Genomic and transcriptomic evidence for scavenging of diverse organic compounds by widespread deep-sea archaea

Meng Li[1,2], Brett J. Baker[1,3], Karthik Anantharaman[1], Sunit Jain[1], John A. Breier[4,5] & Gregory J. Dick[1,6,7]

Microbial activity is one of the most important processes to mediate the flux of organic carbon from the ocean surface to the seafloor. However, little is known about the micro-organisms that underpin this key step of the global carbon cycle in the deep oceans. Here we present genomic and transcriptomic evidence that five ubiquitous archaeal groups actively use proteins, carbohydrates, fatty acids and lipids as sources of carbon and energy at depths ranging from 800 to 4,950 m in hydrothermal vent plumes and pelagic background seawater across three different ocean basins. Genome-enabled metabolic reconstructions and gene expression patterns show that these marine archaea are motile heterotrophs with extensive mechanisms for scavenging organic matter. Our results shed light on the ecological and physiological properties of ubiquitous marine archaea and highlight their versatile metabolic strategies in deep oceans that might play a critical role in global carbon cycling.

[1] Department of Earth and Environmental Sciences, University of Michigan, Ann Arbor, Michigan 48109, USA. [2] Institute for Advanced Study, Shenzhen University, Shenzhen 518060, China. [3] Department of Marine Science, University of Texas Austin, Marine Science Institute, 750 Channel View Drive, Port Aransas, Texas 78373, USA. [4] Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, USA. [5] University of Texas Rio Grande Valley, Brownsville, Texas 78520, USA. [6] Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, Michigan 48109, USA. [7] Center of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, Michigan 48109, USA. Correspondence and requests for materials should be addressed to M.L. (email: limeng848@szu.edu.cn) or to G.J.D. (email: gdick@umich.edu).

Archaea are ubiquitous members of marine microbial communities[1–3]. Four major groups of planktonic archaea have been reported in the global ocean, including Marine Group I Thaumarchaeota (MG-I)[2,3], Marine Group II Euryarchaeota (MG-II)[4], Marine Group III Euryarchaeota (MG-III)[4] and Marine Group IV Euryarchaeota (MG-IV)[5]. While MG-III and MG-IV are predominately found in the deep oceans at relatively low abundance[4,5], qualitative and quantitative studies suggest that MG-II are abundant in surface waters[4,6,7], whereas MG-I dominates at greater depths, sometimes constituting up to nearly 40% of marine microbial plankton[8].

Out of the four major groups of planktonic archaea, only representatives of MG-I have been cultured, which led to the discovery that they oxidize ammonia[9,10]. The MG-I are now generally recognized as the major drivers of nitrification in marine environments[11–13]. Up till now, all MG-I cultures oxidize ammonia and fix carbon, but there is also evidence for heterotrophy or mixotrophy by this group[10,14–16]. In contrast to the relatively well-studied MG-I, the physiology and energy metabolism of MG-II, MG-III and MG-IV remains poorly understood. Recent evidence indicates that MG-II can use organic carbon in the surface oceans[14,17], suggesting that archaea may play an important role in the marine carbon cycle. However, little is known about the heterotrophic metabolism of archaea in the mesopelagic and bathypelagic realms of the ocean[18], which comprise about 70% of ocean volume, account for the majority of marine microbial biomass and productivity[19], and contain huge numbers of archaea[8].

In this study, we reconstructed 59 partial to near-completed genomes and transcriptomes of several ubiquitous uncultured archaea groups from deep-sea hydrothermal plumes and surrounding background seawater at three distinct locations. Hydrothermal vent plumes are hotspots of biogeochemical activity in the deep oceans[20] yet they are composed largely of background deep-sea microorganisms, including archaea[21–23]. Thus, plumes represent a valuable environment for studying deep-sea microorganisms. Our results reveal metabolic characteristics of these ubiquitous marine archaea and suggest that they play critical roles in modulating carbon cycle in deep oceans.

Results

Genomes and transcriptomes of deep-sea archaea. We conducted shotgun metagenomic and metatranscriptomic sequencing on samples from deep-sea hydrothermal vent plumes and surrounding background seawaters at Mid-Cayman Rise in the Caribbean Sea, Guaymas Basin in the Gulf of California and Eastern Lau Spreading Center in the Western Pacific Ocean (Supplementary Table 1). *De novo* assembly of metagenomic reads (Supplementary Table 2) and binning by tetranucleotide signatures revealed 32 archaeal genomic 'bins' containing an estimated total of 59 archaeal genomes (Supplementary Fig. 1 and Supplementary Table 3)[24,25]. Estimates of genome completeness using an inventory of single-copy conserved genes[26] indicate that 26 are more than 70% complete and 18 are 50–70% complete (Supplementary Tables 3 and 4). Phylogenetic analysis revealed the presence of five distinct groups, including 18 genomes from MG-I, 31 from MG-II, 5 from MG-III, 3 from Parvarchaeota, and 2 from putative Deep-sea Hydrothermal Vent Euryarchaeaotic Group-6 (DHVEG-6) (Table 1 and Supplementary Figs 2–4).

Comparative genomics showed that the five MG-I genomic bins had 53 to 59% average amino acid identity to the cultured *Nitrosopumilus maritimus* SCM1 (refs 27,28). One genomic bin (Guaymas69) was the same MG-I population as previously reported[29], while other four bins (Lau19, Guaymas96,

Cayman117 and Cayman91) formed a separate deep-sea clade of MG-I (Supplementary Fig. 5). MG-II genomic bins clustered into three clades, congruent with the phylogeny of 16S rRNA genes (Supplementary Figs 2–5). Consistent with previous results[30,31], the deep-sea MG-II did not contain genes for proteorhodopsin, a light-driven proton pump present in MG-II from the photic zone, indicating that the deep MG-II are not capable of photoheterotrophic metabolism proposed for surface MG-II (refs 17,31–33). Three MG-III genomic bins (Guaymas31/92/93) formed a unique lineage with lower similarity to the fourth bin (Guaymas32, 48%) and two partial single cell genomes (Supplementary Fig. 5). Genomic bins related to Parvarchaeota were distantly related to *Parvarchaeota acidiphitum* ARMAN-4 from acid mine drainage[34] and *Nanoarchaeota* DUSEL1 SCGC-AAA011D5 from the Homestake Mine in South Dakota[26]. We also obtained a genomic bin belonging DHVEG-6, for which there are no publicly available genomes (Supplementary Table 3).

Extracellular peptidases in deep-sea archaea. Our data indicate that these deep-sea archaea are metabolically active as aerobic heterotrophs or, in the case of MG-I, mixotrophs. Genes encoding 28 different families of extracellular peptidases for protein degradation were identified. Four of these families were dominant (S08A, M28A, M14A and M22), comprising 70.5%, 72% and 52% of total extracellular peptidase genes in deep-sea MG-I, -II and -III, respectively (Fig. 1, Supplementary Fig. 6a and Supplementary Table 5). Transcripts of these extracellular peptidase genes were among the most abundant in the transcriptomes of MG-II and MG-III in both Guaymas and Cayman plumes (Supplementary Figs 6a and 7). For MG-I, in addition to the S08A and M22 peptidase families, transcripts for M01, S26A and S26B were also detected (Supplementary Table 5), though the most abundant transcripts in the MG-I transcriptome were related to the transport and oxidation of ammonia (Supplementary Fig. 8). Interestingly, the suite of specific extracellular peptidases for which transcripts were detected varied between archaeal groups (Supplementary Fig. 6a), suggesting diverse protein utilization capabilities by MG-I, -II and -III in the deep oceans[35].

In contrast to MG-I, -II and -III, the genomic bins of Parvarchaeota and DHVEG-6 contained only one gene related to the S01C family (DegP peptidase) and two genes affiliated with M48B family (HtpX peptidase), and no transcripts of these genes were detected (Supplementary Table 5). These results indicate that (1) MG-I, -II, -III, Parvarchaeota and DHVEG-6 are all capable of degrading and utilizing extracellular proteins; (2) protein degradation is a major metabolic pathway for MG-II and -III while a minor metabolic route for MG-I; and (3) the different archaeal groups each have a different suite of protein degradation genes, suggesting niche differentiation on the basis of substrate (Table 1 and Fig. 2).

Although many of the peptidases identified here are divergent from experimentally studied proteins, family-level classifications offer clues to their broad functions. S08A peptidases, one of the largest groups of serine endopeptidases[36], clustered into four major phylogenetic groups: alkaline proteases, oxidant-stable proteases, Bacillopeptidases F-like proteinases and unclassified archaea subtilisin (Supplementary Fig. 9). S08A are N-terminus peptidases containing a catalytic triad sequence of Asp, His and Ser. M28A peptidases are zinc-dependent exopeptidases that selectively cleave N-terminal amino acids from polypeptides[37,38]. M28A-related genes in MG-II and -III fell into six clusters (Supplementary Fig. 10). Clusters 1 and 6 had abundant transcripts but their specific functions are unknown. M14A peptidase is a group of zinc-containing proteolytic

Table 1 | Overview of genomes from five archaeal groups recovered in this study and their ecophysiological characteristics.

Archaea group	Number of genome bins	Number of genomes	Lithoautotrophic metabolisms	Heterotrophic metabolisms
MG-I	6	18	3-Hydroxypropionate/4-hydrobutyrate cycle, ammonia oxidation	Protein degradation, carbohydrate metabolism
MG-II	19	31	Not detected	Protein degradation, carbohydrate metabolism, β-oxidation
MG-III	4	5	Not detected	Protein degradation, carbohydrate metabolism, β-oxidation
Parvarchaeota	2	3	Not detected	Protein degradation, carbohydrate metabolism
Putative DHVEG-6	1	2	Not detected	Protein degradation, carbohydrate metabolism

DHVEG-6, Deep-sea Hydrothermal Vent Euryarchaeaotic Group-6; MG, Marine Group.
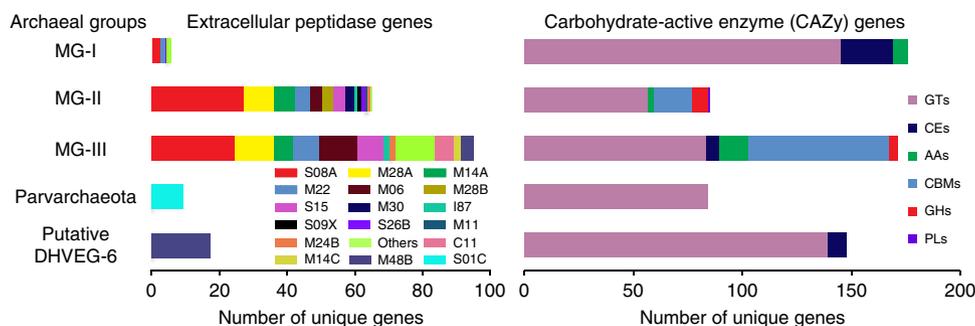
Figure 1 | Genes for extracellular peptidases and carbohydrate-active enzymes. Carbohydrate-active enzymes (CAZYs) include glycoside hydrolases (GHs), glycosyl transferases (GTs), polysaccharide lyases (PLs), carbohydrate esterases (CEs), auxiliary activities (AAs) and carbohydrate-binding modules (CBMs). The two bar graphs indicate the unique genes (60% amino acids identity) for extracellular peptidases and CAZYs identified in five archaeal groups, and the general function description for extracellular peptidases and carbohydrate-active enzymes is listed in Supplementary Tables 5–7. The gene number has been normalized to the total number of genes recovered from each archaeal group.

enzymes (including carboxypeptidase A), which catalyses the hydrolysis of C-terminal peptide bonds in protein and peptide substrates[39]. Genes encoding the M14A family in MG-II and -III were divided into three clusters that are each significantly different from known carboxypeptidase A (Supplementary Fig. 11). This extensive suite of extracellular enzymes that cleave peptides at different sites suggests comprehensive protein degradation by deep-sea MG-II/III and potential use of protein as a growth substrate.

Carbohydrate-active enzymes (CAZYs) in deep-sea archaea. The archaeal genomes also contained diverse and transcriptionally active carbohydrate-active enzymes (Fig. 1, Supplementary Fig. 6b and Supplementary Table 6). In MG-I, Parvarchaeota and DHVEG-6 we identified glycosyl transferases, carbohydrate esterases and other auxiliary activities that may be involved in the modification or creation of glycosidic bonds in various form of organic matter within cells[40]. However, genes related to glycoside hydrolases and carbohydrate-binding modules were more diverse and common in MG-II and -III (Supplementary Table 7), suggesting they are also involved in the hydrolysis of glycosidic bonds in complex sugars[40,41]. Similar to our observations of differential expression of peptidases in MG –II and –III, we observed that α-mannosidase (GH38), amylopullulanase (GH57), 4-α-glucanotransferase (GH77) and chitinases (GH18/CBM5 and GH20/CBM5) were abundant in the MG-II transcriptome, while only chitinases (GH18/CBM5 and GH20/CBM5) were transcriptionally active in MG-III, suggesting

differentiation of carbohydrate use between archaeal groups (Fig. 2 and Supplementary Table 7).

Lipid metabolism in deep-sea archaea. Archaeal genes involved in beta-oxidation of fatty acids were present and abundantly expressed in deep-sea MG-II and -III, suggesting that straight chain lipids may also be organic substrates for these groups (Supplementary Table 8 and Fig. 2). A recent study concluded that MG-II can synthesize crenarchaeol[42], a tetraether lipid widely used in paleoceanography and previously considered to be a unique biomarker for MG-I (refs 43,44). Here, we found the key gene for tetraether lipid biosynthesis, geranylgeranylglyceryl phosphate (GGGP) synthase, in MG-I, -II and -III genomes (Supplementary Table 9). Phylogenetic analysis indicated that GGGP synthase genes from MG-II and MG-III were different from that of MG-I, forming two separate clusters (Supplementary Fig. 12). The genetic basis for tetraether lipid synthesis requires further study in order to better understand the lipid biosynthetic potential of these different groups, and how the pathways are regulated as a function of environmental conditions. In addition, we also identified MG-II genes involved in synthesis of archaeal ether-like lipids, polyunsaturated fatty acids and glycerophospholipids (Supplementary Table 10).

Cell mobility and central metabolism in deep-sea archaea. Genes involved in cell mobility were identified in deep-sea MG-I, –II and –III genomes (Fig. 2 and Supplementary Table 11). Particularly in MG-II and -III, several flagellar genes with different transcriptional patterns were identified as operons in the
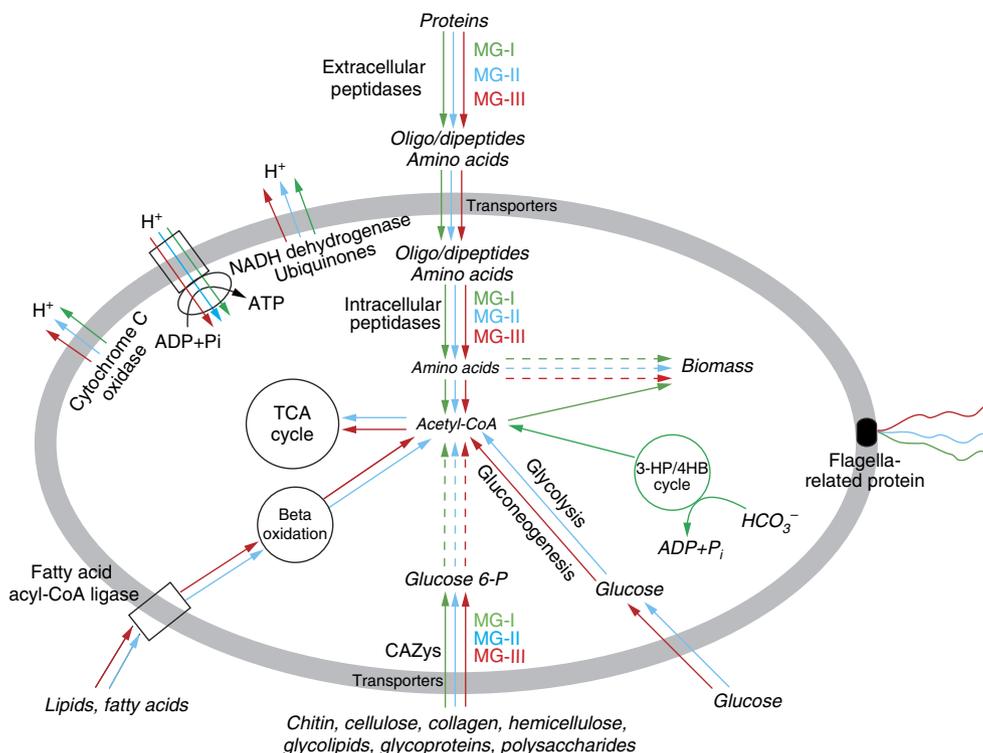
Figure 2 | Proposed transcriptionally active heterotrophic metabolic pathways. Green, blue and red indicate heterotrophic pathways found in MG-I, MG-II and MG-III, respectively.

same contig (Supplementary Table 11 and Supplementary Fig. 13), suggesting that these ubiquitous deep-sea archaea have active flagella, which might facilitate acquisition of organic substrates via chemotaxis[45] or attachment to particles[46].

Finally, genes related to central carbon metabolism were also identified in these genomes and transcriptomes, including the tricarboxylic acid cycle and glycolysis/gluconeogenesis pathway in MG-II and MG-III, and the modified 3-hydroxypropionate/4-hydroxybutyrate (3-HP/4-HB) cycle in deep-sea MG-I (Table 1, Fig. 2 and Supplementary Table 12). These results further highlight that these ubiquitous deep-sea archaea have versatile metabolic pathways that are important for carbon utilization in the deep oceans.

Discussion

Overall, the genomic and transcriptomic results presented here indicate that widespread deep-sea archaea utilize diverse proteins, carbohydrates and straight chain lipids as substrates for heterotrophic growth (Table 1 and Fig. 2). Although our previous work on hydrothermal plumes identified the major groups responsible for production of organic matter through chemosynthesis[21,25,29,47], the results presented here indicate that archaea play a role in consuming that fresh organic matter, which can be regionally significant[20]. Further, because these archaea are widespread throughout the deep ocean, they likely contribute to the broader marine carbon cycle.

Microbial degradation of organic matter in the pelagic ocean-water column strongly influences the export of carbon from the surface oceans, and converts labile carbon into refractory organic matter that can be preserved in deep-sea sediments or as dissolved organic carbon (DOC)[48–52]. The deep-sea DOC reservoir is massive, holding a quantity of carbon that is roughly equivalent to that in the atmosphere. Recent evidence suggests that 30% of the deep DOC reservoir is modern and actively cycled on short time scales[53]. Although the contributions

of deep-sea archaea to such processes remain to be quantified, our findings demonstrate an active archaeal response to pulses of organic matter in the deep sea that may shunt carbon from proteins and carbohydrates to CO_2 and lipids. Because archaea are ubiquitous across varying geochemical regimes and depths of the world's deep oceans[30,54], these results implicate archaea as key players in heterotrophy and provide new mechanistic perspectives on carbon cycling in the deep oceans.

Methods

DNA and concentrated and repurified using the RNeasy MinElute Kit (Qiagen). Total RNA was quantified using RiboGreen (Invitrogen). RNA amplification by random priming and complementary DNA synthesis was performed using the MessageAmp II-Bacteria Kit (Ambion). Shotgun sequencing of DNA and complementary DNA were performed with Illumina HiSeq2000 at the University of Michigan DNA Sequencing Core, and results are shown in Supplementary Table 1.

De novo metagenomic assembly and gene annotation. The raw shotgun sequencing metagenomic reads were depreplicated (100% identity over 100% lengths) and trimmed using Sickle (https://github.com/najoshi/sickle). Samples from the eight vent sites (Guaymas, Von Damm (Cayman Shallow), Beebe (Cayman Deep), Kilo Moana, Abe, Mariner, Tahi Moana, Tui Malila) were each assembled de novo to obtain eight separate assemblies (Supplementary Table 2). Whole genome de novo assemblies were performed using IDBA-UD[59] with the following parameters: −mink 50, −maxk 92, −steps 4 or 6, −min_contig 500. Gene calling and annotations were done using the DOE Joint Genome Institutes (JGI) Integrated Microbial Genomes pipeline (http://img.jgi.doe.gov/cgi-bin/w/main.cgi).

Metagenome binning and metatranscriptomic data analysis. Binning of assembled metagenomic sequences was performed using a combination of tetranucleotide frequencies, contig coverage and %GC content in emergent self-organizing maps[24]. A total of 32 genomic bins from eight assemblies were identified as archaeal bins (Supplementary Fig. 1). Based the inventory of archaeal 162 single-copy conserved genes, 32 archaeal genomic bins were estimated to contain 59 partial to complete genomes (Supplementary Tables 3 and 4). The archaeal 16S and 23S rRNA genes and two ribosomal protein-coding genes from each archaeal genomic bin were further analysed for identifying their taxonomic positions (Supplementary Figs 2–4). Abundance of transcripts for each archaeal genomic bins was determined by mapping all non-rRNA transcripts to the assembled fragments using BWA with default settings[60,61] and normalizing to their sequence length and total number of non-rRNA transcripts in each metatranscriptomic library.

Comparative genomics. Comparative genomics of the deep-sea archaea identified in this study was performed against the representative archaeal genomes within each group (Supplementary Fig. 5). Genomic similarity with known archaea (Nitrosopumilus maritimus (CP000866)[27] in MG-I; surface seawater uncultured marine group II euryarchaeote (CM001443)[17] in MG-II; two single cell genomes of MG-III (SCGC-AAA007-O11 and SCGC-AAA288-E19) in MG-III, Parvarchaea acidiphilium ARMAN-4 (PRJNA38567)[34] and SCGC-AAA011-D5 (ref. 26) in Parvarchaeota) was determined based on reciprocal best BLASTP hits between known genomes and archaeal genomic bins. The average identity (amino acid level) among each archaeal genomic bins and representative archaea genome were generated as Supplementary Fig. 5.

Identification of key functional genes. For genes encoding peptidases and carbohydrate-active enzymes, all annotated genes in deep-sea archaea genomic bins were searched against public databases of peptidases (MEROPS)[62] and CAZYs[63] with E-value $< 10^{-10}$ by BLASTP. All hits were compared against the non-redundant NCBI protein database. Only those that had top hits to peptidases and carbohydrate active enzymes were considered. The extracellular peptidases were further confirmed based on the identification of extracellular transport signals using SignaIP[64], POSRTb[65] or PRED-SIGNAL[66] (Supplementary Table 5). Genes related to carbohydrate active enzymes were further classified into different groups according the prediction of CAZYs (Supplementary Tables 6 and 7). Genes involved in lipid metabolic pathways (Supplementary Tables 8 and 9) and the central metabolism (Supplementary Table 10) in MG-I/II/III were identified using similar procedures based on the published gene databases[17,67].

Sequence alignment and phylogeny. Alignment of amino acid sequences for extracellular peptidase of S08A, M28 and M14A families was performed by MUSCLE using default parameters followed by manual refinement[68]. Phylogenetic analysis of representative peptidase of three families and GGGP synthases were inferred by Maximum Likelihood implemented in Mega 6.0[69] using the Tamura-Nei and passion model after testing by ProTest[70] and bootstrapped 1000 times (Supplementary Figs 9–12).

References
1. Delong, E. F. Microbial domains in the ocean: a lesson from the archaea. *Oceanography* **20**, 124–129 (2007).
2. DeLong, E. F. Archaea in coastal marine environments. *Proc. Natl Acad. Sci. USA* **89**, 5685–5689 (1992).
3. Fuhrman, J. A., McCallum, K. & Davis, A. A. Novel major archaebacterial group from marine plankton. *Nature* **356**, 148–149 (1992).
4. Fuhrman, J. A. & Davis, A. A. Widespread archaea and novel bacteria from the deep sea as shown by 16S rRNA gene sequences. *Mar. Ecol. Prog. Ser.* **150**, 275–285 (1997).
5. Lopez-Garcia, P., Moreira, D., Lopez-Lopez, A. & Rodriguez-Valera, F. A novel haloarchaeal-related lineage is widely distributed in deep oceanic regions. *Environ. Microbiol.* **3**, 72–78 (2001).
6. Murray, A. E. et al. Seasonal and spatial variability of bacterial and archaeal assemblages in the coastal waters near Anvers Island, Antarctica. *Appl. Environ. Microbiol.* **64**, 2585–2595 (1998).
7. Massana, R., DeLong, E. F. & Pedros-Alio, C. A few cosmopolitan phylotypes dominate planktonic archaeal assemblages in widely different oceanic provinces. *Appl. Environ. Microbiol.* **66**, 1777–1787 (2000).
8. Karner, M. B., DeLong, E. F. & Karl, D. M. Archaeal dominance in the mesopelagic zone of the Pacific Ocean. *Nature* **409**, 507–510 (2001).
9. Konneke, M. et al. Isolation of an autotrophic ammonia-oxidizing marine archaeon. *Nature* **437**, 543–546 (2005).
10. Qin, W. et al. Marine ammonia-oxidizing archaeal isolates display obligate mixotrophy and wide ecotypic variation. *Proc. Natl Acad. Sci. USA* **111**, 12504–12509 (2014).
11. Wuchter, C. et al. Archaeal nitrification in the ocean. *Proc. Natl Acad. Sci. USA* **103**, 12317–12322 (2006).
12. Francis, C. A., Roberts, K. J., Beman, J. M., Santoro, A. E. & Oakley, B. B. Ubiquity and diversity of ammonia-oxidizing archaea in water columns and sediments of the ocean. *Proc. Natl Acad. Sci. USA* **102**, 14683–14688 (2005).
13. Church, M. J., Wai, B., Karl, D. M. & DeLong, E. F. Abundances of crenarchaeal amoA genes and transcripts in the Pacific Ocean. *Environ. Microbiol.* **12**, 679–688 (2010).
14. Ouverney, C. C. & Fuhrman, J. A. Marine planktonic archaea take up amino acids. *Appl. Environ. Microbiol* **66**, 4829–4833 (2000).
15. Ingalls, A. E. et al. Quantifying archaeal community autotrophy in the mesopelagic ocean using natural radiocarbon. *Proc. Natl Acad. Sci. USA* **103**, 6442–6447 (2006).
16. Santoro, A. E. et al. Genomic and proteomic characterization of 'Candidatus Nitrosopelagicus brevis': An ammonia-oxidizing archaeon from the open ocean. *Proc. Natl Acad. Sci. USA* **112**, 1173–1178 (2015).
17. Iverson, V. et al. Untangling genomes from metagenomes: revealing an uncultured class of marine Euryarchaeota. *Science* **335**, 587–590 (2012).
18. Schleper, C., Jurgens, G. & Jonuscheit, M. Genomic studies of uncultivated archaea. *Nat. Rev. Microbiol.* **3**, 479–488 (2005).
19. Aristegui, J., Gasol, J. M., Duarte, C. M. & Herndl, G. J. Microbial oceanography of the dark ocean's pelagic realm. *Limnol. Oceanogr.* **54**, 1501–1529 (2009).
20. Dick, G. J. et al. The microbiology of deep-sea hydrothermal vent plumes: ecological and biogeographic linkages to seafloor and water column habitats. *Front. Microbiol* **4**, 124 (2013).
21. Lesniewski, R. A., Jain, S., Anantharaman, K., Schloss, P. D. & Dick, G. J. The metatranscriptome of a deep-sea hydrothermal plume is dominated by water column methanotrophs and lithotrophs. *ISME J.* **6**, 2257–2268 (2012).
22. Reed, D. C. et al. Predicting the response of the deep-ocean microbiome to geochemical perturbations by hydrothermal vents. *ISME J.* **9**, 1857–1869 (2015).
23. Sheik, C. S. et al. Spatially resolved sampling reveals dynamic microbial communities in rising hydrothermal plumes across a back-arc basin. *ISME J.* **9**, 1434–1445 (2014).
24. Dick, G. J. et al. Community-wide analysis of microbial genome sequence signatures. *Genome Biol.* **10**, R85 (2009).
25. Anantharaman, K. et al. Sulfur oxidation genes in diverse deep-sea viruses. *Science* **344**, 757–760 (2014).
26. Rinke, C. et al. Insights into the phylogeny and coding potential of microbial dark matter. *Nature* **499**, 431–437 (2013).
27. Walker, C. B. et al. Nitrosopumilus maritimus genome reveals unique mechanisms for nitrification and autotrophy in globally distributed marine crenarchaea. *Proc. Natl Acad. Sci. USA* **107**, 8818–8823 (2010).
28. Konstantinidis, K. T., Ramette, A. & Tiedje, J. M. The bacterial species definition in the genomic era. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **361**, 1929–1940 (2006).
29. Baker, B. J., Lesniewski, R. A. & Dick, G. J. Genome-enabled transcriptomics reveals archaeal populations that drive nitrification in a deep-sea hydrothermal plume. *ISME J.* **6**, 2269–2279 (2012).
30. 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 Biol. Evol.* **6**, 1549–1563 (2014).
31. Frigaard, N. U., Martinez, A., Mincer, T. J. & DeLong, E. F. Proteorhodopsin lateral gene transfer between marine planktonic bacteria and archaea. *Nature* **439**, 847–850 (2006).
32. Martin-Cuadrado, A. B. et al. A new class of marine Euryarchaeota group II from the mediterranean deep chlorophyll maximum. *ISME J.* **9**, 1619–1634 (2014).

33. Orsi, W. D. et al. Ecophysiology of uncultivated marine euryarchaea is linked to particulate organic matter. ISME J. 9, 1747–1763 (2015).

34. Baker, B. J. et al. Enigmatic, ultrasmall, uncultivated Archaea. Proc. Natl Acad. Sci. USA 107, 8806–8811 (2010).

35. Rawlings, N. D. & Barrett, A. J. Evolutionary families of peptidases. Biochem. J. 290, 205–218 (1993).

36. Siezen, R. J. et al. Homology modelling and protein engineering strategy of subtilases, the family of subtilisin-like serine proteinases. Protein Eng. 4, 719–737 (1991).

37. Jankiewicz, U. & Bielawski, W. The properties and functions of bacterial aminopeptidases. Acta Microbiol. Pol. 52, 217–231 (2003).

38. Gonzales, T. & Robert-Baudouy, J. Bacterial aminopeptidases: properties and functions. FEMS Microbiol. Rev. 18, 319–344 (1996).

39. Quiocho, F. A. & Lipscomb, W. N. Carboxypeptidase A: a protein and an enzyme. Adv. Protein Chem. 25, 1–78 (1971).

40. Cantarel, B. L. et al. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. Nucleic Acids Res. 37, D233–D238 (2009).

41. Davies, G. J., Gloster, T. M. & Henrissat, B. Recent structural insights into the expanding world of carbohydrate-active enzymes. Curr. Opin. Struct. Biol. 15, 637–645 (2005).

42. Lincoln, S. A. et al. Planktonic Euryarchaeota are a significant source of archaeal tetraether lipids in the ocean. Proc. Natl Acad. Sci. USA 111, 9858–9863 (2014).

43. Damste, J. S., Schouten, S., Hopmans, E. C., van Duin, A. C. & Geenevasen, J. A. Crenarchaeol: the characteristic core glycerol dibiphytanyl glycerol tetraether membrane lipid of cosmopolitan pelagic Crenarchaeota. J. Lipid Res. 43, 1641–1651 (2002).

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

45. Stocker, R. Marine microbes see a sea of gradients. Science 338, 628–633 (2012).

46. Donlan, R. M. Biofilms: Microbial life on surfaces. Emerg. Infect. Dis. 8, 881–890 (2002).

47. Anantharaman, K., Breier, J. A. & Dick, G. J. Metagenomic resolution of microbial functions in deep-sea hydrothermal plumes across the Eastern Lau Spreading Center. ISME J. doi:10.1038/ismej.2015.81; e-pub ahead of print.

48. Azam, F. Microbial control of oceanic carbon flux: the plot thickens. Science 280, 694–696 (1998).

49. del Giorgio, P. A. & Duarte, C. M. Respiration in the open ocean. Nature 420, 379–384 (2002).

50. Herndl, G. J. & Reinthaler, T. Microbial control of the dark end of the biological pump. Nat. Geosci. 6, 718–724 (2013).

51. McCarren, J. et al. Microbial community transcriptomes reveal microbes and metabolic pathways associated with dissolved organic matter turnover in the sea. Proc. Natl Acad. Sci. USA 107, 16420–16427 (2010).

52. Jiao, N. et al. Microbial production of recalcitrant dissolved organic matter: long-term carbon storage in the global ocean. Nat. Rev. Microbiol. 8, 593–599 (2010).

53. Follett, C. L., Repeta, D. J., Rothman, D. H., Xu, L. & Santinelli, C. Hidden cycle of dissolved organic carbon in the deep ocean. Proc. Natl Acad. Sci. USA 111, 16706–16711 (2014).

54. Martin-Cuadrado, A. B. et al. Hindsight in the relative abundance, metabolic potential and genome dynamics of uncultivated marine archaea from comparative metagenomic analyses of bathypelagic plankton of different oceanic regions. ISME J. 2, 865–886 (2008).

55. Breier, J. A. et al. A large volume particulate and water multi-sampler with in situ preservation for microbial and biogeochemical studies. Deep Sea Res. Part I Oceanogr. Res. Pap. 94, 195–206 (2014).

56. Breier, J. A. et al. A suspended-particle rosette multi-sampler for discrete biogeochemical sampling in low-particle-density waters. Deep-Sea Res. I 56, 1579–1589 (2009).

57. Anantharaman, K., Breier, J. A., Sheik, C. S. & Dick, G. J. Evidence for hydrogen oxidation and metabolic plasticity in widespread deep-sea sulfur-oxidizing bacteria. Proc. Natl Acad. Sci. USA 110, 330–335 (2013).

58. Dick, G. J. & Tebo, B. M. Microbial diversity and biogeochemistry of the Guaymas Basin deep-sea hydrothermal plume. Environ. Microbiol. 12, 1334–1347 (2010).

59. Peng, Y., Leung, H. C. M., Yiu, S. M. & Chin, F. Y. L. IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. Bioinformatics 28, 1420–1428 (2012).

60. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754–1760 (2009).

61. Li, M. et al. Microbial iron uptake as a mechanism for dispersing iron from deep-sea hydrothermal vents. Nat. Commun. 5, 3192 (2014).

62. Rawlings, N. D., Waller, M., Barrett, A. J. & Bateman, A. MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. Nucleic Acids Res. 42, D503–D509 (2014).

63. Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M. & Henrissat, B. The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res. 42, D490–D495 (2014).

64. Petersen, T. N., Brunak, S., von Heijne, G. & Nielsen, H. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat. Methods 8, 785–786 (2011).

65. Yu, N. Y. et al. PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. Bioinformatics 26, 1608–1615 (2010).

66. Bagos, P. G., Tsirigos, K. D., Plessas, S. K., Liakopoulos, T. D. & Hamodrakas, S. J. Prediction of signal peptides in archaea. Protein Eng. Des. Sel. 22, 27–35 (2009).

67. Brasen, C., Esser, D., Rauch, B. & Siebers, B. Carbohydrate metabolism in archaea: current insights into unusual enzymes and pathways and their regulation. Microbiol. Mol. Biol. Rev. 78, 89–175 (2014).

68. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32, 1792–1797 (2004).

69. Tamura, K., Stecher, G., Peterson, D., Filipski, A. & Kumar, S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol. Bio. Evol. 30, 2725–2729 (2013).

70. Darriba, D., Taboada, G. L., Doallo, R. & Posada, D. ProtTest 3: fast selection of best-fit models of protein evolution. Bioinformatics 27, 1164–1165 (2011).

Acknowledgements

Author contributions

M.L. and G.J.D. designed the study; M.L., B.J.B., A.K., S.J. and G.J.D. performed the genomic analyses and wrote the manuscript. K.A., J.A.B. and G.J.D. collected the samples. All authors contributed to the Supplementary Materials.

Additional information

Accession codes: The nucleotide sequences have been deposited in the DOE JGI-IMG/MER database under the following Taxon Object IDs: 3300001680 (Kilo Moana), 3300001681 (Abe), 3300001678 (Mariner), 3300001679 (Tahi Moana), 3300001676 (Tui Malila), 3300001683 (Guaymas), 3300001781 (Cayman Deep) and 3300001835 (Cayman Shallow).

Supplementary Information accompanies this paper at http://www.nature.com/naturecommunications

Competing financial interests: The authors declare no competing financial interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

How to cite this article: Li, M. et al. Genomic and transcriptomic evidence for scavenging of diverse organic compounds by widespread deep-sea archaea. Nat. Commun. 6:8933 doi: 10.1038/ncomms9933 (2015).