Synchronized dynamics of bacterial niche-specific functions during biofilm development in a cold seep brine pool

Weipeng Zhang¹, Yong Wang², Salim Bougouffa³, Renmao Tian¹, Huiluo Cao¹, Yongxin Li¹, Lin Cai¹, Yue Him Wong¹, Gen Zhang¹, Guowei Zhou¹, Xixiang Zhang³, Vladimir B Bajic³, Abdulaziz Al-Suwailem³, Pei-Yuan Qian^{1,2#}

¹KAUST Global Partnership Program, Division of Life Science, Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong

²Sanya Institute of Deep Sea Science and Engineering, Chinese Academy of Sciences, Sanya, Hai Nan, China

³King Abdullah University of Science and Technology, Thuwal, The Kingdom of Saudi Arabia

Running title: metagenomics of cold seep biofilms

Keywords: microbial adaptation; extreme condition; inter-species interaction; genome binning

#Corresponding author:

Pei-Yuan Qian, PhD, Chair professor, Division of Life Science,

The Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong, China

Phone: 0852-2358-7331, Fax: 0852-2358-1559

E-mail: boqianpy@ust.hk

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/1462-2920.12978

Summary

The biology of biofilm in deep-sea environments is barely being explored. Here, biofilms were developed at the brine pool (characterized by limited carbon sources) and the normal bottom water adjacent to Thuwal cold seeps. Comparative metagenomics based on 50 Gb datasets identified polysaccharide degradation, nitrate reduction, and proteolysis as enriched functional categories for brine biofilms. The genomes of two dominant species: a novel deltaproteobacterium and a novel epsilonproteobacterium in the brine biofilms were reconstructed. Despite rather small genome sizes, the deltaproteobacterium possessed enhanced polysaccharide fermentation pathways, whereas the epsilonproteobacterium was a versatile nitrogen reactor possessing *nar*, *nap* and *nif* gene clusters. These metabolic functions, together with specific regulatory and hypersaline-tolerant genes, made the two bacteria unique compared with their close relatives including those from hydrothermal vents. Moreover, these functions were regulated by biofilm development, as both the abundance and the expression level of key functional genes were higher in later-stage biofilms, and co-occurrences between the two dominant bacteria were demonstrated. Collectively, unique mechanisms were revealed: i) polysaccharides fermentation, proteolysis interacted with nitrogen cycling to form a complex chain for energy generation; ii) remarkably, exploiting and organizing niche-specific functions would be an important strategy for biofilm-dependent adaptation to the extreme conditions.

Introduction

To date, many studies have focused on model laboratory strains to understand biofilm development (Reck *et al.*, 2011; Nagar *et al.*, 2014; Fazli *et al.*, 2014; Le *et al.*, 2014), but biofilms in nature can be highly diverse and complex. Studies focusing on the development of natural biofilms, particularly in extreme marine environment are rarely conducted due to the lack of methods for exploring these communities. Cultivation-independent surveys, such as metagenomics, transcriptomics, proteomics, and single cell genomics have facilitated our understanding of bacterial communities in nature (Lehembre *et al.*, 2013; Adriaenssens *et al.*, 2014; Alcaide *et al.*, 2015). Additionally, biofilms on artificial surfaces have been highlighted as useful models to study biofilm development in natural environments (Guezennec *et al.*, 1998; Jones *et al.*, 2007; Salta *et al.*, 2013; Zhang *et al.*, 2013). For example, using metagenomics and an artificial surface-based biofilm ecosystem, we previously successfully demonstrated that biofilm development in an intertidal zone was driven by oxidative and metal ion stresses (Zhang *et al.*, 2013).

Biofilm development has frequently been linked to microbial survival under adverse, stressful or extreme circumstances. For example, Tseng *et al.* (2013) demonstrated that the extracellular matrix protects *Pseudomonas aeruginosa* biofilms by limiting the penetration of tobramycin. Additionally, Bernier *et al.* (2012) reported that organic carbon fixation by *Acidithiobacillus* spp. was a key mechanism supporting the growth of other microbial taxa within snottite biofilms, and the *Acidithiobacillus* spp. adopted several strategies to adapt to extreme acidity. Moreover, it has been hypothesized that inter-species interactions within biofilm communities was an important mechanism that confers microbial adaptation to extreme environments, such as the hydrothermal vents and deep-sea hypersaline brines (Harrison *et al.*, 2013). However, it remains a challenge to perform *in situ* experiments to understand biofilm development in these habitats due to the difficulty in manipulation and in assessing the function of unculturable microbial groups.

Brine pools are found worldwide and are characterized by hypersalinity and poor nutrient availability, which make them extreme habitats. Thuwal seeps II (Batang *et al.*, 2012) is a newly defined cold seep system located on the seafloor of the Red Sea at a depth of

approximately 850 m, in which hypersaline water has seeped out of the seabed and gradually filled the bottom basin to form a brine pool. As measured during the 2013 cruise, the dissolved oxygen (DO) concentration is 5-240 ml/L from the bottom to the surface. Nitrate can be detected in the brine pool at a concentration of 18.5 mg/L. However, the dissolved organic carbon (DOC) concentration (0.48 mg/L) is extremely low compared to that in the adjacent normal bottom water (NBW, 48.23 mg/L). Moreover, the methane in Thuwal seeps II has been completely consumed. During the sample collection in 2013, we also noticed that the seeping flow was weaker than it had been in 2011. In addition, dead mussels were found at the edge of the brine pool. These observations suggested that the brine pool-associated life might have been exposed to an environment with a limited carbon source.

Functional and genetic features of microbial life in brine pools worldwide remains largely uncharacterized. In the cruise to the Red Sea in 2013, relying on a Remotely Operated Vehicle (ROV) and well-designed sampling devices, we developed biofilms in two locations: the brine pool and NBW adjacent to the Thuwal seeps II system, to understand biofilm development in these environments. Because the structure of biofilm developed in deep-sea environments could be influenced by substrate type (Bellou *et al.*, 2012), here we applied six different substrates to demonstrate the core mechanisms governing biofilm development in the cold seep. Biofilms were developed on aluminium (Al), polyether ether ketone (PEEK), polyvinyl chloride (PVC), polytetrafluoroethylene (PTFE), stainless steel (SS) and titanium (Ti). Subsequently, the 16S rRNA gene-based compositions of the biofilm communities were reported (Zhang *et al.*, 2014a), revealing different community structures between the brine biofilms and the NBW biofilms. There was a strong species sorting effect in the assembly of brine biofilms, as indicated by the highly distinct community structures between the brine water and the brine biofilms. Moreover, the brine biofilms contained a significantly higher biomass compared to the NBW biofilms (Zhang *et al.*, 2014a). These results led us to hypothesize that biofilm development may play an important role in survival of particular species in the extreme conditions of the brine pool. To further test this hypothesis herein, we integrated metagenomic sequencing, single genome binning and molecular methods to illustrate functional foundations underlying biofilm development in this brine pool.

Results

Metagenomic comparison revealed enriched functions in the brine biofilms

Consistent with the 16S rRNA gene amplicon data (Zhang *et al.*, 2014a), the principle coordinate analysis (PCoA) profiles derived from the functional traits suggested different compositions between the brine and NBW biofilms (Fig. S1). A direct comparison between the brine biofilm group (3d-brine-AI, PVC, Ti and 6d-brine-AI, PVC, Ti) and the NBW biofilm group (3d-NBW-AI, PVC, Ti and 6d-NBW-AI, PVC, Ti) using similarity percentage analyses (SIMPER) revealed a set of gene ontology (GOs) (Harris *et al.*, 2004) functions that provided a strong contribution to these two different communities (Fig. 1). The majority of the significantly changed individual functions included polysaccharide metabolism and nitrogen cycling. GOs belonging to polysaccharide catabolic processes were enriched in the 6d-brine biofilms. Nitrogen cycling-related processes, including nitrate reduction, nitrogen fixation, and Mo-molybdopterin cofactor metabolic processes, were also prevalent in the brine biofilms. Furthermore, many genes that participate in proteolysis were identified in the 6d-brine biofilms, suggesting the presence of scavengers in these communities and the recycling of carbon and nitrogen. The significance of the differences in gene abundance was further confirmed using the G-test ($P < 0.01$). Interestingly, most of the energy metabolism-related GOs mentioned above displayed higher abundance in 6d-brine compared with biofilms than 3d-brine biofilms (Fig. 1), as indicated by the enrichment for genes involved in nitrogen fixation, nitrogen compound metabolic processes, and ATP hydrolysis-coupled proton transport. In addition, polysaccharide metabolic and biosynthetic processes as well as nitrogen metabolism could also be identified as enriched functions in the 6d-brine biofilms when compared with the brine water sample (Fig. S2).

Metagenomic sequences were further searched against the Carbohydrate-Active enZYme (CAZy) database (Cantarel *et al.*, 2009), which is particularly useful for the comparison of carbohydrate metabolic pathways. Comparison of the 6d-brine and 6d-NBW biofilms by SIMPER analysis revealed that the two biofilm groups were enriched with highly different CAZy categories (Fig. S3). The CAZys belonging to categories of glycoside hydrolases and glycosyl transferases were enriched in the brine biofilm metagenomes, whereas carbohydrate esterases were enriched in NBW biofilm metagenomes. For instance, GH57, which has alpha-amylase (EC 3.2.1.1) and amylo-pullulanase (EC 3.2.1.41) activity, was enriched in the 6d-brine biofilm for more than 10-fold more than in the NBW samples. In contrast,

carbohydrate esterases were enriched in the NBW biofilms.

In summary, the metagenomic results indicated the following: 1) brine biofilms and NBW biofilms could be differentiated based on polysaccharides biosynthesis and degradation, nitrogen metabolism, and proteolysis; and 2) a functional shift was associated with the developmental process in brine biofilms from day 3 to day 6. To further support the metagenomic results, we compared the type and relative abundance of the transporters present in the brine biofilms and NBW biofilms and found a prevalence of monosaccharide and organic ion absorption in the NBW (Fig. S4). Interestingly, transporters of oligosaccharides, polyols, peptides, and nickel were largely enriched in the brine biofilms.

Genome binning of two dominant bacteria in the brine biofilms

To gain a better understanding of the organization of functions in the brine biofilms, bacterial genomes in the brine biofilms were extracted from the assembled contigs. The brine biofilm metagenomes could be decomposed into 30 single genomes that included heterotrophs and mixotrophs. In the present study, we analyzed the draft genomes of two dominant and representative bacteria from 6d-brine biofilms, which were named *deltaproteobacterium* sp. nov. and *epsilonproteobacterium* sp. nov. Based on phylogenetic trees using concatenated 31 essential genes (Fig. S5), the *deltaproteobacterium* sp. nov. had a close phylogenetic relationship with the bacterial members of the *Pelobacter* and *Geoalkalibacter*, the available genome information for which remains limited. On the other hand, the phylogenetic trees based on 31 essential genes (Fig. S6) demonstrated an evolutionary distinction between the *epsilonproteobacterium* sp. nov. and bacteria belonging to the two classified *Epsilonproteobacterial* orders, Campylobacterales and Nautiliales. In addition, relatively close relationships were observed between the *epsilonproteobacterium* sp. nov. and unclassified *Epsilonproteobacteria* such as *Nitratifractor salsuginis* DSM 16511 and *Sulfurovum* sp. NBC37-1.

Information regarding the two genomes is summarized in Table S1. The completeness of the two genome bins was investigated using a suite of hidden Markov models (HMM) encompassing 107 proteins that are conserved in 95% of all of the sequenced bacteria

(Dupont *et al.*, 2012). The number of expected essential single-copy genes was further investigated by comparison with their close relatives (Table S2). The results showed that the tRNA synthetase class I (PF00750) gene was possibly not to be recovered in the genome bin of *deltaproteobacterium* sp. nov., whereas ribosomal protein S17 (PF00366), ribosomal protein S12 (TIGR00981), ribosomal protein L22 (TIGR01044), and ribosomal protein S10 (TIGR01049) were unlikely to be recovered in the genome bin of the *epsilon*proteobacterium sp. nov. Furthermore, the presence of two GTP-binding protein Era (TIGR00436) in all of the compared genomes indicated that the observed duplication in our genome bins was not due to contamination. Thus, we assumed that the completeness of the two draft genomes was ~99% (106 single-copy genes recovered) and ~96% (103 single-copy genes recovered), respectively (Table S1).

Metabolic reconstruction and unique features of the two novel *Proteobacteria*

Catalytic pathways were constructed based on the genome of the two novel bacteria by BLAST searches against the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (Fig. 2). The categories of Clusters of Orthologous Groups (COGs) were used to show the overall functional patterns (Fig. 3a). The tricarboxylic acid (TCA) cycle and proteases were present in the *deltaproteobacterium* sp. nov. (Fig. 2). In addition, the presence of endoglucanase, starch phosphorylase, xylanase/chitin deacetylase, and proteases supported the potential for the degradation of cellulose, starch, chitin, and proteins. The catabolic capacity appeared to be limited mainly to the utilization of glucose and pentose via glycolysis and the pentose phosphate pathway. The generated pyruvate could be converted to acetyl-coenzyme A and then processed to acetate via acetate kinase and to alcohol via alcohol dehydrogenase. Moreover, a lactate dehydrogenase was detected. Thus, the *deltaproteobacterium* sp. nov. could rely on fermentation for energy production, whereas alcohol, acetate, and lactate were main products. In addition, the identification of *nirBD* genes indicated the potential for the use of nitrite as an electron acceptor. The absence of catalase and superoxide dismutase and the presence of oxygen-sensitive enzymes, such as pyruvate:ferredoxin oxidoreductase, suggested an anaerobic lifestyle. The properties of the representative enzymes are summarized (Table S3). Collectively, the *deltaproteobacterium* sp. nov. possessed complete pathways for anaerobic polysaccharide fermentation and proteolysis.

The metabolic pathways of the *deltaproteobacterium* sp. nov. was compared with 14 relatives belonging to *Deltaproteobacteria*, including bacteria from different genera: *Pelobacter*, *Geoalkalibacter*, *Geobacter*, *Anaeromyxobacter*, *Myxococcus*, *Desulfarculus*, *Desulfurivibrio*, and *Syntrophobacter*. Notably, the genome size of *deltaproteobacterium* sp. nov. (2.5 Mb) was rather smaller than the other epsilonproteobacteria (3.1-5.3Mb with an average of 4.3Mb). The genomes were annotated by BLAST analysis; and the number of genes involved in the carbohydrate metabolism pathways are presented (Table S4). The results suggested similar carbon utilization patterns between *deltaproteobacterium* sp. nov. and several *Deltaproteobacteria* members, such as *Pelobacter propionicus* DSM 2379, *Pelobacter carbinolicus* DSM 2380, *Geobacter lovleyi* SZ, and *Syntrophobacter fumaroxidans* MPOB, which may rely on heterotrophic carbon metabolism for energy production. However, the *deltaproteobacterium* sp. nov. appeared to possess more genes that participate in degradation and fermentation. In contrast, the number of genes classified in the nitrogen metabolism and sulfur metabolism categories in *deltaproteobacterium* sp. nov was significantly lower than those in the genomes used for comparison (Table S4); furthermore, key functional genes, such as *napAB*, *narGH*, and *nifDKH*, were missing in the *deltaproteobacterium* sp. nov genome. The numbers of key functional enzymes involved in carbon and nitrogen metabolism are listed (Fig. 3b). These results confirmed the heterotrophic lifestyle of *deltaproteobacterium* sp. nov. and revealed its unique features.

Similarly, the genes from the epsilonproteobacterium sp. nov. were annotated by KEGG and COG to reconstruct the metabolic pathways and to reveal the overall functional category, respectively (Fig. 2 and 3a). Complete functional pathways involved in carbon and nitrogen metabolism were revealed in the epsilonproteobacterium sp. nov., including the TCA cycle, reductive TCA (rTCA) cycle, nitrate reduction marked by *narGH* and *napAB*, and nitrogen fixation marked by the complete *nif* cluster. Moreover, the peptide transport system was also identified. Consistently, the COG categories of energy production (C), amino acid transport and metabolism (E), and secondary metabolism (Q) were identified as abundant functional categories in the epsilonproteobacterium sp. nov. genome.

In comparison to its relatives, epsilonproteobacterium sp. nov. displayed unique features,

particularly in terms of the genes responsible for nitrate reduction and nitrogen fixation (Table S5 and Fig. 3c). The simultaneous presence of *nar*, *nap*, and *nif* family genes was not identified in other known members within the *Epsilonproteobacteria*. The copy number of *napB* in the *Epsilonproteobacteria* genome was higher than that in the other bacteria. The presence of both functional and regulatory genes and the completeness of the nitrogen cycling pathways suggested that the functional versatility of the epsilonproteobacterium sp. nov. was unlikely due to contig contamination (Fig. 3d). Because specific genes were identified in *Epsilonproteobacteria* inhabiting hydrothermal vents, including formate dehydrogenase H (Fdh-H) and the type II form of pyruvate dehydrogenase (PYDH) (Zhang *et al.*, 2014b), we questioned whether these genes were present in the genome of epsilonproteobacterium sp. nov. However, we did not detect homologs of either of the two genes. Furthermore, we searched the genome of the epsilonproteobacterium sp. nov. and identified several genes that were absent in all of the other *Epsilonproteobacteria*, as exemplified by the cation transport regulator *chaB* (Fig. S7a) and a diguanylate cyclase (Fig. S7b). The phylogenetic analysis suggested that the two genes likely originated from *Gammaproteobacteria* and *Deltaproteobacteria*, respectively. The adjacent genes of *chaB* and diguanylate cyclase (Fig. S7c) in the epsilonproteobacterium sp. nov. genome could be identified in other *Epsilonproteobacteria*, which suggested that the presence of these two genes was not due to contig contamination. Thus, the epsilonproteobacterium sp. nov. was identified as a mixotroph using versatile strategies for dissimilatory nitrogen metabolism, and several unique functional features were revealed.

Microbial co-occurrence patterns

Because the deltaproteobacterium sp. nov. and epsilonproteobacterium sp. nov. dominated the 6d-brine biofilms, we further illustrated their relative abundance in all of the other samples based on the number of 16S rRNA gene amplicons and the metagenome coverage (Fig. 4a). Interestingly, the abundance of the deltaproteobacterium sp. nov. and the epsilonproteobacterium sp. nov. was quite low in brine water, in contrast to that in the brine biofilms, suggesting the occurrence of a species sorting process in biofilm development. Because the dynamic pattern of the deltaproteobacterium sp. nov. and the epsilonproteobacterium sp. nov. in (Fig. 4a) appeared to be synchronized, we questioned whether these two bacteria co-occurred across all water and biofilms. As expected, the

correlation analysis of the 51 biofilm and water samples (16S rRNA gene amplicon samples as listed in our previous study (Zhang *et al.*, 2014a)) revealed a strong association between *deltaproteobacterium* sp. nov. and *epsilonproteobacterium* sp. nov. (Fig. S8).

Polysaccharide fermentation, proteolysis, and nitrogen cycling pathways were activated in later-stage biofilms

To evaluate the functional activities expressed in the brine biofilms, quantitative PCR (qPCR) was performed to examine the active transcription of key enzyme genes related to fermentation in *deltaproteobacterium* sp. nov. and nitrogen cycling in *epsilonproteobacterium* sp. nov. (Fig. 4b) using DNA and RNA extracted from 3d- and 6d-brine biofilms that had developed on different substrates. The results showed an up-regulation of acetate kinase, alcohol dehydrogenase, and protease genes in 6d-brine biofilms of *deltaproteobacterium* sp. nov., regardless of the substrate type. In *epsilonproteobacterium* sp. nov., *narG*, *narH*, *napA*, *napH*, *nifD*, and *nifH* transcripts were identified in all of the tested samples; however, their expression was significantly higher in 6d-brine biofilms. These results suggested that fermentation and nitrogen cycling-related marker genes were activated in 6d-brine biofilms. To validate the qPCR results, the expression of motility-related genes was further determined, and the results showed that the flagellar switch gene *motA* was expressed at a very low level in 6d-brine biofilms, which was consistent with known processes underlying biofilm development.

Molecular characterization of the brine biofilms

Because the metagenomic and genomic comparisons revealed the developmental processes in brine pool biofilms, we further characterized the biofilms using molecular methods, including scanning electron microscopy (SEM) and exopolysaccharide (EPS) concentration measurements. The 6d-brine biofilms contained a significantly higher concentration of EPS and a greater cell density compared with the 6d-NBW biofilms (Fig. 5a). Moreover, SEM observations revealed different phenotypes between 3d- and 6d-brine biofilms. In particular, a well-structured matrix was observed in 6d-brine biofilms (Fig. 5b-e), suggesting the presence of an EPS building process during biofilm development. Furthermore, consistent with the enrichment for polysaccharide fermentation pathways in the brine biofilms,

MALDI-TOF/TOF-mass spectrometry (MALDI-TOF/TOF-MS) confirmed the presence of cellulose in the brine biofilm matrix (Fig. 5f).

Discussion

The mechanisms governing biofilm development in nature are complex. In the present study, based on the potential interactions among enriched ecological processes, we proposed a schematic model of biofilm development in a cold seep brine pool. Furthermore, the correlation between biofilm development and unique features of the two dominant bacteria was highlighted.

Polysaccharide fermentation provided the energy source

The enrichment of polysaccharide biosynthesis pathways in the brine biofilms could control the production of biofilm EPS. This argument was supported by the fact that the brine biofilms were well structured and contained more EPS than the NBW. Intriguingly, polysaccharide degradation and metabolism pathways were also enriched in the brine biofilms, suggesting unique roles of polysaccharides as carbon sources. One characteristic of the growth physiology of many heterotrophs in harsh environments is the ability to use complex polysaccharides as carbon and energy sources (Pysz *et al.*, 2013; Werner *et al.*, 2014). Plant polysaccharides are used as signals for EPS synthesis and biofilm development (Beauregard *et al.*, 2013). In the present study, the presence of beta-glucosidase, starch phosphatase, and chitinase in brine biofilm metagenomes indicated that the bacteria were capable of degrading debris from plants and invertebrates (e.g., tubeworms and clams along the margin of the brine pool (Batang, *et al.*, 2012)). Moreover, the gene expression results suggested a switch from aerobic metabolism to anaerobic polysaccharide fermentation in biofilm from days 3 to 6 of biofilm development, likely resulting from the generation of anaerobic micro-environments. Additional evidence for the generation of oxygen-limited layers in biofilms can be obtained from analyses of the dynamics of the biofilm taxonomic structure. For example, members of the genus *Acinetobacter* are strictly aerobic and non-fermentative (Fournier *et al.*, 2006; Li *et al.*, 2014); the *Acinetobacter* accounted for 5-10% of 3d-brine biofilms but diminished in 6d-brine biofilms (see Fig. 3 in our previous work (Zhang *et al.*, 2014a)). Therefore, the accumulation of particulate carbon in the brine

water likely served as a energy source for brine biofilm development; and anaerobic fermentation is a strategy to circumvent the carbon source limitations.

Electron acceptor shift and nitrogen cycling facilitated energy generation

Nitrate is a preferred electron acceptor under conditions of limited oxygen (Lam *et al.*, 2011). The results of the present study suggested that during biofilm development, the final electron acceptor shifted from oxygen to nitrate as the biofilm became more anaerobic. With nitrate as an electron acceptor, the fermentation products of *deltaproteobacterium* sp. nov. could be reused by the *epsilonproteobacterium* sp. nov. for energy generation. For example, the potential to recycle acetate was observed in *epsilonproteobacterium* sp. nov., as demonstrated by the presence and expression of acetate:CoA ligase. Moreover, the presence of *nirBD* in the *deltaproteobacterium* sp. nov. genome suggested that nitrite produced by nitrate reduction could be further used by *deltaproteobacterium* sp. nov. as an electron acceptor (Fig. 2). In addition to nitrate reduction, nitrogen fixation and proteolysis were also enriched in the brine biofilms. The presence of nitrogen fixation suggested that the biofilm members might be faced with challenges associated with ammonia limitation, as proposed in a previous study (Dekas *et al.*, 2013) of communities in methane seep sediment. Proteolysis was recently reported in polysaccharide-fed bioprocesses (Lü *et al.*, 2013). Proteolysis and amino acids can interact with the central carbon metabolism pathways via several intermediates (Lü *et al.*, 2013), such as pyruvate. In the present study, proteases were identified in the *deltaproteobacterium* sp. nov. genome, while peptide transport systems were identified in *epsilonproteobacterium* sp. nov. (Fig. 2). These results suggested the occurrence of protein degradation, the import activities of peptides and cooperation between different biofilm residents. The Zn-dependent protease was also activated in 6d-brine biofilms (Fig. 4). These results suggested that nitrogen cycling contributed to the energy generation and resource recycling.

A mechanistic model of biofilm development

As mentioned above, the metabolic interactions between *epsilonproteobacterium* sp. nov. and *deltaproteobacterium* sp. nov. were associated with biofilm development. The interactions could be further supported by the co-occurrence pattern (Fig. S8), which has been used to

predict the cooperative potential between bacterial pairs in previous studies (Freilich *et al.*, 2011; Barberán *et al.*, 2012). Coordinating environmental parameters with the inter-species interactions, we therefore propose a model for biofilm development in the brine pool (Fig. 6). During the initiation of biofilm development, species sorting is an important effect, which selects bacterial groups that can use polysaccharide particles or products of polysaccharide degradation. These bacterial groups included the *deltaproteobacterium* sp. nov. and *epsilonproteobacterium* sp. nov. that were characterized in the present study. However, their relative abundance in initial-stage biofilms were quite low, probably due to the inhibition of oxygen and the lack of available carbons. As EPS and simple sugar accumulates, the micro-environment of the biofilm became more anaerobic and rich in carbon, leading to the proliferation of *deltaproteobacterium* sp. nov. and *epsilonproteobacterium* sp. nov. Concomitantly, anaerobic fermentation, proteolysis and nitrogen cycling were activated, leading to coupled chemical reactions that yielded energy and transfer of electrons.

The link between biofilm development and unique functional features of the two dominant species

Genomic information about the *Deltaproteobacteria* and *Epsilonproteobacteria* from cold seeps is limited. The members of these two classes have been reported as key players in biogeochemical processes in hydrothermal environment (Campbell *et al.*, 2009; Vetriani *et al.*, 2014), including the chemolithotrophic oxidation of ammonia, hydrogen and sulfur, and the reduction of sulfate and nitrate. *Epsilonproteobacteria* genomes can rapidly change, which may confer a selective advantage in either pathogenic associations or deep-sea hydrothermal vent environments (Campbell *et al.*, 2009). In contrast to most *Epsilonproteobacteria* (Vetriani *et al.*, 2014), the *epsilonproteobacterium* sp. nov. in the present study possessed both the periplasmic nitrate enzyme *NapA* and the membrane-bound enzyme *NarG* (Fig. 3). They also carried a complete pathway for nitrogen fixation and regulation, making them different from their close relatives, such as the *Sulfurovum* sp. NBC37-1 and *Nitratiruptor* sp. SB155-2, from hydrothermal fields (Vetriani *et al.*, 2014). The lack of specific enzymes for the reported vent-inhabiting *Epsilonproteobacteria* (Zhang *et al.*, 2014b) and the presence of specific genes in *epsilonproteobacterium* sp. nov. (Fig. S7) further supported the functional distinction. In particular, the cation transport regulator *chaB* in the *epsilonproteobacterium* sp. nov. suggested a niche adaptation because this gene is

phylogenetically related to *Thiomicrospira halophila*, a bacterium isolated from hypersaline lakes (Sorokin *et al.*, 2006). The diguanylate cyclase, which synthesizes cyclic di-GMP, would be involved in signal transduction and biofilm development of *epsilon*proteobacterium sp. nov. On the other hand, the *deltaproteobacterium* sp. nov. displayed a close evolutionary relationship with *Pelobacter* and *Geoalkalibacter*; the lack of nitrate and sulfur reduction pathways in *deltaproteobacterium* sp. nov. genome suggested that different electron acceptors were utilized, with a finding illustrated by the genome reduction following the loss of genes in certain pathways. The lack of nitrate reduction pathways may also explain its low abundance in the brine water but proliferation in the brine biofilms through interaction with the *epsilon*proteobacterium sp. nov. However, despite rather small genome size, the *deltaproteobacterium* sp. nov. appeared to possess enhanced pathways for polysaccharide utilization because additional genes that participate in degradation and fermentation were identified. Although only the complete genomes available in NCBI were included for comparison, the results do suggest different features of the two genome bins with many close relatives. Taken together, biofilm development helps bacteria to take advantage of adaptive features and overcome genetic limits through functional cooperation between different species. The unique features of the two dominant species make them fit well with the biofilm lifestyle.

Concluding remarks

Using biofilms in the NBW as a control, we demonstrated in the present study that biofilm development facilitated microbial adaptation to the brine pool. Although the development of biofilms on natural substrates in the brine pool (e.g., biofilms formed on particles in the deep-sea areas (Hou *et al.*, 2004)) was not directly studied, we believe that the principles governing biofilm development in response to *in situ* conditions were successfully revealed, because we evaluated biofilms on six different types of materials, demonstrating common mechanisms for these microbial life on surfaces (Donlan *et al.*, 2002). Therefore, the “artificial-surface-biofilm metagenomics” could be used as a novel model for mechanistic study of natural biofilm development. Although previous biofilm models (Davey *et al.*, 2000; Yoshida *et al.*, 2008) have documented inter-species interactions mediated by biofilm components (e.g., bacterial coaggregation mediated by polysaccharides), we presented different interactive mechanisms from a dynamic perspective. The replacement of oxygen by

nitrate as an electron acceptor and the connection between nitrate reduction and polysaccharide fermentation were important processes in brine biofilm development. Furthermore, exploiting and organizing niche-specific functional features through inter-species interactions and generation of special micro-environments would be important mechanisms of biofilm-dependent adaptation. That is, biofilm development makes the evolutionary advantages of particular bacterial groups become truly advanced.

Experimental procedures

Experimental design, sampling location, and environmental parameter measurements

The details of the experimental design and sampling locations are shown in [Figure S9](#) and a previous study (Zhang *et al.*, 2014a). Briefly, sampling was performed in May 2013 in the Thuwal seeps II (22°16N-38°53E) via the ROV Max Rover, DSSI, USA during the Red Sea exploration cruise organized by King Abdullah University of Science and Technology (KAUST). Biofilms were developed on six different types of materials as mentioned in the Introduction, in the brine pool and NBW for 3 and 6 days. The materials were deployed at a depth of 850 m by operating ROV. The environmental parameters, including temperature, salinity, and DO in the brine were measured using a conductivity-temperature-depth (CTD) plus a SBE43 DO sensor mounted on the ROV. Other environmental parameters, such as the concentration of nitrate and ammonia and the DOC, were determined on-board using a portable data logging colorimeter (DR/890, HACH, USA) according to the manufacturer's protocol.

Metagenomic DNA extraction, Illumina sequencing, assembly, and annotation

DNA samples were extracted from the 3- and 6-day biofilms developed on Al, Ti, and PVC panels in the brine pool and NBW and from the water samples (in total 14 samples). Briefly, cells were recovered from the substrates using sterile cell scrapers and suspended in Tris-HCl buffer. Before nucleic acid extraction, the cells were pelleted by centrifugation at 4000 rpm for 10 min and then lysed with lysozyme, proteinase K, and 10% SDS. Total nucleic acid was extracted and purified using the AllPrep DNA/RNA Mini Kit (Qiagen, Hilden, Germany) following the manufacturer's instructions. The quality and quantity of the DNA were checked by agarose gel electrophoresis and with a Nanodrop device (ND-1000 spectrophotometer,

DiaMed China Limited, Hong Kong) at 260 nm.

The extracted DNA was sequenced using an Illumina HiSeq 2000 platform to generate a dataset of ~50 Gb. The detailed information is summarized in Table S6. Metagenomic analyses were performed according to Mason *et al.* (2012), in which raw and unassembled Illumina reads were searched against a database of molecules that participate in hydrocarbon degradation, and according to Plewniak *et al.* (2013), in which assembled Illumina reads were searched against the GO database (Harris *et al.*, 2004) to identify over-presented categories. In the present study, quality control for the reads was first conducted using the next-generation sequencing (NGS) QC toolkit (2012). After trimming reads containing homopolymers, low-quality reads were removed, and reads containing primer/adaptor contamination were trimmed. Each of these steps was followed by the filtering of reads with a given length cut-off. Raw and unassembled Illumina reads were classified by BLAST searches against the CAZy prokaryote databases on local servers using an e-value of $<10^{-7}$. This database was selected because it contains a nearly complete hydrocarbon degradation pathway. In addition, raw Illumina reads were assembled using a CLC Genomics Workbench (WB). The data were imported into the CLC WB and processed using Illumina 1.5-1.7 quality control specifications, whereby low-quality nucleotides were trimmed. The imported paired reads were then subjected to another round of quality control using the following WB parameters under the "Trim Sequences" function: Ambiguous trim 4, Quality limit 0.03. A list of Illumina adapters and primers was imported into the WB based on the FASTQC list provided. The search for adapters and primers was performed for both forward and reverse sequences. For the assembly, the following parameters were used: word size = 64, bubble size = 75 and a paired distance that was estimated by the software. Annotation by the GO database was conducted using the INDIGO annotation pipeline following the methods described by Alam *et al.* (2013).

PCR amplification of 16S rRNA genes, 454 pyrosequencing, and taxonomic analysis

Briefly, two replicates of DNA were PCR-amplified using primers with an added 6- or 8-nucleotide (nt) barcode for multiplexed pyrosequencing. The barcodes were added to the universal forward primer U515F (5'-GTGYCAGCMGCCGCGGTAA-3') and the reverse primer U1390R (5'-GACGGGCGGTGTGTRCAA -3') to amplify the hypervariable V4-V8

region of the 16S rRNA genes. This pair of primers was selected based on our previous study focusing on the primer design for 16S rRNA gene amplicons in deep-sea metagenomic study (Wang *et al.*, 2009). A 20- μ L PCR reaction mixture contained 0.4 U of Phusion High-Fidelity DNA polymerase (Finnzymes Oy, Espoo, Finland), 1 $\times$ HF reaction buffer, 0.1 mM of each barcoded primer, 0.6 μ L of DMSO, 0.2 mM of dNTPs (TaKaRa, Dalin, China) and 10 ng of purified DNA template. PCR was performed in a thermal cycler (Bio-Rad, USA) under the following conditions: initial denaturation at 98°C for 30 s; 26 cycles of 98°C for 10 s, 60°C for 10 s and 72°C for 15 s; and a final extension at 72°C for 5 min. The PCR products for the same replicate were mixed together and pyrosequenced using the ROCHE 454 FLX Titanium platform, while two replicates were sequenced. In total, 47 biofilm (6 substrates $\times$ 2 replicates $\times$ 2 sites $\times$ 2 time points = 48 while the 3d-brine-SS has 1 replicate) and 4 water samples (brine water and NBW, each with 2 replicates) were obtained. The subsequent taxonomic analysis of the pyrosequencing data was performed using QIIME 1.7.0 (Caporaso *et al.*, 2010). The details of the methods were reported in our previous study (Zhang *et al.*, 2014a).

Statistical analysis of metagenomic data

To confirm that the coverage of the GO categories was high enough to allow a functional comparison, rarefaction curves were drawn using an in-house Perl script. Representative rarefaction curves of 6d-brine-Ti and 6d-NBW-Ti metagenomes are shown (Fig. S10). To calculate the relative abundance of GO categories in the two metagenomes, raw reads obtained after quality control were aligned to ORFs using Bowtie2 (version 2.0.0) (Langmead *et al.*, 2012), and the sequencing coverage of the contigs was calculated with SAMtools (version 0.0.19) (Li *et al.*, 2009).

The normalized counts for the GO and CAZy functional annotations per biofilm sample were standardized to account for unequal sequence coverage between the samples. SIMPER (Clarke *et al.*, 1993) was performed to determine the contribution of each GO and CAZy to similarities between different biofilm groups. First, we compared in parallel the 3d-brine biofilms with the 3d-NBW biofilms and the 6d-brine biofilms with the 6d-NBW biofilms. Because similar results were obtained in the two comparisons, we pooled all of the brine pool biofilms and compared them with the NBW biofilms (Fig. 1). Similarities among different microbial communities were determined based on the composition and relative abundance of

GOs (only GOs with an abundance of more than 0.5% in at least one sample were used) and displayed using PCoA implemented in the Paleontological statistics software package (PAST) (Hammer *et al.*, 2001).

Genome binning, reassembly, and validation

Draft genome binning was conducted according to previously described methods (Albertsen *et al.*, 2013; Tian *et al.*, 2014) in which the genomes could be separated mainly based on the genome coverage, GC content, and tetranucleotide frequency. The Illumina pyrosequencing reads were mapped to the assembled contigs using Bowtie2. The genome coverage was calculated using SAMtools, as described above. PCoA of the tetranucleotide frequency was performed. The open reading frames (ORFs) of the assembled contigs were predicted using Prodigal (Hyatt *et al.*, 2010). The contigs were searched against a set of 107 hidden Markov models (HMMs) of essential proteins (Albertsen *et al.*, 2013) using default cut-off values in the HMM datasets. The identified essential proteins were searched against the NCBI NR database with BLASTP (e-value <1e-05) and assigned taxonomically using MEGAN 5.0 (Huson *et al.*, 2011). Contigs were labeled according to the phylum-level taxonomic affiliation of the essential proteins.

Different metagenome pairs were selected to obtain effective separation and extraction. For instance, the *deltaproteobacterium* sp. nov. and *epsilonproteobacterium* sp. nov. genomes described in the following section were obtained using the 6d-brine-Al/3d-brine-Al and 6d-brine-Ti/6d-NBW-Ti metagenome pairs, respectively. For genome binning of *deltaproteobacterium* sp. nov. (Fig. S11ab), a group of contigs in the 6d-brine-Al showing 2.5 × greater metagenomic coverage than the 3d-brine-Al was selected, which had a GC content of ~50% and included most of the labeled *Deltaproteobacteria* contigs. From these contigs, a core set of *Deltaproteobacteria* contigs was extracted by further PCoA based on the tetranucleotide frequencies. For genome binning of *epsilonproteobacterium* sp. nov. (Fig. S11c-f), a group of contigs in the 6d-brine-Ti showing 100 × greater metagenomic coverage than in the 6d-NBW-Ti was selected and included most of the labeled *Epsilonproteobacteria* contigs. The contigs in this group were clearly separated from the other contigs and further confirmed by PCoA. To increase the completeness of the analysis, genome reassembly was performed: raw reads were BLAST searched against the binned contigs using an e value of

<1e-7 and then subjected to assembly by the SPAdes Genome Assembler (version 3.5.0) (Bankevich *et al.*, 2012).

Genome annotation and pathway analysis

The ORFs in the extracted genomes were annotated by BLAST searches against the KEGG, COG, and CAZy databases using a maximum e-value cut-off of 1e-05. Metabolic pathways, such as denitrification and nitrogen fixation, were analyzed using online tools in KEGGMAPPER (<http://www.genome.jp/kegg/mapper.html>). Key enzymes were rechecked based on the results of the BLAST search against the COG database. Genes of interest that could not be found in the draft genomes were rechecked in the remaining assembled metagenomic contigs.

Phylogenetic tree construction based on essential genes and the 16S rRNA genes

Phylogenetic trees were constructed based on 31 essential genes. AMPHORA (Wu *et al.*, 2008) was used to predict 31 essential marker genes (*tsf*, *smpB*, *rpsS*, *rpsM*, *rpsK*, *rpsJ*, *rpsI*, *rpsE*, *rpsC*, *rpsB*, *rpoB*, *rpmA*, *rplT*, *rplS*, *rplP*, *rplN*, *rplM*, *rplL*, *rplK*, *rplF*, *rplE*, *rplD*, *rplC*, *rplB*, *rplA*, *pyrG*, *pgk*, *nusA*, *infC*, *frr*, and *dnaG*) from the extracted genomes and reference genomes downloaded from the NCBI database. All of the available *Deltaproteobacteria* and *Epsilonproteobacteria* complete genomes were downloaded. Several genomes belonged to the same species (e.g., the *Helicobacter pylori* strains), and in these cases, only selected ones were used for the tree construction. The essential marker genes and BLAST hits collected from NCBI were aligned based on the protein sequences. The aligned protein sequences were then concatenated and imported into Mega (version 6.05; Hall *et al.*, 2013) to construct the ML phylogenetic tree (with 1000 bootstrap replicates). The bootstrap values were calculated with 1000 replicates.

Test of the co-occurrence patterns

Non-random co-occurrence patterns were constructed according to methods described in a previous study (Freilich *et al.*, 2011). The 16S rRNA gene amplicons of 47 biofilm and 4 water samples were used as queries to build a matrix of 97% cutoff OTUs. All of the statistical analyses were performed in the R studio environment (R 2.13) using the vegan,

igraph, and Hmisc packages. A network of co-occurring OTUs was constructed using a Spearman's correlation coefficient >0.8 and statistically significant P-value <0.01 . The OTUs were previously filtered to remove OTUs that were poorly represented or present with a low abundance ($<0.1\%$ in all samples). The matrix was also subjected to multiple test correction using Benjamini-Hochberg standard false discovery rate correction ("FDR-BH") prior to generation of the graph using igraph. Subsequently, the network was visualized in the gephi platform (0.8.2).

RNA extraction and reverse transcription

As described in our previous study (Zhang *et al.*, 2014a), biofilms on different substrates in the brine and NBW were harvested (ten slides per sample) onboard using sterile cotton tips and resuspended in RNAlater buffer. Using the AllPrep DNA/RNA Mini Kit (Qiagen, Hilden, Germany), RNA extraction was performed according to the manufacturer's instructions with modifications. Briefly, bacterial cells from the cotton tips for the biofilm samples were resuspended in RNAlater buffer, pelleted by centrifugation at 4000 g for 10 min and then lysed with lysozyme, proteinase K and 10% SDS. Total nucleic acid was extracted, and DNA was removed using the DNA column provided in the kit. Agarose gel electrophoresis was used to estimate the integrity of the RNA, and a Nanodrop device (ND-1000 spectrophotometer, DiaMed China Limited, Hong Kong) was used to measure the RNA quantity. After a PolyA tailing reaction using *Escherichia coli* PolyA Polymerase I (Austin, TX), the RNA was reverse-transcribed into cDNA using SuperScript III reverse transcriptase (Invitrogen, Carlsbad, CA) and oligo(dT) primers.

qPCR analysis of single-aggregate genes

The fluorogenic PCR method was used to quantify selected genes and reverse-transcribed cDNA as described in the studies reported by Halm *et al.* (2011). Based on the gene sequence information obtained for the samples, primers were designed (Table S7), and the specificities of the sequences were verified by BLAST searches against the NCBI database. The qPCR reactions were conducted using the Kapa SYBR Fast qPCR Kit (Kapa Biosystems, Woburn, MA) for both DNA and cDNA using a Mx3000P qPCR System (Agilent Technologies, Palo Alto, CA). Each 20- μ l qPCR reaction contained 10 μ l of $2 \times$ Master Mix, 1 pmol/ μ l of

forward and reverse primers, and 0.5 µl of either standard or environmental sample. The optimal DNA and cDNA dilution for the qPCR was 1:25 and 1:20, respectively. The cycling parameters were 5 min at 95°C, followed by 40 cycles of 15 s at 95°C, 15 s at 50°C, and 15 s at 60°C. All of the samples were run in triplicate together with no-template controls. The expression of the genes was normalized according to their relative cDNA and DNA abundances (Halm *et al.*, 2011).

Biofilm characterization and measurement of polysaccharide concentrations

The characteristics of the biofilm, including the cell density and surface structures, were assessed using scanning electron microscopy (SEM) as described in our previous work (Zhang *et al.*, 2014a). In the present study, polysaccharides in the biofilms were extracted and quantified according to previously reported procedures (Jiao *et al.*, 2010; Myszka *et al.*, 2009). Biofilm samples were thawed on ice and centrifuged at 15,000 g for 20 min. The pellets were re-suspended in 3 mL of a cold sulfuric acid solution (0.2 M sulfuric acid, pH 1.1), and the biofilm matrix was broken using a glass hand homogenizer mortar and pestle prior to centrifugation at 15,000 g for 20 min. The resulting supernatant was collected and suspended in water, followed by the addition of 1.4 mL of 77% (v/v) H₂SO₄. Subsequently, the samples were mixed with 200 µL of 1% (w/v) cold tryptophan. Finally, the samples were heated at 37°C for 20 min, and the colored product was evaluated at OD500. Calibration curves were prepared using dextran solution. The cellulose in the biofilm matrix was identified using MALDI-TOF/TOF-MS (UltrafleXtreme; Bruker Daltonics, Bremen, Germany).

Nucleotide sequence accession numbers

The major portion of the datasets was deposited in the NCBI Sequence Reads Archive under accession number SRS576191. The draft genomes of the *deltaproteobacterium* sp. nov. and the *epsilonproteobacterium* sp. nov. have been deposited in the WGS submission under accession numbers SUB774613 and SUB774605, respectively.

Conflict of interest

The authors declare no competing financial interests.

Acknowledgements

The authors are grateful to Pui Shuen WONG from Hong Kong University of Science and Technology for the MALDI-TOF/TOF-MS analysis. This study was supported by awards from the Sanya Institute of Deep Sea Science and Engineering, the Chinese Academy of Sciences (SIDSSE, CAS) (SIDSSE-201206), COMRA program of China (COMRRDA12SC01), General Research Fund (661611) from HKSAR government, and King Abdullah University of Science and Technology (SA-C0040/UK-C0016) to Pei-Yuan Qian. In addition, Salim Bougouffa was supported by a SABIC postdoctoral fellowship.

References

- Adriaenssens, E. M., Van Zyl, L., De Maayer, P., Rubagotti, E., Rybicki, E., Tuffin, M., *et al.* (2014) Metagenomic analysis of the viral community in Namib Desert hypoliths. *Environ Microbiol* **17**: 480-495.
- Alam, I., Antunes, A., Kamau, A.A., Kalkatawi, M., Stingl, U., and Bajic, V.B. (2013) INDIGO-integrated data warehouse of microbial genomes with examples from the red sea extremophiles. *PloS One* **8**:e82210.
- Albertsen, M., Hugenholtz, P., Skarszewsk, A., Nielsen, K.L., and Tyson, G.W. (2013) Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. *Nat Biotechnol* **31**:533-538.
- Alcaide, M., Stogios, P. J., Lafraya, Á., Tchigvintsev, A., Flick, R., Bargiela, R., *et al.* (2015) Pressure adaptation is linked to thermal adaptation in salt-saturated marine habitats. *Environ Microbiol* **17**: 332-345.
- Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., *et al.* (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* **19**:455-477.
- Barberán, A., Bates, S.T., Casamayor, E.O., and Fierer, N. (2012) Using network analysis to explore co-occurrence patterns in soil microbial communities. *ISME J* **6**:343-351.
- Batang, Z.B., Papathanassiou, E., Al-Suwailem, A., Smith, C., Salomidi, M., Petihakis, G., *et al.* (2012) First discovery of a cold seep on the continental margin of the central Red Sea. *J Mar Syst* **94**:247-253.
- Beauregard, P.B., Chai, Y., Vlamakis, H., Losick, R., and Kolter, R. (2013) *Bacillus subtilis* biofilm induction by plant polysaccharides. *Proc Natl Acad Sci USA* **110**:1621-1630.

- Bellou, N., Papathanassiou, E., Dobretsov, S., Lykousis, V., and Colijn, F. (2012) The effect of substratum type, orientation and depth on the development of bacterial deep-sea biofilm communities grown on artificial substrata deployed in the Eastern Mediterranean. *Biofouling* **28**: 199-213.
- Campbell, B.J., Smith, J.L., Hanson, T.E., Klotz, M.G., Stein, L.Y., Lee, C.K., *et al.* (2009) Adaptations to submarine hydrothermal environments exemplified by the genome of *Nautilia profundicola*. *PLoS Genet* **5**:e1000362.
- Cantarel, B.L., Coutinho, P.M., Rancurel, C., Bernard, T., and Lombard, V. (2009) The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics. *Nucleic Acids Res* **37**:233-238.
- Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., *et al.* (2010) QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* **7**:335-336.
- Chung, H.C., Lee, O.O., Huang, Y.L., Mok, S.Y., Kolter, R., and Qian, P.Y. (2010) Bacterial community succession and chemical profiles of subtidal biofilms in relation to larval settlement of the polychaete *Hydroides elegans*. *ISME J* **4**:817-828.
- Clarke, K.R. (1993) Non-parametric multivariate analyses of changes in community structure. *Austral Ecol* **18**:117-143.
- Davey, M.E., and O'toole, G.A. (2000) Microbial biofilms: from ecology to molecular genetics. *Microbiol Mol Biol Rev* **64**: 847-867.
- Dekas, A.E., Chadwick, G.L., Bowles, M.W., Joye, S.B., and Orphan, V.J. (2013) Spatial distribution of nitrogen fixation in methane seep sediment and the role of the ANME archaea. *Environ Microbiol* **16**:3012-3029.
- Donlan, R.M. (2002) Biofilms: microbial life on surfaces. *Emerg Infect Dis* **8**:881-890.
- Dupont, C., Rusch, D.B., Yooseph, S., Lombardo, M.J., Richter, R.A., Valas, R., *et al.* (2012) Genomic insights to SAR86, an abundant and uncultivated marine bacterial lineage. *ISME J* **6**:1186-1199.
- Fazli, M., Almblad, H., Rybtke, M. L., Givskov, M., Eberl, L., and Tolker-Nielsen, T. (2014) Regulation of biofilm formation in *Pseudomonas* and *Burkholderia* species. *Environ Microbiol* **16**:1961-1981.
- Fournier, P.E., Vallenet, D., Barbe, V., Audic, S., Ogata, H., Poirel, L., *et al.* (2006) Comparative genomics of multidrug resistance in *Acinetobacter baumannii*. *PLoS Genet* **2**:e7.
- Freilich, S., Zarecki, R., Eilam, O., Segal, E.S., Henry, C.S., Kupiec, M., *et al.* (2011) Competitive and cooperative metabolic interactions in bacterial communities. *Nat Commun* **2**:589.
- Guezennec, J., Ortega-Morales, O., Raguene, G., and Geese, G. (1998) Bacterial colonization of artificial substrate in the vicinity of deep-sea hydrothermal vents. *FEMS Microbiol Ecol* **26**:89-99.

- Hall, B.G. (2013) Building phylogenetic trees from molecular data with MEGA. *Mol Biol Evol* **30**:1229-1235.
- Halm, H., Lam, P., Ferdelman, T.G., Lavik, G., Dittmar, T., LaRoche, J., *et al.* (2011) Heterotrophic organisms dominate nitrogen fixation in the South Pacific Gyre. *ISME J* **6**: 1238-1249.
- Hammer, Ø., Harper, D.A.T., and Ryan, P.D. (2001) PAST: Paleontological statistics software package for education and data analysis. *Palaeontol Electron* **4**:1-9.
- Harris, M.A. (2004) The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res* **32**:D258-D261.
- Harrison, J.P., Gheeraert, N., Tsigelnitskiy, D., and Cocke, C.S. (2013) The limits for life under multiple extremes. *Trends Microbiol* **21**:204-212.
- Hou, S., Saw, J.H., Lee, K.S., Freitas, T.A., Belisle, C., Kawarabayasi, Y., *et al.* (2004) Genome sequence of the deep-sea γ -proteobacterium *Idiomarina loihiensis* reveals amino acid fermentation as a source of carbon and energy. *Proc Natl Acad Sci USA* **101**:18036-18041.
- Huson, D.H., Mitra, S., Ruscheweyh, H.J., Weber, N., and Schuster, S.C. (2011) Integrative analysis of environmental sequences using MEGAN4. *Genome Res* **21**:1552-1560.
- Hyatt, D., Chen, G.L., LoCascio, P.F., Land, M.L., Larimer, F.W., and Hauser, L.J. (2010) Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**:119.
- Jiao, Y., Cody, G.D., Harding, A.K., Wilmes, P., and Schrenk, M. (2010) Characterization of extracellular polymeric substances from acidophilic microbial biofilms. *Appl Environ Microbiol* **76**:2916-2922.
- Jones, D.S., Albrecht, H.L., Dawson, K.S., Schaperdorth, I., Freeman, K.H., Pi, Y., *et al.* (2012) Community genomic analysis of an extremely acidophilic sulfur-oxidizing biofilm. *ISME J* **6**:158-170.
- Jones, P.R., Cottrell, M.T., Kirchman, D.L., and Dexter, S.C. (2007) Bacterial community structure of biofilms on artificial surfaces in an estuary. *Microb Ecol* **53**:153-162.
- Lam, P., and Kuypers, M.M. (2011) Microbial nitrogen cycling processes in oxygen minimum zones. *Annu Rev Mar Sci* **3**:317-345.
- Langmead, B., and Salzberg, S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**:357-359.
- Le, K.Y., Dastgheyb, S., Ho, T.V., and Otto, M. (2014) Molecular determinants of staphylococcal biofilm dispersal and structuring. *Front Cell Infect Microbiol* **4**:167.
- Lee, J.W., Nam, J.H., Kim, Y.H., Lee, K.H., and Lee, D.H. (2008) Bacterial communities in the initial stage of marine biofilm formation on artificial surfaces. *J Microbiol* **46**:174-182.
- Lehembre, F., Doillon, D., David, E., Perrotto, S., Baude, J., Foulon, J., *et al.* (2013) Soil metatranscriptomics for mining eukaryotic heavy metal resistance genes. *Environ Microbiol* **15**:2829-2840.

- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., *et al.* (2009) The sequence alignment/map format and SAMtools. *Bioinformatics* **25**:2078-2079.
- Li, W.G., Zhang, D.Y., Huang, X.F., and Wen, Q. (2014) *Acinetobacter harbinensis* sp. nov., isolated from river water. *Int J Syst Evol Micr* **64**:1507-1513.
- Lü, F., Bize, A., Guillot, A., Monnet, V., Madigou, C., Chapleur, O., *et al.* (2013) Metaproteomics of cellulose methanisation under thermophilic conditions reveals a surprisingly high proteolytic activity. *ISME J* **8**:88-102.
- Mason, O.U., Hazen, T.C., Borglin, S., Chain, P.S., Dubinsk, E.A., Fortney, J.L., *et al.* (2012) Metagenome, metatranscriptome and single-cell sequencing reveal microbial response to Deepwater Horizon oil spill. *ISME J* **6**:1715-1727.
- Myszka, K., and Czaczyk, K. (2009) Characterization of adhesive exopolysaccharide (EPS) produced by *Pseudomonas aeruginosa* under starvation conditions. *Curr Microbiol* **58**:541-546.
- Nagar, E., and Schwarz, R. (2014) To be or not to be planktonic? Self-inhibition of biofilm development. *Environ Microbiol* **17**:1477-1486.
- Patel, R.K., and Jain, M. (2012) NGS QC Toolkit: a toolkit for quality control of next generation sequencing data. *PloS One* **7**:e30619.
- Plewniak, F., Koechler, S., Navet, B., Dugat-Bony, É., Bouchez, O., Peyret, P., *et al.* (2013) Metagenomic insights into microbial metabolism affecting arsenic dispersion in Mediterranean marine sediments. *Mol Ecol* **22**:4870-4883.
- Pysz, M.A., Montero, C.I., Chhabra, S.R., Kelly, R.M., and Rinker, K.D. (2013) Significance of polysaccharides in microbial physiology and the ecology of hydrothermal vent environments, in the seafloor biosphere at mid-ocean ridges (eds W. S.D. Wilcock, E. F. Delong, D. S. Kelley, J. A. Baross and S. Craig Cary), American Geophysical Union, Washington, D. C.. doi:10.1029/144GM14.
- Reck, M., Rutz, K., Kunze, B., Tomasch, J., Surapaneni, S.K. Schulz, S., *et al.* (2011) The biofilm inhibitor carolacton disturbs membrane integrity and cell division of *Streptococcus mutans* through the serine/threonine protein kinase PknB. *J Bacteriol* **20**: 5706
- Salta, M., Wharton, J.A., Blache, Y., Stokes, K.R., and Briand, J.F. (2013) Marine biofilms on artificial surfaces: structure and dynamics. *Environ Microbiol* **15**:2879-2893.
- Sorokin, D.Y., Tourova, T.P., Kolganova, T.V., Spiridonova, E.M., Berg, I.A., Muyzer, G. (2006) *Thiomicrospira halophila* sp. nov., a moderately halophilic, obligately chemolithoautotrophic, sulfur-oxidizing bacterium from hypersaline lakes. *Int J Syst Evol Microbiol* **56**:2375-2380.
- Tian, R.M., Wang, Y., Bougouffa, S., Gao, Z.M., Cai, L., Bajic, V., *et al.* (2014) Genomic analysis reveals versatile heterotrophic capacity of a potentially symbiotic sulfur oxidizing bacterium in sponge. *Environ Microbiol* **16**:3548-61.

Tseng, B. S., Zhang, W., Harrison, J. J., Quach, T. P., Song, J. L., Penterman, J., *et al.* (2013) The extracellular matrix protects *Pseudomonas aeruginosa* biofilms by limiting the penetration of tobramycin. *Environ Microbiol* **15**:2865-2878.

Vetriani, C., Voordeckers, J.W., Crespo-Medina, M., O'Brien, C.E., Giovannelli, D., and Lutz, R.A. (2014) Deep-sea hydrothermal vent Epsilonproteobacteria encode a conserved and widespread nitrate reduction pathway (Nap). *ISME J* **8**:1510-1521.

Wang, Y., and Qian P.Y. (2009) Conservative fragments in bacterial 16S rRNA genes and primer design for 16S ribosomal DNA amplicons in metagenomic studies. *PLoS One* **4**:e7401.

Werner, J., Ferrer, M., Michel, G., Mann, A.J., Huang, S., Juarez, S., *et al.* (2014) Halorhabdus tiamatea: proteogenomics and glycosidase activity measurements identify the first cultivated euryarchaeon from a deep-sea anoxic brine lake as potential polysaccharide degrader. *Environ Microbiol* **16**:2525-37.

Wu, J.A., Kusuma, C., Mond, J.J., and Kokai-Kun, J.F. (2003) Lysostaphin disrupts *Staphylococcus aureus* and *Staphylococcus epidermidis* biofilms on artificial surfaces. *Antimicrob Agents Chemother* **47**:3407-3414.

Wu, M., and Eisen, J.A. (2008) Simple, fast, and accurate method of phylogenomic inference. *Genome Biol* **9**: R151.

Yoshida, Y., Yang, J., Peaker, P.E., Kato, H., Bush, C.A., and Cisar, J.O. (2008) Molecular and antigenic characterization of a *Streptococcus oralis* coaggregation receptor polysaccharide by carbohydrate engineering in *Streptococcus gordonii*. *J Biol Chem* **283**:12654-12664.

Zhang, W., Wang, Y., Lee, O.O., Tian, R., Cao, H., Gao, Z., *et al.* (2013) Adaptation of intertidal biofilm communities is driven by metal ion and oxidative stresses. *Sci Rep* **3**:3180.

Zhang, W.P., Wang, Y., Tian, R.M., Bougouffa, S., Yang, B., Cao, H.L., *et al.* (2014a) Species sorting during biofilm assembly by artificial substrates deployed in a cold seep system. *Sci Rep* **4**:6647.

Zhang, Y., and Sievert, S.M. (2014b) Pan-genome analyses identify lineage-and niche-specific markers of evolution and adaptation in Epsilonproteobacteria. *Front Microbiol* **5**:110.

Figures

Figure 1 Gene ontology (GO) differentiating brine and normal bottom water (NBW) biofilm metagenomes. The biofilms were developed in these two locations for two different durations: 3 days (3d) and 6 days (6d). The brine biofilms and NBW biofilms were compared by SIMPER analysis. GOs are presented in the order of their contribution to the difference (highest to lowest, bottom to top). The significance of the differences in gene abundance were further confirmed by the G-test ($P < 0.01$). Nitrogen cycling, polysaccharide metabolism and proteolysis-related functions are indicated in blue, red and green, respectively.

Figure 2 Overview of the metabolic capacities of *deltaproteobacterium sp. nov.* and *epsilonproteobacterium sp. nov.* based on their genomes and potential interactions. Polysaccharide degradation is coupled to anaerobic fermentation in *deltaproteobacterium sp. nov.*, whereas acetate, alcohol and lactate are the main products. Enzymes related to proteolysis and amino acid metabolism are also present in the *deltaproteobacterium sp. nov.* genome. For *epsilonproteobacterium sp. nov.*, genes responsible for acetate transformation and peptide transport are present. The *epsilonproteobacterium sp. nov.* possesses multiple pathways for nitrate reduction, including the periplasmic nitrate enzyme *NapA*, membrane-bound enzyme *NarG*, and a complete *nif* pathway for nitrogen fixation. Genes associated with the regulation of motility and the assimilation of sulfur were also identified. Abbreviations: APS, adenosine phosphosulfate; PAPS, adenosine 3-phosphate 5-phosphate; lps, waa, gmh and rfa: gene clusters for lipopolysaccharide biosynthesis.

Figure 3 Functional categories for the two novel bacteria and genomic comparison with their close relatives. **(a)** Overview of the cluster of orthologous gene (COG) categories in the epsilonproteobacterium sp. nov. and deltaproteobacterium sp. nov. genomes. Selected COG categories: C, energy production; E, amino acid transport, and metabolism; Q, secondary metabolite biosynthesis, transport and catabolism. **(b)** The numbers of marker genes for carbohydrate metabolism and nitrogen metabolism in 1, deltaproteobacterium sp. nov.; 2, *Pelobacter propionicus* DSM 2379; 3, *Pelobacter carbinolicus* DSM 2380; 4, *Geobacter subterraneus* Red1; 5, *Geobacter bemidjiensis* Bem; 6, *Geobacter daltonii* FRC-32; 7, *Geobacter* sp. M21; 8, *Geobacter uraniireducens* Rf4; 9, *Geobacter lovleyi* SZ; 10, *Anaeromyxobacter* sp. K; 11, *Anaeromyxobacter* sp. Fw109-5; 12, *Myxococcus xanthus* DK 1622; 13, *Desulfarculus baarsii* DSM 2075; 14, *Desulfurivibrio alkaliphilus* AHT2; and 15, *Syntrophobacter fumaroxidans* MPOB. The gene function was annotated by BLAST searches against the Kyoto Encyclopedia of Genes and Genomes (KEGG) database on a local server and checked manually BLAST analysis against the online non-redundant (Nr) database. **(c)** The numbers of marker genes for nitrogen metabolism in 1, epsilonproteobacterium sp. nov.; 2, *Nitratifactor salsuginis* DSM 16511; 3, *Sulfurovum* sp. NBC37-1; 4, *Nitratiruptor* sp. SB 155-2; 5, *Caminibacter mediatlanticus* TB-2; 6, *Nautilia profundicola* AmH; 7, *Lebetimonas* sp. JS170; 8, *Wolinella succinogenes* DSM 1740; 9, *Helicobacterceterum* MIT99-5656; 10, *Arcobacter* sp. L; 11, *Campylobacter coli* CVM N29710; 12, *Campylobacter coli* 15-537360; 13, *Sulfurospirillum barnesii* SES-3; 14, *Sulfuricurvum* sp. IFRC-1; and 15, *Sulfurimonas autotrophica* DSM 16294. **(d)** Organization of nitrate reduction and nitrogen fixation genes in epsilonproteobacterium sp. nov. Individual genes are drawn to scale. The green ORFs represent the *nif* gene family; red represents the *nap* gene family; black is the *nar* gene family; orange shows nitrogen regulators; and white represents genes that are not related to nitrogen metabolism.

Figure 4 Relative abundance of *deltaproteobacterium* sp. nov. and *epsilonproteobacterium* sp. nov. in all of the samples and expression of key functional genes. **(a)** The relative abundance was based on 16S rRNA gene sequences in different samples. Qualified reads were assigned to different classes using the RDP classifier in the QIIME pipeline based on a confidence threshold of 97%, and the mean values calculated from two replicates are shown. The 16S rRNA gene-based relative abundance is consistent with the genome coverage in the metagenomes. **(b)** The expression of genes related to fermentation (beta-glucosidase, acetate kinase, and alcohol dehydrogenase) and proteolysis (Zn-dependent protease) in *deltaproteobacterium* sp. ov. and related to nitrogen cycling (*narG*, *narH*, *napA*, *napB*, *nifD*, and *nifH*), acetate metabolism (acetate: CoA ligase) and motility (*motA*) in *epsilonproteobacterium* sp. nov., were calculated using quantitative PCR. The gene expression level in 6d-brine-Al was set as standard “1” for comparison. All of the samples were analyzed in triplicate. *, Significant difference between 6d-brine and 3d-brine biofilm groups, $P < 0.05$; **, $P < 0.01$.

Figure 5 Molecular characterization of the brine biofilms. **(a)** The exopolysaccharide (EPS) concentration in brine and NBW biofilms. The concentration was calculated based on three replicates. *, Significant difference between brine and NBW biofilms, $P < 0.05$. **(b-e)** Scanning electron microscopy (SEM) images of the Al and PVC surfaces in the brine pool for different durations (3 and 6 days). The 6d-brine biofilms, which contain a well-structured matrix, displayed different phenotypes compared with the 3d-brine biofilms. The biofilms were not homogeneous, and thus, hundreds of SEM pictures were obtained. Representative images are shown. MALDI-TOF/TOF-mass spectrometry detection of cellulose in the brine biofilm matrix. **(f)** The 6d-brine-PVC biofilm is shown as an example. Pure cellulose was used as a positive control, and $[(\text{Glc})_n + \text{Na}]^+$ ions from the cellulose were detected.

Figure 6 A schematic model of biofilm development in the brine pool. Polysaccharide fermentation, nitrate reduction and proteolysis are enriched and activated during biofilm development. Inter-species metabolic interactions between deltaproteobacterium sp. nov. and epsilonproteobacterium sp. nov. facilitate energy generation and resource recycling.