

Metagenomic investigation of the geologically unique Hellenic Volcanic Arc reveals a distinctive ecosystem with unexpected physiology

Anastasis Oulas,¹ Paraskevi N. Polymenakou,^{1†} Rekha Seshadri,^{2†} H. James Tripp,² Manolis Mandalakis,¹ A. David Paez-Espino,² Amrita Pati,² Patrick Chain,³ Paraskevi Nomikou,⁴ Steven Carey,⁵ Stephanos Kiliias,⁴ Christos Christakis,¹ Georgios Kotoulas,¹ Antonios Magoulas,¹ Natalia N. Ivanova² and Nikos C. Kyrpides^{2,6*}

¹*Institute of Marine Biology, Biotechnology and Aquaculture, Hellenic Centre for Marine Research, Gournes Peditados, P.O. Box 2214, Heraklion, Crete 71003, Greece.*

²*Department of Energy, Microbial Genome and Metagenome Program, Joint Genome Institute, Walnut Creek, CA, USA.*

³*Los Alamos National Laboratory, Los Alamos, NM, USA.*

⁴*National and Kapodistrian University of Athens, Faculty of Geology and Geoenvironment, Athens, Greece.*

⁵*Graduate School of Oceanography, University of Rhode Island, Kingston, RI, USA.*

⁶*Department of Biological Sciences, King Abdulaziz, Jeddah, Saudi Arabia.*

Summary

Hydrothermal vents represent a deep, hot, aphotic biosphere where chemosynthetic primary producers, fuelled by chemicals from Earth's subsurface, form the basis of life. In this study, we examined microbial mats from two distinct volcanic sites within the Hellenic Volcanic Arc (HVA). The HVA is geologically and ecologically unique, with reported emissions of CO₂-saturated fluids at temperatures up to 220°C and a notable absence of macrofauna. Metagenomic data reveals highly complex prokaryotic communities composed of chemolithoautotrophs, some methanotrophs, and to our surprise, heterotrophs capable of anaerobic degradation of aromatic hydrocarbons. Our data suggest that aromatic hydrocarbons may indeed be a significant source of carbon in

these sites, and instigate additional research into the nature and origin of these compounds in the HVA. Novel physiology was assigned to several uncultured prokaryotic lineages; most notably, a SAR406 representative is attributed with a role in anaerobic hydrocarbon degradation. This dataset, the largest to date from submarine volcanic ecosystems, constitutes a significant resource of novel genes and pathways with potential biotechnological applications.

Introduction

Hydrothermal vents (HVs) constitute biogeochemical environments characterized by emission of reactive gases, dissolved elements and sharp thermal and chemical gradients (Schrenk *et al.*, 2003; Kristall *et al.*, 2006). Microbes have adapted to survive these extreme conditions, often withstanding extremely high temperatures (e.g. in chimneys reaching maximum temperature bursts of 464°C), salinity, heavy metals and acidic pH (Brock and Freeze, 1969; Pakchung *et al.*, 2006; Perner *et al.*, 2014). HVs provide a wide range of microhabitats for chemolithoautotrophic microorganisms. In these environments, microorganisms are the primary producers by fixing inorganic carbon using chemical energy obtained through the oxidation of reduced inorganic compounds thus converting the geothermally derived energy into microbial biomass (Reysenbach and Shock, 2002).

Despite the vital role of microorganisms in these ecosystems and their potential biotechnological applications, key questions about their diversity and metabolic capabilities remain understudied simply because many hydrothermal vent environments remain largely unexplored (Dando *et al.*, 2000; Reysenbach and Shock, 2002; Campbell and Cary, 2004; Mehta and Baross, 2006; Proskurowski *et al.*, 2008; Brazelton and Baross, 2009; Byrne *et al.*, 2009; Querellou *et al.*, 2010; Xie *et al.*, 2011; Bourbonnais *et al.*, 2012; Kawaichi *et al.*, 2013; Kiliias *et al.*, 2013; Cao *et al.*, 2014; Perner *et al.*, 2014; Stokke *et al.*, 2015). Geologically unique and poorly studied HVs exist in the Hellenic Volcanic Arc (HVA), eastern Mediterranean, where volcanism and hydrothermal activity occur through thinned continental crust associated with the African-Eurasian subduction zone (Dando *et al.*, 2000; Kiliias *et al.*, 2013). Major hydrothermal systems are found

Received 6 May, 2015; accepted 16 October, 2015. *For correspondence. E-mail nckyrpides@lbl.gov; Tel. (925) 2965718; Fax (925) 296-5666. †Equal contribution.

along the HVA at Methana, Milos, Santorini, Nisiros and Kos (Dando *et al.*, 2000). Venting gases in these areas contain substantial amounts of CO₂ (95%), H₂ (9%), H₂S (3%) and CH₄ (1.8%) providing an ideal chemical environment for chemolithoautotrophic primary production (Dando *et al.*, 1995; 2000). The Santorini volcano is world famous because of its recent explosive eruption (~3600 years ago), which was one of the largest known volcanic events in historical time. The Santorini volcanic field extends for 20 km as a line of more than 20 submarine cones. The largest of them is Kolumbo, a 3 km diameter cone with a 1500 m wide crater situated 505 m below sea level.

In 2006, chimneys emitting fluids with varying temperatures of up to 220°C, as well as lower temperature chimneys and vents (with fluids up to 70°C) were discovered in the northern part of the Kolumbo crater (Sigurdsson *et al.*, 2006). This volcanic discharge of CO₂ contributes to local reductions in pH and serve as natural experiments in ocean acidification (Hall-Spencer *et al.*, 2008; Tunnicliffe *et al.*, 2008). Additionally, the enclosed geometry of the crater impedes vertical mixing and results in stably stratified density gradient and accumulation of acidic water (Carey *et al.*, 2013). As a probable consequence, there is an absence of macrofauna (primary consumers and so on) that are typically associated with many HV sites. Indeed, the only thriving biological entities are of bacterial and archaeal origin (Kiliyas *et al.*, 2013).

The exterior of most chimneys and large areas of the seabed around Kolumbo were covered with white/grey and reddish/orange microbial mats. This newly discovered hydrothermal vent field is gaining increasing interest for potential biotechnological exploitation and, recently, a detailed bio-geochemical description of this environment was undertaken (Kiliyas *et al.*, 2013). 16S rRNA gene pyrosequencing analysis demonstrated the presence of highly diverse microbial communities and revealed the dominance of archaeal sequences closely related to *Nitrosopumilus maritimus* strongly suggesting that nitrification should play a key role in this environment (Kiliyas *et al.*, 2013).

In the present study, we used shotgun metagenomic sequencing of microbial communities from two different sites located within the same volcanic complex (HVA; Fig. 1) in order to decipher their metabolic potential and environmental adaptation strategies. This comparison reveals the magnitude of phylogenetic and metabolic diversity that can be found at the Santorini volcanic complex. The first sample (hereafter white mat) was taken from the CO₂-rich environment of the Kolumbo volcano with temperatures varying from 60 to 70°C (Fig. 1C), whereas the second sample (hereafter red mat) was taken from the CO₂-poor environment of the Santorini caldera with temperatures close to 22°C (Fig. 1B). We

present a comparative metagenomic analysis of the two samples, as well as comparison with other publicly available datasets from related environments. Insights from taxonomic and functional observations from these microbial communities suggest unexpected physiological diversity including, notably, the use of aromatic hydrocarbons as an important carbon and energy source.

Results and discussion

White mat and red mat samples were collected from the HVA approximately 10 km apart, proximal to the Kolumbo and Santorini volcanic chambers, respectively, and processed as discussed in Supplementary Methods. Assembly and annotation yielded 2 928 214 and 6 508 507 contigs totalling 1445 and 3305 Gb of sequence encoding 3 874 122 and 8 540 156 coding sequences (CDS) for white mat and red mat, respectively (Table S1).

Community composition and binning

The taxonomic composition of assembled metagenomic sequences was assessed by phylogenetic analysis of 16S rDNA sequences and protein-coding marker genes (Supplementary Methods). Five hundred forty-two contigs containing 16S rRNA genes (Fig. S1) and 427 contigs containing phylogenetic marker proteins (Table S2) were used. 16S rRNA gene analysis of the two samples identified representatives of 58 archaeal and bacterial phyla and candidate divisions, whereas 30 phyla were detectable by protein phylogeny (Table S3). The results of phylogenetic analysis based on 16S rRNA genes and protein markers (Figs 2 and 3, Datasets S1, S2, S3, Table S3) were largely concordant and indicated that both samples were dominated by representatives of four bacterial phyla, Proteobacteria, Bacteroidetes, Chloroflexi and Planctomycetes. A clear difference in community structure between the two mats could be seen in the dominant taxa. The most abundant lineages in the white mat were Proteobacteria, including methylotrophic γ -proteobacteria from Methylococcales order and ϵ -proteobacteria from Campylobacteriales order (Figs 2 and 3, Datasets S1, S2, S3, Table S3). Both lineages were nearly undetectable in the red mat. Among the most abundant individual ϵ -proteobacterial taxa in the white mat were some typical inhabitants of hydrothermal vents, such as the nitrate-reducing chemoautotroph *Nitratiruptor*, the sulfur-oxidizing chemoautotroph *Sulfurimonas* and the CO-oxidizing heterotroph *Sulfurospirillum* (Nakagawa *et al.*, 2005; Callac *et al.*, 2015) (Datasets S2, S3). In contrast to well-characterized ϵ -proteobacterial lineages, little is known about thermophilic γ -proteobacterial methylotrophs living in hydrothermal vent systems (Hirayama *et al.*, 2014). The functional implications of

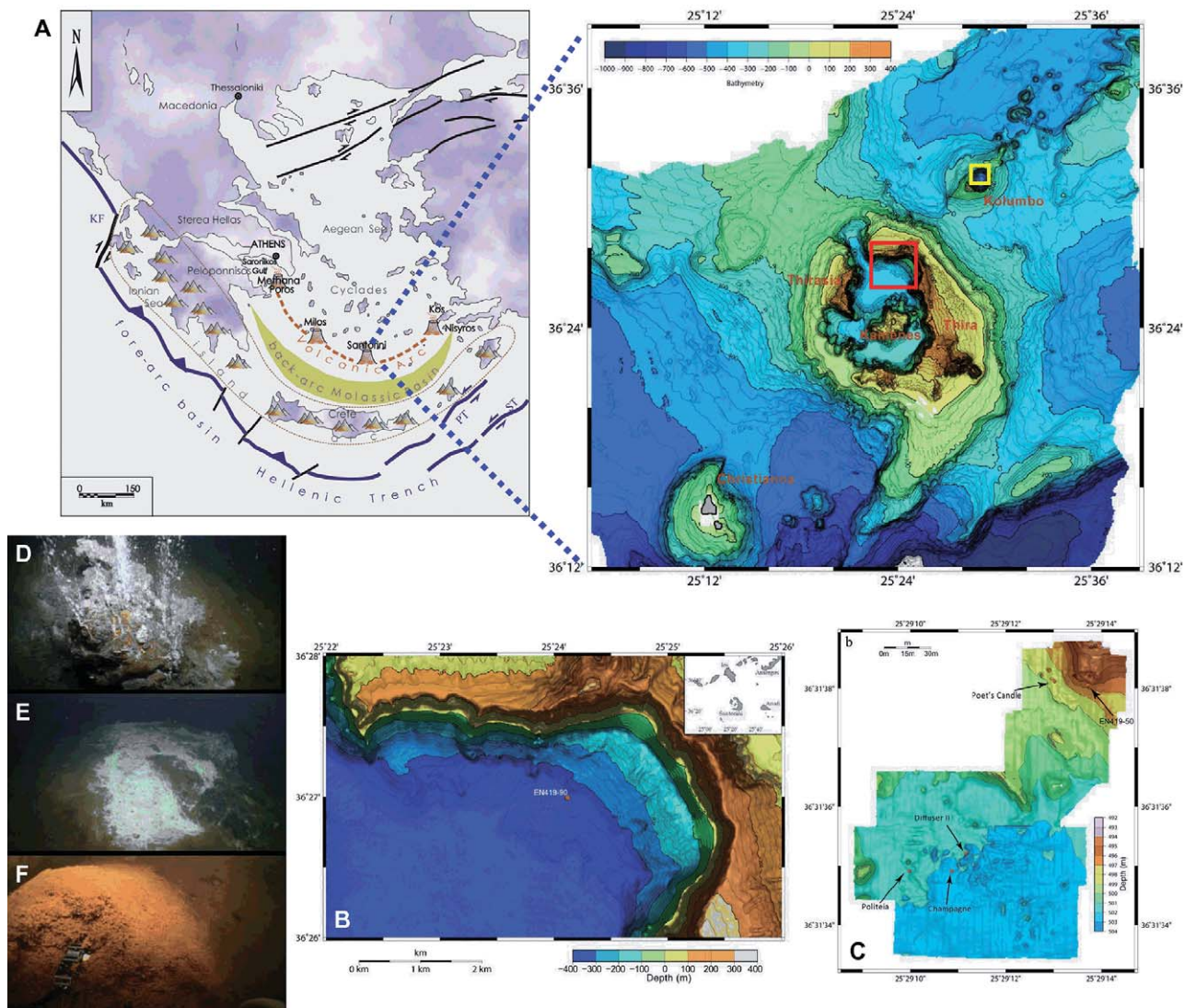


Fig. 1. Maps and bottom photographs of the sampling locations.

A. Swath bathymetry map of Santorini-Kolumbo volcanic field.

B. Detailed bathymetry map of the north part of Santorini crater (red square in a). The location of the red mat sample as encoded as (EN419-90) is indicated by a red dot.

C. Detailed bathymetry map of Kolumbo hydrothermal vent field located in the north part of the crater floor (yellow square in a). The location of the most active vents (i.e. Champagne, Diffuser II), the less active vents (i.e. Politeia, Poet's Candle) and the white mat sample also encoded as (EN419-50) are indicated by red dots. (D, E, F) Bottom photographs of the active Champagne vent.

their presence are discussed below. As expected, other thermophilic lineages, such as *Thermotoga*, *Caldiserica* and *Acetothermia* (formerly OP1) were found only in the white mat living at 60–70°C.

The red mat was dominated by Bacteroidetes (order Bacteroidales), Marinimicrobia (formerly candidate division SAR406) and Chloroflexi (Figs 2 and 3, Datasets S1, S2, S3, Table S3). Although the representatives of Chloroflexi have been recovered from several environments including hydrothermal vent sediments and hot springs (Yamada *et al.*, 2006; Kawaichi *et al.*, 2013), this phylum was largely absent from the white mat.

Several bacterial lineages were detected in both samples, albeit at different abundances. These include iron-oxidizing ζ -proteobacteria, sulfate-reducing δ -proteobacteria of Desulfobacterales order and representatives of phyla Planctomycetes, Verrucomicrobia, Ignavibacteria, Bacteroidetes (order Flavobacteriales), as well as candidate phylum Parcubacteria (previously known as OD1). The representatives of these shared lineages from both the white and red mats were quite divergent (Figs 2 and 3, Datasets S1, S2, S3, Table S3), underscoring the difference in community composition between the two samples.

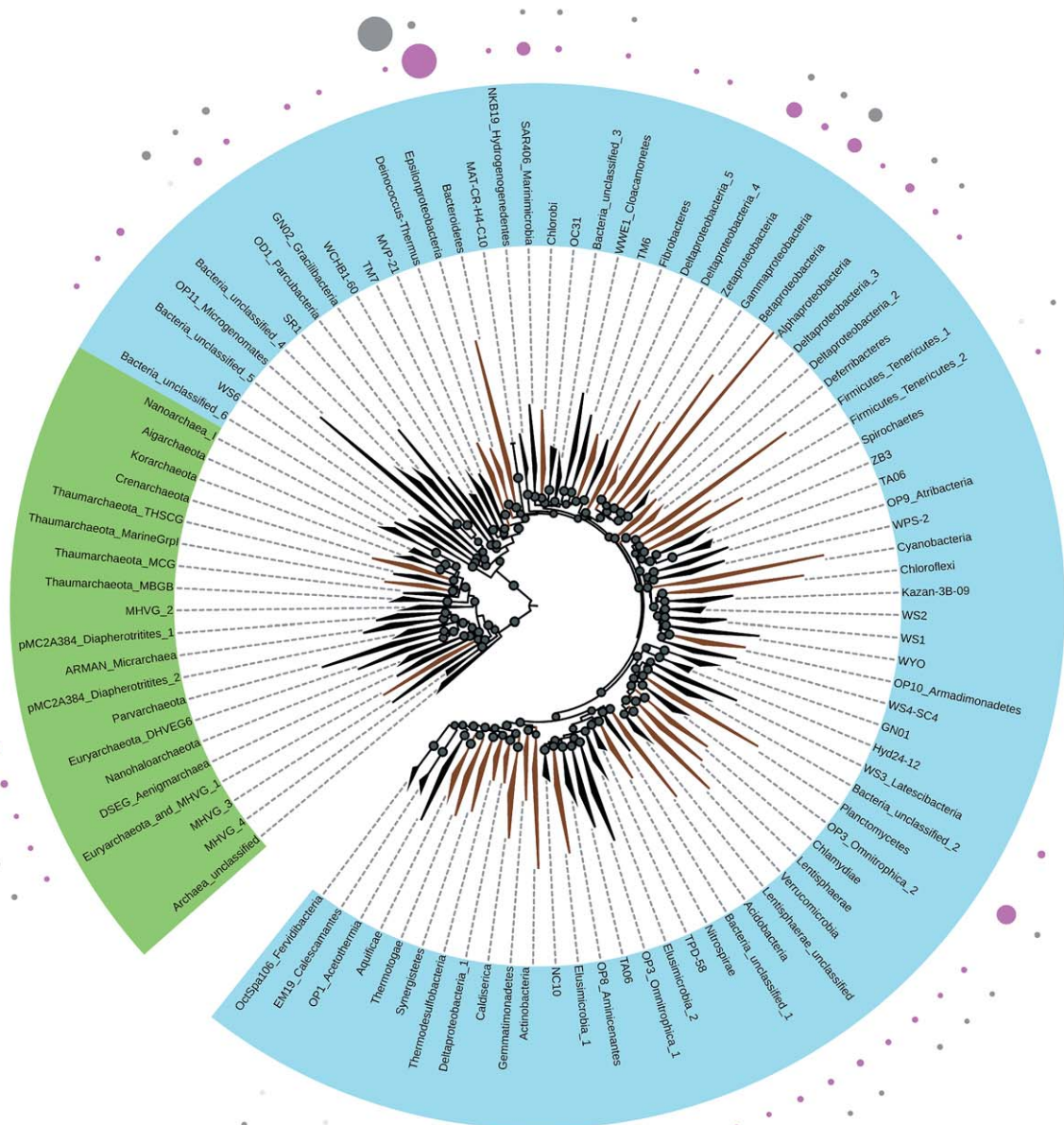


Fig. 2. Phylogenetic tree of 16S assemblies and protein bins from red and white mats. Phyla, phylum-level candidate divisions, and major subdivisions of large phyla form the leaves of this 'near maximum likelihood' tree of 16S rRNA assemblies from each mat. Nodes with >80% bootstrap are indicated by a small black circle. Archaeal phyla are coloured green; Bacterial phyla are blue. Branches coloured in brown mean that the associated taxon has at least one cultured representative. The four concentric rings of coloured circles on the outside of the tree indicate the normalized number of 16S assemblies and protein bins recovered for the adjacent taxon. Pink, red mat 16S assemblies; light grey, white/grey mat 16S assemblies; dark red, red mat protein bins; dark grey, white/grey mat protein bins.

Archaeal lineages were not abundant in either sample (Figs 2 and 3). Among the Archaea, 16S rRNA genes from thermophilic Euryarchaea (mainly Thermoplasmata and Thermococci) were identified only in the white mat, which is consistent with the higher temperatures reported for this sample. Sequences similar to the ammonia-oxidizing *Nitrosopumilus* (Thaumarchaeota

Marine Group I) were detected in both mats but at low abundances. Contrarily, a high proportion of sequences akin to *Nitrosopumilus* has been previously reported in polymetallic spires located in the same area (Kiliyas *et al.*, 2013). This can be attributed to differences in sample type (i.e. loosely floating seafloor white mat versus firmly attached mat on gas chimney surface) and

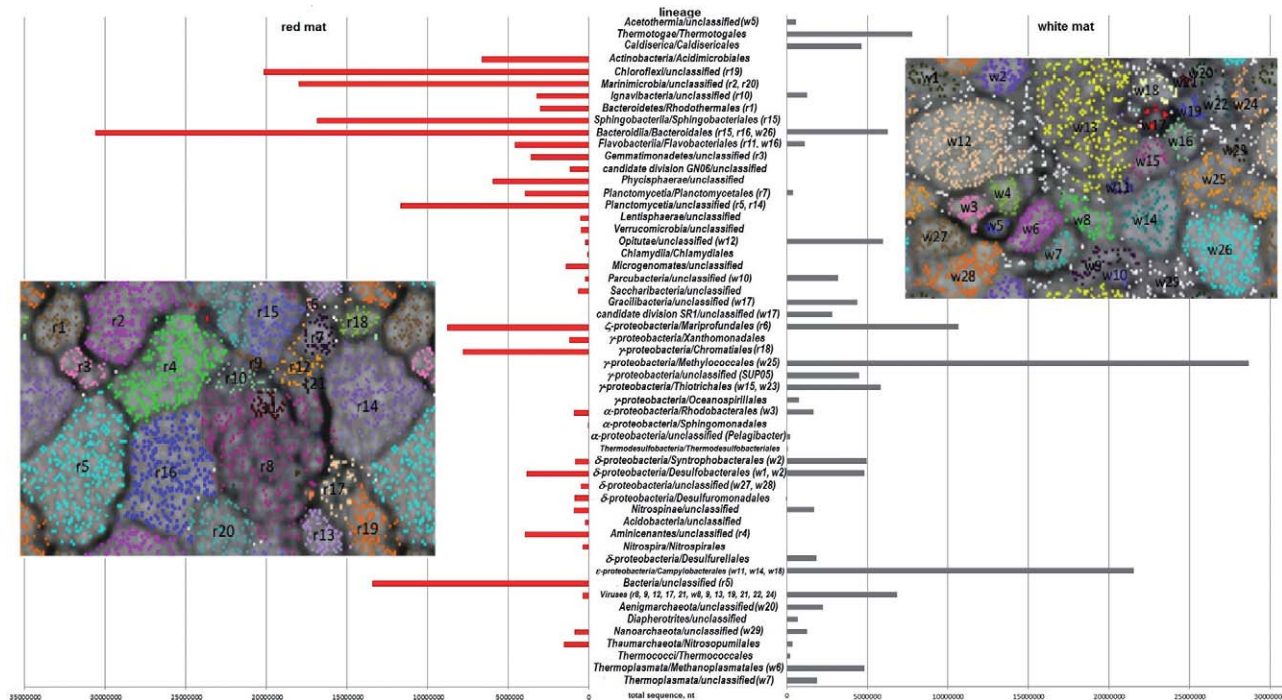


Fig. 3. ESOM of genomic sequence fragments based on tetranucleotide frequency of the (A) red mat and (B) white mat scaffold dataset. In both panels results are shown as ESOM generated visualizations, as well as bar charts (red – red mat and grey – white mat) showing the total number of mat specific sequences corresponding to a specific lineage identified based on phylogenetic analysis of 16S rRNA gene sequences or concatenated protein markers. For ESOM plots each point represents a 10 000 bp segment derived from scaffolds longer than 50 Kbp and 20 Kbp for red mat and white mat respectively. Data points are coloured based on their scaffold's best hit from all genomes in the IMG nucleotide database (white data points have no significant hit). In (A) clusters are: r1. Bacteroidetes (Rhodothermus); r2. SAR 406 (40×_48 GC); r3. Gemmatimonas; r4. OP8; r5. Planctomycetes (30×); r6. ζ -proteobacteria (Mariprofundus); r7. Planctomycetes (30×_40 GC); r8. Bacteroidetes/phages; r9. Bacteroidetes/phages; r10. Chlorobi (Ignavibacteria); r11. Bacteroidetes (Flavobacteria); r12. Phages; r13. γ -proteobacteria (Halorhodospira); r14. Planctomycetes (20×); r15. Bacteroidetes (150×); r16. Bacteroidetes (20×); r17. Phages; r18. *rm_chromatiales* r19. Chloroflexi (100×); r20. SAR 406 (40×_50×); r21. Phage. For (B) clusters are: w1. Desulfobacteriaceae; w2. δ -proteobacteria (NaphS2 purple; deep; orange); w3. Rhodobacteriaceae; w4. Planctomycetes; w5. OP1; w6. Euryarchaeota (Methanomicrobia); w7. Euryarchaeota; w8. Phages; w9. Phages; w10. OD1; w11. Sulfurovum; w12. Coraliomargarita (); w13. Phages; w14. Sulfurospirillum (Campylobacter); w15. Francisella; w16. Flavobacteria; w17. GN02; w18. ϵ -proteobacteria (candidate JS1); w19. Proteobacteria/Phage; w20. Euryarchaeota; w21. Phages; w22. Phages; w23. Methylophaga; w24. Phages; w25. Methylococcaceae (38_GC, high_GC); w26. Bacteroidetes vc21; w27. Desulfobacteriaceae; w28. δ -proteobacteria; w29. Nano (low GC). Note that the map is displayed as 'tiles' (from top to bottom and side to side).

adopted sequencing methods (i.e. PCR amplicon versus metagenomic).

The contigs from most abundant and well assembled lineages were used as seeds to recruit additional sequences by analysis of their tetranucleotide composition using ESOM (Ultsch and Moerchen, 2005). A depiction of these results can be viewed in Fig. 3 and Dataset S3 while general binning statistics for white mat and red mat bins is provided in Table S4. Sizable bins were generated for several candidate lineages with no cultured representatives [e.g., Marinimicrobia, Aminicenantes (formerly OP8) and Parcubacteria (formerly OD1)], and it was possible to assign novel physiology to some of these, as discussed below.

Finally, the viral fraction of both metagenomes was investigated, and 247 and 729 viral/phage contigs

comprising 7.5 Mb and 8 Mb of sequence were identified in the red and white mat, respectively (Datasets S4, S5). Viruses from the three families of the order Caudovirales (Myoviridae, Siphoviridae and Podoviridae) as well as eukaryotic viruses from the families Baculoviridae and Mimiviridae were detected in both mats according to the IMG lineage assignment that uses contig gene composition (Dataset S6). Additionally, contigs of the Marseillevirus family were detected only in the red mat and contigs of the Herpesviridae (macaviruses) family were found exclusively in the white mat (Fig. S2, Dataset S6). Connectivity between host and viral contigs could be established in the white mat through BLASTn searches of the CRISPR-Cas system spacers identified (when both contigs were long enough to be assigned taxonomically; Dataset S5, Supplementary Methods).

Functional composition

Functional properties of the white mat and red mat communities were investigated and contrasted using KEGG Orthology (KO) or Pfam assignments. Significant differences in abundance were observed between the two samples for broad categories like cell motility and chemotaxis, replication and repair, signal transduction and more (data not shown). More specifically, the white mat sample displayed higher abundance of genes associated with mismatch repair and homologous recombination, which is consistent with previous findings (Xie *et al.*, 2011) showing that microbial communities inhabiting extreme environments encode extensive DNA repair mechanisms to cope with environmental stress like high temperature, or high concentrations of H₂S and other sulfur compounds, trace metals, etc. (Pruski and Dixon, 2003). The white mat is also characterized by greater abundance of genes associated with signal transduction (two-component systems), chemotaxis and motility, akin to previous findings for hydrothermal vent systems (Xie *et al.*, 2011). The relative abundance of these functions in the white mat could be the result of adaptation to steep chemical and thermal gradients due to diffuse venting from the Kolumbo volcano chimneys.

As submarine hydrothermal vents are typically dominated by chemolithoautotrophic communities that fix inorganic carbon using energy derived from the oxidation of reduced inorganic compounds that are plentiful in hydrothermal vent fluids (e.g. reduced sulfur compounds, hydrogen, methane) (Corliss *et al.*, 1979; Alain *et al.*, 2002), we examined these processes by analyzing **marker genes** for key enzymatic steps for CO₂ fixation, sulfur-, nitrogen- and methane metabolism. Where possible, metabolic processes were attributed to specific taxa or lineages in the red mat or white mat samples, based on detection of marker genes in Emergent Self-Organizing Maps (ESOM)-delineated bins. **Furthermore, our datasets were contrasted with publicly available metagenomic datasets from related environments – such as other hydrothermal vents, hydrothermal vent plumes or deep subsurface samples (Table S5). Normalization was performed as described in Materials and Methods.**

Carbon source

Autotrophic carbon fixation. Carbon fixation is a keystone biosynthetic process used by hydrothermal vent microbes to assimilate inorganic carbon (*i.e.* carbon dioxide) into biomass. In this study, we analyzed all six pathways known to perform CO₂ fixation (Cody *et al.*, 2001; Berg, 2011). Of these, evidence for the four major pathways were detected in both samples: (i) reductive tricarboxylic acid (rTCA) cycle, (ii) reductive acetyl coenzyme A (CoA) pathway (or Wood–Ljungdahl pathway), (iii) reductive

pentose phosphate or Calvin–Benson–Bassham (CBB) cycle, and (iv) 3-hydroxypropionate/malyl-coA bicycle. Two additional pathways, which are only described in hyperthermophilic archaea to date, were not identified (Huber *et al.*, 2008; Berg *et al.*, 2010). These six pathways differ in terms of their energy and reducing equivalent requirements, oxygen sensitivity and more (Berg, 2011).

The rTCA cycle is known to play an important role in CO₂ fixation in hydrothermal vent environments (Campbell and Cary, 2004; Hugler *et al.*, 2005; Nakagawa and Takai, 2008). Markers of the rTCA cycle [ATP citrate lyase (*acl*) (EC:2.3.3.8) and pyruvate:ferredoxin oxidoreductase (EC:1.2.7.1)] were detected in both metagenome samples but overrepresented in the white mat (Fig. 4A, Fig. S3b). This is a likely reflection of the dominant sulfur- and hydrogen oxidizing ϵ -proteobacterial members in the white mat, where diffuse venting results in > 99 weight % CO₂ (Fig. 4A). Hydrothermal vent ϵ -proteobacteria are known to employ the rTCA cycle for CO₂-fixation (Hugler *et al.*, 2005). Correspondingly, a putative ATP-citrate lyase gene was detected in a white mat bin for an unknown ϵ -proteobacteria from the order Campylobacteriales. The rTCA cycle requires significantly less ATP and reducing equivalent compared with other cycles, which is of potential advantage in an energy limiting environment.

Markers of the reductive acetyl-CoA (Wood–Ljungdahl) pathway were also detected in both samples including carbon monoxide dehydrogenase (CODH – EC:1.2.99.2) and acetyl-CoA synthase [EC:2.3.1.169] (*acsB*). These markers were found in higher proportions (>2 fold) in the red mat sample compared with the white mat (Fig. 4A, Fig. S3c). Carbon monoxide dehydrogenase is the key enzyme of the reductive acetyl-CoA pathway, which may couple CO oxidation to sulfate reduction, or reduce CO₂ to either acetate (acetogenesis) or methane (methanogenesis) (Wu *et al.*, 2005; King and Weber, 2007). This pathway is previously reported in anaerobic ammonia-oxidizing Planctomycetes, sulfate-reducing bacteria and in autotrophic Archaeoglobales (Berg, 2011). In this study, markers of Wood–Ljungdahl pathway were identified in red mat bin for candidate phylum OP8 (Aminicenantes), as well as in a white mat bin for unclassified δ -proteobacteria (NaphS2-like) (Fig. 4B). Due to high demand for metals, coenzymes (tetrahydropterin and cobalamin) and strict anaerobiosis requirements by key enzymes (CO dehydrogenase/acetyl-CoA synthase), occurrence of this pathway is restricted to specific niches despite its high energetic efficiency (Berg, 2011). In the expanded comparison, this pathway is clearly absent from all the plume samples (likely due to oxygen sensitivity) and may be a dominant autotrophic CO₂ fixation pathway in the subsurface samples (Fig. S4). The reductive

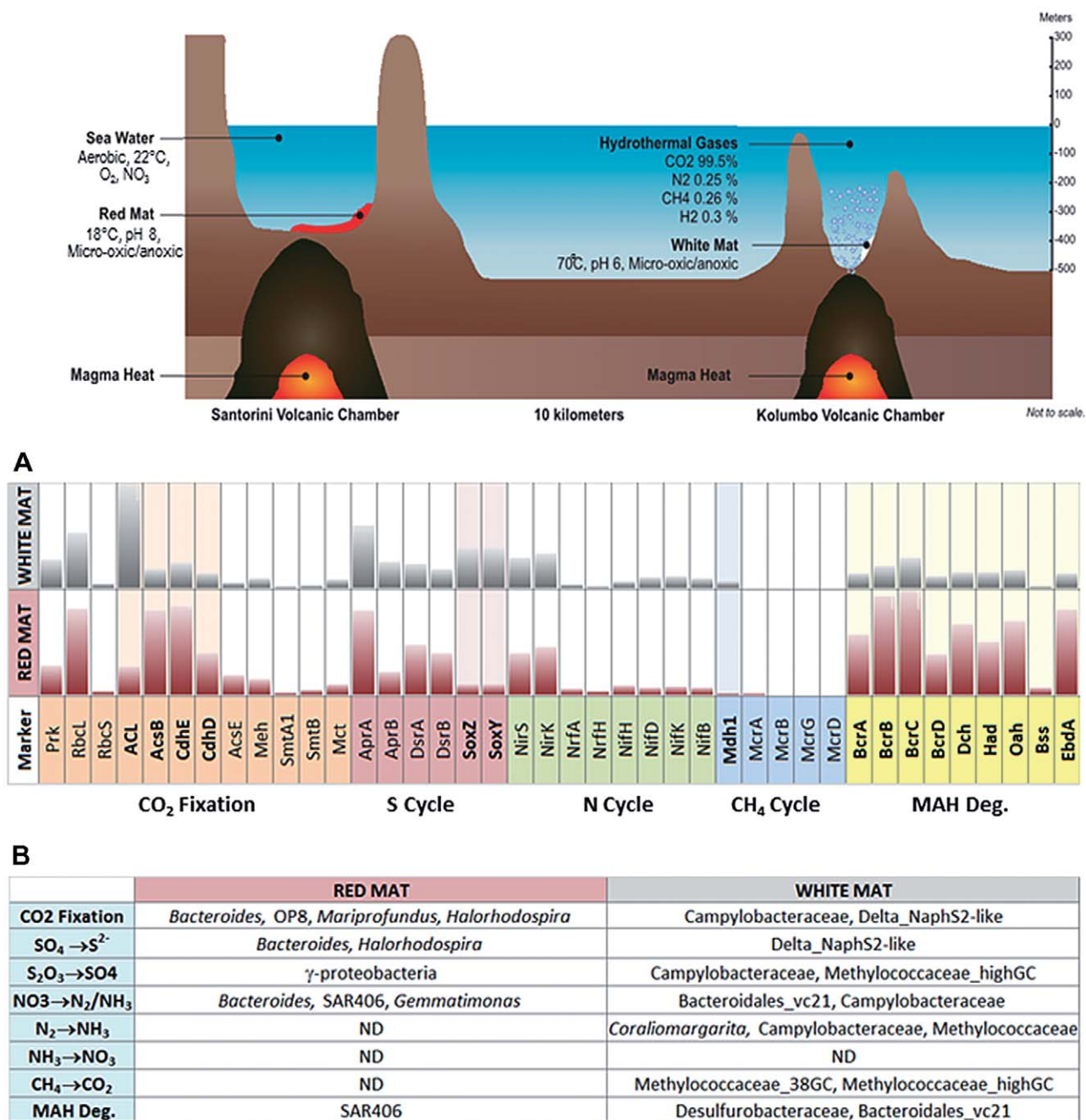


Fig. 4. Schematic depiction of the two volcanic chambers located in the Hellenic Volcanic arc and associated metadata.

A. Contrasting normalized counts of key biomarkers in the red mat and white mat sample for dominant processes examined in this study.

MAH deg. refers to anaerobic degradation of monocyclic aromatic hydrocarbons (B). Who does what – assignment of metabolic role to major taxonomic groups based on detection of biomarkers in ESOM bins.

acetyl-CoA pathway may also serve in energy conservation by generating an electrochemical gradient (Berg, 2011).

Although CBB pathway was not the most prevalent mode of CO₂ fixation in the studied environment, marker genes of this pathway, such as ribulose-bisphosphate carboxylase (RubisCO – EC:4.1.1.39) and phospho-

ribulokinase (PRK – EC:2.7.1.19), were detected in both samples (Fig. S3a). Some of these markers were captured in ζ- (*Mariprofundus*-like) and γ-proteobacterial (*Halorhodospira*-like) bins from red mat (Fig. 4), although these taxonomic assignments are tentative as they are not colocalized with cognate genes. Details of specific RubisCo genes examined using high scoring BLASTp hits

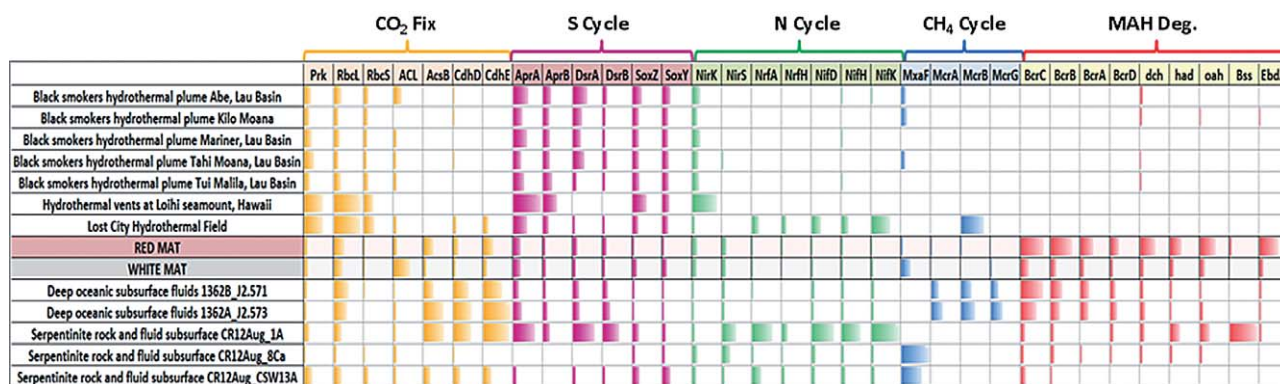


Fig. 5. Contrasting normalized counts of key biomarkers in HVA samples with 12 additional samples from related environments. Metabolic markers are colour-coded according to their overarching role category: Orange: carbon fixation; pink: sulfur metabolism; green: nitrogen metabolism; blue: methane metabolism; red: anaerobic degradation of monoaromatic hydrocarbons (MAH).

can be found in Supplementary Materials. The CBB cycle is known to be ubiquitous and energetically expensive with an upper temperature limit of ~70 to 75°C due to heat instability of intermediate products. Comparing the expanded set of metagenomic samples, the CBB cycle is seen as the predominant autotrophic CO₂ fixation pathway in the vent samples (Spillway iron-oxidizing mat and Lost City) and the plume samples (Fig. 5, Fig. S4), but less-so in the subsurface samples. The occurrence of this pathway in the higher temperature Lost City hydrothermal vent samples (with a reported temperature of ~90°C) is a curiosity; however, unknown specifics about the sample site might provide a suitable explanation.

Some relatively minor evidence of the 3-hydroxypropionate bicycle [genes involved include mesaconyl-C4 CoA hydratase, succinyl-CoA:(S)-malate CoA-transferase (EC:2.8.3.22), malyl-CoA lyase (EC:4.1.3.24)] was seen in the red and white mat samples, although not captured in any bins.

Methanotrophy. In addition to autotrophy discussed above, evidence of methanotrophy (methane as sole carbon and energy source) is presented by functional marker data (Fig. 4B), primarily in the white mat sample, and possibly restricted to a few γ -proteobacterial Methylococcaceae lineages. Specifically, the marker gene, methanol dehydrogenase (Mdh1) (Lau *et al.*, 2013), was found predominantly in the Methylococcaceae bins from the white mat (Fig. 4B). Correspondingly, a small amount of methane (0.26%) was detected in gas samples from the Kolumbo hydrothermal vents (Carey *et al.*, 2013). Although γ -proteobacterial methanotrophs are known to be aerobic and the sampled environment is primarily anaerobic, anaerobic oxidation of methane may be proposed using oxygen produced from the dismutation of nitric oxide into nitrogen and oxygen gas during denitrification, as previously described for

Methylomirabilis oxyfera (Ettwig *et al.*, 2010). Evidence for another type of anaerobic methanotrophy (using reverse methanogenesis) has been found in the Guaymas Basin hydrothermal vents as represented by two archaeal methane oxidizers (ANME-1 and ANME-2) (Teske *et al.*, 2003). However, 16S rRNA gene data do not support their presence in the Kolumbo/Santorini samples. Marker genes for methanogenesis (McrABG) were negligible (concurring with absence of acetoclastic and hydrogenotrophic methanogens in either sample) as methanogens are known to be inhibited by sulfate (Winfrey and Zeikus, 1977).

Anaerobic degradation of aromatic hydrocarbons. Unexpectedly, both samples showed evidence for chemoheterotrophy based on presence of markers for anaerobic degradation of low molecular weight aromatic hydrocarbons. These include enzymes for anaerobic transformation of aromatic compounds by radical addition of fumarate to methyl or ethyl groups on the aromatic ring (e.g. benzylsuccinate synthase), aromatic hydroxylation (e.g. ethylbenzene dehydrogenase) and conversion of the central intermediate, benzoyl-coenzyme A, to acetyl-CoA, by reductive ring cleavage (e.g. benzoyl-CoA reductase) (Heider, 2007; An *et al.*, 2013a,b). Examining their occurrence in all available metagenomes within the IMG/M database (Markowitz *et al.*, 2014), these anaerobic hydrocarbon degradation marker genes were detected primarily in oil- or coal-impacted environments, natural oil seeps, peatlands and contaminated sites. Besides these highly impacted areas, some marker genes were also detected in the deep marine and terrestrial subsurface samples in our subset, but not in the other hydrothermal vent or plume environments (Fig. 5, Fig. S5). Interestingly, the occurrence of these markers in the red mat was four times higher compared with the white mat (Fig. 4, Fig. S5).

A complete operon for the anaerobic degradation of benzoyl-CoA was identified in the red mat and attributed to the candidate phylum SAR406 [Marinimicrobia or Marine Group A; (Rinke *et al.*, 2013)] (Fig. S6), for which little information is available. SAR406 representatives are prevalent in oxygen-deficient marine systems (e.g. minimum zones, anoxic basins) (Gordon and Giovannoni, 1996; Allers *et al.*, 2013), and they were recently proposed to be involved in the marine sulfur cycle (Wright *et al.*, 2014). Our findings propose a previously undescribed ecological role for members of this candidate phylum. Although no markers were identified, the δ -proteobacterial lineage NaphS2 detected in the white mat sample (Dataset S3) is a known anaerobic degrader of naphthalene (DiDonato *et al.*, 2010).

Additional complementary evidence supporting the presence of anaerobic hydrocarbon degradation markers comes from the observation of shared CRISPR-Cas system spacers between the HVA samples and engineered environments. Specifically, we have observed that a significant percentage of the spacer hits from both red and white mat (6.1 % and 11.8 % respectively) were against hydrocarbon-impacted environments including oil seeps, subsurface oil reservoirs and anthropogenically contaminated sites (Dataset S7). At this stage no phage genomes were reconstructed and due to the inability of CRISPRs to assemble well, we are forced to work with tiny scaffolds with limited taxonomic information. Thus, we were unable to clearly define the source of the spacers identified in this study.

Given these observations, discovering a potential source of these compounds in the Santorini-Kolumbo volcanic field is important. In a previous study at a Guaymas Basin hydrothermal vent, seabed hydrocarbons were attributed to the hydrothermal alteration of organic sediment originating from highly productive surface waters [i.e. organic matter is pyrolyzed to petroleum-like aliphatic and aromatic hydrocarbons, short-chain fatty acids, ammonia, and methane; (Teske *et al.*, 2014)]. Other studies suggest an abiotic origin of hydrocarbons from thermogenesis occurring deep in the earth's crust (using CO₂ and mantle carbon) that could subsequently be dissolved and vented out with hydrothermal fluids (Proskurowski *et al.*, 2008; Konn, 2009; Konn *et al.*, 2011; Li *et al.*, 2012).

Both of these theories could explain the source of monoaromatics in the Santorini hydrothermal area. However, our post-hoc attempts to detect alkylated monoaromatic hydrocarbons in mat samples after 6 years of cold storage were unsuccessful, though the absence of such volatile compounds could be due to their evaporation losses during sampling and/or storage. It should be also stressed that a source of monoaromatics cannot be attributed to anthropogenic pollution as this should also

result in high levels of polycyclic aromatic hydrocarbons (PAHs). Our measurements in the red mat samples revealed total PAHs concentrations (sum of 16 PAH members) to be lower than 100 ng g⁻¹ dry weight (unpublished results), suggesting that the impact of pollution was rather limited. Clearly, a comprehensive, accurate sampling/analysis of monoaromatic compounds and their metabolites, along with transcriptomic characterization of the marker genes would be the most judicious approach to assess whether their presence and activities are germane in Santorini-Kolumbo volcanic field.

Energy source

Sulfur cycle. Sulfur compound oxidation and reduction are dominant strategies for energy production by hydrothermal vent microbes, and presumably hydrogen sulfide is a major electron donor used by resident chemolithotrophs. Active sulphide-sulphate structures have been mapped at the north part of the Kolumbo hydrothermal field where the streams of white mat are located (Kiliass *et al.*, 2013). Sulfur oxidation (Sox) genes indicative of neutrophilic sulfur oxidizing chemoautotrophs, capable of oxidizing various reduced sulfur compounds to sulfate (Kelly *et al.*, 1997; Friedrich *et al.*, 2001), were at least four times more prevalent in the ϵ -proteobacteria-dominated white mat (Fig. 4A, Fig. S7) – where they can likely make quick use of electron donor supplied by reduced vent fluids or from the chemical oxidation of metal sulfides (Schipper and Sand, 1999). Sox genes were identified in white mat bins for Campylobacteraceae and Methylococcaceae, as well as in a red mat bin for an unknown γ -proteobacterium (Fig. 4B, Supplementary Results and Conclusions, Fig. S7).

Marker genes for the dissimilatory anaerobic sulfate reduction to sulfide were found in similar proportions in both samples. These genes were captured in red mat bins for Bacteroidetes, Halorhodospira and in a white mat bin for delta-NaphS2-like (Fig. 4). Overall, the presence of marker genes for sulfur oxidation and dissimilatory sulfate reduction suggest the potential that such processes are occurring in all the metagenomic samples contrasted in our analyses (Fig. 5).

Nitrogen cycle. Previous reports from the Kolumbo crater hydrothermal vent field have shown that the nitrogen cycle plays a crucial role in the establishment of the geochemical environment of the field (Kiliass *et al.*, 2013). High ammonium and nitrate concentrations were previously detected in the Kolumbo hydrothermal vent field (Kiliass *et al.*, 2013), and have also been reported in the Guaymas Basin, the Okinawa Trough Backarc Basin and the Endeavor segment of the Juan de Fuca Ridge

(Ishibashi *et al.*, 1995; Mehta *et al.*, 2003; Bourbonnais *et al.*, 2012). These inorganic nitrogen compounds are important terminal electron acceptors for chemolithoautotrophic growth in hydrothermal vents (Ishibashi *et al.*, 1995; Mehta *et al.*, 2003; Nakagawa and Takai, 2008). The rapid consumption of ammonium by chemoautotrophs has been demonstrated in the hydrothermal vent plumes of the Juan de Fuca Ridge (Lam *et al.*, 2004) as well as in Kolumbo geothermal gradient (Kiliyas *et al.*, 2013). Evidence of nitrification, denitrification and nitrogen fixation in hydrothermal vents systems have all been previously reported (Mehta *et al.*, 2003; Bourbonnais *et al.*, 2012; Voss *et al.*, 2013).

Both red and white mat samples are replete with genes required for denitrification (Figs 4, 5, S8, S9), suggesting that nitrate could be an important electron acceptor, possibly coupled to sulfur-oxidation, as a significant energy-generating pathway for chemolithoautotrophic growth in these environments. The dominant fate of nitrate in the red and white mat is likely denitrification (based on proportions of the NirKS markers), although some minimal evidence of dissimilatory nitrate reduction to ammonium (DNRA) (NrfAH markers) is also captured in both mat samples (Fig. 4, Fig. S9). NirKS genes were found in bins for Bacteroidales, Gemmatimonadetes and SAR406 in the red mat, and Bacteroidales, Campylobacteriales and SUP05-like γ -proteobacteria were found in the white mat (Figs 4B, S9) (see Supplementary Results and Conclusions for gene details).

A small number of genes encoding putative ammonia monooxygenases (amoABC) for nitrification were detected in both red and white samples, although unassigned to bins. Nonetheless, *Nitrosopumilus* spp. (Thaumarchaeota) detected in both samples (Fig. 4) is a known chemoautotrophic ammonia oxidizer (Martens-Habbena *et al.*, 2009; Walker *et al.*, 2010). Specifically, amoA sequences showed high identity to those from *N. maritimus* SCM1, Candidatus *Nitrosopumilus koreensis* AR1 and Candidatus *Nitrosopumilus* spp. AR2.

Other electron donors. In addition to the oxidation of reduced sulfur and nitrogen compounds, there is evidence based on phylogenetic marker-based identification of taxa known to be involved in hydrogen oxidation (by way of Ni-Fe hydrogenases), iron oxidation (protein marker-based evidence of potentially iron-oxidizing members within Mariprofundales and *Nitrospira* spp.; Fig. 3) and potential manganese or other donor oxidation (by multicopper oxidases) in both red and white mat samples (data not shown). Notably, potential iron respiration genes were found in Aminicenantes (OP8) bin in the red mat (Supplementary Results and Conclusions).

Therefore, a large and diverse species of electron donors is likely in these samples.

Overall, this investigation provides new insights into the spectrum of possible physiological roles of hydrothermal vent microorganisms. Evidence from the analysis of our HV samples reveals that autotrophic carbon fixation is mainly a product of the rTCA and CBB cycle, which appeared to be primarily driven by the oxidation of reduced sulfur compounds through SOX-dependent pathway, using either oxygen or nitrate (derived from ammonia oxidation) as terminal electron acceptors. Thus, the availability of reduced sulfur compounds, as well as oxygen and nitrate, appear to be key parameters structuring the microbial community as reflected by the community composition of the 16S and binning results presented here. On the basis of the genomic organization of the key genes of the carbon fixation and sulfur oxidation and ammonia oxidation pathways contained in the metagenomic sequences, both obligate and facultative autotrophs appear to be present and contributing to biomass production. Moreover, the observation of a complete operon for the anaerobic degradation of benzoyl-CoA in the red mat, and attributed to the candidate phylum SAR406 not only proposes a previously undescribed ecological role for members of this candidate phylum, but also opens up new avenues of exploration into the origin of aromatic hydrocarbons in the HVA ecosystem and their role in shaping microbial community structure and function.

Experimental procedures

Samples collection

Microbial mat samples were collected during the sampling expedition EN419 undertaken in May–June 2006 in Santorini volcanic complex by the R/V Endeavor of the Rhode Island University in collaboration with the Hellenic Centre for Marine Research. The white mat sample (36.5273°N, 25.4872°E, depth: 494 m) was collected from the Kolumbo crater, whereas the red mat sample (36.4499°N, 25.4021°E, depth: 336 m) was retrieved from the Santorini caldera (see Supplementary Information 1.1, 1.2).

DNA isolation and library construction

Approximately 50 μ g of high molecular weight DNA (>23 kb) was isolated from each microbial mat following the detailed protocol described by Brady (Brady, 2007). DNA samples were shipped on dry ice to Joint Genome Institute, Department of Energy, USA, for library construction (see Supplementary Information 1.3).

Sequencing, assembly and annotation

Sequencing was performed using two platforms, Roche 454 and Illumina (GAIIIX generating 114 bp reads and HiSeq2000

generating 151 bp reads). Roche 454 pyrosequencing generated 1 429 091 reads (507 Mb) for the red mat and 299 975 reads (114 Mb) for the white/grey mat. The Illumina platform produced 45 337 178 GAIIIX reads and 304 325 806 HiSeq2000 reads for the red mat with a total of 51 Gb. For the white mat 49 227 146 GAIIIX reads and 304 325 806 HiSeq2000 reads totaling 53 Gb were generated.

The assembly of the 454 data was performed with Newbler v2.4, whereas the Illumina reads were assembled with IDBA_UD v1.1.1 (Peng *et al.*, 2012). Contigs from both assemblies were merged together using MeGAMerge (Scholz *et al.*, 2014) (see Supplementary Information 1.4). Assembled sequences were annotated using the IMG/M metagenome analysis system (Markowitz *et al.*, 2014) (see Supplementary Information 1.5 – Table S1). Both datasets are available in the IMG/M system (accessions 3300002231 and 3300002242 for red and white mat respectively).

Tree construction for 16S rRNA gene sequences found on contigs

Full-length 16S rRNA gene sequences (>1250 bp) were merged with the representative set of all 16S rRNA gene sequences used by Rinke and colleagues (2013) and treed using the SILVA-aligner and FastTree programs; shorter 16S rRNA gene sequences (>500 bp) were 'Quick-Added' using ARB (see Supplementary Information 1.6).

Tree construction for concatenated protein marker genes

A reference set of 56 single copy protein marker genes from 689 isolate and single cell genomes with known taxonomy were used to create Hidden Markov Models (HMMs) that enabled extraction, alignment and phylogenetic tree construction of these protein marker genes found on metagenomic contigs (see Supplementary Information 1.6).

Binning

We used Emergent Self Organizing Maps (ESOM) (Ultsch and Moerchen, 2005) based on tetranucleotide frequency as a binning method (see Supplementary Information 1.7).

Functional abundance comparisons

Gene counts for individual KEGG Orthology terms and Pfam domains for each metagenome were retrieved from the IMG/M v4.0 management system (Markowitz *et al.*, 2014). To enable comparison, these counts were normalized by dividing with the mean count of 10 universal (conserved in all three domains of life) ribosomal proteins (Table S2), as an estimate of the number of populations for each sample. We also performed normalization using individual ribosomal protein markers, as well as length-normalized counts (by dividing the sums of the lengths of the proteins by the expected length of the marker protein) to account for any possible inflation in counts due to fragmentation of ORFs. All approaches yielded identical results and inferences.

CRISPR-Cas system spacer information

We used a combination of three CRISPR-Cas spacer detection tools: piler-CR (Edgar, 2007); CRT (Bland *et al.*, 2007); and CRASS (Skenneron *et al.*, 2013). Both piler-CR and CRT tools were run against metagenomic assembled data, whereas the CRASS tool was run against unassembled reads. Using these tools, we generated a total spacer database of 44 402 and 252 238 spacers for the red mat and the white mat metagenomes respectively. After removing redundant spacers, we used BLASTn to map these spacers to the red and white mat samples as well as an additional 3695 metagenomes found in IMG/M (see Supplementary Information 1.8, 1.9).

Acknowledgements

The officers and the crew of the R/V Endeavor of the Rhode Island University are gratefully acknowledged for their important contribution to the field work during the Thera 2006 Expedition (http://oceanexplorer.noaa.gov/explorations/06blacksea/logs/summary_thera/summary_thera.html). This work was performed under the auspices of the Hellenic Initiative of MikroBioKosmos, the US Department of Energy's Office of Science, Biological and Environmental Research Program, the University of California, Lawrence Berkeley National Laboratory under contract No. DE-AC02-05CH11231, and the European program MARBIGIN FP7-REGPOT-2010-1 (grant to A.M.). DNA extraction was supported by the Hellenic Centre for Marine Research – Crete Department, Greece. We thank Dr. Susannah Tringe from the Department of Energy Joint Genome Institute for guidance and advice on analyses. We are also grateful to Matthew Schrenk from Michigan State University and Michael Rappe from Hawaii Institute of Marine Biology at University of Hawaii at Manoa for allowing us the use of their metagenomic data, as well as clarification on metadata for those samples.

The authors declare that no conflict of interest exists.

References

- Alain, K., Querellou, J., Lesongeur, F., Pignet, P., Crassous, P., Raguene, G., *et al.* (2002) *Caminibacter hydrogeniphilus* gen. nov., sp. nov., a novel thermophilic, hydrogen-oxidizing bacterium isolated from an East Pacific Rise hydrothermal vent. *Int J Syst Evol Microbiol* **52**: 1317–1323.
- Allers, E., Wright, J.J., Konwar, K.M., Howes, C.G., Beneze, E., Hallam, S.J., and Sullivan, M.B. (2013) Diversity and population structure of Marine Group A bacteria in the Northeast subarctic Pacific Ocean. *ISME J* **7**: 256–268.
- An, D., Brown, D., Chatterjee, I., Dong, X., Ramos-Padron, E., Wilson, S., *et al.* (2013a) Microbial community and potential functional gene diversity involved in anaerobic hydrocarbon degradation and methanogenesis in an oil sands tailings pond. *Genome* **56**: 612–618.
- An, D., Caffrey, S.M., Soh, J., Agrawal, A., Brown, D., Budwill, K., *et al.* (2013b) Metagenomics of hydrocarbon resource environments indicates aerobic taxa and genes to be unexpectedly common. *Environ Sci Technol* **47**: 10708–10717.

- Berg, I.A. (2011) Ecological aspects of the distribution of different autotrophic CO₂ fixation pathways. *Appl Environ Microbiol* **77**: 1925–1936.
- Berg, I.A., Ramos-Vera, W.H., Petri, A., Huber, H., and Fuchs, G. (2010) Study of the distribution of autotrophic CO₂ fixation cycles in Crenarchaeota. *Microbiology* **156**: 256–269.
- Bland, C., Ramsey, T.L., Sabree, F., Lowe, M., Brown, K., Kyrpides, N.C., and Hugenholtz, P. (2007) CRISPR recognition tool (CRT): A tool for automatic detection of clustered regularly interspaced palindromic repeats. *BMC Bioinformatics* **8**: 209.
- Bourbonnais, A., Lehmann, M.F., Butterfield, D.A., and Juniper, S.K. (2012) Subseafloor nitrogen transformations in diffuse hydrothermal vent fluids of the Juan de Fuca Ridge evidenced by the isotopic composition of nitrate and ammonium. *Geochem Geophys Geosy* **13**: Q02T01. doi:10.1029/2011gc003863.
- Brady, S.F. (2007) Construction of soil environmental DNA cosmid libraries and screening for clones that produce biologically active small molecules. *Nat Protoc* **2**: 1297–1305.
- Brazelton, W.J., and Baross, J.A. (2009) Abundant transposases encoded by the metagenome of a hydrothermal chimney biofilm. *ISME J* **3**: 1420–1424.
- Brock, T.D., and Freeze, H. (1969) *Thermus aquaticus* gen. n. and sp. n., a non-sporulating extreme thermophile. *J Bacteriol* **98**: 289–297.
- Byrne, N., Strous, M., Crepeau, V., Kartal, B., Birrien, J.L., Schmid, M., et al. (2009) Presence and activity of anaerobic ammonium-oxidizing bacteria at deep-sea hydrothermal vents. *ISME J* **3**: 117–123.
- Callac, N., Rouxel, O., Lesongeur, F., Liorzou, C., Bollinger, C., Pignet, P., et al. (2015) Biogeochemical insights into microbe-mineral-fluid interactions in hydrothermal chimneys using enrichment culture. *Extremophiles* **19**: 597–617.
- Campbell, B.J., and Cary, S.C. (2004) Abundance of reverse tricarboxylic acid cycle genes in free-living microorganisms at deep-sea hydrothermal vents. *Appl Environ Microbiol* **70**: 6282–6289.
- Cao, H., Wang, Y., Lee, O.O., Zeng, X., Shao, Z., and Qian, P.Y. (2014) Microbial sulfur cycle in two hydrothermal chimneys on the Southwest Indian Ridge. *MBio* **5**: e980–e913.
- Carey, S., Nomikou, P., Croff-Bell, K., Lilliey, M., Lupton, J., Roman, C., et al. (2013) CO₂ degassing from hydrothermal vents at Kolumbo submarine volcano, Greece, and the accumulation of acidic crater water. *Geology* **41**: 1035–1038.
- Cody, G.D., Boctor, N.Z., Hazen, R.M., Brandes, J.A., Morowitz, H.J., and Yoder, H.S.J. (2001) Geochemical roots of autotrophic carbon fixation: Hydrothermal experiments in the system citric acid, H₂O-(±FeS)-(±NiS). *Geochim Cosmochim Acta* **65**: 3357–3576.
- Corliss, J.B., Dymond, J., Gordon, L.I., Edmond, J.M., von Herzen, R.P., Ballard, R.D., et al. (1979) Submarine thermal springs on the Galapagos Rift. *Science* **203**: 1073–1083.
- Dando, P.R., Hughes, J.A., Leahy, Y., Niven, S.J., and Smith, C. (1995) Gas venting rates from submarine hydrothermal areas around the island of Milos, Hellenic Volcanic Arc. *Cont Shelf Res* **15**: 913–929.
- Dando, P.R., Aliani, S., Arab, H., Bianchi, C.N., Brehmer, M., Cocito, S., et al. (2000) Hydrothermal studies in the Aegean Sea. *Phys Chem Earth B* **25**: 1–8.
- DiDonato, R.J., Jr, Young, N.D., Butler, J.E., Chin, K.J., Hixson, K.K., Mouser, P., et al. (2010) Genome sequence of the deltaproteobacterial strain NaphS2 and analysis of differential gene expression during anaerobic growth on naphthalene. *PLoS ONE* **5**: e14072.
- Edgar, R.C. (2007) PILER-CR: Fast and accurate identification of CRISPR repeats. *BMC Bioinformatics* **8**: 18.
- Ettwig, K.F., Butler, M.K., Le Paslier, D., Pelletier, E., Mangenot, S., Kuypers, M.M., et al. (2010) Nitrite-driven anaerobic methane oxidation by oxygenic bacteria. *Nature* **464**: 543–548.
- Friedrich, C.G., Rother, D., Bardischewsky, F., Quentmeier, A., and Fischer, J. (2001) Oxidation of reduced inorganic sulfur compounds by bacteria: Emergence of a common mechanism? *Appl Environ Microbiol* **67**: 2873–2882.
- Gordon, D.A., and Giovannoni, S.J. (1996) Detection of stratified microbial populations related to *Chlorobium* and *Fibrobacter* species in the Atlantic and Pacific oceans. *Appl Environ Microbiol* **62**: 1171–1177.
- Hall-Spencer, J.M., Rodolfo-Metalpa, R., Martin, S., Ransome, E., Fine, M., Turner, S.M., et al. (2008) Volcanic carbon dioxide vents show ecosystem effects of ocean acidification. *Nature* **454**: 96–99.
- Heider, J. (2007) Adding handles to unhandy substrates: Anaerobic hydrocarbon activation mechanisms. *Curr Opin Chem Biol* **11**: 188–194.
- Hirayama, H., Abe, M., Miyazaki, M., Nunoura, T., Furushima, Y., Yamamoto, H., and Takai, K. (2014) *Methylomarinovum caldicuralii* gen. nov., sp. nov., a moderately thermophilic methanotroph isolated from a shallow submarine hydrothermal system, and proposal of the family Methylothermaceae fam. nov. *Int J Syst Evol Microbiol* **64**: 989–999.
- Huber, H., Gallenberger, M., Jahn, U., Eylert, E., Berg, I.A., Kockelkorn, D., et al. (2008) A dicarboxylate/4-hydroxybutyrate autotrophic carbon assimilation cycle in the hyperthermophilic Archaeon *Ignicoccus hospitalis*. *Proc Natl Acad Sci USA* **105**: 7851–7856.
- Hugler, M., Wirsen, C.O., Fuchs, G., Taylor, C.D., and Sievert, S.M. (2005) Evidence for autotrophic CO₂ fixation via the reductive tricarboxylic acid cycle by members of the epsilon subdivision of proteobacteria. *J Bacteriol* **187**: 3020–3027.
- Ishibashi, J., Sano, Y., Wakita, H., Gamo, T., Tsutsumi, M., and Sakai, H. (1995) Helium and carbon geochemistry of hydrothermal fluids from the Mid-Okinawa Trough Back Arc Basin, southwest of Japan. *Chem Geol* **123**: 1–15.
- Kawaichi, S., Ito, N., Kamikawa, R., Sugawara, T., Yoshida, T., and Sako, Y. (2013) *Ardenticatena maritima* gen. nov., sp. nov., a ferric iron- and nitrate-reducing bacterium of the phylum ‘Chloroflexi’ isolated from an iron-rich coastal hydrothermal field, and description of *Ardenticatena* classis nov. *Int J Syst Evol Microbiol* **63**: 2992–3002.
- Kelly, D.P., Shergill, J.K., Lu, W.P., and Wood, A.P. (1997) Oxidative metabolism of inorganic sulfur compounds by bacteria. *Antonie Van Leeuwenhoek* **71**: 95–107.

- Kilias, S.P., Nomikou, P., Papanikolaou, D., Polymenakou, P.N., Godelitsas, A., Argyraki, A., *et al.* (2013) New insights into hydrothermal vent processes in the unique shallow-submarine arc-volcano, Kolumbo (Santorini), Greece. *Sci Rep* **3**: 2421.
- King, G.M., and Weber, C.F. (2007) Distribution, diversity and ecology of aerobic CO-oxidizing bacteria. *Nat Rev Microbiol* **5**: 107–118.
- Konn, C. (2009) Hydrocarbons and oxidized organic compounds in hydrothermal fluids from Rainbow and Lost City ultramafic-hosted vents. *Chem Geol* **258**: 299–314.
- Konn, C., Testemale, D., Querellou, J., Holm, N.G., and Charlou, J.L. (2011) New insight into the contributions of thermogenic processes and biogenic sources to the generation of organic compounds in hydrothermal fluids. *Geobiology* **9**: 79–93.
- Kristall, B., Kelley, D.S., Hannington, M.D., and Delaney, J.R. (2006) Growth history of a diffusely venting sulfide structure from the Juan de Fuca Ridge: A petrological and geochemical study. *Geochem Geophys Geosy* **7**: Q07001. doi:10.1029/2005GC001166.
- Lam, P., Cowen, J.P., and Jones, R.D. (2004) Autotrophic ammonia oxidation in a deep-sea hydrothermal plume. *FEMS Microbiol Ecol* **47**: 191–206.
- Lau, E., Fisher, M.C., Steudler, P.A., and Cavanaugh, C.M. (2013) The methanol dehydrogenase gene, *mxhF*, as a functional and phylogenetic marker for proteobacterial methanotrophs in natural environments. *PLoS ONE* **8**: e56993.
- Li, J., Peng, X., Zhou, H., Li, J., Chen, S., Wu, Z., and Yao, H. (2012) Characteristics and source of polycyclic aromatic hydrocarbons in the surface hydrothermal sediments from two hydrothermal fields of the Central Indian and Mid-Atlantic Ridges. *Geochem J* **46**: 31–43.
- Markowitz, V.M., Chen, I.M., Chu, K., Szeto, E., Palaniappan, K., Pillay, M., *et al.* (2014) IMG/M 4 version of the integrated metagenome comparative analysis system. *Nucleic Acids Res* **42**: D568–D573.
- Martens-Habbena, W., Berube, P.M., Urakawa, H., de la Torre, J.R., and Stahl, D.A. (2009) Ammonia oxidation kinetics determine niche separation of nitrifying Archaea and Bacteria. *Nature* **461**: 976–979.
- Mehta, M.P., and Baross, J.A. (2006) Nitrogen fixation at 92 degrees C by a hydrothermal vent archaeon. *Science* **314**: 1783–1786.
- Mehta, M.P., Butterfield, D.A., and Baross, J.A. (2003) Phylogenetic diversity of nitrogenase (*nifH*) genes in deep-sea and hydrothermal vent environments of the Juan de Fuca Ridge. *Appl Environ Microbiol* **69**: 960–970.
- Nakagawa, S., and Takai, K. (2008) Deep-sea vent chemoautotrophs: Diversity, biochemistry and ecological significance. *FEMS Microbiol Ecol* **65**: 1–14.
- Nakagawa, S., Takai, K., Inagaki, F., Horikoshi, K., and Sako, Y. (2005) *Nitratiruptor tergaricus* gen. nov., sp. nov. and *Nitratifactor salsuginis* gen. nov., sp. nov., nitrate-reducing chemolithoautotrophs of the epsilon-Proteobacteria isolated from a deep-sea hydrothermal system in the Mid-Okinawa Trough. *Int J Syst Evol Microbiol* **55**: 925–933.
- Pakchung, A.A.H., Philippa, S., Simpson, J.L., and Codd, R. (2006) Life on earth. Extremophiles continue to move the goal posts. *Environ Chem* **3**: 77–93.
- Peng, Y., Leung, H.C., Yiu, S.M., and Chin, F.Y. (2012) IDBA-UD: A de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* **28**: 1420–1428.
- Perner, M., Gonnella, G., Kurtz, S., and LaRoche, J. (2014) Handling temperature bursts reaching 464 degrees C: Different microbial strategies in the sisters peak hydrothermal chimney. *Appl Environ Microbiol* **80**: 4585–4598.
- Proskurowski, G., Lilley, M.D., Seewald, J.S., Fruh-Green, G.L., Olson, E.J., Lupton, J.E., *et al.* (2008) Abiogenic hydrocarbon production at lost city hydrothermal field. *Science* **319**: 604–607.
- Pruski, A.M., and Dixon, D.R. (2003) Toxic vents and DNA damage: First evidence from a naturally contaminated deep-sea environment. *Aquat Toxicol* **64**: 1–13.
- Querellou, J., Børresen, T., Boyen, C., Dobson, A., Höfle, M., Ianora, A., *et al.* (2010) Marine biotechnology: A new vision and strategy for Europe. *ESF Marine Board Position Paper* **15**: 1–93.
- Reysenbach, A.L., and Shock, E. (2002) Merging genomes with geochemistry in hydrothermal ecosystems. *Science* **296**: 1077–1082.
- Rinke, C., Schwientek, P., Sczyrba, A., Ivanova, N.N., Anderson, I.J., Cheng, J.F., *et al.* (2013) Insights into the phylogeny and coding potential of microbial dark matter. *Nature* **499**: 431–437.
- Schippers, A., and Sand, W. (1999) Bacterial leaching of metal sulfides proceeds by two indirect mechanisms via thiosulfate or via polysulfides and sulfur. *Appl Environ Microbiol* **65**: 319–321.
- Scholz, M., Lo, C.C., and Chain, P.S. (2014) Improved assemblies using a source-agnostic pipeline for MetaGenomic Assembly by Merging (MeGAMerge) of contigs. *Sci Rep* **4**: 6480.
- Schrenk, M.O., Kelley, D.S., Delaney, J.R., and Baross, J.A. (2003) Incidence and diversity of microorganisms within the walls of an active deep-sea sulfide chimney. *Appl Environ Microbiol* **69**: 3580–3592.
- Sigurdsson, H., Carey, S., Alexandri, M., Vougioukalakis, G., Croff, K., Roman, C., *et al.* (2006) Marine investigations of Greece's Santorini volcanic field. *Eos (Washington DC)* **87**: 337–342.
- Skenneron, C.T., Imelfort, M., and Tyson, G.W. (2013) Crass: Identification and reconstruction of CRISPR from unassembled metagenomic data. *Nucleic Acids Res* **41**: e105.
- Stokke, R., Dahle, H., Roalkvam, I., Wissuwa, J., Daae, F.L., Tooming-Klunderud, A., *et al.* (2015) Functional interactions among filamentous Epsilonproteobacteria and Bacteroidetes in a deep-sea hydrothermal vent biofilm. *Environ Microbiol* doi:10.1111/1462-2920.12970.
- Teske, A., Dhillon, A., and Sogin, M.L. (2003) Genomic markers of ancient anaerobic microbial pathways: Sulfate reduction, methanogenesis, and methane oxidation. *Biol Bull* **204**: 186–191.
- Teske, A., Callaghan, A.V., and LaRowe, D.E. (2014) Biosphere frontiers of subsurface life in the sedimented hydrothermal system of Guaymas Basin. *Front Microbiol* **5**: 362.
- Tunncliffe, V., Rose, J.M., Bates, A.E., and Kelly, N.E. (2008) Parasitization of a hydrothermal vent limpet (Lepetodrilidae, Vetigastropoda) by a highly modified

- copepod (Chitonophilidae, Cyclopoida). *Parasitology* **135**: 1281–1293.
- Ultsch, A., and Moerchen, F. (2005) ESOM-Maps: tools for clustering, visualization, and classification with Emergent SOM Technical Report Dept of Mathematics and Computer Science, University of Marburg, Germany 46.
- Voss, M., Bange, H.W., Dippner, J.W., Middelburg, J.J., Montoya, J.P., and Ward, B. (2013) The marine nitrogen cycle: Recent discoveries, uncertainties and the potential relevance of climate change. *Philos Trans R Soc Lond B Biol Sci* **368**: 20130121.
- Walker, C.B., de la Torre, J.R., Klotz, M.G., Urakawa, H., Pinel, N., Arp, D.J., *et al.* (2010) Nitrosopumilus maritimus genome reveals unique mechanisms for nitrification and autotrophy in globally distributed marine crenarchaea. *Proc Natl Acad Sci USA* **107**: 8818–8823.
- Winfrey, M.R., and Zeikus, J.G. (1977) Effect of sulfate on carbon and electron flow during microbial methanogenesis in freshwater sediments. *Appl Environ Microbiol* **33**: 275–281.
- Wright, J.J., Mewis, K., Hanson, N.W., Konwar, K.M., Maas, K.R., and Hallam, S.J. (2014) Genomic properties of Marine Group A bacteria indicate a role in the marine sulfur cycle. *ISME J* **8**: 455–468.
- Wu, M., Ren, Q., Durkin, A.S., Daugherty, S.C., Brinkac, L.M., Dodson, R.J., *et al.* (2005) Life in hot carbon monoxide: The complete genome sequence of *Carboxydotherrmus hydrogenoformans* Z-2901. *PLoS Genet* **1**: e65.
- Xie, W., Wang, F., Guo, L., Chen, Z., Sievert, S.M., Meng, J., *et al.* (2011) Comparative metagenomics of microbial communities inhabiting deep-sea hydrothermal vent chimneys with contrasting chemistries. *ISME J* **5**: 414–426.
- Yamada, T., Sekiguchi, Y., Hanada, S., Imachi, H., Ohashi, A., Harada, H., and Kamagata, Y. (2006) *Anaerolinea thermolimos* sp. nov., *Levilinea saccharolytica* gen. nov., sp. nov. and *Leptolinea tardivitalis* gen. nov., sp. nov., novel filamentous anaerobes, and description of the new classes Anaerolineae classis nov. and Caldilineae classis nov. in the bacterial phylum Chloroflexi. *Int J Syst Evol Microbiol* **56**: 1331–1340.

Supporting information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Fig. S1. Work flow for analysis of assembled 16S rRNA gene sequences.

Fig. S2. Viral contig assignments based on gene composition for (a) red mat and (b) white mat.

Fig. S3. Enrichment of genes normalized using IMG estimation utility for major carbon fixation pathways. (a) CBB cycle, (b) rTCA cycle, and (c) reductive acetyl-CoA pathway. Diagrams are based on KEGG pathway maps. Enzyme (EC) or COG classification numbers for each step were included in boxes. Box color indicated total gene counts of each enzyme with darker red displaying higher gene count and green color representing lower gene count according to the legend provided. The red mat metagenome is always shown on the left and the white mat metagenome on the right.

Fig. S4. Comparison of normalized counts for marker genes of major types of autotrophic carbon fixation. Markers of the CBB cycle include phosphoribulokinase [EC:2.7.1.19] (PRK, prkB), ribulose-bisphosphate carboxylase large chain [EC:4.1.1.39] (rbcL) and ribulose-bisphosphate carboxylase small chain [EC:4.1.1.39] (rbcS). rTCA cycle marker is ATP citrate (pro-S)-lyase [EC:2.3.3.8] (ACLY). Markers for the reductive acetyl-coA or Wood-Ljungdahl pathway include acetyl-CoA synthase [EC:2.3.1.169] (acsB), 5-methyltetrahydrofolate corrinoid/iron sulfur protein methyltransferase [EC:2.1.1.258] (acsE), acetyl-CoA decarbonylase/synthase complex subunit delta [EC:2.1.1.245] (cdhD, acsD), and acetyl-CoA decarbonylase/synthase complex subunit gamma [EC:2.1.1.245] (cdhE, acsC).

Fig. S5. Comparison of normalized counts for marker genes of anaerobic aromatic hydrocarbon degradation. Red and white mat samples are highlighted in the red box. Marker genes are benzoyl-CoA reductase subunit A [EC:1.3.7.8] (bcrA, badF), benzoyl-CoA reductase subunit B [EC:1.3.7.8] (bcrB, badE), benzoyl-CoA reductase subunit C [EC:1.3.7.8] (bcrC, badD), benzoyl-CoA reductase subunit D [EC:1.3.7.8] (bcrD, badG), cyclohexa-1,5-dienecarbonyl-CoA hydratase [EC:4.2.1.100] (dch), 6-hydroxycyclohex-1-ene-1-carbonyl-CoA dehydrogenase [EC:1.1.1.368] (had), 6-oxo-cyclohex-1-ene-carbonyl-CoA hydrolase [EC:3.7.1.-] (oah), benzylsuccinate synthase [EC:4.1.99.11] (E4.1.99.11), and ethylbenzene hydroxylase subunit alpha [EC:1.17.99.2] (ebdA).

Fig. S6. Schematic depicting of the complete benzoyl-coA degradation operon discovered in the red mat bin for SAR406.

Fig. S7. Sox-dependent and Sox-independent sulfur oxidation pathways identified in the present study. Genes are normalized using IMG estimation utility. SO_4^{2-} : sulfate; SO_3^{2-} : sulfite; S^0 : sulfur; $\text{S}_2\text{O}_3^{2-}$, thiosulfate; S^{2-} : sulfide; APS: adenylylsulfate; SOX: sulfur oxidation multi-enzyme complex. Box color indicates the total gene counts for each enzyme with colour coding as in Fig. S5. White mat gene counts are shown on the left and red mat gene counts are displayed on the right. Genes encoding the putative sulfide quinone oxidoreductase (sqr) and the flavocytochrome *c*/sulfide dehydrogenase (FccAB) were not identified. These enzymes catalyze the oxidation of sulfide to sulfur, with FccAB often associated with low sulfide concentrations (Mussmann *et al.*, 2007; Mussmann *et al.*, 2011). On the other hand sulfite reductase (EC:1.8.1.2) was found in high frequencies in both the white and red mat metagenomes. This enzyme is known to catalyze the oxidation of sulphide to sulfite. This may reflect the adaptation of different microbial mat communities to varying sulfide concentrations. The elemental sulfur formation reaction catalyzed by sulfide quinone oxidoreductase or FccAB can be by-passed and direct oxidation to sulfite can be performed. There was further evidence for the reverse dissimilatory sulfate reductase reaction for sulfur oxidization due to presence of key enzymes involved in this pathway [dissimilatory sulfate reductase (EC:1.8.99.3), adenosine-5'-phosphosulfate reductase (EC:1.8.99.2) and ATP sulfurylase (EC:2.7.7.4)].

Fig. S8. Nitrogen metabolism pathway. Components of the nitrogen metabolism pathways identified in this study. Box

color indicated the gene counts for each enzyme with color coding as previously shown in Fig. S7.

Fig. S9. Comparison of normalized counts for marker genes of the nitrogen cycle. Dissimilatory nitrate reduction to ammonia or DNRA markers are nitrite reductase (cytochrome *c*-552) [EC:1.7.2.2] (*nrfA*) and cytochrome *c* nitrite reductase small subunit (*nrfH*). Denitrification markers are nitrite reductase (NO-forming) [EC:1.7.2.1] (*nirK*) and nitrite reductase (NO-forming) / hydroxylamine reductase [EC:1.7.2.1 1.7.99.1] (*nirS*). Nitrification markers are ammonia monooxygenase subunit A [1.14.99.39] (*amoA*), ammonia monooxygenase subunit B (*amoB*) and ammonia monooxygenase subunit C (*amoC*). Nitrogen fixation markers shown are nitrogenase molybdenum-iron protein alpha chain [EC:1.18.6.1] (*nifD*), nitrogenase iron protein NifH [EC:1.18.6.1] (*nifH*) and nitrogenase molybdenum-iron protein beta chain [EC:1.18.6.1] (*nifK*).

Fig. S10. Schematic showing the presence of the minimal gene set for nitrogen fixation in white mat bins for Campylobacteraceae, Coraliomargarita and Methylococcaceae_38GC.

Fig. S11. Principal component analysis PC1 vs. PC2 plot of metagenomes chosen for comparison in this study using normalized KO counts. Grey circles are hydrothermal vent plume samples, red circles are red and white mat from Hellenic Volcanic Arc, orange circles are serpentinite rock and fluid samples, yellow circles are deep oceanic subsurface samples, pink circle is Lost City hydrothermal vent, and purple circle is hydrothermal vent from Loihi Seamount, Hawaii (Spillway mat).

Table S1. Assembly and annotation statistics for white and red mat samples.

Table S2. Protein markers used for inference of contig taxonomy and function count normalization.

Table S3. Assemblies of rRNA by Taxon for Red and White Mats. Depth of coverage of assembled 16S rRNA gene sequences and counts of binned protein-containing contigs are reported by taxon on the table below. White mat counts were normalized to the raw red mat counts by multiplying by 2.29, which is the ratio of the red mat to white mat metagenome size.

Table S4. Bin statistics: General statistics for bins in red and white mat metagenomes.

Table S5. Samples metadata for publicly available metagenomes from IMG.

Table S6. Pfams related to iron respiration pathways and their counts in white and red mat samples.

Dataset S1. ARB Tree of 542 16S Assemblies from the Red and White Mats. Sequences in red were assembled from the red mat; sequences in grey from the white/grey mat. Sequences in black are reference sequences from the ARB database used by Rinke *et al.* (2013). The data elements shown for the black sequences are: phylum/division assignment, this study; phylum/division assignment Rinke *et al.*; GenBank nucleotide accession; sequence length; % Identity to Least Common Ancestor (LCA) in SILVA; taxonomy of LCA in SILVA. The data elements shown for the red and white/grey sequences are: phylum/division assignment, this study; IMG scaffold ID; sequence length; literal 'no trim' (final sequences were not trimmed during load to ARB); % Identity to Least Common Ancestor (LCA) in SILVA; taxonomy of LCA in SILVA. Groups of sequences having no 16S assemblies and which were collapsed for this figure are suffixed with 'noKV.' All other groupings are shown in the Fig. 2 of the main text.

Dataset S2. Phylogenetic analysis of metagenomic contigs containing protein markers genes for red and white mat samples.

Dataset S3. Taxonomic assignment and abundance estimation of red mat and white mat contigs. (a) Taxonomic assignment of contigs containing one or more protein marker genes has been inferred based on their placement in phylogenetic tree. (b) Total sequences of contigs assigned to the lineage has been determined by multiplying contig length by respective coverage (*Excel File Supplementary_Dataset_S3.xlsx*).

Dataset S4. Viral contigs detected in the red mat.

Dataset S5. Viral contigs detected in the white mat. It is indicated the host contig connection when possible.

Dataset S6. Viral contig assignments based on gene composition detected in both mats.

Dataset S7. Metagenomes containing CRISPR spacers from the red mat and the white mat. We used the Genomes On Line Database (GOLD) sample project ID from IMG/M to categorized the ecosystems. (*) Spacer counts refer to the number of total hits that a spacer from the red mat (R_) or white mat (W_) in any given metagenome.

Dataset S8. List of spacer sequences.