131	190	4.8E-8	IPR007123	Gelsolin-like domain	
Lokiarch_14340	335	Pfam	PF00626	Gelsolin repeat	19	79	3.0E-15	IPR007123	Gelsolin-like domain	
Lokiarch_14340	335	Gene3D	G3DSA:3.40.20.10		129	216	1.9E-15	IPR029006	ADF-H/Gelsolin-like domain	
Lokiarch_14340	335	SMART	SM00262	Gelsolin homology domain	119	209	1.2E-8	IPR007122	Villin/Gelsolin	GO:0003779
Lokiarch_14340	335	SMART	SM00262	Gelsolin homology domain	2	85	8.1E-15	IPR007122	Villin/Gelsolin	GO:0003779
Lokiarch_14340	335	SUPERFAMILY	SSF55753		128	215	9.58E-16			
Lokiarch_14340	335	Gene3D	G3DSA:3.40.20.10		11	82	1.3E-17	IPR029006	ADF-H/Gelsolin-like domain	
Lokiarch_14340	335	PANTHER	PTHR11977		19	178	1.2E-24	IPR007122	Villin/Gelsolin	GO:0003779
Lokiarch_49620	335	Gene3D	G3DSA:3.40.20.10		11	83	4.5E-18	IPR029006	ADF-H/Gelsolin-like domain	
Lokiarch_49620	335	Pfam	PF00626	Gelsolin repeat	131	190	5.2E-8	IPR007123	Gelsolin-like domain	
Lokiarch_49620	335	Pfam	PF00626	Gelsolin repeat	19	79	4.2E-16	IPR007123	Gelsolin-like domain	
Lokiarch_49620	335	SUPERFAMILY	SSF55753		118	215	2.8E-16			
Lokiarch_49620	335	SMART	SM00262	Gelsolin homology domain	119	209	4.3E-10	IPR007122	Villin/Gelsolin	GO:0003779
Lokiarch_49620	335	SMART	SM00262	Gelsolin homology domain	2	85	1.3E-15	IPR007122	Villin/Gelsolin	GO:0003779
Lokiarch_49620	335	SUPERFAMILY	SSF55753		10	85	1.31E-19			
Lokiarch_49620	335	PANTHER	PTHR11977		19	178	2.1E-25	IPR007122	Villin/Gelsolin	GO:0003779
Lokiarch_49620	335	Gene3D	G3DSA:3.40.20.10		129	216	7.5E-16	IPR029006	ADF-H/Gelsolin-like domain	
Lokiarch_16460	141	Pfam	PF00626	Gelsolin repeat	19	81	3.6E-6	IPR007123	Gelsolin-like domain	
Lokiarch_16460	141	Gene3D	G3DSA:3.40.20.10		2	87	2.8E-13	IPR029006	ADF-H/Gelsolin-like domain	
Lokiarch_16460	141	SUPERFAMILY	SSF55753		19	88	2.57E-11			
Lokiarch_42060	199	SMART	SM00262	Gelsolin homology domain	2	89	0.0042	IPR007122	Villin/Gelsolin	GO:0003779
Lokiarch_42060	199	SUPERFAMILY	SSF55753		14	80	1.24E-12			
Lokiarch_42060	199	Pfam	PF00626	Gelsolin repeat	14	79	6.6E-14	IPR007123	Gelsolin-like domain	
Lokiarch_42060	199	Gene3D	G3DSA:3.40.20.10		14	81	9.7E-13	IPR029006	ADF-H/Gelsolin-like domain	
Lokiarch_06290	199	Gene3D	G3DSA:3.40.20.10		14	81	1.7E-13	IPR029006	ADF-H/Gelsolin-like domain	
Lokiarch_06290	199	Pfam	PF00626	Gelsolin repeat	14	79	2.8E-14	IPR007123	Gelsolin-like domain	
Lokiarch_06290	199	SUPERFAMILY	SSF55753		14	80	5.14E-13			
Lokiarch_35550	317	SUPERFAMILY	SSF55753		9	112	1.01E-8			
Lokiarch_35550	317	Gene3D	G3DSA:1.10.720.30		279	309	3.8E-4	IPR003034	SAP domain	GO:0003676
Lokiarch_35550	317	Gene3D	G3DSA:3.40.20.10		13	115	2.5E-9	IPR029006	ADF-H/Gelsolin-like domain	
Lokiarch_33820	329	SUPERFAMILY	SSF55753		22	118	2.22E-7			
Lokiarch_33820	329	Gene3D	G3DSA:3.40.20.10		39	119	6.2E-9	IPR029006	ADF-H/Gelsolin-like domain	
Lokiarch_32420	180	Gene3D	G3DSA:3.40.20.10		26	93	1.0E-6	IPR029006	ADF-H/Gelsolin-like domain	
Lokiarch_32420	180	SUPERFAMILY	SSF55753		9	91	3.97E-8			
Lokiarch_14920	235	SUPERFAMILY	SSF55753		35	107	4.91E-8			
Lokiarch_14920	235	Gene3D	G3DSA:3.40.20.10		33	114	3.5E-7	IPR029006	ADF-H/Gelsolin-like domain	
hypothetical protein										
Lokiarch_16740	650	Coils	Coil		619	640	-			
Lokiarch_16740	650	ProSiteProfiles	PS51312	Steadiness box (SB) domain profile.	568	636	8.583	IPR017916	Steadiness box	
hypothetical proteins with MON1 domains										
Lokiarch_21780	130	Pfam	PF03164	Trafficking protein Mon1	5	90	2.3E-6	IPR004353	Vacuolar fusion protein MON1	
Lokiarch_01670	137	Pfam	PF03164	Trafficking protein Mon1	6	95	1.3E-6	IPR004353	Vacuolar fusion protein MON1	
Lokiarch_15160	128	Pfam	PF03164	Trafficking protein Mon1	5	94	2.2E-5	IPR004353	Vacuolar fusion protein MON1	
hypothetical proteins with longin domains										
Lokiarch_01890	169	SUPERFAMILY	SSF64356		1	127	9.16E-7	IPR011012	Longin-like domain	GO:0006810
Lokiarch_01890	169	Gene3D	G3DSA:3.30.450.60		1	99	4.1E-6			
Lokiarch_13110	132	Gene3D	G3DSA:3.30.450.50		51	121	1.4E-4	IPR010908	Longin domain	
Lokiarch_13110	132	SUPERFAMILY	SSF64356		43	122	7.88E-7	IPR011012	Longin-like domain	GO:0006810
Lokiarch_03280	252	Gene3D	G3DSA:3.30.450.60		3	102	1.4E-5			
Lokiarch_03280	252	SUPERFAMILY	SSF64356		3	106	1.16E-7	IPR011012	Longin-like domain	GO:0006810
Lokiarch_22790	130	SUPERFAMILY	SSF64356		6	118	3.89E-5	IPR011012	Longin-like domain	GO:0006810
Lokiarch_04850	167	SUPERFAMILY	SSF64356		1	126	6.41E-9	IPR011012	Longin-like domain	GO:0006810
Lokiarch_04850	167	Gene3D	G3DSA:3.30.450.60		1	90	4.9E-7			
hypothetical proteins with FAM92 domains										
Lokiarch_46220	216	ProSiteProfiles	PS51021	BAR domain profile.	1	213	14.541	IPR004148	BAR domain	GO:0005515|GO:0005737
Lokiarch_46220	216	Coils	Coil		139	174	-			
Lokiarch_46220	216	Coils	Coil		92	130	-			
Lokiarch_46220	216	SUPERFAMILY	SSF103657		8	202	3.14E-12			
Lokiarch_46220	216	Pfam	PF06730	FAM92 protein	40	211	2.5E-6	IPR009602	FAM92 protein	
Lokiarch_08900	216	Coils	Coil		139	174	-			
Lokiarch_08900	216	SUPERFAMILY	SSF103657		7	202	5.36E-14			
Lokiarch_08900	216	Coils	Coil		92	130	-			
Lokiarch_08900	216	Pfam	PF06730	FAM92 protein	9	211	1.5E-10	IPR009602	FAM92 protein	
hypothetical proteins with ubiquitin-like domains										
Lokiarch_29280	133	SUPERFAMILY	SSF54236		45	125	9.44E-10	IPR029071	Ubiquitin-related domain	
Lokiarch_29280	133	ProSiteProfiles	PS50053	Ubiquitin domain profile.	53	131	10.658	IPR000626	Ubiquitin-like	GO:0005515
Lokiarch_29280	133	Gene3D	G3DSA:3.10.20.90		66	124	8.3E-9			
Lokiarch_29290	266	Phobius	CYTOPLASMIC_DOMAIN	Region of a membrane-bound protein predicted to be outside the membrane, in the cytoplasm.	55	266	-			
Lokiarch_29290	266	TMHMM	TMhelix	Region of a membrane-bound protein predicted to be embedded in the membrane.	32	54	-			
Lokiarch_29290	266	Phobius	CYTOPLASMIC_DOMAIN	Region of a membrane-bound protein predicted to be outside the membrane, in the cytoplasm.	1	6	-			
Lokiarch_29290	266	SUPERFAMILY	SSF54236		179	263	9.21E-7	IPR029071	Ubiquitin-related domain	
Lokiarch_29290	266	Phobius	NON_CYTOPLASMIC_DOMAIN	Region of a membrane-bound protein predicted to be outside the membrane, in the extracellular region.	27	31	-			
Lokiarch_29290	266	Phobius	TRANSMEMBRANE	Region of a membrane-bound protein predicted to be embedded in the membrane.	32	54	-			
Lokiarch_29290	266	Phobius	TRANSMEMBRANE	Region of a membrane-bound protein predicted to be embedded in the membrane.	7	26	-			
Lokiarch_29290	266	TMHMM	TMhelix	Region of a membrane-bound protein predicted to be embedded in the membrane.	7	28	-			

Lokiarch_29310	113	SUPERFAMILY	SSF54236		40	109	7.66E-10	IPR029071	Ubiquitin-related domain	
Lokiarch_29310	113	Gene3D	G3DSA:3.10.20.90		47	106	5.0E-7			
Lokiarch_29310	113	Pfam	PF00240	Ubiquitin family	49	105	3.9E-6	IPR000626	Ubiquitin-like	GO:0005515
Lokiarch_29310	113	ProSiteProfiles	PS50053	Ubiquitin domain profile.	52	113	9.33	IPR000626	Ubiquitin-like	GO:0005515
Lokiarch_37670	88	ProSiteProfiles	PS50053	Ubiquitin domain profile.	20	82	10.811	IPR000626	Ubiquitin-like	GO:0005515
Lokiarch_37670	88	Pfam	PF00240	Ubiquitin family	20	77	1.7E-6	IPR000626	Ubiquitin-like	GO:0005515
Lokiarch_37670	88	SUPERFAMILY	SSF54236		12	85	9.42E-8	IPR029071	Ubiquitin-related domain	
Lokiarch_37670	88	Gene3D	G3DSA:3.10.20.90		20	81	5.4E-7			
Lokiarch_28320	1687	Phobius	TRANSMEMBRANE	Region of a membrane-bound protein predicted to be embedded in the membrane.	412	435	-			
Lokiarch_28320	1687	TMHMM	TMhelix	Region of a membrane-bound protein predicted to be embedded in the membrane.	1011	1033	-			
Lokiarch_28320	1687	SUPERFAMILY	SSF54236		1064	1134	1.8E-7	IPR029071	Ubiquitin-related domain	
Lokiarch_28320	1687	ProSiteProfiles	PS50053	Ubiquitin domain profile.	1064	1137	11.028	IPR000626	Ubiquitin-like	GO:0005515
Lokiarch_28320	1687	Phobius	NON_CYTOPLASMIC_DOMAIN	Region of a membrane-bound protein predicted to be outside the membrane, in the extracellular region.	844	1687	-			
Lokiarch_28320	1687	Phobius	CYTOPLASMIC_DOMAIN	Region of a membrane-bound protein predicted to be outside the membrane, in the cytoplasm.	436	818	-			
Lokiarch_28320	1687	Phobius	NON_CYTOPLASMIC_DOMAIN	Region of a membrane-bound protein predicted to be outside the membrane, in the extracellular region.	1	411	-			
Lokiarch_28320	1687	Gene3D	G3DSA:3.10.20.90		1077	1137	5.8E-5			
Lokiarch_28320	1687	Phobius	TRANSMEMBRANE	Region of a membrane-bound protein predicted to be embedded in the membrane.	819	843	-			
Lokiarch_28320	1687	TMHMM	TMhelix	Region of a membrane-bound protein predicted to be embedded in the membrane.	821	843	-			
hypothetical proteins with ZINC-finger, RING-type domains (excluding those with additional von Willebrand factor domains)										
Lokiarch_38540	192	ProSiteProfiles	PS50089	Zinc finger RING-type profile.	4	46	8.934	IPR001841	Zinc finger, RING-type	GO:0005515|GO:0008270
Lokiarch_38540	192	Pfam	PF13639	Ring finger domain	3	45	4.5E-7	IPR001841	Zinc finger, RING-type	GO:0005515|GO:0008270
Lokiarch_38540	192	SUPERFAMILY	SSF48452		96	190	6.13E-10			
Lokiarch_38540	192	Gene3D	G3DSA:3.30.40.10		3	53	3.8E-9	IPR013083	Zinc finger, RING/FYVE/PHD-type	
Lokiarch_38540	192	SUPERFAMILY	SSF57850		3	52	1.9E-8			
Lokiarch_38540	192	Gene3D	G3DSA:1.25.40.10		96	190	6.1E-9	IPR011990	Tetratricopeptide-like helical	GO:0005515
Lokiarch_14290	675	Coils	Coil		450	471	-			
Lokiarch_14290	675	Gene3D	G3DSA:3.30.40.10		609	659	1.0E-7	IPR013083	Zinc finger, RING/FYVE/PHD-type	
Lokiarch_14290	675	Coils	Coil		561	582	-			
Lokiarch_14290	675	SUPERFAMILY	SSF57850		601	662	9.77E-9			
Lokiarch_14290	675	Coils	Coil		97	118	-			
Lokiarch_14290	675	ProSiteProfiles	PS50089	Zinc finger RING-type profile.	613	656	9.583	IPR001841	Zinc finger, RING-type	GO:0005515|GO:0008270
Lokiarch_49580	678	Coils	Coil		450	471	-			
Lokiarch_49580	678	Coils	Coil		561	582	-			
Lokiarch_49580	678	Gene3D	G3DSA:3.30.40.10		608	661	7.9E-8	IPR013083	Zinc finger, RING/FYVE/PHD-type	
Lokiarch_49580	678	SUPERFAMILY	SSF57850		601	662	1.02E-8			
Lokiarch_49580	678	ProSiteProfiles	PS50089	Zinc finger RING-type profile.	613	656	9.583	IPR001841	Zinc finger, RING-type	GO:0005515|GO:0008270
Lokiarch_36980	67	ProSiteProfiles	PS50089	Zinc finger RING-type profile.	9	53	8.89	IPR001841	Zinc finger, RING-type	GO:0005515|GO:0008270
Lokiarch_36980	67	SUPERFAMILY	SSF57850		5	60	6.53E-7			
Lokiarch_36980	67	Gene3D	G3DSA:3.30.40.10		1	59	1.6E-6	IPR013083	Zinc finger, RING/FYVE/PHD-type	
Lokiarch_35850	178	Phobius	CYTOPLASMIC_DOMAIN	Region of a membrane-bound protein predicted to be outside the membrane, in the cytoplasm.	178	178	-			
Lokiarch_35850	178	Phobius	NON_CYTOPLASMIC_DOMAIN	Region of a membrane-bound protein predicted to be outside the membrane, in the extracellular region.	151	155	-			
Lokiarch_35850	178	TMHMM	TMhelix	Region of a membrane-bound protein predicted to be embedded in the membrane.	128	150	-			
Lokiarch_35850	178	Phobius	TRANSMEMBRANE	Region of a membrane-bound protein predicted to be embedded in the membrane.	156	177	-			
Lokiarch_35850	178	Phobius	TRANSMEMBRANE	Region of a membrane-bound protein predicted to be embedded in the membrane.	127	150	-			
Lokiarch_35850	178	Gene3D	G3DSA:3.30.40.10		41	101	2.0E-4	IPR013083	Zinc finger, RING/FYVE/PHD-type	
Lokiarch_35850	178	TMHMM	TMhelix	Region of a membrane-bound protein predicted to be embedded in the membrane.	160	177	-			
Lokiarch_35850	178	Phobius	CYTOPLASMIC_DOMAIN	Region of a membrane-bound protein predicted to be outside the membrane, in the cytoplasm.	1	126	-			
Lokiarch_30770	150	Gene3D	G3DSA:3.30.40.10		58	81	2.0E-4	IPR013083	Zinc finger, RING/FYVE/PHD-type	
Lokiarch_30770	150	Gene3D	G3DSA:3.30.40.10		126	149	2.0E-4	IPR013083	Zinc finger, RING/FYVE/PHD-type	
Lokiarch_30770	150	Phobius	TRANSMEMBRANE	Region of a membrane-bound protein predicted to be embedded in the membrane.	31	51	-			
Lokiarch_30770	150	TMHMM	TMhelix	Region of a membrane-bound protein predicted to be embedded in the membrane.	31	53	-			
Lokiarch_30770	150	Phobius	CYTOPLASMIC_DOMAIN	Region of a membrane-bound protein predicted to be outside the membrane, in the cytoplasm.	1	30	-			
Lokiarch_30770	150	Phobius	NON_CYTOPLASMIC_DOMAIN	Region of a membrane-bound protein predicted to be outside the membrane, in the extracellular region.	52	150	-			
Lokiarch_45630	100	SUPERFAMILY	SSF57903		28	80	1.15E-5	IPR011011	Zinc finger, FYVE/PHD-type	
Lokiarch_45630	100	Gene3D	G3DSA:3.30.40.10		19	82	2.1E-7	IPR013083	Zinc finger, RING/FYVE/PHD-type	
Lokiarch_46790	149	Gene3D	G3DSA:3.30.40.10		57	81	4.1E-4	IPR013083	Zinc finger, RING/FYVE/PHD-type	
Lokiarch_46790	149	Gene3D	G3DSA:3.30.40.10		124	147	4.1E-4	IPR013083	Zinc finger, RING/FYVE/PHD-type	

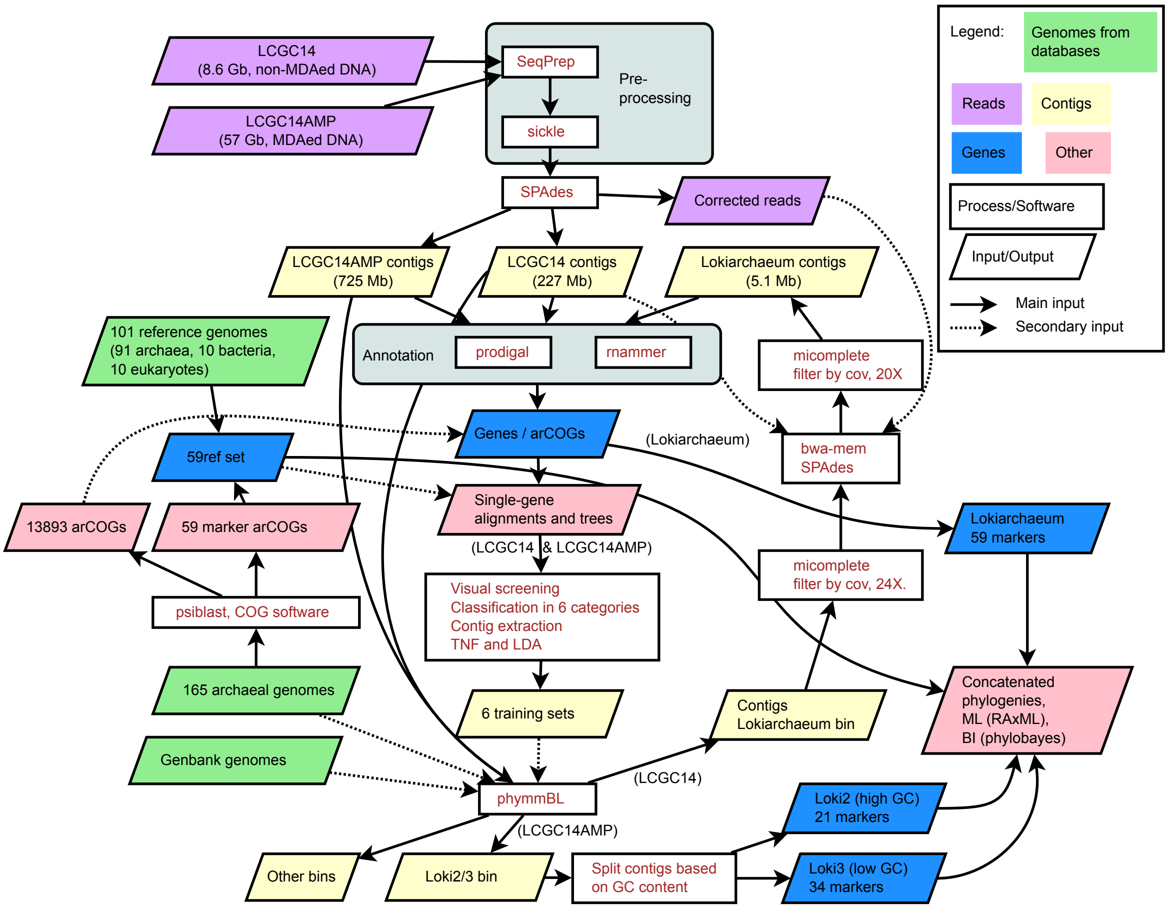
Lokiarch_46790	149	Phobius	TRANSMEMBRANE	Region of a membrane-bound protein predicted to be embedded in the membrane.	31	54	-			
Lokiarch_46790	149	Phobius	CYTOPLASMIC_DOMAIN	Region of a membrane-bound protein predicted to be outside the membrane, in the cytoplasm.	1	30	-			
Lokiarch_46790	149	TMHMM	TMhelix	Region of a membrane-bound protein predicted to be embedded in the membrane.	31	53	-			
Lokiarch_46790	149	Phobius	NON_CYTOPLASMIC_DOMAIN	Region of a membrane-bound protein predicted to be outside the membrane, in the extracellular region.	55	149	-			
Lokiarch_07660	320	Gene3D	G3DSA:3.30.40.10		248	308	1.1E-4	IPR013083	Zinc finger, RING/FYVE/PHD-type	
Lokiarch_07660	320	Coils	Coil		34	58	-			
Lokiarch_07660	320	SUPERFAMILY	SSF57850		247	300	4.24E-5			
Lokiarch_34010	175	Gene3D	G3DSA:3.30.40.10		115	164	7.2E-8	IPR013083	Zinc finger, RING/FYVE/PHD-type	
Lokiarch_34010	175	SUPERFAMILY	SSF57850		115	164	1.8E-5			
roadblock domain proteins										
Lokiarch_02330	153	SUPERFAMILY	SSF103196		8	128	6.8E-19			
Lokiarch_02330	153	Pfam	PF03259	Roadblock/LC7 domain	16	102	3.3E-9	IPR004942	Dynein light chain-related	
Lokiarch_02330	153	SMART	SM00960	Roadblock/LC7 domain	11	103	5.0E-6	IPR004942	Dynein light chain-related	
Lokiarch_02330	153	Gene3D	G3DSA:3.30.450.30		8	129	9.6E-20			
Lokiarch_03940	142	SMART	SM00960	Roadblock/LC7 domain	17	106	3.9E-6	IPR004942	Dynein light chain-related	
Lokiarch_03940	142	Gene3D	G3DSA:3.30.450.30		7	135	1.5E-13			
Lokiarch_03940	142	Coils	Coil		7	28	-			
Lokiarch_03940	142	Pfam	PF03259	Roadblock/LC7 domain	20	94	1.8E-9	IPR004942	Dynein light chain-related	
Lokiarch_03940	142	SUPERFAMILY	SSF103196		16	134	2.79E-12			
Lokiarch_06200	131	SMART	SM00960	Roadblock/LC7 domain	5	98	7.7E-4	IPR004942	Dynein light chain-related	
Lokiarch_06200	131	Pfam	PF03259	Roadblock/LC7 domain	7	83	7.4E-6	IPR004942	Dynein light chain-related	
Lokiarch_06200	131	SUPERFAMILY	SSF103196		3	98	2.09E-8			
Lokiarch_06200	131	Gene3D	G3DSA:3.30.450.30		3	98	2.2E-8			
Lokiarch_06410	135	SUPERFAMILY	SSF103196		15	134	4.45E-29			
Lokiarch_06410	135	SMART	SM00960	Roadblock/LC7 domain	22	111	4.3E-26	IPR004942	Dynein light chain-related	
Lokiarch_06410	135	Gene3D	G3DSA:3.30.450.30		17	135	2.6E-27			
Lokiarch_06410	135	Pfam	PF03259	Roadblock/LC7 domain	22	110	4.5E-20	IPR004942	Dynein light chain-related	
Lokiarch_06430	117	SMART	SM00960	Roadblock/LC7 domain	8	96	0.0015	IPR004942	Dynein light chain-related	
Lokiarch_06430	117	Gene3D	G3DSA:3.30.450.30		7	107	2.3E-6			
Lokiarch_06430	117	SUPERFAMILY	SSF103196		1	99	5.93E-9			
Lokiarch_06430	117	Pfam	PF03259	Roadblock/LC7 domain	10	95	7.3E-8	IPR004942	Dynein light chain-related	
Lokiarch_06490	100	Pfam	PF03259	Roadblock/LC7 domain	10	81	1.4E-11	IPR004942	Dynein light chain-related	
Lokiarch_06490	100	Gene3D	G3DSA:3.30.450.30		7	99	4.8E-15			
Lokiarch_06490	100	SUPERFAMILY	SSF103196		7	99	5.1E-14			
Lokiarch_06490	100	SMART	SM00960	Roadblock/LC7 domain	9	100	2.1E-8	IPR004942	Dynein light chain-related	
Lokiarch_08020	117	SMART	SM00960	Roadblock/LC7 domain	5	93	5.9E-14	IPR004942	Dynein light chain-related	
Lokiarch_08020	117	SUPERFAMILY	SSF103196		3	117	1.69E-18			
Lokiarch_08020	117	Pfam	PF03259	Roadblock/LC7 domain	7	91	2.4E-10	IPR004942	Dynein light chain-related	
Lokiarch_08020	117	Gene3D	G3DSA:3.30.450.30		3	116	2.4E-18			
Lokiarch_25010	117	SUPERFAMILY	SSF103196		3	117	1.96E-19			
Lokiarch_25010	117	Gene3D	G3DSA:3.30.450.30		3	117	1.1E-18			
Lokiarch_25010	117	Pfam	PF03259	Roadblock/LC7 domain	7	91	1.7E-11	IPR004942	Dynein light chain-related	
Lokiarch_25010	117	SMART	SM00960	Roadblock/LC7 domain	5	93	2.1E-14	IPR004942	Dynein light chain-related	
Lokiarch_41940	125	Gene3D	G3DSA:3.30.450.30		8	125	3.3E-27			
Lokiarch_41940	125	SMART	SM00960	Roadblock/LC7 domain	11	101	8.1E-26	IPR004942	Dynein light chain-related	
Lokiarch_41940	125	Pfam	PF03259	Roadblock/LC7 domain	13	100	6.5E-20	IPR004942	Dynein light chain-related	
Lokiarch_41940	125	SUPERFAMILY	SSF103196		4	124	4.97E-29			
Lokiarch_43650	123	SUPERFAMILY	SSF103196		1	122	5.41E-23			
Lokiarch_43650	123	Gene3D	G3DSA:3.30.450.30		2	122	1.4E-22			
Lokiarch_43650	123	SMART	SM00960	Roadblock/LC7 domain	8	98	7.6E-15	IPR004942	Dynein light chain-related	
Lokiarch_43650	123	Pfam	PF03259	Roadblock/LC7 domain	20	97	3.8E-14	IPR004942	Dynein light chain-related	
Lokiarch_44990	249	Coils	Coil		1	22	-			
Lokiarch_44990	249	Gene3D	G3DSA:3.30.450.30		149	236	9.0E-5			
Lokiarch_44990	249	SUPERFAMILY	SSF103196		146	240	3.31E-7			
Lokiarch_44990	249	SMART	SM00960	Roadblock/LC7 domain	141	226	5.4E-4	IPR004942	Dynein light chain-related	
Lokiarch_54060	142	SUPERFAMILY	SSF103196		22	133	4.19E-11			
Lokiarch_54060	142	SMART	SM00960	Roadblock/LC7 domain	17	118	6.3E-7	IPR004942	Dynein light chain-related	
Lokiarch_54060	142	Pfam	PF03259	Roadblock/LC7 domain	20	94	1.5E-9	IPR004942	Dynein light chain-related	
Lokiarch_54060	142	Gene3D	G3DSA:3.30.450.30		7	135	3.7E-12			



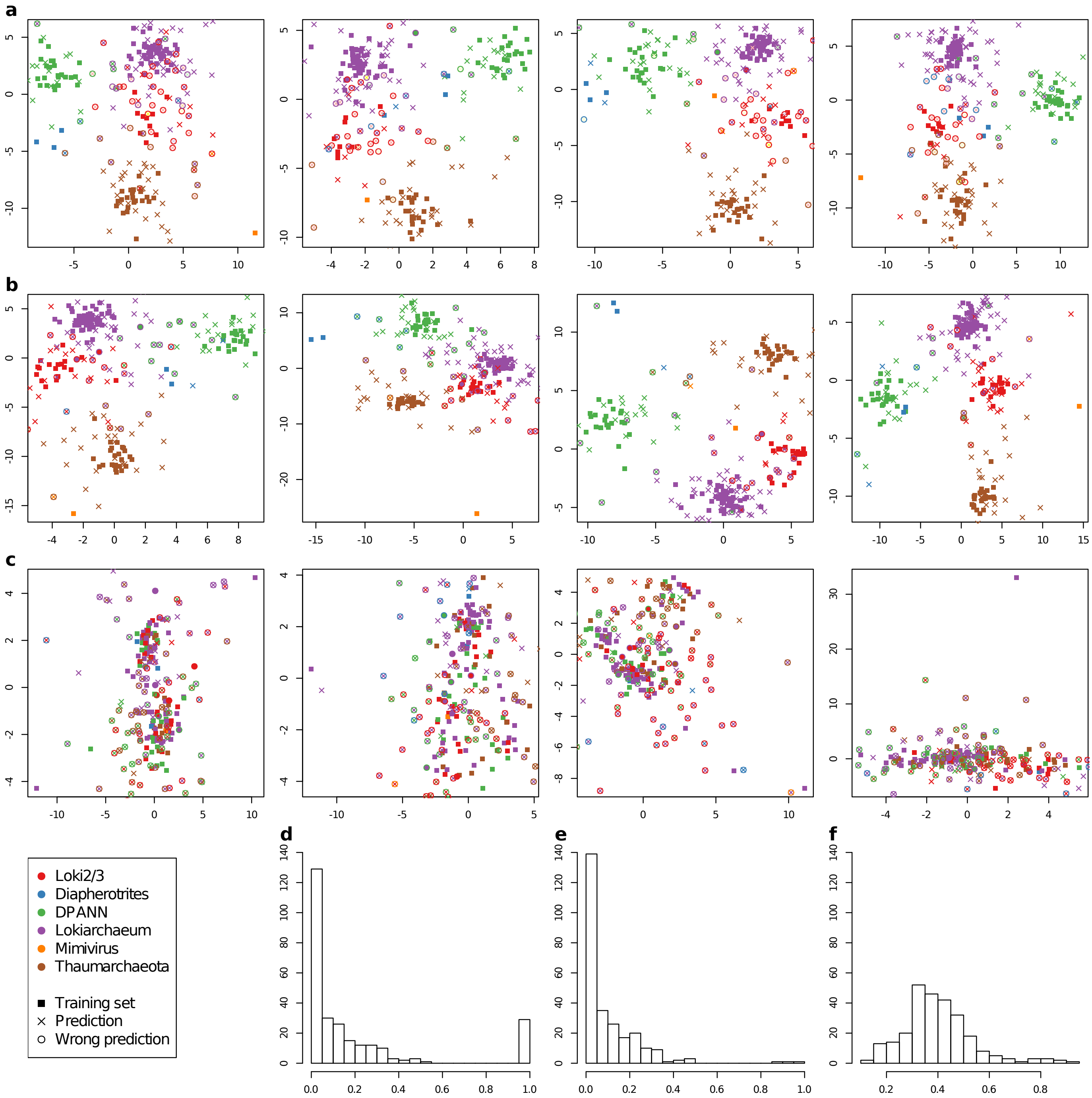
Suppl. Figure S1: Diversity of archaeal taxa identified in LCGC14 sediments. Maximum likelihood phylogeny of OTUs derived from the LCGC14 16S rDNA amplicon sequences. Major archaeal groups are coloured as follows: grey, Bacteria; green, DPANN, red, Euryarchaeota; pink, Korarchaeota; light orange, DSAG; light blue, MHVG; dark blue, Crenarchaeota; brown, Aigarchaeota; teal, MCG and C3; dark orange, Thaumarchaeota.



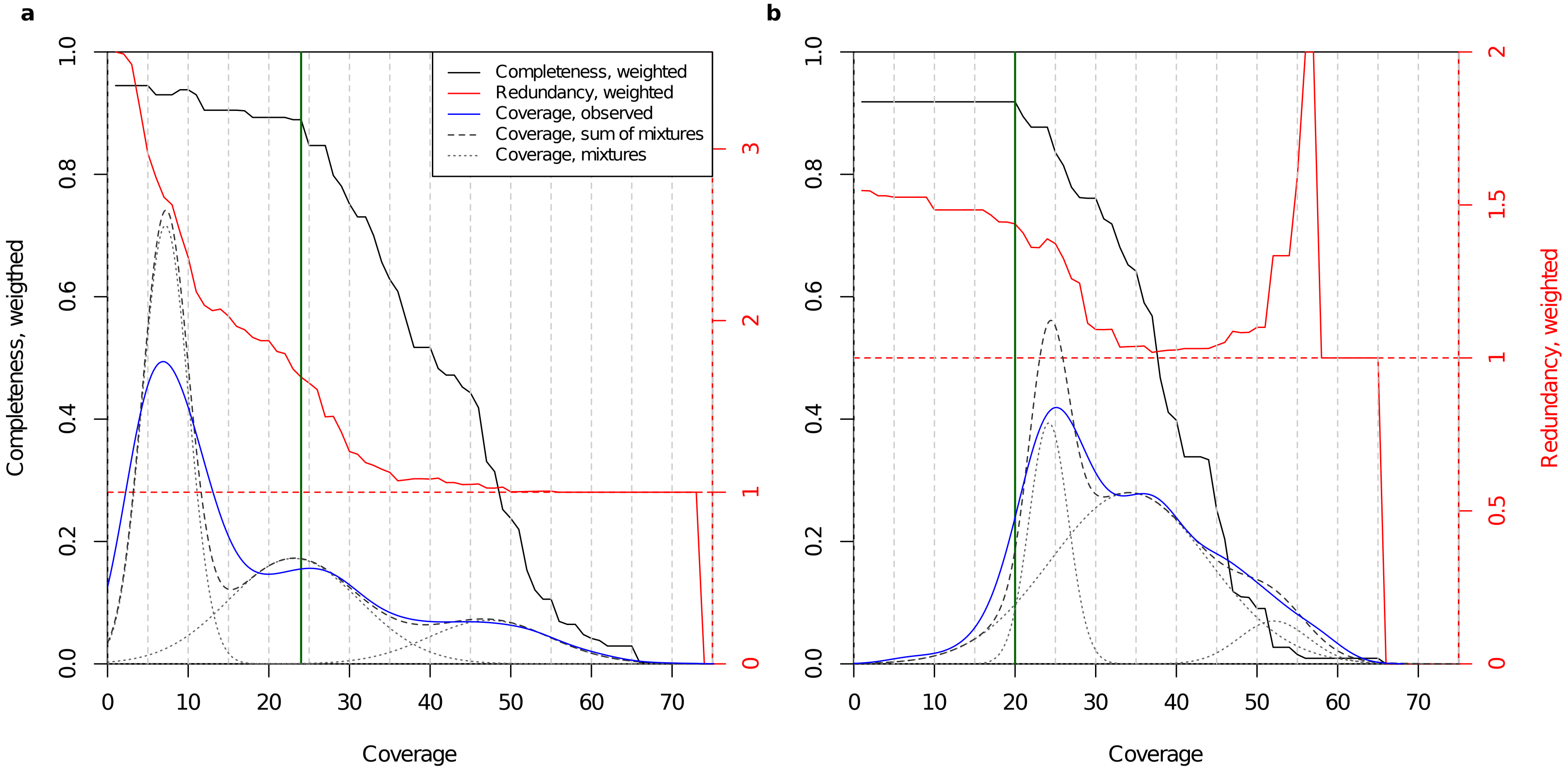
Suppl. Figure S2: Maximum likelihood phylogeny of DSAG diversity in LCGC14 sediments. The phylogenetic tree shows the full phylogenetic diversity of DSAG currently available in the SILVA database (release 119). Outgroup sequences representing all other archaeal phyla are given in grey. Sequences that were classified as alpha, beta and gamma in are given in blue, red and green respectively. Related sequences identified with BLAST versus the SILVA database are given in black. The single sequence from the OTU that was classified as DSAG in this study is depicted in purple.



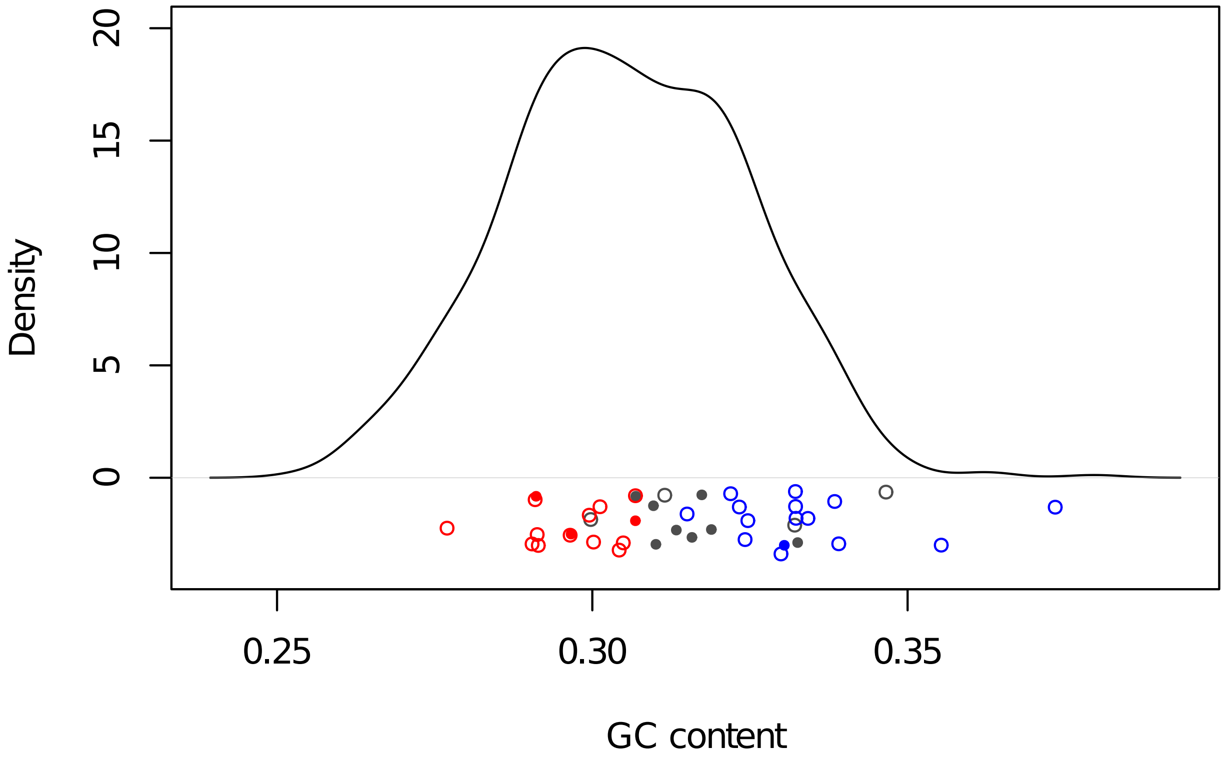
Suppl. Figure S3: Full schematic overview of metagenomics approach. This flowchart summarizes which data and software were used, starting from raw reads, to obtain the marker genes used throughout this contribution, including in phylogenetic reconstructions. See the legend in the top-right corner for more details on the meaning of colours and symbols.



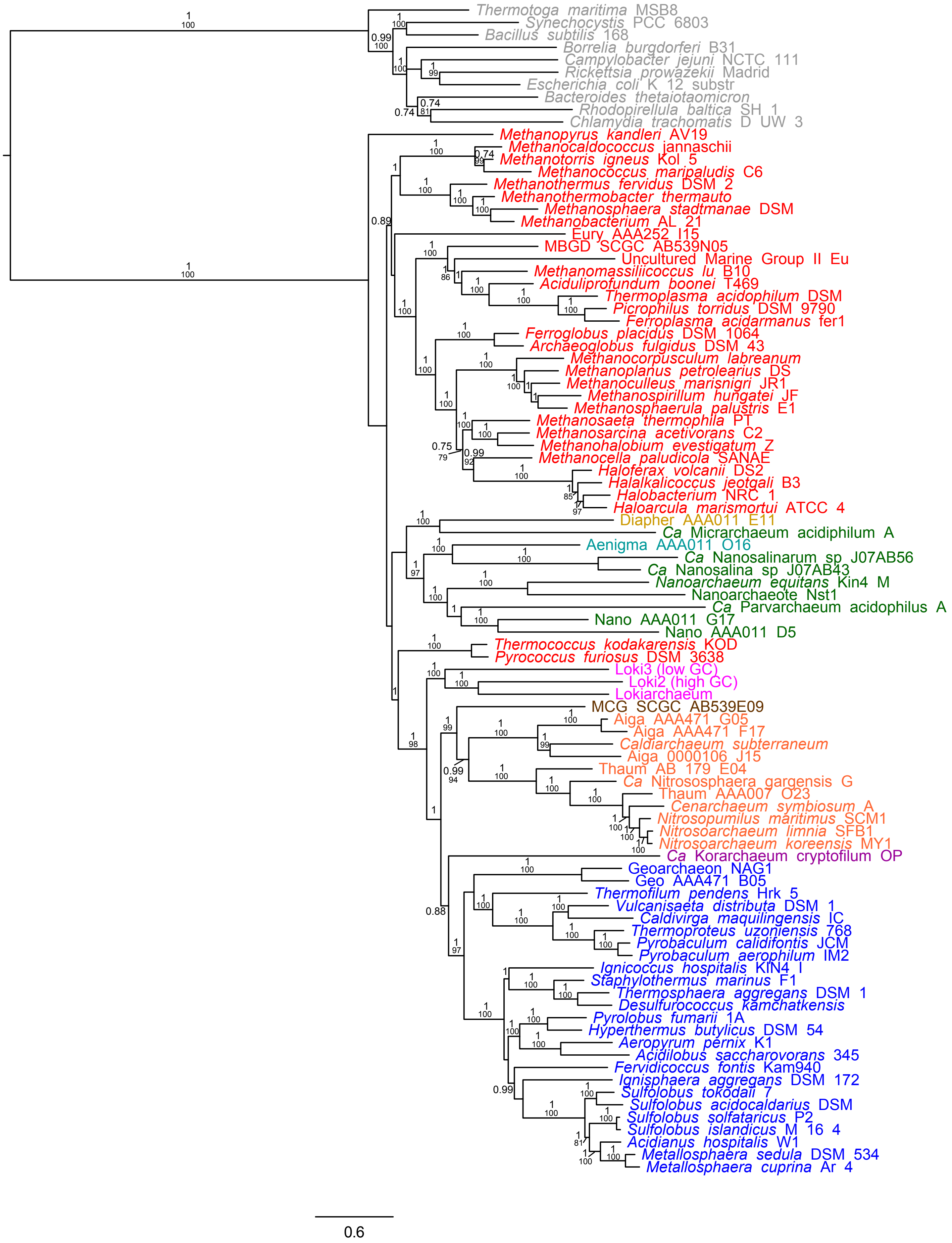
Suppl. Figure S4: Quality control of the training sets used in the binning procedure. a-c: Results of the linear discriminant analysis (LDA) for the initial (a) and second (b) rounds of training set benchmarking, and for a round where categories have been randomized (c). In each panel, four out of the 100 repetitions of contig selection, LDA calculation and LDA prediction are represented. The x-axis and y-axis are the first and second axis of the LDA, respectively. Each point represents a contig from the training sets, colour-coded by category: red, Loki2/3; blue, Diapherotrites; green, DPANN; purple, Lokiarchaeum; orange, Mimivirus; brown, Thaumarchaeota (see legend in the bottom left corner of the figure). In a, the putative classifications are marked with a lighter colour. Contigs used for LDA calculation are shown as squares, predictions as crosses. Whenever the classification and the LDA prediction disagree, the dot is circled with the colour of the LDA prediction. By definition, all putative classifications are show as incorrect, since putative categories were not included in the LDA calculation. d-h: histograms showing the fraction of wrong predictions for the initial (d), second (e) and randomized (f) rounds of LDA.



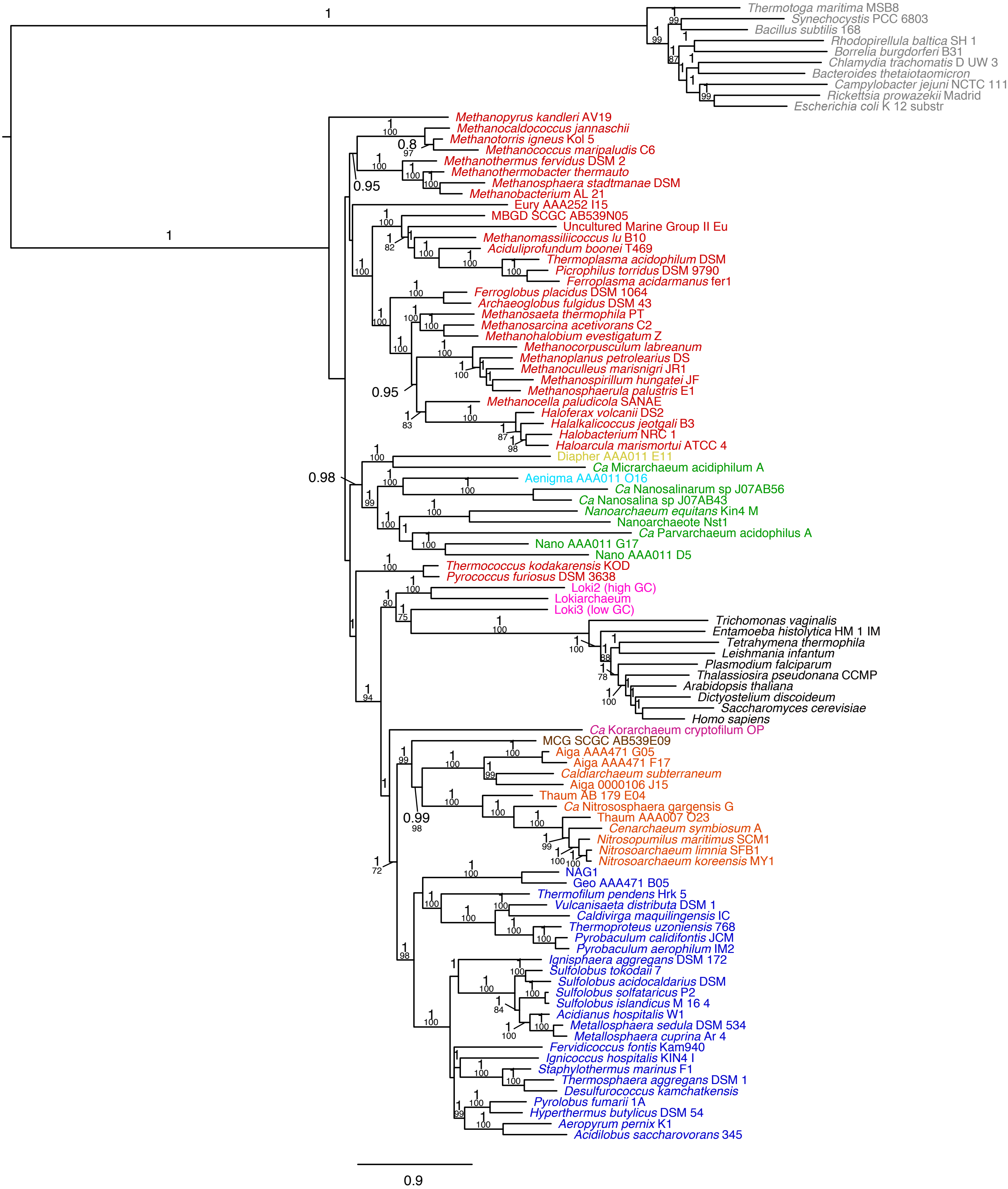
Suppl. Figure S5: Completeness and redundancy analyses of contigs of the Lokiarchaeum bin. Completeness and redundancy of lokiarchaeal contigs was assessed before (a) and after reassembly of the corresponding reads (b), with increasing coverage thresholds. For each value of the threshold (x-axis), the weighted completeness and redundancy of the contigs with only this coverage and above was estimated (see methods for details). Completeness is shown in solid black lines (y-axis to the left of the plot) and redundancy with solid red lines (y-axis to the right of the plot). In addition, the distribution of coverage on contigs larger than 5 kb is shown with a blue, solid line. This distribution was fitted with three normal distributions (shown in grey, dotted lines) using a mixture model. The sum of the three fitted distributions is shown with a black, solid line. The selected cut-offs (24X and 20X for a and b, respectively) are indicated with a vertical, dark green line.



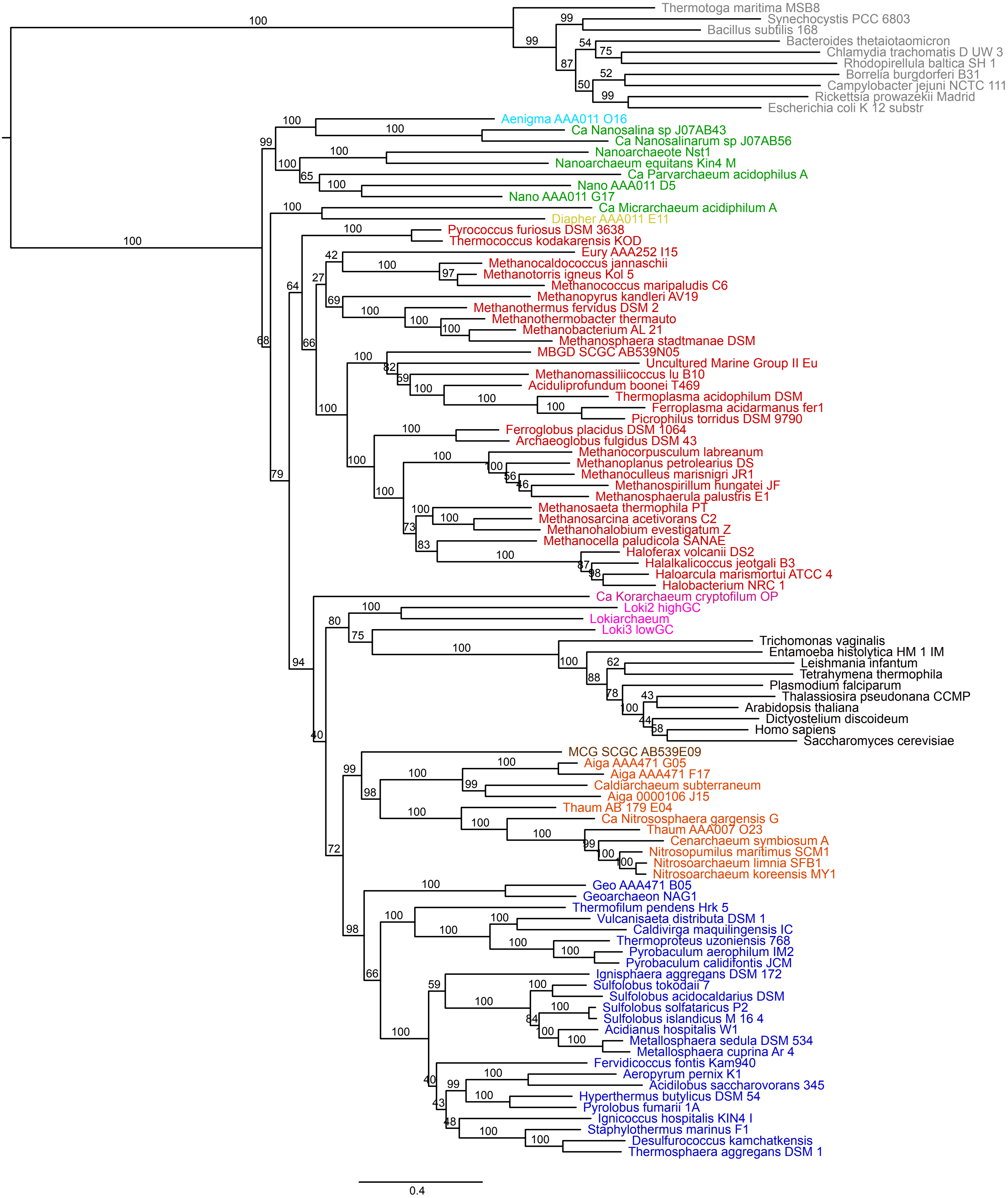
Suppl. Figure S6: **GC-content-based separation of Loki2 and Loki3 contigs**. The density curve shows the GC content distribution of contigs larger than 5 kbp in the Loki2/3 bin. Contigs carrying taxonomic markers of the '59ref set' identified during the analysis of LCGC14AMP and present in two copies are represented by open circles, contigs with single-copy taxonomic markers by closed circles. Contigs attributed to Loki2 (high GC) and Loki3 (low GC) are marked in blue and red, respectively; non-attributed contigs in grey. Position of circles on the y-axis was jittered to improve readability.



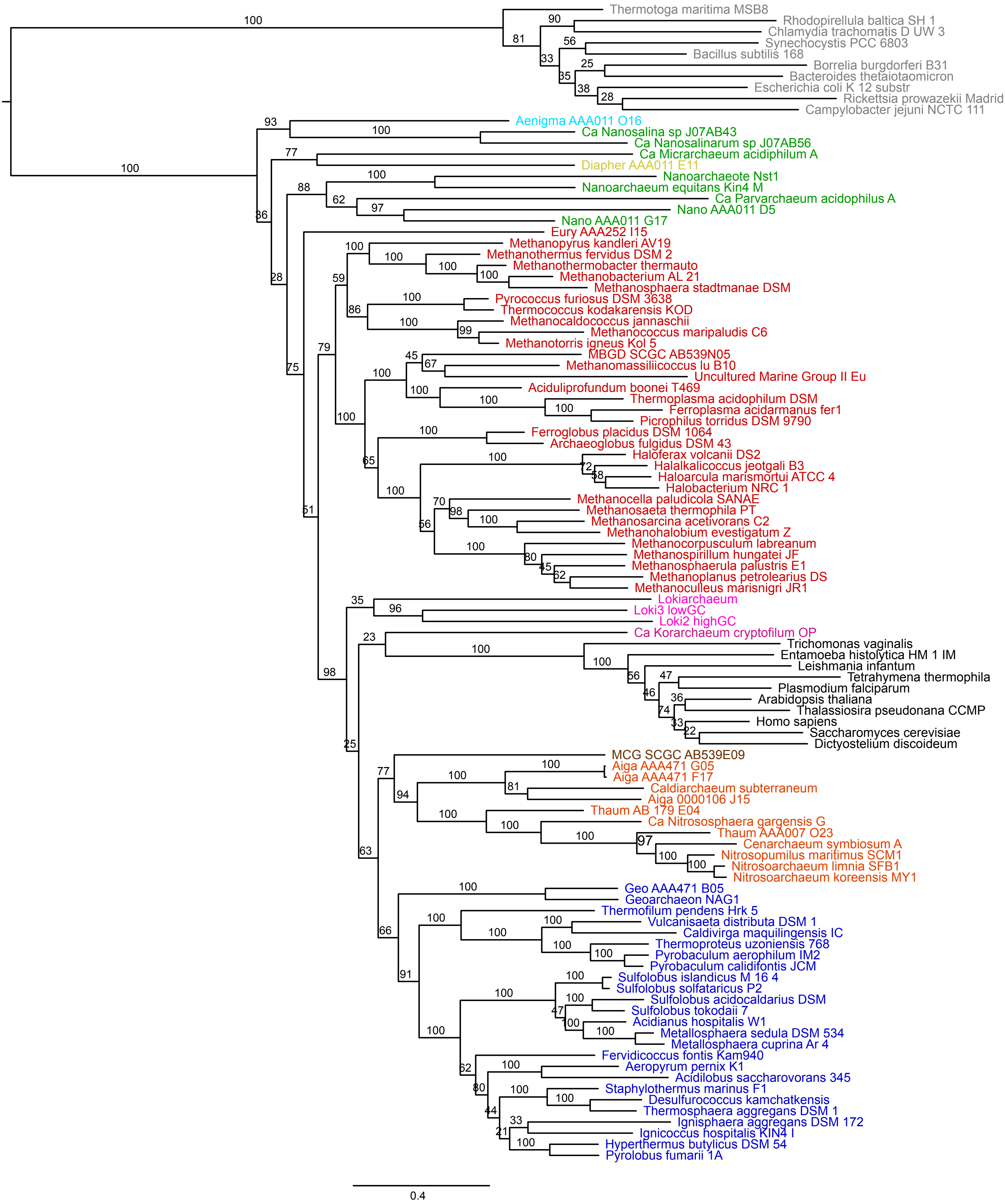
Suppl. Figure S7: Lokiarchaeota represent a deeply-rooting clade of the TACK superphylum. Bayesian and Maximum likelihood analysis based on a concatenation of the alignment of 36 highly conserved bacterial and archaeal proteins. The phylogenetic tree shows high support for the proposed Lokiarchaeota phylum (comprising Lokiarchaeum, Loki2 and Loki3), representing a deeply-rooting clade of the TACK superphylum. Sequences were aligned with Mafft-LINSi v. 7.130b, positions with >50% gaps were removed. The phylogeny was inferred with PhyloBayes MPI 1.5a25, using the CAT model and a GTR substitution matrix. Numbers above branches show posterior probabilities (PP) of Bayesian phylogeny only if they were > 0.8, and maximum-likelihood bootstrap supports (BS) > 70. The scale represents the number of substitutions per site. Higher taxonomic groups are colour-coded: grey, Bacteria; blue, Crenarchaeota; red, Euryarchaeota; orange, Aig- and Thaumarchaeota; green, Nano- and Nanohaloarchaeota; yellow, Diapherotrites; teal, Aenigmarchaeota; brown, Miscellaneous Crenarchaeal Group (MCG); purple, Korarchaeota; pink, Lokiarchaeota.



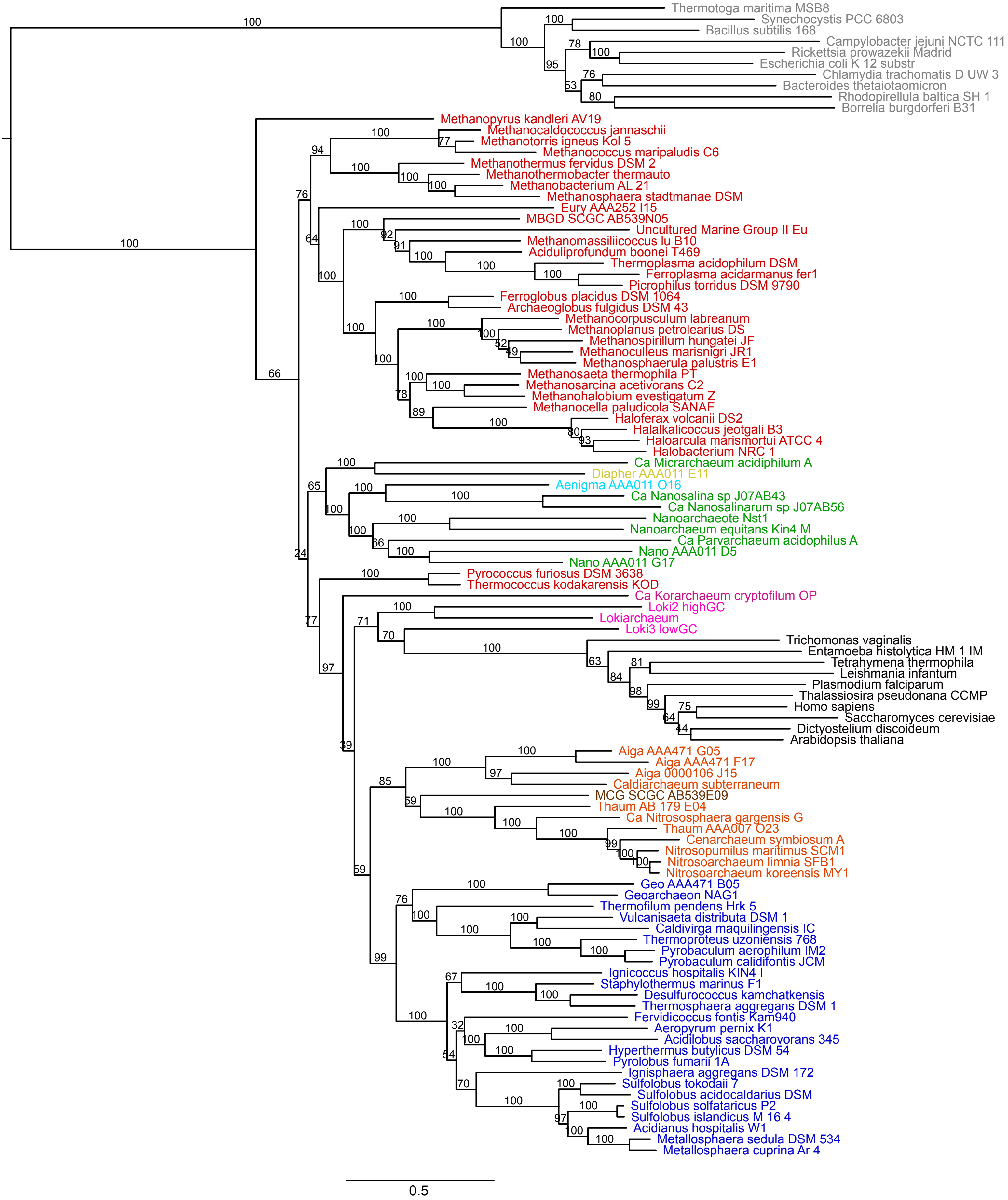
Suppl. Figure S8: Eukaryotes are placed within the Lokiarchaeota in the Tree of Life. Bayesian phylogeny based on a concatenation of the alignment of 36 highly conserved proteins. In the phylogenetic tree, eukaryotes are placed within the Lokiarchaeota with high support (PP=1; BS=80). The scale represents the number of substitutions per site. Methods, colour coding and scale as described in Suppl. Figure S7.



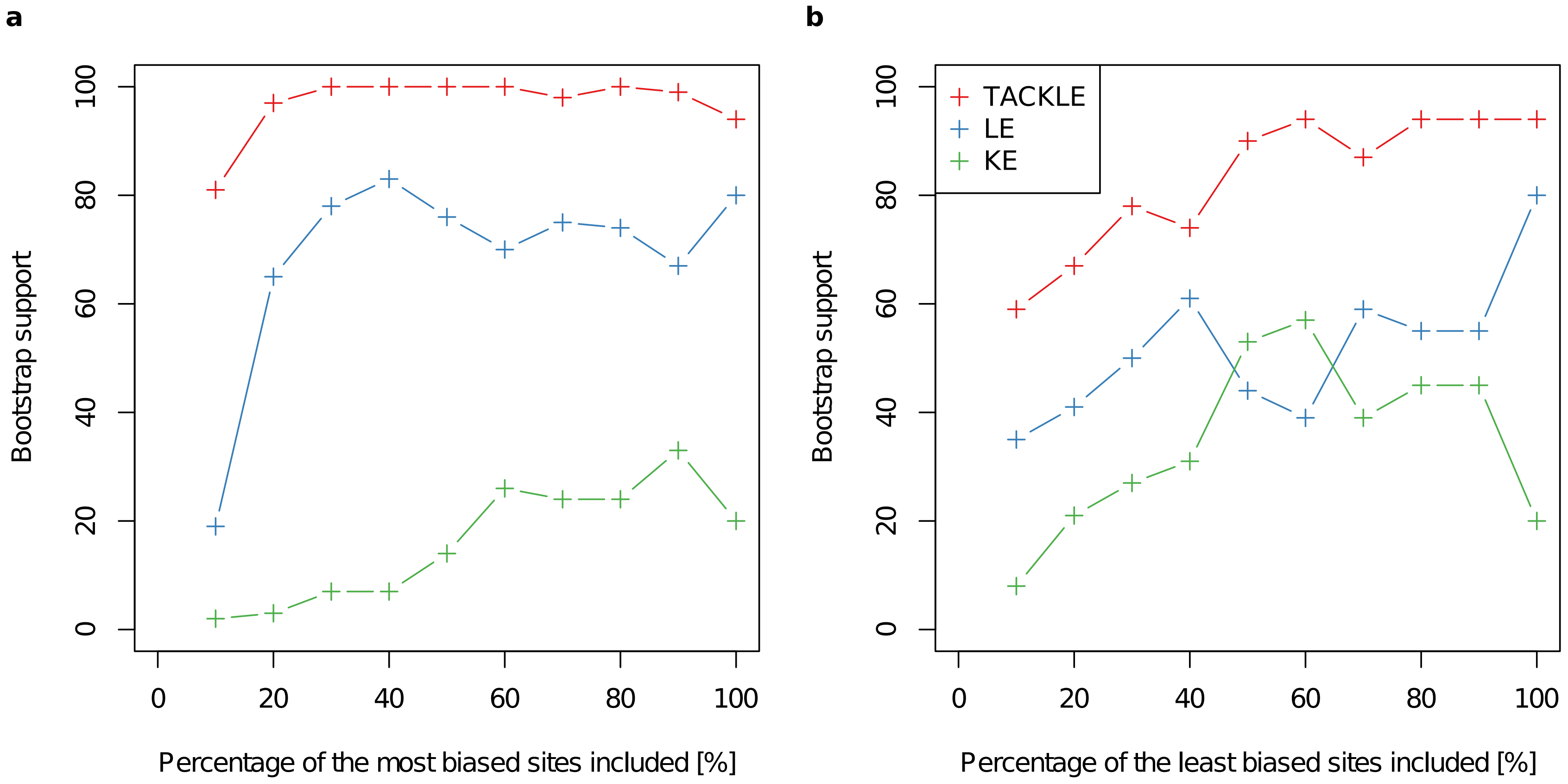
Suppl. Figure S9: Maximum likelihood phylogeny based on datasets and analysis described for Suppl. Figure S8. Numbers on branches indicate bootstrap support. Colour coding, bootstraps and scale as described in Suppl. Figure S7.



Suppl. Figure S10: Maximum likelihood phylogeny of ribosomal protein markers (21 markers, Suppl. Table S2). Methods, colour coding and scale as described in Suppl. Figure S7.



Suppl. Figure S11: Maximum likelihood phylogeny of non-ribosomal protein markers (15 markers; Suppl. Table S2). Methods, colour coding and scale as described in Suppl. Figure S7.

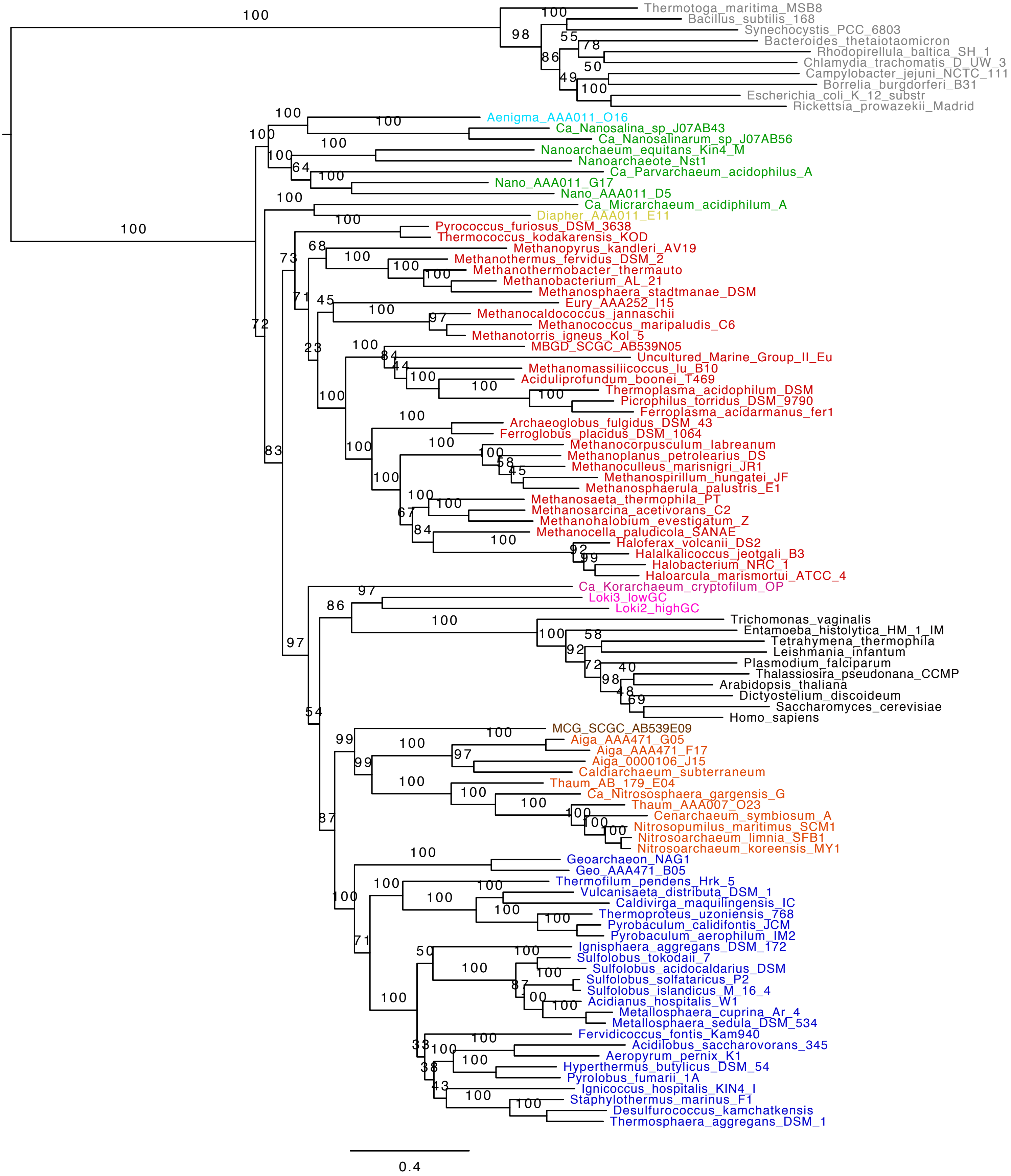


Suppl. Figure S12: Compositional bias analyses support the Lokiarchaeota-Eukarya affiliation. Effect of removal of compositional bias on the phylogenomic affiliation of Lokiarchaeum and eukaryotes is shown (a-b). The sites of the concatenated alignment of 36 conserved proteins were ranked according to their compositional bias, and increasing amounts (x-axis) of either the most (a) or the least (b) biased sites were included in maximum likelihood phylogeny. Alignment, position trimming and phylogenies were performed as in Suppl. Figure S7. In each phylogeny, the bootstrap support of three bipartitions was recorded (y-axis). The three tested bipartitions correspond to (i) the TACK superphylum, Lokiarchaeota and Eukaryotes forming a clade (TACKLE, red); (ii) Lokiarchaeota being associated with Eukaryotes (LE, blue); (iii) Korarchaeota being associated with eukaryotes (KE, green).

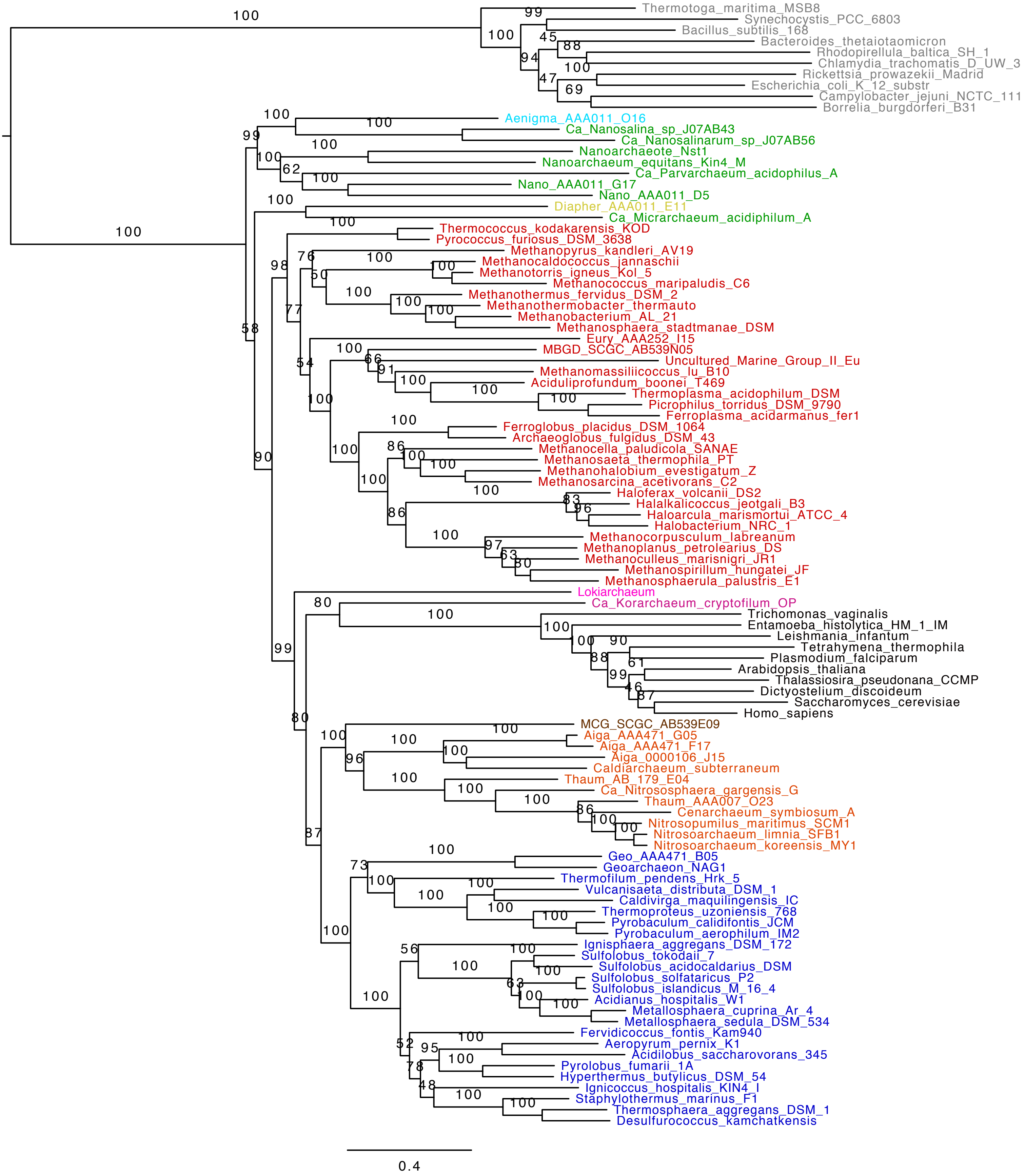
Supplementary Figure S13

The figure panels can be found on the next pages

Suppl. Figure S13: Effect of taxon removal on the phylogenomic affiliation between Lokiarchaeum and eukaryotes. Maximum likelihood phylogenies based on a concatenation of the alignment of 36 highly conserved marker proteins, as in Suppl. Figure S7, from which various taxa have been removed. See Suppl. Table S5 for a summary. Taxa removed were: Lokiarchaeum **(a)**, Loki2 and Loki3 **(b)**, all Lokiarchaeota **(c)**, Bacteria **(d)**, Eukarya **(e)**, Bacteria and Eukarya **(f)**, DPANN archaea **(g)**, Euryarchaeota **(h)**, Thaumarchaeota **(i)**, Crenarchaeota **(j)**, Aigarchaeota **(k),** Korarchaeota **(l)** and SAGs **(m)**. Colour-coding, bootstraps and scale as in Suppl. Figure S7.

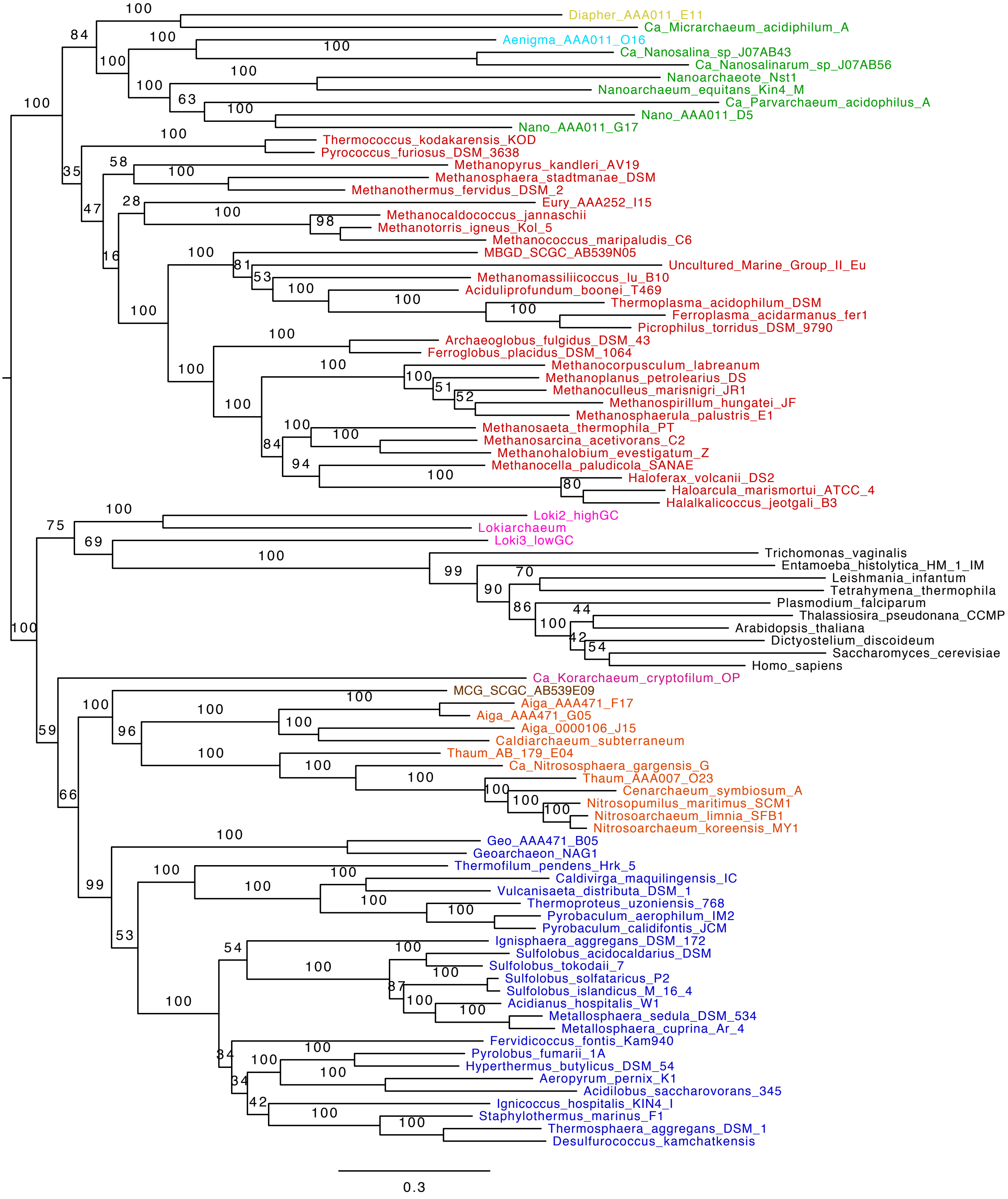


Supp. Fig. S13a: 36 conserved arCOGs, without Lokiarchaeum / RAxML / PROTGAMMALG



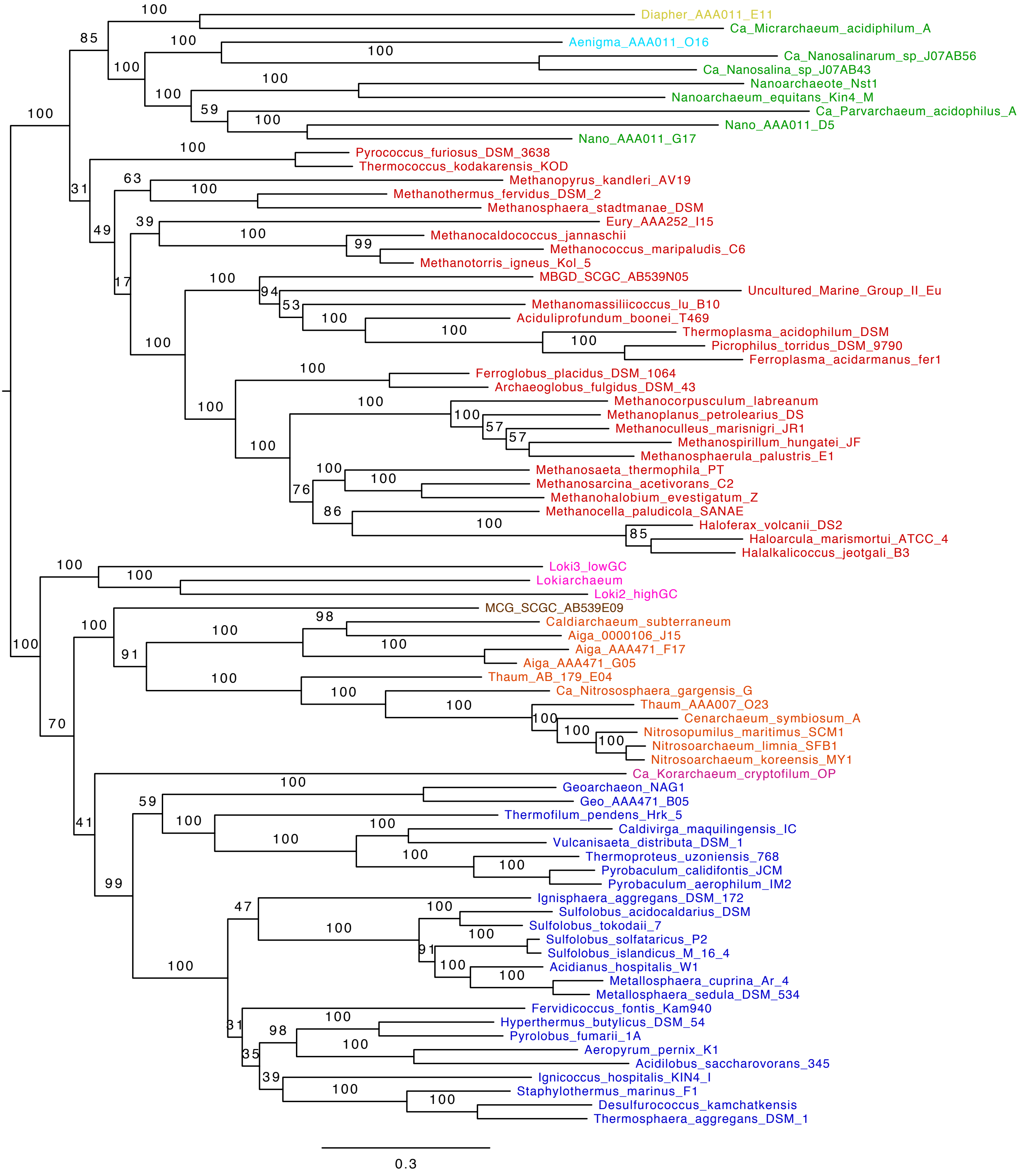
Supp. Fig. S13b: 58 conserved arCOGs, without Loki2/Loki3 / RAxML / PROTGAMMALG

Supp. Fig. S13c: 58 conserved arCOGs, without Lokiarchaeum/Loki2/Loki3 / RAxML / PROTGAMMALG



Supp. Fig. S13d: 36 conserved arCOGs, without Bacteria / RAxML / PROTGAMMALG

Su**pp. Fig. S13e**: 36 conserved arCOGs, without Eukaryotes / RAxML / PROTGAMMALG



Supp. Fig. S13f: 36 conserved arCOGs, without Bacteria/Eukayotes / RAxML / PROTGAMMALG

Supp. Fig. S13g: 36 conserved arCOGs, without DPANN / RAxML / PROTGAMMALG

Supp. Fig. S13h: 36 conserved arCOGs, without Euryarchaeota / RAxML / PROTGAMMALG

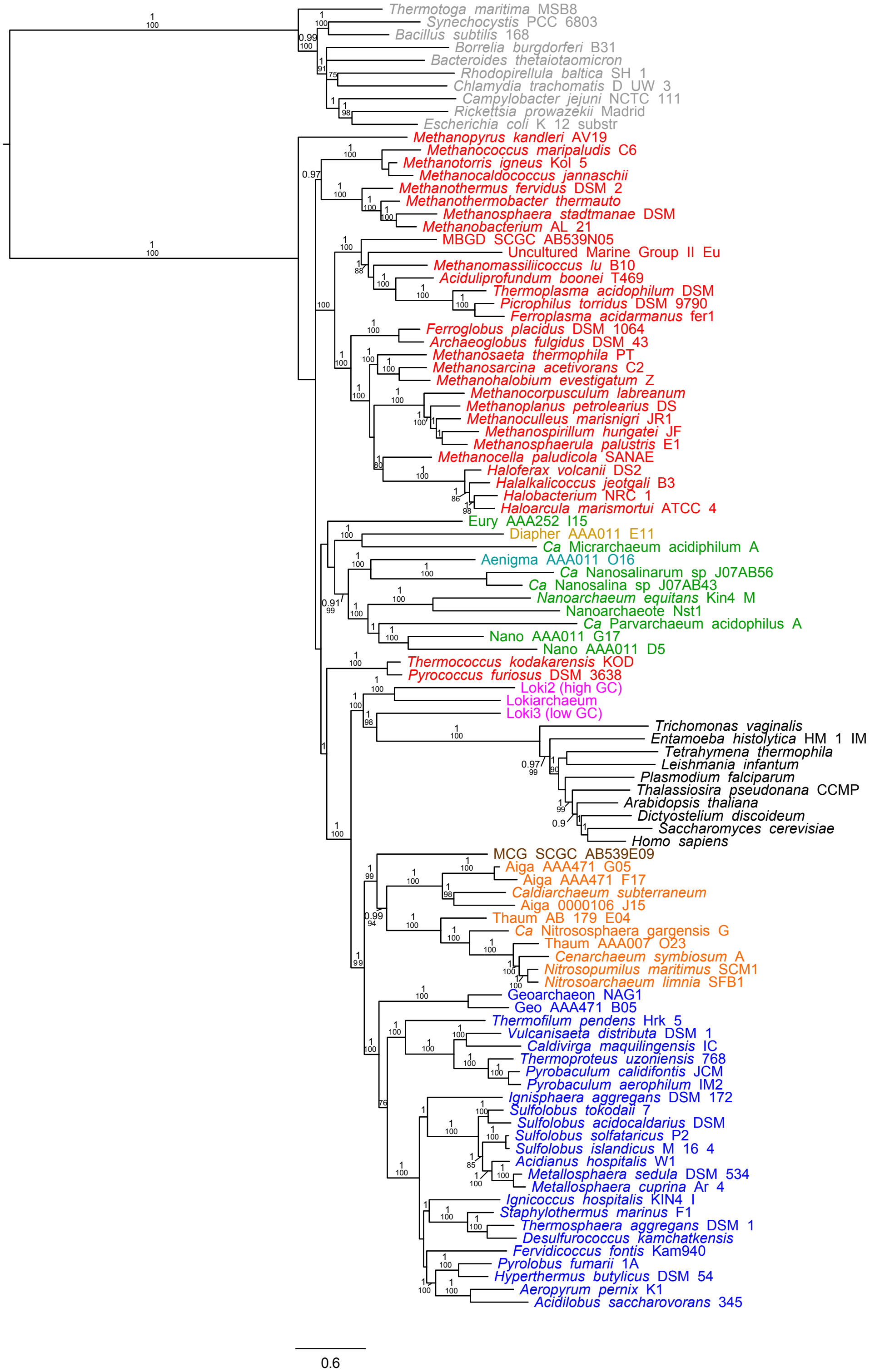
Supp. Fig. S13i: 36 conserved arCOGs, without Thaumarchaeota / RAxML / PROTGAMMALG

Supp. Fig. S13j: 36 conserved arCOGs, without Crenarchaeota / RAxML / PROTGAMMALG

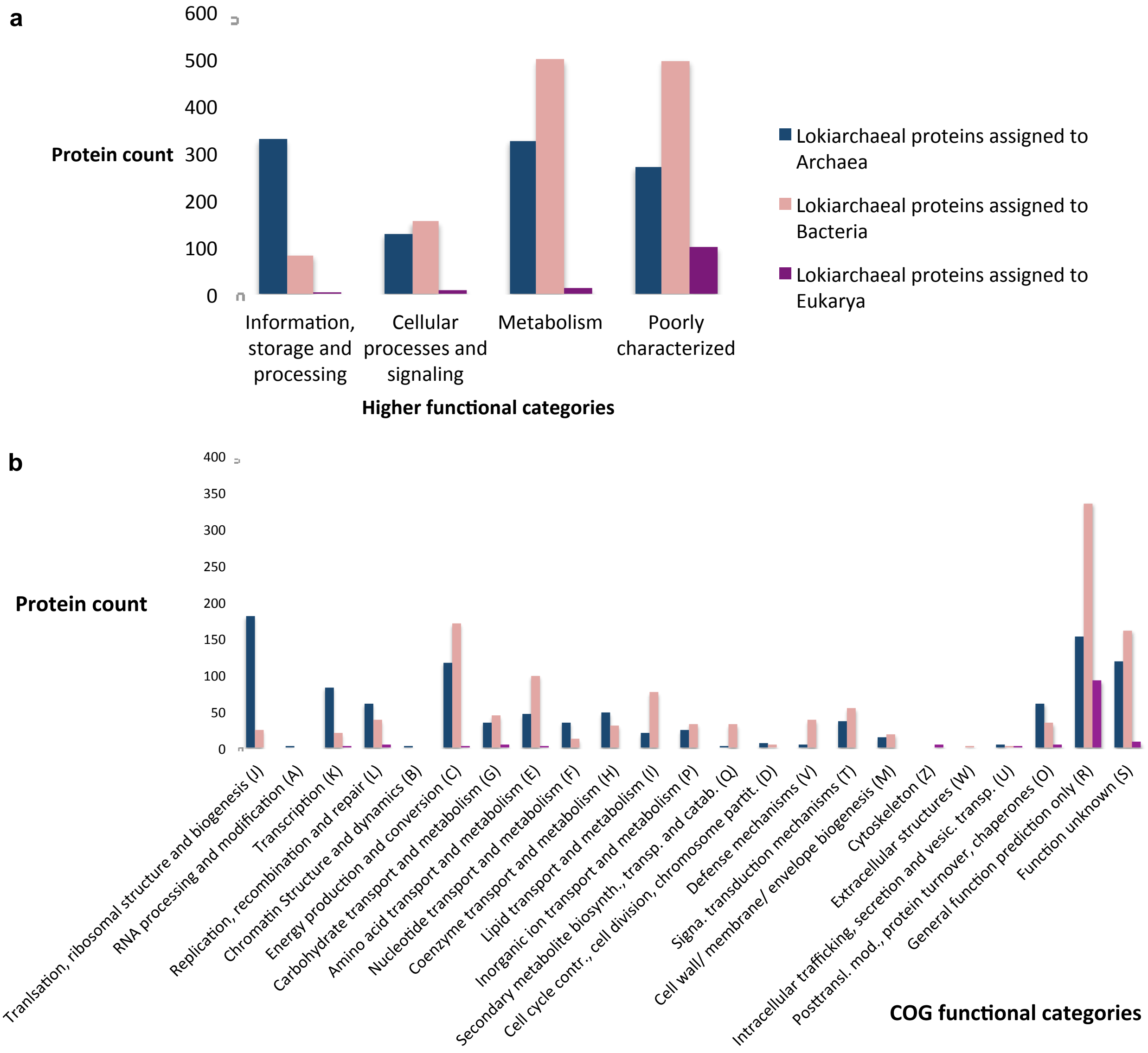
Supp. Fig. S13k: 36 conserved arCOGs, without Aigarchaeota / RAxML / PROTGAMMALG

Supp. Fig. S13l: 36 conserved arCOGs, without Korarchaeota / RAxML / PROTGAMMALG

Supp. Fig. S13m: 36 conserved arCOGs, without SAGs / RAxML / PROTGAMMALG

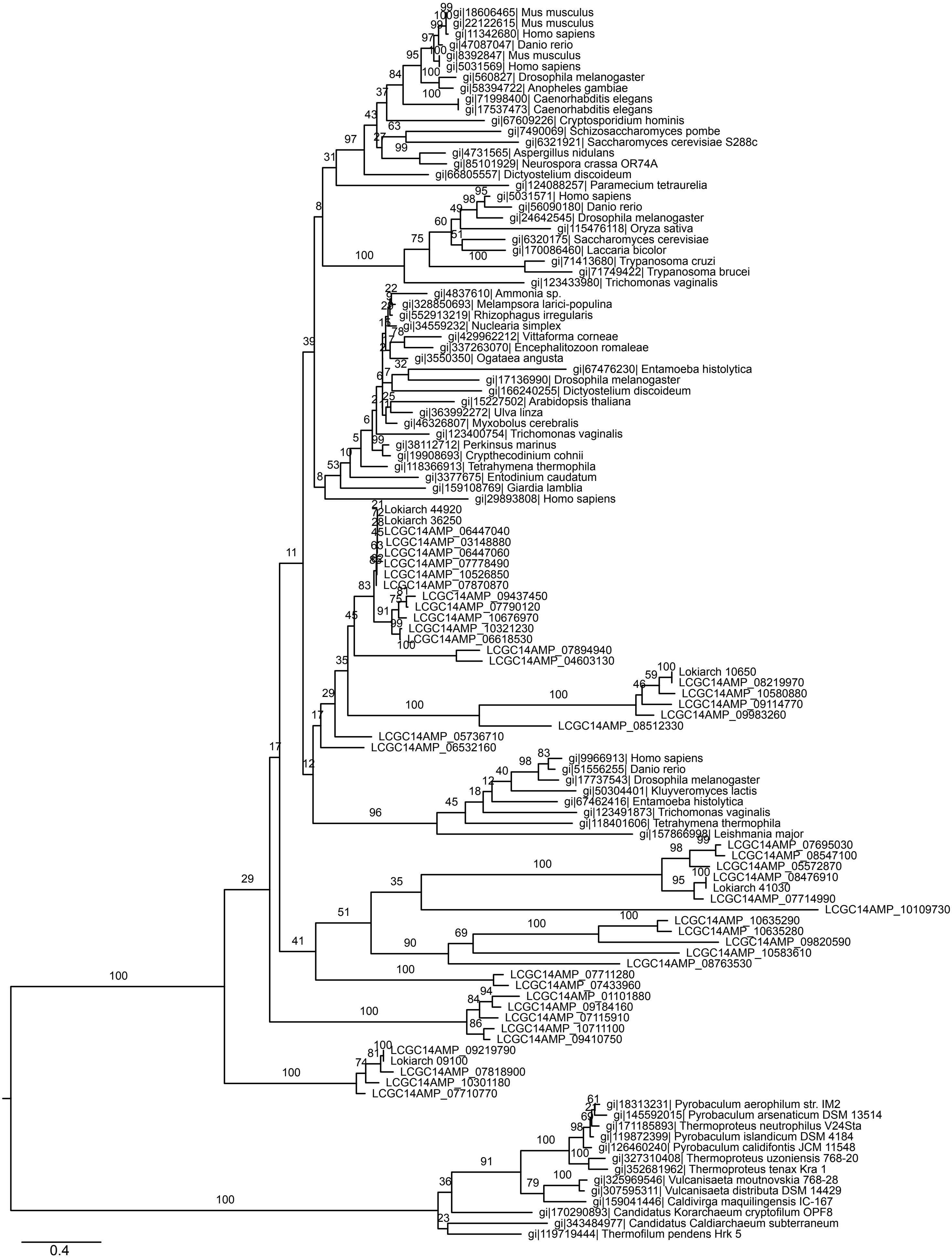


Suppl. Figure S14: Bayesian and Maximum likelihood analysis based on a concatenation of the alignment of 36 highly conserved bacterial and archaeal proteins excluding Korarchaeota. Sequences were aligned with Mafft L-INS-i v. 7.130b and positions with >50% gaps were removed. The phylogeny was inferred with PhyloBayes MPI 1.5a25 using the CAT model and a GTR substitution matrix. Numbers above branches represent posterior probabilities (PP) of Bayesian phylogeny (>0.8) and maximum likelihood (ML) bootstrap support (BS) (> 70). Scale represents the number of substitutions per site based on ML tree. Higher taxonomic groups are color-coded: grey, Bacteria; black, eukaryotes; blue, Crenarchaeota; red, Euryarchaeota; orange, Aig- and Thaumarchaeota; green, Nano- and Nanohaloarchaeota; yellow, Diapherotrites; teal, Aenigmarchaeota; brown, Miscellaneous Crenarchaeal Group (MCG); pink, Lokiarchaeota.



Suppl. Figure S15: Overview of functional classification (COG category) of lokiarchaeal proteins assigned to major taxonomic categories. a) Protein counts based on higher functional categories. **b)** Protein counts based on all COG functional categories. Taxonomic categories are color-coded, e.g. Archaea are shown in dark blue; Bacteria in pink and Eukarya in purple.

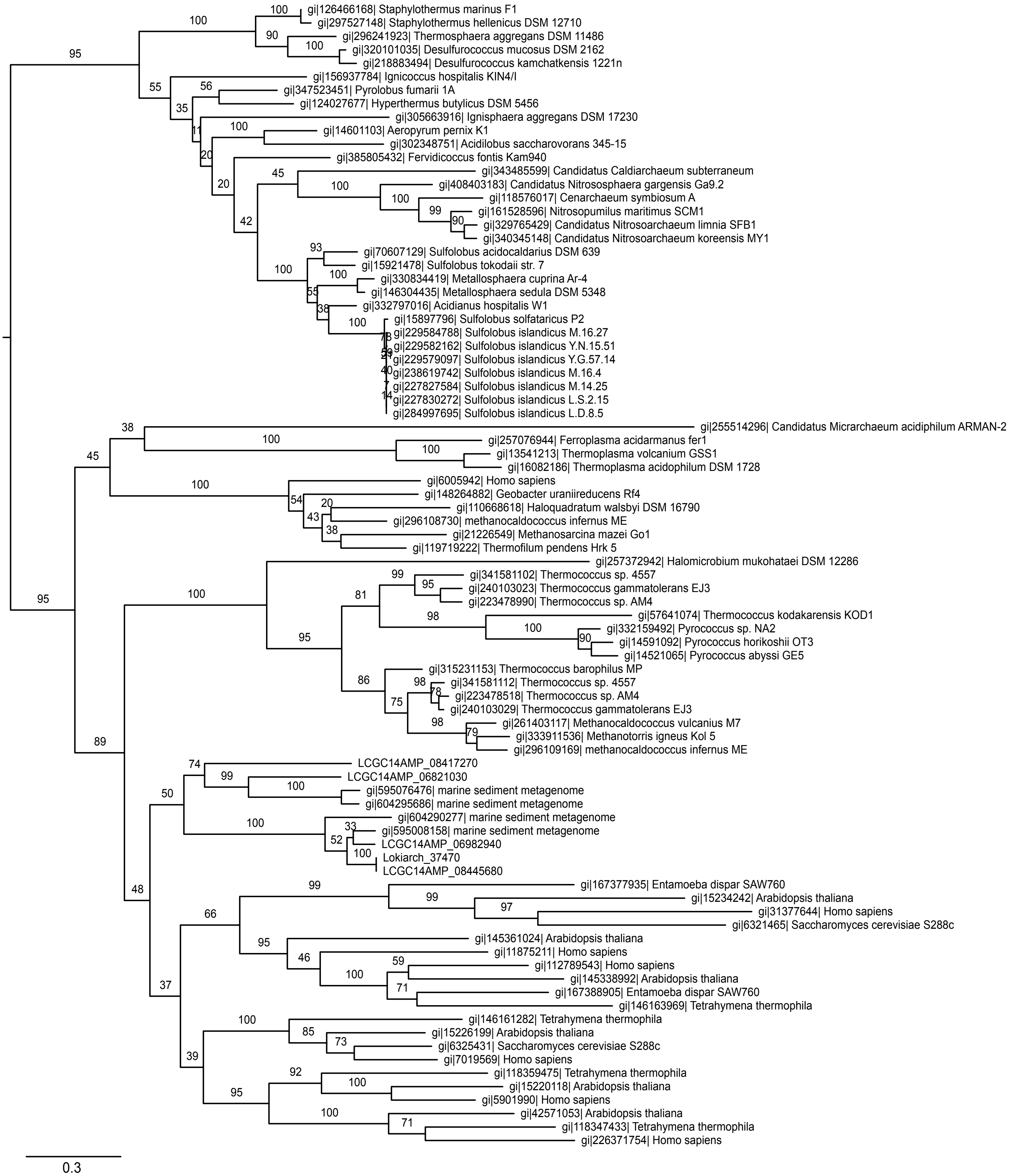
Suppl. Figure S16: Location and genomic context of ESPs, other proteins affiliated to eukaryotes as well as additional proteins potentially involved in extended archaeal cellular complexity (EACC) in the Lokiarchaeum composite genome. Each contig containing at least one protein of these two sets is represented, in no particular order. The contig number (shortened for the sake of space) is shown above the contig. Proteins taxonomically affiliated to Eukaryota are shown in purple, with a larger dot. Proteins potentially invovled in EACC are shown with small orange circles inside. Colouring according to the taxonomic affiliation obtained from MEGAN (see section 1.12 of the Methods section), simplified at domain level. Green, cellular organisms; blue, Archaea; red, Bacteria; grey, Other; dark grey, No hits.



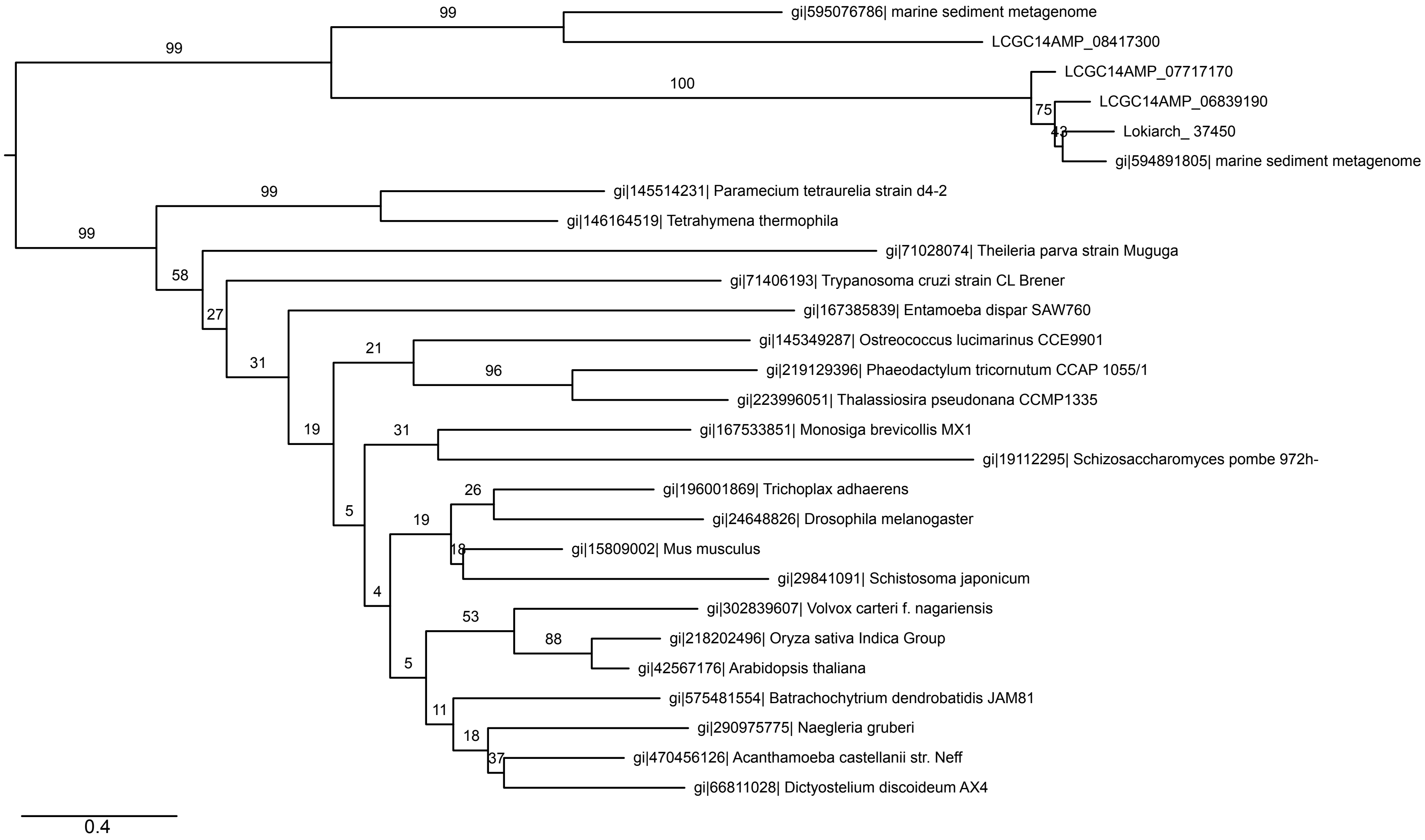
Suppl. Figure S17: Un-collapsed maximum likelihood phylogeny based on 378 aligned amino acid positions of eukaryotic, and archaeal actin homologs as well as of homologs identified in Lokiarchaeum and LCGC14AMP (shown in Fig. 3a). Only eukaryotic actins and the actin-related protein families 1-3 were included in this reconstruction. (A maximum likelihood phylogeny including also the more distant eukaryotic Arp families can be provided upon request). Sequences were aligned using MAFFT L-INS-i and 100 slow maximum likelihood bootstraps were calculated and are shown at their respective branches. Crenactin was used as out-group. The scale bar represents number of amino acid substitutions per site.

Suppl. Figure S18: Un-collapsed maximum likelihood phylogeny based on 150 aligned amino acid positions of small Ras- and Arf-type GTPases (IPR1806 and IPR6689) including all homologs identified in Lokiarchaeum and other archaea, and as well as representative sets of eukaryotic and bacterial sequences (see Methods). Sequences were aligned using MAFFT L-INS-i and 100 slow bootstraps were calculated. Scale bar indicates number of amino acid substitutions per site.

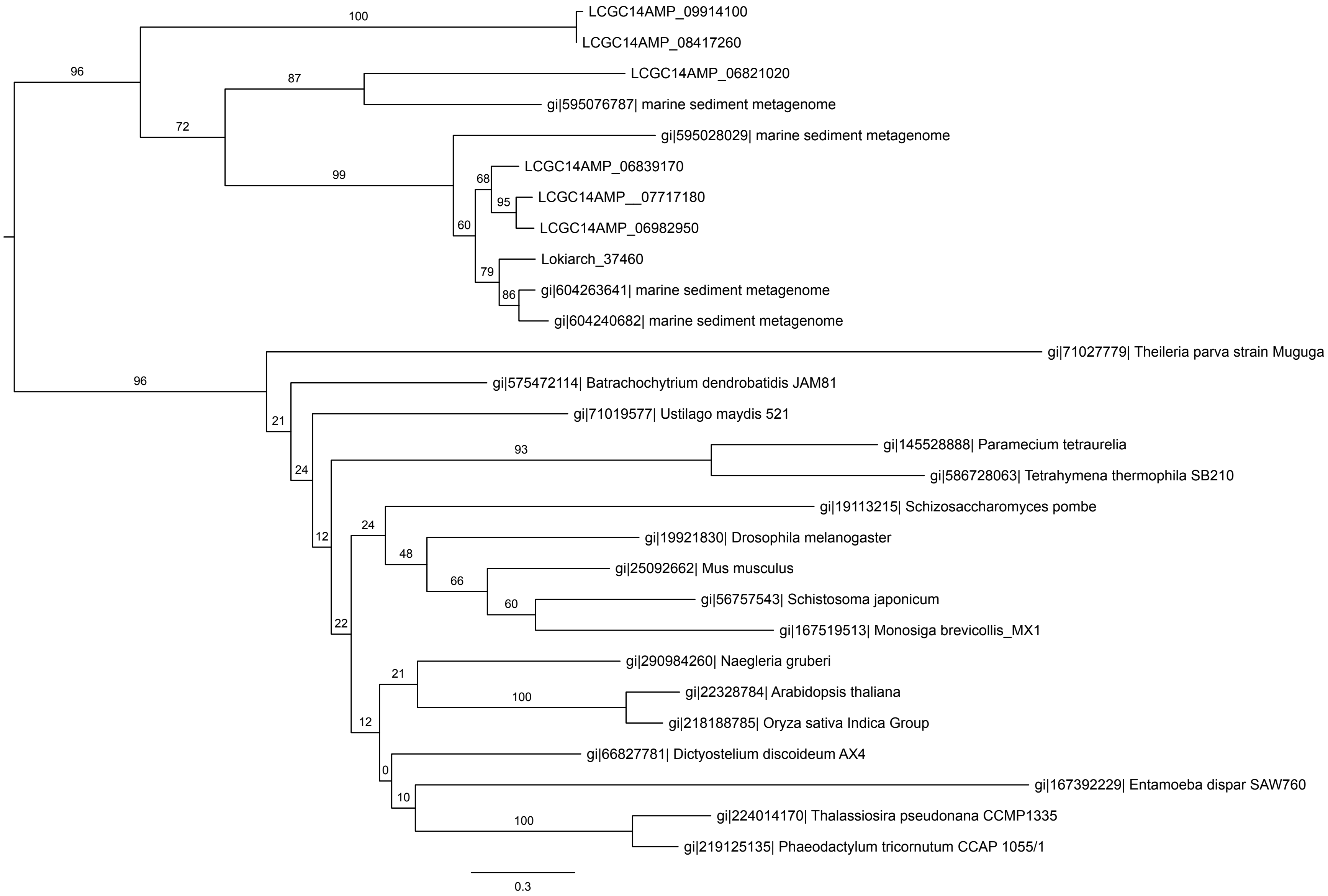
Suppl. Figure S19: Un-collapsed maximum likelihood phylogeny based on 207 aligned amino acid positions of archaeal ESCRT-III homologs, eukaryotic Vps2/24/46 and Vps20/32/60 family proteins as well as homologs present in Lokiarchaeum as well as LCGC14AMP (shown in Fig. 4b). Sequences were aligned using MAFFT L-INS-i and 100 slow bootstraps were calculated. Scale bar indicates number of amino acid substitutions per site.



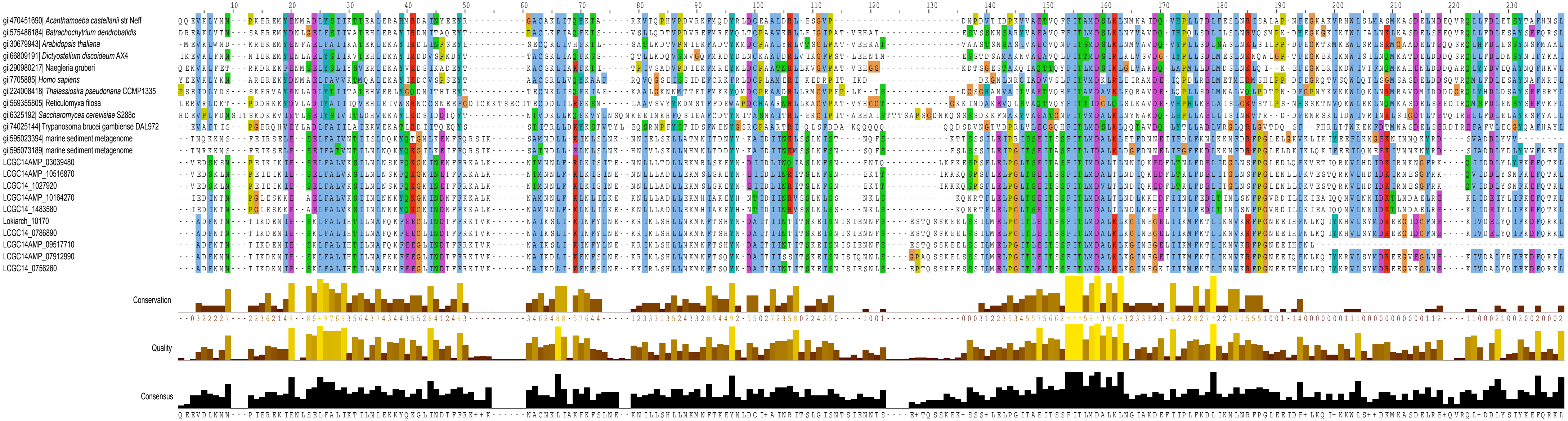
Suppl. Figure S20: Un-collapsed maximum likelihood phylogeny based on 388 aligned amino acid positions of AAA-type ATPase Vps4 homologs including homologs present in Lokiarchaeum as well as LCGC14AMP (shown in Fig. 4c). Sequences were aligned using MAFFT L-INS-i and 100 slow bootstraps were calculated. Scale bar indicates number of amino acid substitutions per site.



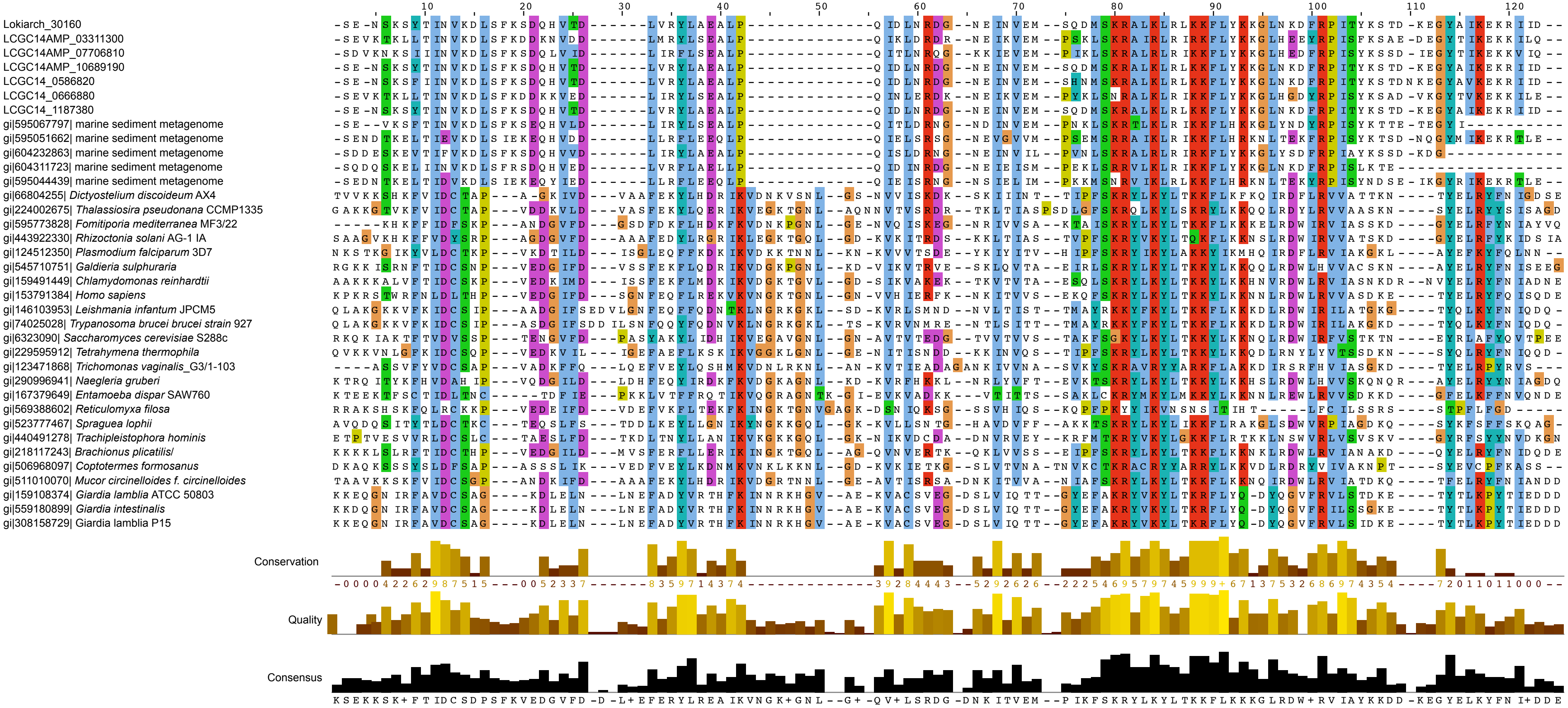
Suppl. Figure S21: Maximum likelihood phylogeny based on 226 aligned amino acid positions of lokiarchaeal EAP30 domain proteins including homologs present in Lokiarchaeum as well as LCGC14AMP. Sequences were aligned using MAFFT L-INS-i and 100 slow bootstraps were calculated. Scale bar indicates number of amino acid substitutions per site.



Suppl. Figure S22: Maximum likelihood phylogeny based on 168 aligned amino acid positions of Vps25 homologs including sequences of Lokiarchaeum as well as the LCGC14AMP metagenome. Sequences were aligned using MAFFT L-INS-i and 100 slow bootstraps were calculated. Scale bar indicates number of amino acid substitutions per site.



Suppl. Figure S23: Multiple sequence alignment of eukaryotic and lokiarchaeal Vps28 homologs. The alignment contains homologs of Lokiarchaeota and homologous sequences identified in publicly available metagenomes and was generated with MAFFT G-INS-i using default settings and the ends were manually trimmed.



Suppl. Figure S24: Multiple sequence alignment of eukaryotic and lokiarchaeal ribosomal protein L22 homologs. The alignment contains homologs of Lokiarchaeota as well as homologous sequences present in publicly available metagenomes. The alignment was generated with MAFFT L-INS-i using default settings and the ends were manually trimmed.