Institut Pasteur

Research in Microbiology 164 (2013) 425–438

www.elsevier.com/locate/resmic

Metagenomics of Kamchatkan hot spring filaments reveal two new major (hyper)thermophilic lineages related to Thaumarchaeota

Laura Eme [a,1], Laila J. Reigstad [b,c], Anja Spang [d], Anders Lanzén [b,c,e], Thomas Weinmaier [f], Thomas Rattei [f], Christa Schleper [b,c,d,*], Céline Brochier-Armanet [a,**]

[a] Aix-Marseille Université, Laboratoire de Chimie Bacterienne – UMR CNRS 7283, IFR88, 31 chemin Joseph Aiguier, F-13402 Marseille, France
[b] Centre for Geobiology, Department of Biology, Allégaten 41, University of Bergen, N-5020 Bergen, Norway
[c] Centre for Geobiology, Department of Earth Science, Allégaten 41, University of Bergen, N-5020 Bergen, Norway
[d] Department of Genetics in Ecology, University of Vienna, Althanstrasse 14, A-1090 Vienna, Austria
[e] Computational Biology Unit, Bergen Center for Computational Science, Thormøhlensgate 55, University of Bergen, N-5020 Bergen, Norway
[f] Department of Computational Systems Biology, University of Vienna, Althanstrasse 14, A-1090 Vienna, Austria

Received 8 November 2012; accepted 11 February 2013
Available online 5 March 2013

Abstract

Based on phylogenetic analyses and gene distribution patterns of a few complete genomes, a new distinct phylum within the Archaea, the Thaumarchaeota, has recently been proposed. Here we present analyses of six archaeal fosmid sequences derived from a microbial hot spring community in Kamchatka. The phylogenetic analysis of informational components (ribosomal RNAs and proteins) reveals two major (hyper-) thermophilic clades ("Hot Thaumarchaeota-related Clade" 1 and 2, HTC1 and HTC2) related to Thaumarchaeota, representing either deep branches of this phylum or a new archaeal phylum and provides information regarding the ancient evolution of Archaea and their evolutionary links with Eukaryotes.
© 2013 Institut Pasteur. Published by Elsevier Masson SAS. All rights reserved.

Keywords: Archaea; Phylogeny; Hyperthermophily; Ammonia oxidation; *Caldiarchaeum subterraneum*; Hot Thaumarchaeota-related Clade

1. Introduction

Soon after its discovery, the domain Archaea has been divided into two major phyla, the Euryarchaeota and the Crenarchaeota (Woese et al., 1990). Though members of this domain were initially considered to be restricted to extreme and/or anaerobic environments, molecular environmental surveys have later demonstrated that Archaea are also found in moderate environments including soils, sediments as well as marine and freshwater habitats (Robertson et al., 2005; Schleper et al., 2005). Furthermore, these surveys revealed new major archaeal lineages, including putative phyla such as Korarchaeota (Barns et al., 1996), Nanoarchaeota (Huber et al., 2002), Thaumarchaeota (initially referred to as mesophilic Crenarchaeota or group I, and encompassing all currently known Ammonia-Oxidizing Archaea (AOA)) (Brochier-Armanet et al., 2008; DeLong, 1992; Fuhrman et al., 1992), as well as the Hot Water Crenarchaeotic Group I (HWCGI)/'Aigarchaeota' (Nunoura et al., 2011). Comparative genomic and phylogenetic analyses have confirmed the initial proposal of Korarchaeota as a separate phylum within the Archaea (Elkins et al., 2008), as well as of Thaumarchaeota (Spang et al., 2010), whereas Nanoarchaeota have been

* Corresponding author. Department of Genetics in Ecology, University of Vienna, Althanstrasse 14, A-1090 Vienna, Austria. Tel.: +43 1427757800; fax: +43 142779578.
** Corresponding author. Present address: Université de Lyon, Université Lyon 1, CNRS, UMR5558, Laboratoire de Biométrie et Biologie Evolutive, 43 boulevard du 11 novembre 1918, F-69622 Villeurbanne, France. Tel.: +33 4 26 23 44 76; fax: +33 4 72 43 13 88.
E-mail addresses: christa.schleper@univie.ac.at (C. Schleper), celine.brochier-armanet@univ-lyon1.fr (C. Brochier-Armanet).
[1] Present address: Centre for Comparative Genomics and Evolutionary Bioinformatics, Dalhousie University, Halifax, Canada.

0923-2508/$ - see front matter © 2013 Institut Pasteur. Published by Elsevier Masson SAS. All rights reserved.
http://dx.doi.org/10.1016/j.resmic.2013.02.006

suggested to represent a fast evolving lineage within the Euryarchaeota (Brochier et al., 2005; Makarova and Koonin, 2005). Finally, the taxonomic status of 'Aigarchaeota' as a separate archaeal phylum remains under debate (Brochier-Armanet et al., 2012; Gribaldo and Brochier-Armanet, 2012).

The diversity and abundance of SSU rRNA and *amoA* genes (encoding subunit A of ammonia monooxygenase) indicate that Thaumarchaeota range among the most widespread and prolific microorganisms on our planet (Gubry-Rangin et al., 2011; Karner et al., 2001; Leininger et al., 2006; Wuchter et al., 2006) and encompass diverse sub-lineages – group I.1a, group I.1b or, HWCG III/Thermophilic AOA (ThAOA) – represented by cultivated members (Konneke et al., 2005; Tourna et al., 2011) or enriched cultures (de la Torre et al., 2008; French et al., 2012; Hatzenpichler et al., 2008; Jung et al., 2011; Kim et al., 2012; Lehtovirta-Morley et al., 2011; Mosier et al., 2012b,c). Further uncultivated lineages with a putative affiliation to Thaumarchaeota have been identified by environmental SSU rRNA surveys (e.g. SCG, FSCG, group I.1c, group 1A/psL12, SAGMCG-I) (see Brochier-Armanet et al., 2012; Pester et al., 2011 and references therein). Finally, genome sequences are currently only available for a few members of the groups I.1a and I.1b within the Thaumarchaeota (Blainey et al., 2011; Hallam et al., 2006; Kim et al., 2011; Mosier et al., 2012a,b; Spang et al., 2012; Walker et al., 2010). Inclusion of more sequence data is thus needed to better characterize the diversity and the boundaries of this phylum.

Although widely distributed across the archaeal tree, hyperthermophiles constitute the first diverging lineages of the currently described archaeal phyla (Crenarchaeota, Euryarchaeota, Korarchaeota) indicating that the last common ancestor of Archaea might have been a hyperthermophile (Forterre et al., 2000, 2002; Gribaldo and Brochier-Armanet, 2006; Stetter, 1996; Woese, 1987; Woese et al., 1990). Thaumarchaeota are no exception to this observation as testified by the discovery of deeply branching thaumarchaeotal SSU rRNA and *amo*A genes in terrestrial hot springs (Dodsworth et al., 2011; Reigstad et al., 2008; Zhang et al., 2008), and the enrichment culture of '*Ca*. Nitrosocaldus yellowstonii' (a representative of ThAOA/HWCG III group) which grows at 74 °C (de la Torre et al., 2008). This suggests that mesophilic representatives of this phylum are derived from a thermophilic or hyperthermophilic ancestor and adapted secondarily to colder habitats (Barns et al., 1996; Brochier-Armanet et al., 2012; Hatzenpichler et al., 2008; Lopez-Garcia et al., 2004; Preston et al., 1996; Reigstad et al., 2008; Schleper et al., 1997). However, because of the scarcity of biological and genomic data, these thermophilic thaumarchaeotal lineages are poorly known.

Here we investigate a fosmid library derived from a terrestrial hot spring (85 °C and pH 5.5) from Uzon Caldera in Kamchatka. The phylogenetic analyses of SSU and LSU rRNA and informational proteins carried by six large genomic fragments revealed two major groups, HTC1 (Hot Thaumarchaeota-related Clade 1) and HTC2 that branched-off before the diversification of the currently recognized thaumarchaeotal lineages. The analysis of the 249 genes encoded by these genomic fragments provides first insights into these two groups. The evolutionary relationships between HTC1, HTC2, Thaumarchaeota and other archaeal lineages, as well as between Eucarya and Archaea, are discussed.

2. Materials and methods

2.1. Site description and sampling

The terrestrial hot spring, referred to as Kam37, is located close to the ranger hut in the Central thermal field of Uzon Caldera in Kamchatka, Russia. It had a small bottom opening of 10 × 15 cm with a steady discharge and very little rim overflow. The sample taken in August 2005 was harvested with a sterile syringe well below the surface. It consisted of finger-long greyish filaments continuously flushed and attached to the edges of the small spring opening. The temperature within these filaments was 85 °C, while the spring fluid had a temperature of 91 °C and a pH of 5.5. The sample was taken from the bottom of the accessible part of the hot spring to avoid material that had been exposed to temporary dryness or longer periods of direct air exposure. Chemical analyses showed the thermal fluid to be mainly composed of Na (225 mg/l), K (16 mg/l), Ca (35 mg/l), and Mg (3.2 mg/l) with high contents of Fe (789 µg/l), Mn (370 µg/l) and B (17 mg/l). These values were similar to the chemical data obtained from the springs referred to as Jenn's Pools located close to Kam37 (Kyle et al., 2007).

2.2. Extraction of high molecular weight DNA

To obtain high molecular weight (hmw) DNA for the metagenomic analyses, 15.4 g (wet weight) of filaments were ground in a sterile mortar. The material was briefly centrifuged (10 s at 6000× g) in a tabletop centrifuge to remove remnants of ungrounded material. The supernatant was then centrifuged at 10,000× g for 30 min at 4 °C to pellet the cells. The cells were embedded in low-melt agarose plugs (1% agarose) and lysed chemically by lysozyme and Proteinase K as described elsewhere (Quaiser et al., 2002; Reigstad et al., 2011). The agarose plugs were stored at 4 °C in 1× TE buffer with 200 mM EDTA until construction of the metagenomic library.

2.3. Construction and screening of the metagenomic fosmid library

The hmw DNA from the agarose plugs was purified and size-analysed using a two-phase pulse-field gel electrophoresis (200 V, 5–50 s, 20 h, 14 °C) as described elsewhere (Quaiser et al., 2002; Reigstad et al., 2011). Approximately 0.5 µg of hmw DNA was used to construct the metagenomic library using the fosmid vector pEpiFOS-5 (EpiFOS™ Fosmid Library Production Kit, Epicentre) as recommended by the manufacturer. Approximately 36,000 recombinant clones were transferred to 384-well microtitre plates containing 50 µl of LB^+ medium and 7% glycerol (v/v) (Reigstad et al., 2011).

The plates were incubated at 37 °C for 24 h, and were subsequently stored at −80 °C. To facilitate PCR screening of the fosmid library, a DNA pool representing all fosmids of one 384-well microtitre plate was made by transferring a print of the plate onto an LB plate and subsequently pooling all colonies as described previously (Schleper et al., 1998). The metagenomic library was screened for archaeal SSU rRNA genes using primers 20F/958R as described below.

2.4. Construction of SSU rRNA clone libraries and taxonomic affiliation of SSU rRNA sequences

Two SSU rRNA gene libraries were made: one archaeal-specific library with the primers 20F/958R (DeLong, 1992), and one prokaryotic library using the primers 515F (Moyer et al., 1998) and 1408R (Amann et al., 1995). All libraries (including the fosmid library) were made from the same input DNA harvested from the agarose plug. 108 clones obtained from products with prokaryotic primers and 149 clones from the archaeal-specific library were analysed using RDPII (v8.1) Chimera Check (Cole et al., 2007) and the Bellerophon Chimera Detection Program (Huber et al., 2004), and their affiliation was determined by using the Maximum Likelihood and Neighbour Joining methods implemented in ARB (v07.12.07org) (Ludwig et al., 2004) on the data from the ARB-compatible SILVA database (Release104) (Pruesse et al., 2007).

2.5. Sequencing of fosmids and gene annotation

In order to expand our knowledge on the two clusters of interest (HTC1 and HTC2) revealed by the phylogenetic analyses of SSU rRNA sequences (cf. Results and discussion), we screened the fosmid library using the primers 20F/958R. Nine clones belonging to these clusters were identified, six of which were selected for full-length sequencing (1M19, 2C9, 1C18, 34P11, 1C23, and 1N15).

The six fosmid inserts selected for sequencing were purified (Qiagen Plasmid Mini Kit, Qiagen) and sequenced using the 454 pyrosequencing FLX technology (Norwegian High-Throughput Sequencing Centre, University of Oslo, Norway). Post-sequencing, the DNA reads (between 8600 and 12,800 reads per fosmid with approximately 250 bp read length) were assembled into contigs using the Newbler Assembler software v 2.3 (454 Life Sciences). Default settings were used except for setting the minimum read identity to 96% and the minimum read overlap to 50 nucleotides. The annotation of the contigs was performed using the Pedant Genome processing pipeline and annotation tool (Rattei et al., 2008; Walter et al., 2009). The analyses included predictions of: (i) ORFs, (ii) protein function by sequence similarity searches against the non-redundant protein sequence database and the FunCat catalogue, (iii) motif and conserved domains based on searches against several protein domain databases, and (iv) tRNA. Small (SSU) and large (LSU) subunit rRNA genes were manually identified using sequence similarity searches. The annotation was manually investigated and ORFs were assigned to COG categories (Cluster of Orthologous Group, www.ncbi.nlm.nih.gov/COG/) (Tatusov et al., 1997) using a BlastP/SIMAP *E*-value threshold of 10^{-5}.

2.6. Phylogenetic analyses

189 thaumarchaeotal SSU rRNA sequences were retrieved from the Ribosomal Database Project (RDPII, http://rdp.cme.msu.edu/; Release 10) (Cole et al., 2009). This dataset was enriched with the 1753 archaeal SSU rRNA sequences longer than 1200 nucleotides available at the RDPII. Additional archaeal SSU rRNA sequences of Thaumarchaeota, Korarchaeota, 'Aigarchaeota' and deeply branching uncultured archaeal lineages were retrieved from GenBank using BlastN at the NCBI. The retrieved sequences were aligned together with the Thaumarchaeota-related SSU rRNA sequences from our two clone libraries and from fosmids of this study (16 and 9 sequences, respectively) using MAFFT (default parameters) (Katoh et al., 2002). The resulting alignment was manually inspected and refined when necessary with ED, the alignment editor of the MUST package (Philippe, 1993). Regions of doubtful homology were removed with NET from the MUST package. A preliminary distance tree was inferred using the neighbour joining method implemented in PHYLIP (Felsenstein, 2004). Based on the resulting phylogeny, a subset of 105 SSU rRNA sequences reflecting the genetic diversity of Thaumarchaeota and other archaeal lineages was selected. These sequences were aligned and trimmed as described above. 1064 nucleic acid positions were conserved for phylogenetic analyses.

In a second step, a supermatrix was constructed by concatenating archaeal SSU and LSU rRNA alignments from complete archaeal genomes and fosmids available in GenBank and from the draft genome of '*Ca*. Nitrososphaera viennensis' (Tourna et al., 2011), and from our four Thaumarchaeota-related fosmids (1M19, 1C23, 1C18 and 34P11) that contained both genes. SSU and LSU rRNA datasets were aligned and trimmed separately as described above and concatenated leading to a supermatrix containing 3462 nucleic acid positions.

Finally, using the same approach we constructed four protein supermatrices (referred to as F1a, F2a, F1ae, and F2ae). The first supermatrix (F1a, 55 sequences, 653 amino acid positions) contained five proteins involved in information processing: L31e, L39e and S19e, RPR2 and EF-1α from 2C9, 1N15 and 34P11 fosmids and from complete archaeal genomes available at the NCBI. The L31e and L39e are ribosomal proteins of the 50S unit, whereas S19e is a part of the 30S unit. The RPR2 protein is a subunit of ribonuclease P, a protein complex generating mature tRNA from their precursors, while EF-1α is a subunit of the elongation factor-1 complex involved in protein biosynthesis. The second supermatrix (F2a, 53 sequences, 1193 amino acid positions) included the topoisomerase-primase Toprim and the polypeptide chain release factor aRF1 in addition to the five proteins mentioned above. This allowed the construction of a longer alignment but the inclusion of sequences from a single fosmid only (1N15). Two additional supermatrices F1ae (66 sequences, 618 amino acid positions) and F2ae (64 sequences,

1139 amino acid positions) were constructed by including eukaryotic homologues as outgroup to F1a and F2a.

Phylogenetic trees were computed using the maximum likelihood (ML) and the Bayesian methods implemented in Treefinder (Jobb et al., 2004) and MrBayes 3.1.2 (Ronquist and Huelsenbeck, 2003), respectively. According to the model selection tool implemented in Treefinder (AICc criterion), we used the Global Time Reversible (GTR) (Rodriguez et al., 1990) and the Le and Gascuel models (Le and Gascuel, 2008) for ML analyses of the nucleic and amino acid datasets, respectively. In both cases we included a gamma distribution (Γ_4) (four discrete categories and an estimated alpha parameter) to take into account evolutionary rate variations across sites. Branch robustness of ML phylogenetic trees was estimated by the non-parametric bootstrap procedure implemented in Treefinder (100 replicates). Bayesian analyses were performed using the GTR+ Γ_4 model (for the two rRNA datasets) or a mixed model $+\Gamma_4$ (for protein supermatrices). MrBayes was run with four chains for 1 million generations and trees were sampled every 100 generations. To construct the consensus tree, the first 1500 trees were discarded as "burn-in". Additional Bayesian phylogenetic analyses of the four protein supermatrices were performed using PhyloBayes 3.3b (Lartillot et al., 2009) with the CAT+ Γ_4 model (Lartillot and Philippe, 2004). Two chains were run for at least 10,000 cycles, saving one tree in ten. The first 300 trees were discarded as "burn-in" and the remaining trees from each chain were used to test for convergence and compute the 50% majority rule consensus tree. Posterior probabilities computed by MrBayes will be referred as PP_{MB} whereas those computed by PhyloBayes will be referred as PP_{PB}.

2.7. Comparison of stem G + C content of SSU rRNA sequences

The locations of stems in SSU rRNA sequences were retrieved from the RNA secondary STRucture and statistical ANalysis Database (RNA STRAND v2.0, http://www.rnasoft.ca/strand/) (Andronescu et al., 2008) for 20 Archaea with a known optimal growth temperature (OGT). Based on these data, we identified the homologous regions in the 105 SSU rRNA sequences used for phylogenetic analyses. When available, the OGTs of Archaea included in our SSU rRNA tree were retrieved from the German National Resource Centre for Biological Material (DSMZ, http://www.dsmz.de/) or from the literature. A Pearson correlation test between the stem G + C content and the OGT was performed. Comparisons between the OGT means of various archaeal clades were performed using the Student's *t*-test. The hypotheses of normality and homoscedasticity were verified through the Shapiro−Wilk test of normality and the *F*-test. All statistical tests were done using R (version 2.14.1) (Team, 2011).

2.8. Sequence accession numbers

All sequences have been deposited in the EMBL-bank under accession numbers HE574566−71 for the fosmids complete sequences; under JF739547−49 for the SSU rRNA genes of the fosmids 5C14, 5D21 and 6C13; under JF305824−972 for the 149 SSU rRNA sequences of the archaeal clone library and under JF317811−918 for the 108 SSU rRNA sequences of the prokaryotic clone library.

3. Results and discussion

3.1. Sampling site and material

Thick finger-long greyish filamentous samples were taken from the edges of a terrestrial hot spring in the Central thermal field of Uzon Caldera in Kamchatka. The temperature within these filaments was 85 °C, while the spring fluid had a temperature of 91 °C and a pH of 5.5. An SSU rRNA gene survey suggested that the bacterial community was dominated by uncultivated members of the Aquificales (Table S1), as often found in (sub)neutral terrestrial hot springs (Hugenholtz et al., 1998; Reigstad et al., 2010; Reysenbach et al., 2000; Skirnisdottir et al., 2000; Spear et al., 2005). Archaeal SSU rRNA sequences were affiliated with Euryarchaeota, Crenarchaeota, and MCG (Miscellaneous Crenarchaeotic Group), whereas sixteen sequences were most similar to Thaumarchaeota (Table S1).

3.2. Analyses of SSU rRNA sequences reveal two novel (hyper)thermophilic lineages related to Thaumarchaeota

Bayesian and ML phylogenetic analyses of the sixteen SSU rRNA sequences most similar to Thaumarchaeota (clones 1−15 and clone 35) revealed that they formed two well-supported monophyletic clusters (PP_{MB} = 1.00 for both groups, and bootstrap values (BV) = 87% and 76%, Fig. 1). These two clusters together formed a monophyletic group (PP_{MB} = 1.00 and BV = 63%) which branched-off before the divergence of HWCG II and the currently recognized thaumarchaeotal lineages (PP_{MB} = 1.00 and BV = 71%, and PP_{MB} = 0.99 and BV = 83%, Fig. 1). At this stage, we would like to stress that even these two lineages are closely related to Thaumarchaeota, they cannot be formally affiliated to this phylum (see below). However, because of their close relationship with Thaumarchaeota, the two clusters will be subsequently referred to as HTC1 (Hot Thaumarchaeota-related Clade 1, including clones 1, 2, 3, 5, 12 and 13) and HTC2 (Hot Thaumarchaeota-related Clade 2, including clones 4, 6, 7, 8, 9, 10, 11, 14, 15 and 35) (Fig. 1). Although these clades contained also sequences from uncultivated archaea belonging the previously proposed group THSC1 (Terrestrial Hot Spring Crenarchaeota 1) or THSCG (Terrestrial Hot Spring Crenarchaeotic group) (Takai and Sako, 1999; Takai et al., 2001), we think that the names HTC1 and HTC2 are more appropriated because they take into account the evolution of the archaeal taxonomy and underline the evolutionary link between these archaeal lineages and Thaumarchaeota.

The (hyper)thermophilic nature of HTC1 and HTC2 was also supported by the G + C content of their SSU rRNA sequences. The stem G + C content of SSU rRNA is strongly correlated to the optimal growth temperature (OGT) in

EURYARCHAEOTA
Aciduliprofundum boonei T469 CP001941
Methanococcus voltae A3 CP002057
Methanocaldococcus infernus ME CP002009
Methanothermobacter thermautotrophicus AB020530
Thermococcus gammatolerans AY206705
Ferroglobus placidus AJ299217
KORARCHAEOTA
korarchaeote SRI-306 AF255604
'Ca. Korarchaeum cryptofilum OPF8' CP000968
HWCG IV/UCII
unidentified archaeon AB019733
uncultured archaeon AB301983
uncultured archaeon AY166110
uncultured archaeon AY166102
uncultured archaeon AB302002
uncultured archaeon AB302033
MCG
uncultured archaeon AB302001
uncultured archaeon AB305332
uncultured archaeon AB302031
uncultured archaeon AB301988
uncultured archaeon AB301987
uncultured archaeon AB301992
uncultured archaeon AB301991
HWCG I ('AIGARCHAEOTA')
uncultured archaeon EU635916
uncultured archaeon EU635914
uncultured archaeon DQ490009
uncultured archaeon AB329825
uncultured archaeon AB302032
uncultured crenarchaeote EF100631
'Ca. Caldiarchaeum subterraneum' 315427393
CRENARCHAEOTA
Pyrobaculum islandicum DSM 4184 CP000504
Caldivirga maquilingensis IC-167 CP000852
Thermofilum pendens Hrk 5 CP000505
Sulfolobus islandicus AY247898
Sulfolobus metallicus EU419200
Acidianus brierleyi D26489
Caldisphaera draconis EF057392
Fervidicoccus fontis EF552404
Hyperthermus butylicus DSM 5456 CP000493
HWCG II
uncultured crenarchaeote AF361212
uncultured archaeon AB302019
uncultured archaeon AB302013
THAUMARCHAEOTA
uncultured archaeon AY885072
uncultured archaeon AB293209
Group 1A/psL12
uncultured crenarchaeote psL12 U63343
uncultured archaeon AB302036
uncultured archaeon APA2-17cm AF119135
uncultured marine group 1 crenarchaeote GQ267148
uncultured marine crenarchaeote AD1000-23-H12 EU686635
unidentified archaeon Y08984
uncultured crenarchaeote FFSC1 AJ006919
uncultured marine group 1 crenarchaeote GQ267146
uncultured archaeon AB301997
HWCG III
'Ca. Nitrosocaldus yellowstonii' EU239960
Unclutured 'Ca. Nitrosocaldus sp.' 45-H-12 AB201308
Group I.1b
'Ca. Nitrososphaera viennensis'
'Ca. Nitrososphaera gargensis' EU281336
Uncultured crenarchaeote 34 kb genomic fragment 29i4 AJ496176
Uncultured crenarchaeote genomic fragment 54d9 AY278106
Group I.1a
uncultured archaeon AB305330
uncultured archaeon AB301984
crenarchaeote symbiont of Axinella sp. AF421159
Cenarchaeum symbiosum AF083071
uncultured marine group 1 crenarchaeote DQ299274
Nitrosopumilus maritimus SCM1 CP000866
HTC1
uncultured archaeon DQ834113
34P11 HE574571
1C18 HE574566
clone 12 JF317811
unidentified archaeon AB007307
unidentified archaeon AB007308
clone 3 JF305826
uncultured archaeon EF156497
uncultured archaeon EF156580
uncultured archaeon DQ924800
uncultured archaeon EF558692
uncultured archaeon DQ834107
uncultured archaeon EF156496
uncultured archaeon FJ797321
clone 2 JF305825
clone 5 JF305828
1C23 HE574567
clone 13 JF317812
1N15 HE574569
clone 1 JF305824
HTC2
uncultured archaeon AB302008
unidentified archaeon AB019732
uncultured archaeon EU559690
uncultured archaeon EF156582
uncultured archaeon DQ833964
2C9 HE574570
uncultured archaeon EF558698
clone 15 JF317814
clone 14 JF317813
clone 35 JF317834
uncultured archaeon FJ797318
clone 6 JF305829
clone 10 JF305833
clone 11 JF305834
clone 4 JF305827
clone 9 JF305832
1M19 HE574568
6C13 JF739547
5D21 JF739548
5C14 JF739549
clone 8 JF305831
clone 7 JF305830
0.1

Fig. 1. Archaeal SSU rRNA phylogeny. Unrooted Bayesian phylogenetic tree of 105 SSU rRNA sequences representative of the archaeal diversity (1064 unambiguously aligned nucleic acid positions). SSU rRNA sequences from this study are in bold. They have been sequenced either from fosmids (underlined names) or from SSU rRNA clone libraries. The scale bar represents the average number of substitutions per site. Numbers at branch represent, respectively, posterior probabilities inferred by MrBayes and bootstrap values computed by Treefinder. For clarity, only values greater than 0.50 and 50%, respectively, are shown. HTC = Hot Thaumarchaeota-related Clade; HWCG = Hot Water Crenarchaeotic Group; UCII = Uncultured Crenarchaea group II; MCG = Miscellaneous Crenarchaeotic Group.

prokaryotes, and is higher in organisms living in hot environments because G:C pairs are more stable than A:T pairs due to their additional hydrogen bond (Galtier and Lobry, 1997). This property has been used to predict OGT in ancestral sequences (Boussau et al., 2008; Galtier et al., 1999), but it could also be used to infer the OGT of present-day prokaryotes. Using secondary structure information from the RNA STRAND Database (Andronescu et al., 2008), we determined the stem G + C content of the SSU rRNA sequences used to build the tree shown in Fig. 1 (Table S2). In agreement with previous studies, the correlation between the stem G + C content of SSU rRNA and the OGT was strong ($R = 0.93$; $R^2 = 0.8663$, see also Figure S1A). The average stem G + C contents of HTC1 (75.8%) and HTC2 (80.2%) were significantly higher than those of archaeal lineages that live in temperate or moderately hot environments (i.e. group I.1a, I.1b, psL12/group 1A, and MCG) (p-values <0.02, see also Figure S1B) and was compatible with life in hot environments. Interestingly, the stem G + C contents of HTC1 were significantly lower compared to HTC2 (p-value $<10^{-8}$), suggesting that the latter might be adapted to higher temperatures than the former, even if they were detected in the same pool (Figure S1B).

Altogether, these results are compatible with the hypothesis that the ancestor of HTC1, HTC2, HWCG II, and Thaumarchaeota lived in hot environments and that adaptation to moderate habitats occurred secondarily during the diversification of Thaumarchaeota (Barns et al., 1996; Brochier-Armanet et al., 2012; Hatzenpichler et al., 2008; Lopez-Garcia et al., 2004; Nunoura et al., 2011; Preston et al., 1996; Reigstad et al., 2008; Schleper et al., 1997).

3.3. First genomic insights into HTC

In order to expand our knowledge on HTC1 and HTC2, we investigated a large-insert metagenomic fosmid library built from the same DNA preparation. Out of the nine clones that belonged to HTC1 (34P11, 1C18, 1C23, and 1N15) and HTC2 (1M19, 2C9, 5C14, 5D21, and 6C13) (Fig. 1), six were selected for full-length sequencing (1M19, 2C9, 1C18, 34P11, 1C23, and 1N15). The general features of the six fosmids are given in Table 1A.

The two HTC2 fosmids (excluding rRNA coding genes) showed higher overall G + C content (61% and 68%, respectively) than the four HTC1 fosmids (46–54%) (Table 1A). The annotation of the six fosmids revealed compact gene organization with short intergenic regions (Fig. 2). In four of the six fosmids (1C18, 34P11, 1C23, and 1M19), the LSU rRNA was located downstream to the SSU rRNA whereas the 5S rRNA was not present. In the two remaining fosmids the SSU rRNA gene was located at the 3′ extremity of the genomic fragment, possibly explaining the absence of LSU rRNA coding genes. In this context it is interesting to mention that SSU and LSU rRNA (but not 5S rRNA) genes are clustered and present in only one copy in Thaumarchaeota genomes (Blainey et al., 2011; Hallam et al., 2006; Kim et al., 2011; Mosier et al., 2012a,b; Spang et al., 2012; Walker

Table 1
(A) General features of the six fully sequenced HTC fosmids. (B) Taxonomic distribution and functional assignment (according to COG categories) of the best BlastP hit of HTC ORFs. For each taxonomic group, the number (and the percentage) of gene families belonging to each functional category is indicated.

(A)

Fosmids	HTC1[a]				HTC2[a]	
	1N15	1C18	34P11	1C23	2C9	1M19
Accession number	HE574569	HE574566	HE574571	HE574567	HE574570	HE574568
Size (bp)	36,379	36,885	39,723	19,222	44,176	17,060
Total fosmid G + C content (%)[b]	46	54	53	50	68	61
SSU rRNA G + C content (%)	62.2	63.6	63.6	62.0	65.7	63.3
Number of ORFs	42	42	47	21	58	21
Average ORF length (bp)	789	745	687	691	721	450
ORFs with assigned function	30	23	25	11	38	9
Conserved hypothetical proteins	4	4	5	2	9	1
Hypothetical proteins	4	3	5	2	1	7
Proteins with no hits in any public database	1	7	8	2	3	2
RNA (tRNA and rRNA)	3	5	5	4	7	2

(B)

	Thaumarchaeota	C. subterraneum	Crenarchaeota	Euryarchaeota	Korarchaeota	Bacteria	Eucarya
Informational processing genes[c]	32 (74.4%)	6 (60%)	13 (26.5%)	11 (25.6%)	1 (50%)	3 (16.7%)	0 (0%)
Amino acids/nucleotides/coenzymes/ions[d]	3 (7%)	0 (0%)	8 (16.3%)	8 (18.6%)	1 (50%)	3 (16.7%)	1 (100%)
Energy metabolism[d]	1 (2.3%)	3 (30%)	10 (20.4%)	4 (9.3%)	0 (0%)	5 (27.8%)	0 (0%)
Cell compartment, traffic, communication[d]	0 (0%)	0 (0%)	7 (14.3%)	0 (0%)	0 (0%)	4 (22.2%)	0 (0%)
Unknowns	7 (16.3%)	1 (10%)	11 (22,4%)	20 (46.5%)	0 (0%)	3 (16.7%)	0 (0%)
Sum	43 (100%)	10 (100%)	49 (100%)	43 (100%)	2 (100%)	18 (100%)	1 (100%)

[a] Hot Thaumarchaeota-related Clade.
[b] SSU and LSU rRNA genes were not taken into account for G + C content estimations.
[c] COG categories J, K, L, and O.
[d] COG categories E, F, H, I, P, C, G, M, N, T, U, R, and S.

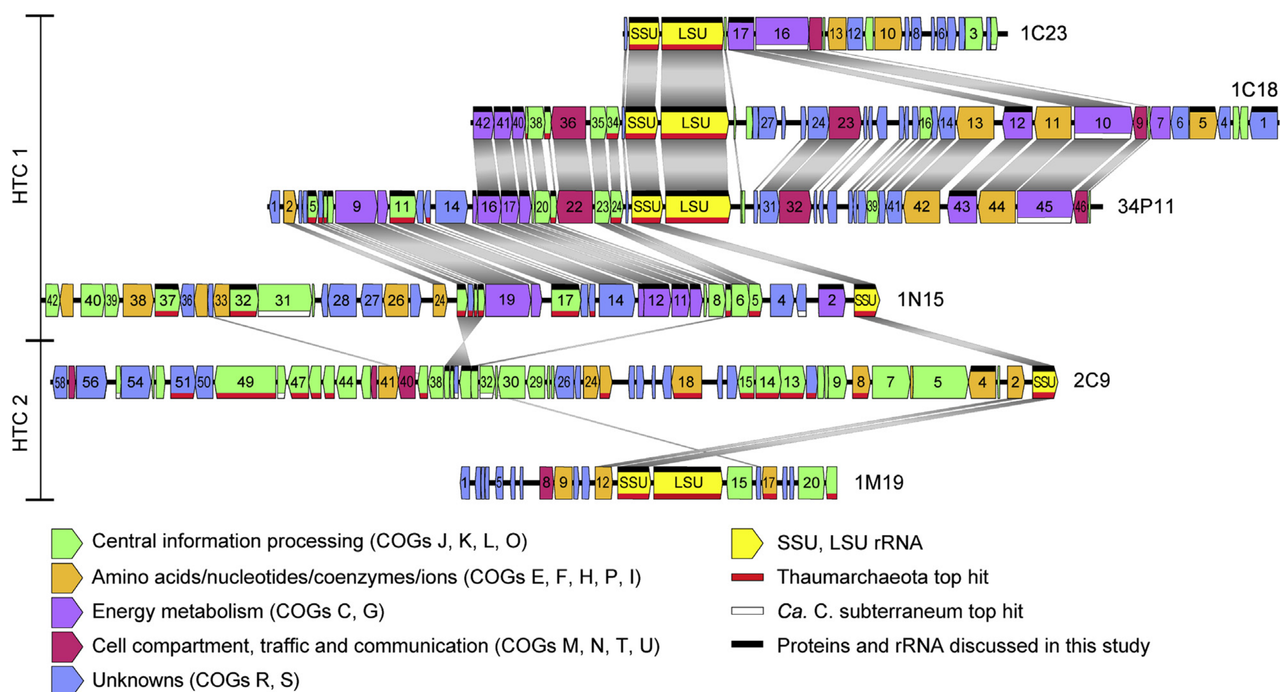

Fig. 2. Genomic organisation of the six HTC fosmids 1C23, 1C18, 34P11, 1N15, 2C9, and 1M19. The colour code of the open reading frames (ORF) refers to a higher-level functional classification of COGs (Makarova et al., 2007; Tatusov et al., 2001): Green: Central information processing (J: Translation, ribosomal structure and biogenesis. K: Transcription. L: DNA replication, recombination and repair. O: Posttranslational modification, protein turnover, chaperones). Orange: Amino acids/nucleotides/coenzymes/ions (E: Amino acid transport and metabolism. F: Nucleotide transport and metabolism, H: Coenzyme transport and metabolism, I: Lipid transport and metabolism, P: Inorganic ion transport and metabolism). Purple: Energy metabolism (C: Energy production and conversion. G: Carbohydrate transport and metabolism). Red: Cell compartment, traffic and communication (M: Cell wall/membrane/envelope biogenesis, outer membrane. N: Cell motility. T: Signal transduction mechanisms. U: Intracellular trafficking, secretion, and vesicular transport.) Blue: Unknowns (R: General function prediction only. S: Function unknown). ORFs labelled with a red rectangle at the bottom have a top BlastP hit to Thaumarchaeota. Hits returned by BlastP searches against *nr* were considered as potential homologues of the query ORF only if the associated E-value was lower than 10^{-5}. Homologous ORFs (i.e. showing more than 40% amino acid identity and an *E*-value $<10^{-5}$ when blasted against each other) belonging to different fosmids are connected with a grey shading.

et al., 2010), whereas they are located in different genomic regions in '*Ca.* Caldiarchaeum subterraneum', the only genomic sequence available for 'Aigarchaeota' (Nunoura et al., 2011). Three out of the four fosmids belonging to HTC1 (1C18, 34P11 and 1N15) were highly syntenic (Fig. 2), whereas fosmids 1C23 (HTC1), 2C9 and 1M19 (both HTC2) showed little gene order conservation (Fig. 2). Consistent with their close proximity in the SSU rRNA tree, fosmids 34P11 and 1C18 (HTC1) showed an overall nucleotide sequence identity of up to 98–99% over approximately 28 kb. In contrast, the syntenic ORFs of 1N15 showed only 42% amino acid identity on average to their homologues found in 34P11 and 1C18. This was congruent with the fact that 1N15 and 34P11/1C18 represented two distant sublineages within HTC1 (Fig. 2).

The six fully sequenced fosmids together encoded 249 ORFs, representing 173 gene families of which 109 (63%) could be assigned a putative function and covered 17 functional Cluster of Orthologous Groups (COG) categories (Fig. 2 and Table S3). The total number of gene families showing highest similarity with proteins from Thaumarchaeota, Crenarchaeota or Euryarchaeota was in the same range (42, 49, and 43, respectively) (Table 1B), whereas only ten and two ORFs were most similar to proteins from '*Ca*. C. subterraneum' and 'Ca. Korarchaeum cryptophilum', respectively (Table 1B and Table S3). While ORFs with highest similarity to euryarchaeotal and crenarchaeotal proteins were evenly distributed over the diverse COG categories, the majority of proteins more similar to proteins from Thaumarchaeota belonged to COG categories involved in informational processing (69.2%, Table 1B and Table S3). The overrepresentation of informational processing genes with highest similarities to Thaumarchaeota is in agreement with the SSU rRNA phylogeny which suggests that HTC1 and HTC2 share a closer evolutionary link with this phylum compared to other archaeal phyla.

Interestingly, despite the apparent evolutionary relationship between HTC1, HTC2 and, Thaumarchaeota, the survey of genes involved in central energy metabolism highlighted some differences (Table 2). For instance, a four ORF-comprising gene cluster is found on the two HTC1 fosmids, 34P11 (ORF15-18) and 1N15 (ORF10-13). It encodes the α, β, δ and γ subunits of a 2-oxoacid:ferredoxin oxidoreductase (OFOR) (Table S3). A gene cluster encoding the α, β, and γ subunits is also found on 1C18. But it is located at the 3′ extremity of the fosmid, possibly explaining the absence the δ subunit. OFORs

Table 2
Presence and absence of central carbon metabolism genes encoding fumarate hydratase, aspartate ammonia-lyase, and the different subunits of 2-oxoacid:ferredoxin oxidoreductases (OFOR) in fosmids 1C23, 1C18, 34P11, and 1N15 from Hot Thaumarchaeota-related Clades compared to ammonia-oxidizing Thaumarchaeota and other archaeal groups.

arCOG	COG	Annotation	1C23	1C18	34P11	1N15	Thaum	C. sub	Kor	Cren	Arch	Halo	Mbac	Mcoc	Mmic	Mcel	Msar	Mpyr	Thec	Thep	Nano
01749	0114	Fumarate hydratase	+	+	+		−	+	−	+	−	+	−	−	−	−	+	−	−	−	−
01750	1027	Aspartate ammonia-lyase					+	−	−	−	−	−	−	−	−	−	−	−	−	+	−
01599	1013	OFOR, beta subunit		+	+	+	+	+	+	+	+	+[a]	+	+	+	−	+	+	+	+	−
01600	1014	OFOR, beta and gamma subunits					−	−	−	−	−	−	+	+	−	−	+	−	−	−	−
01601	1013	OFOR, beta subunit					−	+	+	+	+	−	+	+	+	+	+	+	+	+	−
01602	1014	OFOR, gamma subunit		+	+	+	−	−	+	+	+	−	+	+	+	+	+	+	+	+	−
01603	1014	OFOR, gamma subunit					−	+	+	+	+	−	+	+	+	+	+	+	+	+	−
01604	1144	OFOR, gamma and delta subunits					−	+	−	+	−	−	+	−	−	−	−	−	−	−	−
01605	01144	OFOR, delta subunit			+	+	−	+	+	+	+	−	+	+	+	+	+	+	+	+	−
01606	0674	OFOR, alpha and gamma subunits					+	+	−	+	+	+[a]	+	+	+	−	+	−	+	+	−
01607	0674	OFOR, alpha subunit		+	+	+	−	−	+	−	+	−	+	+	+	−	+	+	+	−	−
01608	0674	OFOR, alpha subunit					−	+	+	+	+	−	+	+	+	+	+	+	+	+	−

Abbreviations: Thaum: Thaumarchaeota; C. sub: '*Ca.* C. subterraneum'; Kor: korarchaeota; Cren: Crenarchaeota; Arch: Archaeglobales; Halo: Halobacteriales; Mbac: Methanobacteriales; Mcoc: Methanococcales; Mmic: Methanomicrobiales; Mcel: Methanocellales; Msar: Methanosarcinales; Mpyr: Methanopyrales; Thec: Thermococcales; Thep: Thermoplasmatales; Nano: Nanoarchaeota.
[a] Two copies per genome.

are iron–sulfur cluster proteins that catalyse the decarboxylation of different 2-oxoacids including pyruvate and 2-oxoglutarate, two intermediates of the central carbon metabolism (Kerscher and Oesterhelt, 1982). These enzymes have been classified into different sub-families according to their substrate specificity (Fukuda and Wakagi, 2002; Mai and Adams, 1996; Tersteegen et al., 1997; Uyeda and Rabinowitz, 1971). Irrespective of their substrates, OFORs can be further distinguished by the number and the type of their subunits (e.g. α_2, $\alpha\beta$, $\alpha_2\beta_2$, $\alpha\beta\delta\gamma$) that are encoded by gene clusters (Kletzin and Adams, 1996; Zhang et al., 1996). Based on sequence similarity, the OFOR subunits have been assigned to different arCOG-families (Table 2) (Makarova et al., 2007). Euryarchaeota, Crenarchaeota, Korarchaeota, and '*Ca.* C. subterraneum' usually encode several types of OFORs, whereas Thaumarchaeota contain a single gene cluster composed of two genes belonging to arCOGs 01599 and 01606 (Table 2). In contrast, the four OFOR subunits encoded by the HTC1 fosmids were assigned to arCOGs 01599, 01602, 01605, and 01607. Furthermore, the sequences of the corresponding α and β subunits are most similar to those found in anaerobic bacteria and archaea (not shown). This suggests that the OFORs harboured by HTC1 are significantly different from those found in Thaumarchaeota from groups I.1a and I.1b.

Secondly, according to Nunoura et al. (2011), Crenarchaeota, some Euryarchaeota (Halobacteriales and Methanosarcinales) as well as '*Ca.* C. subterraneum' encode a fumarate hydratase (FH) (COG0114/arCOG01749, Table 2), which has been predicted to catalyse the hydration of fumarate to malate in the citric acid cycle of various bacteria and archaea. In contrast, all available thaumarchaeotal genomes and some Euryarchaeota (Thermoplasmatales, ARMAN, and Methanobacterium strain AL-21) encode a homologous enzyme that belongs to a separate (ar)COG family (COG1027/arCOG01750, Table 2) often termed aspartate ammonia-lyase (or aspartase) due to its high sequence similarity to the respective lyase gene of *Escherichia coli* of the same COG. The phylogenetic analysis of FH and aspartate ammonia-lyase revealed two distinct horizontal gene transfer (HGT) events in Archaea (Figure S2). More precisely, HTC1 members seem to have acquired their enzyme from Crenarchaeota whereas an HGT occurred between a euryarchaeotal lineage related to Thermoplasmatales or Methanobacterium strain AL-21 and the ancestor of Thaumarchaeota (represented here by groups I.1a and HWCG III). Indeed, according to the phylogeny of Archaea, the grouping of Thaumarchaeota with Thermoplasmatales and *Methanobacterium* sp. AL-21 is quite unexpected and in contradiction with the phylogeny of species. If this grouping reflects a vertical inheritance, the gene would have been present in the ancestor of Archaea and secondarily lost in all lineages, except the three mentioned above. Moreover, this does not explain why *Methanobacterium* sp. AL-21 appears more closely related to Thaumarchaeota than to Thermoplasmatales (given that Methanobacterium and Thermoplasmatales are both Euryarchaeota). In contrast, a single HGT between a euryarchaeotal lineage (related to Methanobacterium or Thermoplasmatales) and the

ancestor of Thaumarchaeota (e.g. the ancestor of group I.1a and HWCG III) can easily explain this branching pattern.

Although these findings represent only a glimpse into the putative energy metabolism of HTCs they suggest that differences with the (so far described) Thaumarchaeota might exist.

3.4. HTC1 and HTC2 represent two sublineages of a new major and ancient archaeal group

To further explore the position of the HTC1 and HTC2 lineages in the Archaea domain, a supermatrix gathering the SSU and LSU rRNA sequences from fosmids 1C18, 34P11, 1C23, and 1M19 and complete archaeal genomes was analysed (Fig. 3). The ML and Bayesian resulting trees were in agreement with the SSU rRNA tree (Fig. 1). More precisely, the monophyly of HTC1 was recovered (PP_{MB} = 1.00 and BV = 100%), with fosmids 34P11 and 1C18 being more closely related to each other than to fosmid 1C23 (PP_{MB} = 1.00 and BV = 99%) and the sister-ship between HTC1 and HTC2 being again well supported (PP_{MB} = 0.99 and BV = 74%). Finally, the close relationship between HTC and Thaumarchaeota was recovered with maximal supports. The main difference between the two trees relied on the robust clustering of Thaumarchaeota and HTC with '*Ca.* C. subterraneum' (PP_{MB} = 1.00 and BV = 99%, Fig. 3), whereas the phylogenetic position of the latter was unresolved in the SSU rRNA tree (Fig. 1). The grouping of Thaumarchaeota and '*Ca.* C. subterraneum' was in agreement with recent phylogenetic analyses (Brochier-Armanet et al., 2011; Nunoura et al., 2011).

The relationship between HTC, Thaumarchaeota and '*Ca.* C. subterraneum' was confirmed by the analysis of informational proteins encoded in the HTC fosmids (L31e, L39e, S19e, RPR2, and EF-1α, F1a supermatrix). More precisely, the inferred Bayesian and ML trees supported the monophyly of HTC1 (PP_{MB} = 1.00, BV = 100%, and PP_{PB} = 1.00), the

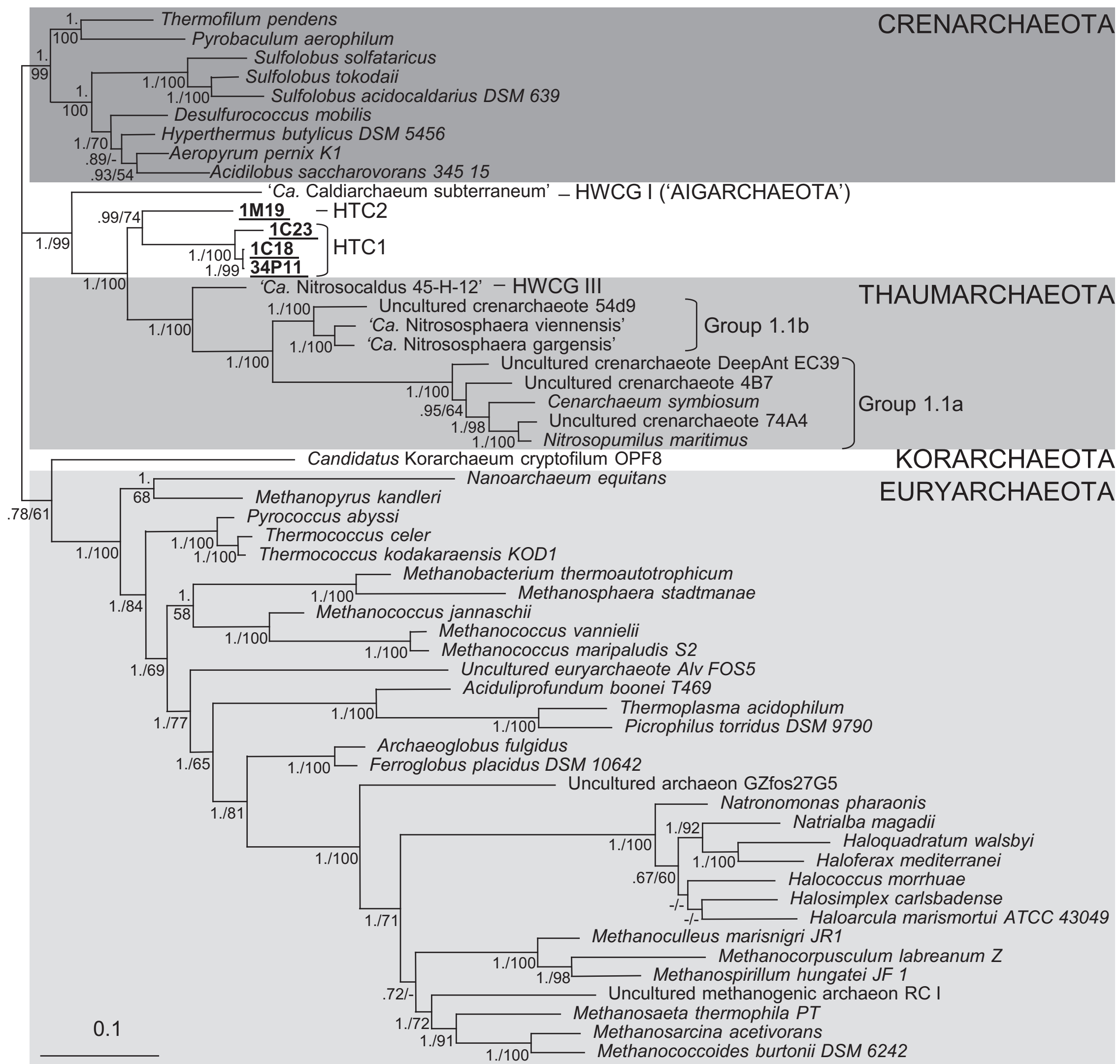

Fig. 3. Unrooted Bayesian phylogenetic tree based on a concatenation of 55 SSU and LSU rRNA sequences (3462 unambiguously aligned nucleic acid positions). Names of fosmid sequences from this study are underlined and in bold. The scale bar represents the average number of substitutions per site. Numbers at branch represent, respectively, posterior probabilities inferred by MrBayes and bootstrap values computed by Treefinder. For clarity, only values greater than 0.50 and 50%, respectively, are shown. HTC = Hot Thaumarchaeota-related Clade; HWCG = Hot Water Crenarchaeotic Group.

Fig. 4. (A) Unrooted Bayesian tree based on the F1a supermatrix (five proteins, 55 archaeal sequences, 653 unambiguously aligned amino acid positions); (B) Bayesian tree based on the F1ae supermatrix that includes eleven eukaryotic sequences as outgroup (66 sequences, 618 unambiguously aligned amino acid positions). Names of fosmid sequences from this study are underlined and in bold. The scale bars represent the average number of substitutions per site. Numbers at branch represent, respectively, posterior probabilities inferred by MrBayes, bootstrap values computed by Treefinder, and posterior probabilities estimated by PhyloBayes. For clarity, only values greater than 0.50, 50%, and 0.50, respectively, are shown. HTC = Hot Thaumarchaeota-related Clade; HWCG = Hot Water Crenarchaeotic Group.

sister-ship of HTC1 and HTC2 (PP_{MB} = 1.00, BV = 50% and PP_{PB} = 0.97), and the grouping of '*Ca.* C. subterraneum' with HTC and Thaumarchaeota (PP_{MB} = 1.00, BV = 50% and PP_{PB} = 1.00), albeit the relationships between these lineages were not resolved (Fig. 4A). Finally, the addition of two central information processing proteins present only on fosmid 1N15 (supermatrix F2a) led to a more resolved tree, supporting, among others, the clustering of HTC1 with Thaumarchaeota with maximal supports and the sister-ship of these two lineages with '*Ca.* C. subterraneum'/'Aigarchaeota' (PP_{MB} = 1.00, BV = 89% and PP_{PB} = 1.00) (Figure S3A).

The grouping of HTC1 and HTC2 with Thaumarchaeota raises the question of the taxonomic status of these two new lineages. Indisputably, HTC is genetically diverse and only distantly related to currently recognized thaumarchaeotal lineages, as exemplified by the divergence observed in SSU rRNA sequences (Fig. 1), thus representing a new major archaeal clade of high-taxonomic rank. This new major clade likely corresponds to an early-branching thaumarchaeotal lineage (e.g. a class or at least an order), even if we cannot fully exclude that they represent a lineage distinct from Thaumarchaeota, thus a new phylum. Given the scarcity of genomic, ecological and physiological data, the absence of any cultivated HTC representatives, and the opened question regarding the status of 'Aigarchaeota', a definitive answer cannot yet be reached regarding the precise taxonomic status of HTC and of its two sublineages, HTC1 and HTC2.

3.5. *Informational proteins from HTC fosmids provide insights regarding the early diversification of Archaea*

Early phylogenetic analyses based on SSU rRNA sequences suggested a sister-relationship between Thaumarchaeota and Crenarchaeota (DeLong, 1992). In contrast, phylogenetic analyses of protein data from the first thaumarchaeotal genome sequence (i.e. *Cenarchaeum symbiosum*) supported Thaumarchaeota as the first diverging lineage within Archaea, i.e. before the split of Euryarchaeota and Crenarchaeota (Brochier-Armanet et al., 2008). However, following the additional sequence data from other thaumarchaeotal representatives, the early emergence of this phylum was either less strongly supported (Spang et al., 2010; Walker et al., 2010), or challenged by their grouping with Korarchaeota and Crenarchaeota (Groussin and Gouy, 2011; Guy and Ettema, 2011; Nunoura et al., 2011). Resolving the relationships among archaeal phyla has thus become an important question in evolutionary biology research. Here we had the opportunity to investigate the phylogenetic position of Thaumarchaeota by including data from a deep-branching-related lineage. Therefore, we added eukaryotic homologues as outgroup in the F1a and F2a alignments, leading to the F1ae (five proteins, 618 positions) and F2ae (seven proteins, 1139 positions) supermatrices. The F1ae Bayesian and ML trees (including fosmids 2C9, 1N15, and 34P11) were less resolved than those based on F2ae (including only fosmid 1N15) (Fig. 4B and Figure S3B), as expected given the smallest number of analysed sites. However, both F1ae and F2ae Bayesian and ML trees supported the branching of Thaumarchaeota, HTCs, and '*Ca.* C. subterraneum' before the divergence of Crenarchaeota, Euryarchaeota and Korarchaeota (PP_{MB} = 1.0, BV < 50%, and PP_{PB} = 0.96, Fig. 4B; PP_{MB} = 1.0, BV = 84% and PP_{PB} = 1.0, Figure S3B). These results are in agreement with early analyses suggesting that Thaumarchaeota (and relatives) could indeed represent the first diverging archaeal phylum. But, according to the small number of HTC1 and HTC2 proteins that could be used in this analysis, we cannot fully exclude the risk of biases or tree reconstruction artefacts. Therefore, these results have to be confirmed when additional protein sequences from Thaumarchaeota and related lineages become available.

The basal branching of Thaumarchaeota, HTCs and '*Ca.* C. subterraneum' could also be interpreted differently. Indeed, a number of hypotheses postulate that the first eukaryote derived from the association (symbiosis, engulfment, or other) between a *bona fide* archaeon (i.e. a modern archaeon) and a bacterium. Recent phylogenomics analyses aiming at revisiting these hypotheses were reported (Cox et al., 2008; Foster et al., 2009; Guy and Ettema, 2011; Kelly et al., 2011; Pisani et al., 2007; Rivera and Lake, 2004; Thiergart et al., 2012; Williams et al., 2012) but failed to reach a consensus despite of the use of very similar datasets, indicating that additional data and approaches are required to tackle this issue (Gribaldo et al., 2010). Under the assumption that Eucarya derive indeed from Archaea, our results favour the hypothesis that Thaumarchaeota, HTC, and 'Aigarchaeota' could be the closest present-day archaeal relatives of Eucarya.

4. Conclusions

Our metagenomic analyses of a terrestrial hot spring in Kamchatka provided the first genomic data from two novel major (hyper)thermophilic archaeal lineages, HTC1 and HTC2, related to Thaumarchaeota. Our analyses suggest that HTC1 and HTC2 represent either an ancient lineage of high-taxonomic rank within Thaumarchaeota or less likely a distinct phylum. Additional data will be needed to precisely characterize the energy metabolism and metabolic capacities of the respective organisms. Further studies of these two groups will help to shed light on the major question of the origin of ammonia oxidation in Archaea, and through their key position in the archaeal tree, will help to uncover important aspects of ancient evolution.

Acknowledgements

We would like to thank the organisers of the Microbial Observatory of Kamchatka, especially Elisaveta A. Bonch-Osmolovskaya, for facilitating the sampling of terrestrial hot springs in Uzon Caldera and Mt. Mutnovsky in the Kamchatka Peninsula during a workshop in 2005. We would also like to acknowledge Rym Agrebi, Celine Petitjean, Mathieu Groussin and Marc Bailly-Bechet for stimulating discussions and helpful comments. We thank Michelle Leger for careful reading of the manuscript. The work was supported by the

Norwegian Research Council, Grant 172206 to CS and LJR, by Grant P23000 of the Austrian Science Fund (FWF) given to CS. CB-A is supported by the ANR grant (ANR-08-GENM-024-002), by the Investissement d'Avenir grant (ANR-10-BINF-01-01) and is member of the Institut Universitaire de France. AS is supported by a DOC-Forte fellowship of the Austrian Academy of Sciences. LE was the recipient of a grant from the French Ministere de l'Enseignement Superieur et de la Recherche and is currently supported by a Dalhousie University Centre for Comparative Genomics and Evolutionary Bioinformatics postdoctoral fellowship from the Tula Foundation.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.resmic.2013.02.006.

References

Amann, R.I., Ludwig, W., Schleifer, K.H., 1995. Phylogenetic identification and in situ detection of individual microbial cells without cultivation. Microbiol. Rev. 59, 143–169.

Andronescu, M., Bereg, V., Hoos, H.H., Condon, A., 2008. RNA strand: the RNA secondary structure and statistical analysis database. BMC Bioinform. 9, 340.

Barns, S.M., Delwiche, C.F., Palmer, J.D., Pace, N.R., 1996. Perspectives on archaeal diversity, thermophily and monophyly from environmental rRNA sequences. Proc. Natl. Acad. Sci. U S A 93, 9188–9193.

Blainey, P.C., Mosier, A.C., Potanina, A., Francis, C.A., Quake, S.R., 2011. Genome of a low-salinity ammonia-oxidizing archaeon determined by single-cell and metagenomic analysis. PLOS One 6, e16626.

Boussau, B., Blanquart, S., Necsulea, A., Lartillot, N., Gouy, M., 2008. Parallel adaptations to high temperatures in the Archaeaneon. Nature 456, 942–945.

Brochier, C., Gribaldo, S., Zivanovic, Y., Confalonieri, F., Forterre, P., 2005. Nanoarchaea: representatives of a novel archaeal phylum or a fast-evolving euryarchaeal lineage related to Thermococcales? Genome Biol. 6, R42.

Brochier-Armanet, C., Boussau, B., Gribaldo, S., Forterre, P., 2008. Mesophilic Crenarchaeota: proposal for a third archaeal phylum, the Thaumarchaeota. Nat. Rev. Microbiol. 6, 245–252.

Brochier-Armanet, C., Forterre, P., Gribaldo, S., 2011. Phylogeny and evolution of the Archaea: one hundred genomes later. Curr. Opin. Microbiol. 14, 274–281.

Brochier-Armanet, C., Gribaldo, S., Forterre, P., 2012. Spotlight on the Thaumarchaeota. ISME J. 6, 227–230.

Cole, J.R., Chai, B., Farris, R.J., Wang, Q., Kulam-Syed-Mohideen, A.S., McGarrell, D.M., Bandela, A.M., Cardenas, E., Garrity, G.M., Tiedje, J.M., 2007. The ribosomal database project (RDP-II): introducing myRDP space and quality controlled public data. Nucleic Acids Res. 35, D169–D172.

Cole, J.R., Wang, Q., Cardenas, E., Fish, J., Chai, B., Farris, R.J., Kulam-Syed-Mohideen, A.S., McGarrell, D.M., Marsh, T., Garrity, G.M., Tiedje, J.M., 2009. The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. Nucleic Acids Res. 37, D141–D145.

Cox, C.J., Foster, P.G., Hirt, R.P., Harris, S.R., Embley, T.M., 2008. The archaebacterial origin of eukaryotes. Proc. Natl. Acad. Sci. U S A 105, 20356–20361.

DeLong, E.F., 1992. Archaea in coastal marine environments. Proc. Natl. Acad. Sci. U S A 89, 5685–5689.

Dodsworth, J.A., Hungate, B.A., Hedlund, B.P., 2011. Ammonia oxidation, denitrification and dissimilatory nitrate reduction to ammonium in two US Great Basin hot springs with abundant ammonia-oxidizing Archaea. Environ. Microbiol. 13, 2371–2386.

Elkins, J.G., Podar, M., Graham, D.E., Makarova, K.S., Wolf, Y., Randau, L., Hedlund, B.P., Brochier-Armanet, C., et al., 2008. A korarchaeal genome reveals insights into the evolution of the Archaea. Proc. Natl. Acad. Sci. U S A 105, 8102–8107.

Felsenstein, J., 2004. PHYLIP (Phylogeny Inference Package) Version 3.6.

Forterre, P., Bouthier De La Tour, C., Philippe, H., Duguet, M., 2000. Reverse gyrase from hyperthermophiles: probable transfer of a thermoadaptation trait from archaea to bacteria. Trends Genet. 16, 152–154.

Forterre, P., Brochier, C., Philippe, H., 2002. Evolution of the Archaea. Theor. Popul. Biol. 61, 409–422.

Foster, P.G., Cox, C.J., Embley, T.M., 2009. The primary divisions of life: a phylogenomic approach employing composition-heterogeneous methods. Philos. Trans. R. Soc. Lond. B, Biol. Sci. 364, 2197–2207.

French, E., Kozlowski, J.A., Mukherjee, M., Bullerjahn, G., Bollmann, A., 2012. Ecophysiological characterization of ammonia-oxidizing archaea and bacteria from freshwater. Appl. Environ. Microbiol. 78, 5773–5780.

Fuhrman, J.A., McCallum, K., Davis, A.A., 1992. Novel major archaebacterial group from marine plankton. Nature 356, 148–149.

Fukuda, E., Wakagi, T., 2002. Substrate recognition by 2-oxoacid:ferredoxin oxidoreductase from Sulfolobus sp. strain 7. Biochim. Biophys. Acta 1597, 74–80.

Galtier, N., Lobry, J.R., 1997. Relationships between genomic G+C content, RNA secondary structures, and optimal growth temperature in prokaryotes. J. Mol. Evol. 44, 632–636.

Galtier, N., Tourasse, N., Gouy, M., 1999. A nonhyperthermophilic common ancestor to extant life forms. Science 283, 220–221.

Gribaldo, S., Brochier-Armanet, C., 2006. The origin and evolution of Archaea: a state of the art. Philos. Trans. R. Soc. Lond. B, Biol. Sci. 361, 1007–1022.

Gribaldo, S., Brochier-Armanet, C., 2012. Time for order in microbial systematics. Trends Microbiol. 20, 209–210.

Gribaldo, S., Poole, A.M., Daubin, V., Forterre, P., Brochier-Armanet, C., 2010. The origin of eukaryotes and their relationship with the Archaea: are we at a phylogenomic impasse? Nat. Rev. Microbiol. 8, 743–752.

Groussin, M., Gouy, M., 2011. Adaptation to environmental temperature is a major determinant of molecular evolutionary rates in archaea. Mol. Biol. Evol. 28, 2661–2674.

Gubry-Rangin, C., Hai, B., Quince, C., Engel, M., Thomson, B.C., James, P., Schloter, M., Griffiths, R.I., et al., 2011. Niche specialization of terrestrial archaeal ammonia oxidizers. Proc. Natl. Acad. Sci. U S A 108, 21206–21211.

Guy, L., Ettema, T.J., 2011. The archaeal 'TACK' superphylum and the origin of eukaryotes. Trends Microbiol. 19, 580–587.

Hallam, S.J., Konstantinidis, K.T., Putnam, N., Schleper, C., Watanabe, Y., Sugahara, J., Preston, C., de la Torre, J., et al., 2006. Genomic analysis of the uncultivated marine crenarchaeote Cenarchaeum symbiosum. Proc. Natl. Acad. Sci. U S A 103, 18296–18301.

Hatzenpichler, R., Lebedeva, E.V., Spieck, E., Stoecker, K., Richter, A., Daims, H., Wagner, M., 2008. A moderately thermophilic ammonia-oxidizing crenarchaeote from a hot spring. Proc. Natl. Acad. Sci. U S A 105, 2134–2139.

Huber, H., Hohn, M.J., Rachel, R., Fuchs, T., Wimmer, V.C., Stetter, K.O., 2002. A new phylum of Archaea represented by a nanosized hyperthermophilic symbiont. Nature 417, 63–67.

Huber, T., Faulkner, G., Hugenholtz, P., 2004. Bellerophon: a program to detect chimeric sequences in multiple sequence alignments. Bioinformatics 20, 2317–2319.

Hugenholtz, P., Pitulle, C., Hershberger, K.L., Pace, N.R., 1998. Novel division level bacterial diversity in a Yellowstone hot spring. J. Bacteriol. 180, 366–376.

Jobb, G., von Haeseler, A., Strimmer, K., 2004. TREEFINDER: a powerful graphical analysis environment for molecular phylogenetics. BMC Evol. Biol. 4, 18.

Jung, M.Y., Park, S.J., Min, D., Kim, J.S., Rijpstra, W.I., Sinninghe Damste, J.S., Kim, G.J., Madsen, E.L., Rhee, S.K., 2011. Enrichment and characterization of an autotrophic ammonia-oxidizing archaeon of mesophilic crenarchaeal group I.1a from an agricultural soil. Appl. Environ. Microbiol. 77, 8635–8647.

Karner, M.B., DeLong, E.F., Karl, D.M., 2001. Archaeal dominance in the mesopelagic zone of the Pacific Ocean. Nature 409, 507–510.

Katoh, K., Misawa, K., Kuma, K., Miyata, T., 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 30, 3059–3066.

Kelly, S., Wickstead, B., Gull, K., 2011. Archaeal phylogenomics provides evidence in support of a methanogenic origin of the Archaea and a thaumarchaeal origin for the eukaryotes. Proc. Biol. Sci. 278, 1009–1018.

Kerscher, L., Oesterhelt, D., 1982. Pyruvate:ferredoxin oxidoreductase — new findings on an ancient enzyme. Trends Biochem. Sci. 7, 371–374.

Kim, B.K., Jung, M.Y., Yu, D.S., Park, S.J., Oh, T.K., Rhee, S.K., Kim, J.F., 2011. Genome sequence of an ammonia-oxidizing soil archaeon, "Candidatus Nitrosoarchaeum koreensis" MY1. J. Bacteriol. 193, 5539–5540.

Kim, J.G., Jung, M.Y., Park, S.J., Rijpstra, W.I., Sinninghe Damste, J.S., Madsen, E.L., Min, D., Kim, J.S., et al., 2012. Cultivation of a highly enriched ammonia-oxidizing archaeon of Thaumarchaeotal group I.1b from an agricultural soil. Environ. Microbiol. 14, 1528–1543.

Kletzin, A., Adams, M.W., 1996. Molecular and phylogenetic characterization of pyruvate and 2-ketoisovalerate ferredoxin oxidoreductases from Pyrococcus furiosus and pyruvate ferredoxin oxidoreductase from Thermotoga maritima. J. Bacteriol. 178, 248–257.

Konneke, M., Bernhard, A.E., de la Torre, J.R., Walker, C.B., Waterbury, J.B., Stahl, D.A., 2005. Isolation of an autotrophic ammonia-oxidizing marine archaeon. Nature 437, 543–546.

Kyle, J., Schroeder, P., Wiegel, J., 2007. Microbial silicification in sinters from two terrestrial hot springs in the Uzon Caldera, Kamchatka, Russia. Geomicrobiol. J. 24, 627–641.

Lartillot, N., Philippe, H., 2004. A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. Mol. Biol. Evol. 21, 1095–1109.

Lartillot, N., Lepage, T., Blanquart, S., 2009. PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. Bioinformatics 25, 2286–2288.

Le, S.Q., Gascuel, O., 2008. An improved general amino acid replacement matrix. Mol. Biol. Evol. 25, 1307–1320.

Lehtovirta-Morley, L.E., Stoecker, K., Vilcinskas, A., Prosser, J.I., Nicol, G.W., 2011. Cultivation of an obligate acidophilic ammonia oxidizer from a nitrifying acid soil. Proc. Natl. Acad. Sci. U S A 108, 15892–15897.

Leininger, S., Urich, T., Schloter, M., Schwark, L., Qi, J., Nicol, G.W., Prosser, J.I., Schuster, S.C., Schleper, C., 2006. Archaea predominate among ammonia-oxidizing prokaryotes in soils. Nature 442, 806–809.

Lopez-Garcia, P., Brochier, C., Moreira, D., Rodriguez-Valera, F., 2004. Comparative analysis of a genome fragment of an uncultivated mesopelagic crenarchaeote reveals multiple horizontal gene transfers. Environ. Microbiol. 6, 19–34.

Ludwig, W., Strunk, O., Westram, R., Richter, L., Meier, H., Yadhukumar, Buchner, A., Lai, T., et al., 2004. ARB: a software environment for sequence data. Nucleic Acids Res. 32, 1363–1371.

Mai, X., Adams, M.W., 1996. Characterization of a fourth type of 2-keto acid-oxidizing enzyme from a hyperthermophilic archaeon: 2-ketoglutarate ferredoxin oxidoreductase from Thermococcus litoralis. J. Bacteriol. 178, 5890–5896.

Makarova, K.S., Koonin, E.V., 2005. Evolutionary and functional genomics of the Archaea. Curr. Opin. Microbiol. 8, 586–594.

Makarova, K.S., Sorokin, A.V., Novichkov, P.S., Wolf, Y.I., Koonin, E.V., 2007. Clusters of orthologous genes for 41 archaeal genomes and implications for evolutionary genomics of archaea. Biol. Direct 2, 33.

Mosier, A.C., Allen, E.E., Kim, M., Ferriera, S., Francis, C.A., 2012a. Genome sequence of "Candidatus Nitrosoarchaeum limnia" BG20, a low-salinity ammonia-oxidizing archaeon from the San Francisco Bay estuary. J. Bacteriol. 194, 2119–2120.

Mosier, A.C., Allen, E.E., Kim, M., Ferriera, S., Francis, C.A., 2012b. Genome sequence of "Candidatus Nitrosopumilus salaria" BD31, an ammonia-oxidizing archaeon from the San Francisco Bay estuary. J. Bacteriol. 194, 2121–2122.

Mosier, A.C., Lund, M.B., Francis, C.A., 2012c. Ecophysiology of an ammonia-oxidizing archaeon adapted to low-salinity habitats. Microb. Ecol. 64, 955–963.

Moyer, C., Tiedje, J., Dobbs, F., Karl, D., 1998. Diversity of deep-sea hydrothermal vent Archaea from Loihi Seamount. Hawaii Deep Sea Res. Part II 45, 303–317.

Nunoura, T., Takaki, Y., Kakuta, J., Nishi, S., Sugahara, J., Kazama, H., Chee, G.J., Hattori, M., et al., 2011. Insights into the evolution of Archaea and eukaryotic protein modifier systems revealed by the genome of a novel archaeal group. Nucleic Acids Res. 39, 3204–3223.

Pester, M., Schleper, C., Wagner, M., 2011. The Thaumarchaeota: an emerging view of their phylogeny and ecophysiology. Curr. Opin. Microbiol. 14, 300–306.

Philippe, H., 1993. MUST, a computer package of management utilities for sequences and trees. Nucleic Acids Res. 21, 5264–5272.

Pisani, D., Cotton, J.A., McInerney, J.O., 2007. Supertrees disentangle the chimerical origin of eukaryotic genomes. Mol. Biol. Evol. 24, 1752–1760.

Preston, C.M., Wu, K.Y., Molinski, T.F., DeLong, E.F., 1996. A psychrophilic crenarchaeon inhabits a marine sponge: Cenarchaeum symbiosum gen. nov., sp. nov. Proc. Natl. Acad. Sci. U S A 93, 6241–6246.

Pruesse, E., Quast, C., Knittel, K., Fuchs, B.M., Ludwig, W., Peplies, J., Glockner, F.O., 2007. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. Nucleic Acids Res. 35, 7188–7196.

Quaiser, A., Ochsenreiter, T., Klenk, H.P., Kletzin, A., Treusch, A.H., Meurer, G., Eck, J., Sensen, C.W., Schleper, C., 2002. First insight into the genome of an uncultivated crenarchaeote from soil. Environ. Microbiol. 4, 603–611.

Rattei, T., Tischler, P., Arnold, R., Hamberger, F., Krebs, J., Krumsiek, J., Wachinger, B., Stumpflen, V., Mewes, W., 2008. SIMAP – structuring the network of protein similarities. Nucleic Acids Res. 36, D289–D292.

Reigstad, L.J., Richter, A., Daims, H., Urich, T., Schwark, L., Schleper, C., 2008. Nitrification in terrestrial hot springs of Iceland and Kamchatka. FEMS Microbiol. Ecol. 64, 167–174.

Reigstad, L.J., Jorgensen, S.L., Schleper, C., 2010. Diversity and abundance of Korarchaeota in terrestrial hot springs of Iceland and Kamchatka. ISME J. 4, 346–356.

Reigstad, L.J., Bartossek, R., Schleper, C., 2011. Preparation of high-molecular weight DNA and metagenomic libraries from soils and hot springs. Meth. Enzymol. 496, 319–344.

Reysenbach, A.L., Ehringer, M., Hershberger, K., 2000. Microbial diversity at 83 degrees C in Calcite Springs, Yellowstone National Park: another environment where the Aquificales and "Korarchaeota" coexist. Extremophiles 4, 61–67.

Rivera, M.C., Lake, J.A., 2004. The ring of life provides evidence for a genome fusion origin of eukaryotes. Nature 431, 152–155.

Robertson, C.E., Harris, J.K., Spear, J.R., Pace, N.R., 2005. Phylogenetic diversity and ecology of environmental Archaea. Curr. Opin. Microbiol. 8, 638–642.

Rodriguez, F., Oliver, J.L., Marin, A., Medina, J.R., 1990. The general stochastic model of nucleotide substitution. J. Theor. Biol. 142, 485–501.

Ronquist, F., Huelsenbeck, J.P., 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19, 1572–1574.

Schleper, C., Swanson, R.V., Mathur, E.J., DeLong, E.F., 1997. Characterization of a DNA polymerase from the uncultivated psychrophilic archaeon Cenarchaeum symbiosum. J. Bacteriol. 179, 7803–7811.

Schleper, C., DeLong, E.F., Preston, C.M., Feldman, R.A., Wu, K.Y., Swanson, R.V., 1998. Genomic analysis reveals chromosomal variation in natural populations of the uncultured psychrophilic archaeon Cenarchaeum symbiosum. J. Bacteriol. 180, 5003–5009.

Schleper, C., Jurgens, G., Jonuscheit, M., 2005. Genomic studies of uncultivated Archaea. Nat. Rev. Microbiol. 3, 479–488.

Skirnisdottir, S., Hreggvidsson, G.O., Hjorleifsdottir, S., Marteinsson, V.T., Petursdottir, S.K., Holst, O., Kristjansson, J.K., 2000. Influence of sulfide and temperature on species composition and community structure of hot spring microbial mats. Appl. Environ. Microbiol. 66, 2835–2841.

Spang, A., Hatzenpichler, R., Brochier-Armanet, C., Rattei, T., Tischler, P., Spieck, E., Streit, W., Stahl, D.A., Wagner, M., Schleper, C., 2010. Distinct gene set in two different lineages of ammonia-oxidizing archaea supports the phylum Thaumarchaeota. Trends Microbiol. 18, 331–340.

Spang, A., Poehlein, A., Offre, P., Zumbragel, S., Haider, S., Rychlik, N., Nowka, B., Schmeisser, C., et al., 2012. The genome of the ammonia-oxidizing Candidatus Nitrososphaera gargensis: insights into metabolic

versatility and environmental adaptations. Environ. Microbiol.. http://dx.doi.org/10.1111/j.1462-2920.2012.02893.x.

Spear, J.R., Walker, J.J., McCollom, T.M., Pace, N.R., 2005. Hydrogen and bioenergetics in the Yellowstone geothermal ecosystem. Proc. Natl. Acad. Sci. U S A 102, 2555–2560.

Stetter, K.O., 1996. Hyperthermophiles in the history of life. Ciba Found. Symp. 202, 1–10.

Takai, K., Sako, Y., 1999. A molecular view of archaeal diversity in marine and terrestrial hot water environments. FEMS Microbiol. Ecol. 28, 177–188.

Takai, K., Moser, D.P., DeFlaun, M., Onstott, T.C., Fredrickson, J.K., 2001. Archaeal diversity in waters from deep South African gold mines. Appl. Environ. Microbiol. 67, 5750–5760.

Tatusov, R.L., Koonin, E.V., Lipman, D.J., 1997. A genomic perspective on protein families. Science 278, 631–637.

Tatusov, R.L., Natale, D.A., Garkavtsev, I.V., Tatusova, T.A., Shankavaram, U.T., Rao, B.S., Kiryutin, B., Galperin, M.Y., et al., 2001. The COG database: new developments in phylogenetic classification of proteins from complete genomes. Nucleic Acids Res. 29, 22–28.

Team, R.D.C., 2011. R: A Language and Environment for Statistical Computing. Vienna, Austria.

Tersteegen, A., Linder, D., Thauer, R.K., Hedderich, R., 1997. Structures and functions of four anabolic 2-oxoacid oxidoreductases in Methanobacterium thermoautotrophicum. Eur. J. Biochem. 244, 862–868.

Thiergart, T., Landan, G., Schenk, M., Dagan, T., Martin, W.F., 2012. An evolutionary network of genes present in the eukaryote common ancestor polls genomes on eukaryotic and mitochondrial origin. Genome Biol. Evol. 4, 466–485.

de la Torre, J.R., Walker, C.B., Ingalls, A.E., Konneke, M., Stahl, D.A., 2008. Cultivation of a thermophilic ammonia oxidizing archaeon synthesizing crenarchaeol. Environ. Microbiol. 10, 810–818.

Tourna, M., Stieglmeier, M., Spang, A., Konneke, M., Schintlmeister, A., Urich, T., Engel, M., Schloter, M., et al., 2011. Nitrososphaera viennensis, an ammonia oxidizing archaeon from soil. Proc. Natl. Acad. Sci. U S A 108, 8420–8425.

Uyeda, K., Rabinowitz, J.C., 1971. Pyruvate-ferredoxin oxidoreductase. 3. Purification and properties of the enzyme. J. Biol. Chem. 246, 3111–3119.

Walker, C.B., de la Torre, J.R., Klotz, M.G., Urakawa, H., Pinel, N., Arp, D.J., Brochier-Armanet, C., Chain, P.S., et al., 2010. Nitrosopumilus maritimus genome reveals unique mechanisms for nitrification and autotrophy in globally distributed marine crenarchaea. Proc. Natl. Acad. Sci. U S A 107, 8818–8823.

Walter, M.C., Rattei, T., Arnold, R., Guldener, U., Munsterkotter, M., Nenova, K., Kastenmuller, G., Tischler, P., et al., 2009. PEDANT covers all complete RefSeq genomes. Nucleic Acids Res. 37, D408–D411.

Williams, T.A., Foster, P.G., Nye, T.M., Cox, C.J., Embley, T.M., 2012. A congruent phylogenomic signal places eukaryotes within the Archaea. Proc. Biol. Sci.. http://dx.doi.org/10.1098/rspb.2012.1795.

Woese, C.R., Kandler, O., Wheelis, M.L., 1990. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. Proc. Natl. Acad. Sci. U S A 87, 4576–4579.

Woese, C.R., 1987. Bacterial evolution. Microbiol. Rev. 51, 221–271.

Wuchter, C., Abbas, B., Coolen, M.J., Herfort, L., van Bleijswijk, J., Timmers, P., Strous, M., Teira, E., et al., 2006. Archaeal nitrification in the ocean. Proc. Natl. Acad. Sci. U S A 103, 12317–12322.

Zhang, Q., Iwasaki, T., Wakagi, T., Oshima, T., 1996. 2-oxoacid:ferredoxin oxidoreductase from the thermoacidophilic archaeon, Sulfolobus sp. strain 7. J. Biochem. 120, 587–599.

Zhang, C.L., Ye, Q., Huang, Z., Li, W., Chen, J., Song, Z., Zhao, W., Bagwell, C., et al., 2008. Global occurrence of archaeal amoA genes in terrestrial hot springs. Appl. Environ. Microbiol. 74, 6417–6426.