

Received Date : 26-Apr-2015

Revised Date : 22-Nov-2015

Accepted Date : 24-Nov-2015

Article type : Original Article

Environmental switching during biofilm development in a cold seep system and functional determinants of species sorting

Authors

Weipeng Zhang¹, Renmao Tian¹, Yang Bo¹, Huiluo Cao¹, Lin Cai¹, Lianguo Chen¹, Guowei Zhou¹, Jin Sun¹, Xixiang Zhang², Abdulaziz Al-Suwailem², Pei-Yuan Qian^{1#}

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

²King Abdullah University of Science and Technology, Thuwal, The Kingdom of Saudi Arabia

Running title: Functional bases for species sorting

Keywords: species sorting; biofilm; brine pool; *Marinobacter*; *Oleispira*; metagenome

Corresponding author:

Pei-Yuan Qian

Email: 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/mec.13501

This article is protected by copyright. All rights reserved.

Telephone: (852) 23587331

Fax: (852) 23581559

Abstract

The functional basis for species sorting theory remains elusive, especially for microbial community assembly in deep sea environments. Using artificial surface-based biofilm models, our recent work revealed taxonomic succession during biofilm development in a newly defined cold seep system, the Thuwal cold seeps II, which comprises a brine pool and the adjacent normal bottom water (NBW) to form a metacommunity via the potential immigration of organisms from one patch to another. Here, we designed an experiment to investigate the effects of environmental switching between the brine pool and the NBW on biofilm assembly, which could reflect environmental filtering effects during bacterial immigration to new environments. Analyses of 16S rRNA genes of 71 biofilm samples suggested that the microbial composition of biofilms established in new environments was determined by both the source community and the incubation conditions. Moreover, a comparison of 18 metagenomes provided evidence for biofilm community assembly that was based primarily on functional features rather than taxonomic identities; metal ion resistance and amino acid metabolism were the major species sorting determinants for the succession of biofilm communities. Genome binning and pathway reconstruction of two bacterial species (*Marinobacter* sp. and *Oleispira* sp.) further demonstrated metal ion resistance and amino acid metabolism as functional traits conferring the survival of habitat generalists in both the brine pool and NBW. The results of the present study sheds new light on microbial community assembly in special habitats and bridges a gap in species sorting theory.

Introduction

A mechanistic understanding of microbial community assembly in nature is of great interest. In the framework of species sorting theory, particular species are selected from a pool of bacteria to form a community as a result of selection by local abiotic and biotic environmental conditions (Jones *et al.* 2009; Langenheder *et al.* 2011; Stegen *et al.* 2012; Székely *et al.* 2012; Zhang *et al.*

2014). The species sorting concept in the metacommunity framework has been applied to explain bacterial immigration into new environments (Peay *et al.* 2012; Sullam *et al.* 2012; Székely *et al.* 2012; Orsini *et al.* 2013), and it assumes that microbial communities have the potential to rapidly adapt to new environments by adjusting their species compositions. The establishment of new communities has been proposed to be influenced by both the source community and the incubation conditions. Furthermore, the filtering effect by environmental parameters is relatively weak for generalists, which possess a high level of functional plasticity. In contrast, habitat specialists who have a low functional plasticity respond primarily to environmental factors. Thus, the dynamics of the habitat specialists in natural microbial communities are modulated according to environmental processes, whereas habitat generalists are governed mainly by dispersal processes.

Although studies have demonstrated the roles of species sorting in the assembly of natural microbial communities and proposed general principles as mentioned above, most of the available evidence is confined to the taxonomic level (Jones *et al.* 2009; Langenheder *et al.* 2011; Stegen *et al.* 2012; Székely *et al.* 2012; Zhang *et al.* 2014). The functional bases for species sorting effects remain largely unknown; and the functional features of specialists and generalists are elusive, especially for microbial communities in deep sea environments. Significant progress has been achieved over the past few years in revealing the functions of microorganisms in natural subsurface systems. Although the majority of these organisms could not be obtained in pure culture, a greater awareness of their activities could be achieved through cultivation-independent surveys such as metagenomics. Because the microbiomes of an increasing number of environments have undergone deep sequencing, the diversity of microbial activities continues to rise (Öztürk *et al.* 2013; Li *et al.* 2014). However, microbial lifestyles in deep sea environments remain mostly undiscovered. Moreover, the application of metagenomics to explain the effects of ecological theories would lead to interesting findings from a functional perspective.

The microbial ecology of cold seeps is a new topic. In recent studies (Zhang *et al.* 2014; Zhang *et al.* 2015), biofilms were developed at two locations in the Red Sea, namely, a brine pool located at a water depth of 840 and 850 m and adjacent normal bottom water (NBW) next to the Thuwal cold seeps II (Batang *et al.* 2012). The brine pool and NBW are separated by a thin interface. The brine pool is the coldest temperature (21.7°C) and least amount of saline (7.4%) among the brine

Accepted Article

pools reported in the Red Sea. The brine water is enriched for metals at ten-fold or higher concentrations relative to the NBW, including aluminum (0.24 ppm), arsenic (0.34 ppm), copper (0.20 ppm), and iron (0.10 ppm) (Batang *et al.* 2012). The NBW contains more than 100-fold higher concentrations of dissolved organic carbons (48.23 mg/L and 0.48 mg/L in the NBW and brine pool, respectively). Based on 16S rRNA gene and metagenomics analyses, we observed different community structures between the brine biofilms and NBW biofilms. These spatially distinct communities are connected to form a metacommunity via the potential dispersal of organisms from one patch to another. However, the local conditions of the water differ from one another, and thus, the species sorting and bacterial functional plasticity might be important in shaping the structures of the communities in this cold seep system.

In the present study, we designed an experiment to study microbial community assembly in the Thuwal cold seeps II system in the framework of the metacommunity concept, particularly the environmental filtering effect on community establishment after bacterial immigration to a new environment. The biofilms were allowed to develop separately in brine and NBW for three days before being moved to NBW and brine, respectively, and then incubated for an additional three days (hereafter referred to as switched biofilms: “S-brine” for biofilms switched from NBW to brine and “S-NBW” for biofilms switched from brine to NBW). Subsequently, the 16S rRNA genes and metagenomes of the switched biofilms were sequenced and compared with the biofilms that were developed continuously in the brine pool or NBW. Thus, we were able to evaluate the influence of species sorting by particular environmental parameters in the assembly of the immigrated communities and identify the functional bases for these processes.

Materials and Methods

Experimental design

The sampling locations were introduced in a previous study (Zhang *et al.* 2014). The experimental design is shown in Figure 1. Briefly, a particular experimental set-up was built to guarantee that i) the substrates were completely immersed in water; ii) the biofilms were stable during movement by the remotely operated vehicle (ROV); and iii) the experimental set-up could be easily manipulated to reduce variations in the microbial community produced during launch and recovery. Based on a previous study showing that the structure of biofilms developed in deep-sea

environments could be influenced by the type of substrate (Meier *et al.* 2013), in the present study, we applied six different substrates to demonstrate common mechanisms of microbial life on surfaces. First, three steel frames, each containing six different types of substrate, were placed in the brine pool. In parallel, three steel frames were placed in NBW and incubated for three days. Each frame contained three carousels, each carousel contained six different types of substrate, and each type of substrate in the same carousel contained 10 slides, which were embedded in the slots. Thus, each frame contained 30 slides of the same substrate type: 10 slides for metagenomic DNA extraction, 10 for the two replicates of 16S rRNA amplicons, and 10 for other analysis such as cell density measurements which are not shown in the present work; RNA was extracted using the same samples that were used for the DNA analysis. The six different substrates included aluminum (Al), polyether ether ketone (PEEK), polyvinyl chloride (PVC), polytetrafluoroethene (PTFE), stainless steel (SS), and titanium (Ti). After three days, one frame from the brine pool and another from the NBW were recovered. Concomitantly, one frame was transferred from the brine pool to the NBW, and one was transferred from the NBW to the brine pool. The remaining four frames were further incubated in the two environments for three additional days (a total of six days) and then recovered. Collectively, we obtained three groups of biofilms: the switched biofilms included S-brine and S-NBW biofilms; the brine biofilm included 3d-brine and 6d-brine biofilms; and the NBW biofilms included 3d-NBW and 6d-NBW biofilms.

DNA extraction, PCR amplification of 16S rRNA genes and 454 pyrosequencing

Immediately after recovery from the sea, the biofilms were harvested using sterile cotton tips and resuspended in Tris-HCl buffer. The samples were stored at -80°C until DNA extraction. DNA extraction was performed as previously described (Zhang *et al.* 2014). Briefly, bacterial cells from cotton tips in Tris-HCl buffer were pelleted by centrifugation at 4,000 × g for 10 min. Lysozyme, proteinase K and 10% SDS were added to the suspended cells before further lysis and purification using the AllPrep DNA/RNA Mini Kit (Qiagen, Hilden, Germany) according to 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 V4-V8 region of the 16S rRNA genes in Bacteria and Archaea were amplified using the biofilm samples as template and the universal forward primer U515F (5'-GTGYCAGCMGCCGCGGTAA-3') and the reverse primer U1390R

(5'-GACGGGCGGTGTGTRCAA-3'). The PCR reaction of 26 thermal cycles was performed using the same conditions as described in our previous study. Pyrosequencing of the PCR products and two replicates (excluding the 3d-brine-SS) was conducted using the ROCHE 454 FLX Titanium platform in CHGB (Shanghai, China). In total, 71 samples of the 16S rRNA amplicons, including 24 switched biofilms, 23 brine biofilms and 24 NBW biofilms were analyzed in the present study.

Sequence quality filtering and taxonomic classification of the 16S rRNA

Quality control was performed to remove reads with ambiguous nucleotides: reads <150 bp and reads containing homopolymers ≥ 6 bp. Subsequently, the QIIME 1.7.043 software package (Caporaso *et al.*, 2010) was used to analyze the 16S rRNA pyrosequencing data following steps described in our previous study (Lee *et al.*, 2012). The reads were assigned to their corresponding samples according to their barcodes, clustered, and then assigned to operational taxonomic units (OTUs) with an identity of 97%. Representatives of the most abundant reads were selected from each OTU for subsequent analysis. *De novo* alignment using MUSCLE (Edgar *et al.* 2004) available in QIIME and reference-based alignment against the Silva108 database were performed for representative OTUs. Chimeras in the aligned reads were identified by ChimeraSlayer and then removed from the dataset. Finally, the reads were assigned to different taxa using the Ribosomal Database Project (RDP) classifier version 2.2 (Wang *et al.* 2007) against Silva108 with a bootstrap confidence level of 50%. The read information for the switched biofilms is shown in Table S1. Rarefaction curves detailing the number of genera given the number of sequenced 16S rRNA gene amplicons are displayed in Figure S1. The microbial composition of the samples at the class and genus levels are displayed, and a cut-off of 1% relative abundance for at least one replicate was used; that is, the class or genus that did not achieve an abundance greater than 1% in at least one of the two replicates was defined as “minor”.

Statistical analysis of 16S rRNA datasets

Bray-Curtis distances using genera with a relative abundance of >0.1% as input were generated and visualized using principal coordinate analysis (PCoA) implemented in the Paleontological statistics software package (PAST) (Hammer *et al.* 2001). To confirm the PCoA results, cluster analysis based on Bray-Curtis distance was performed. Furthermore, one-way permutational

multivariate analysis of variance (PERMANOVA) implemented in PAST was used to assess the effect of switching on the community composition. The biofilms on all the six substrates were grouped together as replicates, and the S-brine biofilms were compared with the 3d-NBW and 6d-NBW biofilms separately. Similarly, the S-NBW biofilms were compared with the 3d-brine and 6d-brine biofilms separately. In addition, to identify bacterial genera differentiating the S-brine and NBW biofilm communities, comparisons were performed by SIMPER (Clarke *et al.* 1993) analysis based on the Bray-Curtis distance using relative abundance of genera as input.

Sequence quality filtering, assembly and annotation of metagenomic sequences

The DNA samples that were used for the metagenomic sequencing were extracted from the switched biofilms, brine biofilms and NBW biofilms developed on Al, Ti, and PVC (18 samples in total; and the six switched biofilms were documented in the present study). The Al biofilm sample had the greatest biomass, the Ti biofilm sample had the lowest biomass, whereas PVC represented a plastic material (in contrast to metal). Therefore, the biofilms on these three materials were selected for the metagenomic analysis. The sequencing was performed using an Illumina HiSeq 2000 platform to generate a metagenomic dataset of ~70 Gb, which consisted of 100-bp paired-end reads. The detailed information for the switched biofilms is summarized in Table S2. Metagenomic analyses were performed by referring to previous studies (Mason *et al.* 2012). Briefly, quality control was conducted using the next-generation sequencing (NGS) QC toolkit (Patel *et al.* 2012), and reads containing homopolymers and low-quality reads were removed. Each of these steps was followed by the filtering of reads with a specific cut-off length (reads of <100 bp were removed). Subsequently, qualified Illumina reads were assembled as described in our previous study (Zhang *et al.* 2015). For the annotation, ORFs were predicted using the Prokaryotic Dynamic Programming Genefinding Algorithm (Prodigal, v2.50) (Hyatt *et al.* 2010), and BLAST searches against the clusters of orthologous groups (COGs) (Tatusov *et al.* 2000) and the Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa *et al.* 2000) database were conducted for the functional annotation.

Statistical analysis of the metagenomic datasets

To calculate the relative abundance of COG 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 of COG functional annotations per biofilm sample were standardized to account for the unequal sequence coverage between the samples. Bray-Curtis distances/similarities among different microbial communities were determined based on the composition and relative abundance of COGs (only COGs with an abundance of more than 0.05% in at least one sample were used) and visualized using PCoA implemented in PAST. To confirm the PCoA results, cluster analysis was performed based on the Bray-Curtis distance. SIMPER was performed to determine the contribution of each COG to the differences between the switched biofilms and the source biofilms.

Genome binning and validation and pathway analysis

Draft genome binning was performed according to previously described methods (Albertsen *et al.* 2013; Tian *et al.* 2014). Briefly, the genomes were separated based on the genome coverage, GC content, and tetranucleotide frequency. The Illumina pyrosequencing reads were mapped to the assembled contigs using Bowtie2 (Langmead *et al.* 2012); the genome coverage was calculated using SAMtools (Li *et al.* 2009); whereas the contigs were searched against a set of 107 hidden Markov models (HMM) 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 5.046. Finally, the contigs were labeled according to the phylum-level taxonomic affiliation based on the essential proteins.

All of the switched biofilm datasets were assembled together. Genome binning resulted in an over representation of two bacterial species. *Marinobacter* sp. and *Oleispira* sp. genomes described in the following section were obtained using the S-brine-Ti/S-NBW-Ti metagenome pairs. For the genome binning of *Marinobacter* sp., a group of contigs in the T-NBW-Ti that exhibited 10.0 × greater metagenomic coverage than the S-brine-Ti was selected, which contained ~52% GC

content and included most of the labeled *Gammaproteobacterial* contigs. Correspondence analysis (CA) of the tetranucleotide frequencies was used to further separate the target genome from the contaminating genomes, resulting in a core set of contigs. The ORFs were obtained, and the ORFs in each contig were taxonomically assigned using the BLASTP and MEGAN programs according to the method described above. Finally, a draft genome of the *Marinobacter* sp. was obtained. The findings were confirmed via a taxonomical assignment using the BLASTP and MEGAN programs to reveal a draft genome. Similar methods were used to obtain the *Oleispira* sp. draft genome.

The ORFs in the extracted genomes were annotated by BLAST against the KEGG and COG databases with a maximum e-value cut-off of 1e-05. To construct the metabolism pathways, the online tool KEGGMAPPER (<http://www.genome.jp/kegg/mapper.html>) was applied. The annotation of key enzymes was validated based on the results of the BLAST search against the non-redundant (Nr) database.

Phylogenetic tree construction based on essential genes

To identify the phylogeny of the extracted genomes, phylogenetic trees were constructed based on 31 essential genes. The 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*) were predicted by AMPHORA (Wu *et al.* 2008) from the extracted genomes and the reference genomes downloaded from the NCBI database. The essential marker genes were aligned based on the protein sequences separately and then linked together. The concatenated aligned protein sequences were then imported into Mega (version 6.05) (Hall *et al.* 2013) to construct the ML phylogenetic tree. The bootstrap values were calculated with 1000 replicates.

The phylogenetic tree of 2-hydroxy-6-oxonona-2, 4-dienedioate hydrolase from *Oleispira* sp. was constructed using protein sequences. The reference sequences were obtained from the NCBI-Nr database using BLASTP. Aalignment was conducted with Muscle in the software molecular evolutionary genetics analysis (MEGA 6.0) (Tamura *et al.*, 2013) using the following parameters: gap open -50, cluster method UPGMB and min diag length 24. The Gblock software (Castresana *et al.*, 2000) was applied to eliminate the less informative sites in the alignments.

The construction of maximum likelihood trees of was conducted using MEGA 6.06 with the Tamura-Nei model, the Nearest-Neighbor-Interchange (NNI) method and 1000 bootstrap value.

RNA extraction and reverse transcription

For RNA extraction, biofilms on different substrates in the brine pool and in the NBW were harvested on a board using sterile cotton tips and stored in RNAlater buffer. RNA extraction was performed according to the manufacturer's instructions using the AllPrep DNA/RNA Mini Kit (Qiagen, Hilden, Germany), which allowed the extraction of DNA and RNA from the same sample. Briefly, bacterial cells from cotton tips (for the biofilm samples) and filters (for the water samples) were resuspended in RNAlater buffer, pelleted by centrifugation at $4,000 \times g$ for 10 min and then lysed with lysozyme, proteinase K and 10% SDS. The 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 concentration. The RNA was subjected to a PolyA tailing reaction using *Escherichia coli* PolyA Polymerase I (Austin, TX) and then reverse-transcribed into cDNA using SuperScript III reverse transcriptase (Invitrogen, Carlsbad, CA) and oligo(dT) primers. The three replicates of the RNA samples (10 slides of the same substrate type were pooled together as one replicate) were extracted for the quantitative real-time PCR (q-RT-PCR) assay.

q-RT-PCR detection of key functional genes

To examine the expression of selected functional genes, fluorogenic PCR was applied to quantify the genes in the DNA samples and reverse-transcribed the cDNA as described in the study by Halm *et al.* (2011). Both DNA and RNA were extracted from one sample at the same time using ALLPREP RNA/DNA kit, and this allowed us to use q-RT-PCR to qualify the abundance of a target gene in RNA (cDNA) and DNA. Primers were designed (Table S3) using Primer5 software, and the specificities of the sequences were verified by BLAST analysis against the NCBI database. The PCR reactions were conducted using the Kapa SYBR Fast qPCR Kit (Kapa Biosystems, Woburn, MA) for both DNA and cDNA with a Mx3000P qPCR machine (Agilent Technologies, Palo Alto, CA). The 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 μ L of template. The optimal dilutions of DNA and cDNA were determined prior to performing the qPCR. 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. Each sample was assessed in triplicate, and the relative expression levels of the genes were normalized based on their cDNA/DNA ratios (Halm *et al.* 2011).

Results

Taxonomic structure of the switched biofilms

Rarefaction curves of 16S rRNA gene amplicons of the microbial communities are shown in Figure S1. Lines reaching asymptotes suggested that the reads were sufficient to characterize the communities, which lacked a high level of diversity. Assignment of the reads at the class level revealed distinct taxonomic profiles between the S-brine and S-NBW biofilm samples (Fig. S2; the corresponding phyla are shown as well). First, the bacterial reads from the S-NBW biofilms were assigned mostly to the *Gammaproteobacteria* (85% to 95% within each sample), whereas the abundance of the *Gammaproteobacteria* was lower in the S-brine biofilms (25%-62%). Moreover, the proportions of several bacteria (namely Verrucomicrobia, TM6, TA06, Planctomycetes, Hyd24-12, Deferribacteria, Candidate division WS3, OP9, OP11, OD1 and Acidobacteria) differed between the switched biofilm and the biofilms that were developed continuously in the brine pool or NBW. For example, Firmicutes (consisted of Bacilli and Clostridia) accounted for 15%-20% in the 6d-brine-PVC and 6d-brine-Ti biofilms (Zhang *et al.* 2014), whereas they were diminished in the S-brine-PVC and S-brine-Ti biofilms.

To further illustrate the effects of environmental switching on biofilm community composition, reads from the biofilms were further classified at the genus level, and genera with a relative abundance >1.0% are summarized in Figure 2. The S-brine biofilms displayed different patterns at the genus level compared with the source biofilms, namely the 3d-NBW and 6d-NBW biofilms. For example, *Marinobacter* and *Oleispira* were enriched in the S-brine biofilms compared with the 3d-NBW and 6d-NBW biofilms, as illustrated by the SIMPER analysis of the community dissimilarity between core taxa groups (Fig. S3). Interestingly, these two genera were also enriched in S-NBW biofilms compared with the 3d-brine and 6d-brine biofilms (Fig. S4); they accounted for ~10% and ~50% of the S-NBW biofilms, respectively. Significant differences in community structures at the genus level between the switched biofilms and the biofilms that were developed continuously in the brine pool or NBW were confirmed using one-way

PERMANOVA (S-brine versus 3d-NBW: Pseudo- $F_{212,4899} = 6.59$, $P = 0.0009$; S-brine versus 6d-NBW: Pseudo- $F_{194,4485} = 41.4$, $P = 0.0001$; S-NBW versus 3d-brine: Pseudo- $F_{196,4531} = 73.83$, $P = 0.0001$; S-NBW versus 6d-brine: Pseudo- $F_{165,3818} = 75.91$, $P = 0.0001$).

Similarities based on taxonomic and functional contents

To study the taxonomic structural similarity between the switched biofilms and the biofilms that were developed continuously in the brine pool or NBW, PCoA based on the relative genera abundance was performed for all of the 71 samples (all of the biofilms consisted of two biological replicates excluding 3d-brine-SS). As a result (Fig. 3A), similarity between the S-brine biofilms and the S-NBW biofilms were observed, which was consistent with the above-mentioned observation that the switched biofilms were significantly different from the brine and NBW biofilms. Furthermore, the S-NBW biofilms were more similar to the 6d-NBW biofilms than to the 6d-brine biofilms, while S-brine biofilms were located between the 6d-NBW and 6d-brine biofilms, suggesting increased colonization pressure in NBW than in brine water. However, according to the taxonomic PCoA plot (Fig. 3A), there was no clear border between the switched biofilms and the biofilms that were developed continuously in the brine pool or NBW, which could be support by the cluster analysis showing two major groups based on taxonomic composition (Fig. S5A).

To further characterize the switched biofilms, metagenomic analyses were conducted for the switched biofilms developed on Al, PVC and Ti. The metagenome information is summarized in Table S2. Based on the PCoA, we observed substantial differences between the taxonomic profile based on the 16S rRNA genes and the functional profile based on the COG annotation. The functional content clearly distinguished the switched biofilms from the brine and NBW biofilms by PC2 (explaining 35.67% of the variance; Fig. 3B). These results could be supported by the cluster analysis (Fig. S5B) because the dendrograms based on the COG abundance positioned the S-brine and S-NBW biofilms on the same branch.

Metagenomic comparison and enriched functions for the switched biofilms

To identify the effects of switching on functional profile changes in the biofilms, we compared the metagenomes of the switched biofilm with their source biofilms. A direct comparison of the S-brine and NBW (3d-NBW and 6d-NBW) biofilm metagenomes using SIMPER revealed a set of COG functions that strongly contributed to these two different types of communities (Fig. 4). The majority of the significantly changed individual functions was related to metal ion stress resistance, amino acid metabolism and protein biosynthesis. For example, the ABC-type metal ion transport system (COG1135) (Bakker *et al.* 1987) was prevalent in S-brine biofilms but displayed a very low abundance in NBW biofilms, regardless of the substrate type (Fig. 4). Student's t-test was performed to check the significant changes of selected COG functions among different biofilm groups. For example, significant differences were demonstrated for COG1135 ($P < 0.005$), COG0471 ($P < 0.005$), COG0553 ($P < 0.001$), and COG2206 ($P < 0.05$).

Similarly, a comparison between the S-NBW and brine biofilms was performed. COG functions contributing to the differences between the two biofilm groups are listed in Figure 5. The S-NBW biofilms were enriched largely for metal ion resistance-related functions, as well as some functions involved in amino acid metabolism. For example, the enriched functions included cation efflux pumps (COG0841, COG0569 and COG3696), aminopeptidase N (COG0308), branched-chain amino acid permeases (COG1114), and dipeptidyl aminopeptidases/acylaminoacyl-peptidases (COG1506). Moreover, the comparison also revealed signal transduction-related functions, such as transcriptional regulators (COG5001 and COG2909) enriched in the S-NBW biofilms.

Pathway reconstruction of two representative and dominant bacteria

To further elucidate the organization of enriched functions in the switching functions, we applied genome binning to obtain single bacterial genomes from the metagenomic datasets. As mentioned above, the abundances of *Oleispira* in the S-brine biofilms and of *Marinobacter* in the S-NBW were significantly higher than those in the brine and NBW biofilms. Thus, it can be speculated that these two bacteria possess a functional adaptation to both the brine pool and NBW. We reconstructed the genomes of the two bacteria (*Marinobacter* sp. and *Oleispira* sp.) to illustrate the functional organization of the switched biofilms (Fig. 6). The information for the two

genomes is displayed in Table S4. Following the methods described in previous publications (Albertsen *et al.* 2013; Tian *et al.* 2014), the completeness of the two genome bins was investigated using a suite of HMM encompassing 107 proteins that are conserved in 95% of all sequenced bacteria. The number of expected essential single-copy genes was further investigated by comparison with their close relatives (Table S5). Based on the results, one copy of the genes ribosomal protein L35 (TIGR00001), ribosomal protein S5 (TIGR01021), and ribosomal protein L20 (TIGR01032) may not have been recovered in the genome bin of *Marinobacter* sp. because two copies of these genes were present in the complete genome of *M. hydrocarbonoclasticus* strATCC49840. Furthermore, the presence of two GTP-binding Era proteins (TIGR00436) in all of the compared genomes indicated that its duplication in our genome bins was not due to contamination. However, 70 out of 107 essential single-copy genes were identified in the draft genome of *Oleispira* sp. Thus, we assumed that the completeness of the two draft genomes was >97% and ~65%, respectively (Table S4).

Based on the phylogenetic trees using the concentrated 31 essential genes (Fig. S6), *Marinobacter* sp. was closely related (clustered into the same branch in the phylogenetic tree) to the reported marine bacteria *M. aquaeolei* VT8, which is an “opportunotroph” that can survive under various environmental conditions (Singer *et al.* 2011). Based on the analysis of the nearly complete genome, the *Marinobacter* sp. identified in the present study possesses numerous genes for heavy metal transport, zinc, and iron uptake (Fig. 6A). Complete tricarboxylic acid (TCA) cycle and glycolysis pathways were present, suggesting a heterotrophic lifestyle. Many genes involved in resistance to oxidative stresses were identified, including catalase, superoxide dismutase, ferrochelatase, ferric-chelate reductase, and cytochrome genes. The AntiSMASH (Antibiotics & Secondary Metabolite Analysis Shell) (Medema *et al.* 2011) search also revealed a complete biosynthetic cluster of ectoine. Moreover, amino acid metabolism and utilization pathways were identified, as well as many two-component regulators, suggesting the presence of a complicated regulation network.

The *Oleispira* sp. was closely related to *O. antarctica* strain RB-8, which has been reported to be an oil-degrading bacterium (Kube *et al.* 2013). The metabolic pathways were reconstructed based on the genomic annotation of this bacterium. The results suggested this bacterium is a

heterotrophic bacterial species that utilizes a large proportion of genes to perform amino acid and lipid metabolism (Fig. 6B). For example, complete pathways for the transformation of L-glutamate to L-arginine and L-glutamine were identified. Moreover, pathways involved in dissimilatory nitrate reduction were present, suggesting the potential to use nitrate as an electron acceptor. Furthermore, heavy metal export transporters were also identified. Collectively, because of the functional versatility, the *Oleispira* sp. would also be a heterotroph that can adapt to both the brine pool and the NBW.

We further compared the potential ability to utilize different amino acids in *Marinobacter* sp., *Marinobacter aquaeolei* VT8, *Oleispira* sp., and *Oleispira antarctica* RB-8 (Table S6). These bacteria appeared to prefer different amino acids, which was consistent with the notion that different amino acids have different metabolic costs (Behrends *et al.* 2013). Moreover, other special characteristics were also identified in the genome bins of *Oleispira* sp. and *Marinobacter* sp., such as genes involved in arachidonic acid metabolism, suggesting the occurrence of gene transfer events. For example, the phylogeny of 2-hydroxy-6-oxonona-2, 4-dienedioate hydrolase (K05714, an enzyme that catalyzes a step in phenylpropanoid compound degradation) from *Oleispira* sp. is shown in Figure S7.

Expression of key functional genes

To evaluate the functional activities expressed in the switched biofilms, q-RT-PCR was performed to examine the active transcription of key genes encoding enzyme related to amino acid metabolism and metal ion stress resistance in *Marinobacter* sp. and *Oleispira* sp. (Fig. S8). Intriguingly, the results revealed an up-regulation of amino acid metabolism-related genes, such as glutamine amidotransferase and serine O-acetyltransferase, in the S-NBW biofilms for both bacteria (the number of transcripts per gene was significantly higher in the switched biofilms than in the source biofilms). Glutamine amidotransferases utilize the amide of glutamine in the biosynthesis of amino acids, coenzymes and amino sugars, and they are important for microbial nitrogen metabolism and environmental adaptation (McDonagh B *et al.* 2012). The serine O-acetyltransferase participates in cysteine metabolism and sulfur metabolism (Takagi *et al.* 2003). In contrast, metal ion stress resistance-related genes, such as the Mn²⁺/Zn²⁺ transporter and ABC-type Fe³⁺-siderophore transporter, were up-regulated in S-brine biofilms. Catalase is an

important enzyme that protects cells against oxidative damage by reactive oxygen species (ROS), which can be induced by metal ions (Scharf *et al.* 2014). Consistently, catalases were also up-regulated in the S-brine biofilms for both bacteria.

Discussion

The major aim of the present study was to test the species sorting effect in response to environmental conditions. Environment switching between the brine pool and NBW during biofilm development revealed that the composition of the biofilm community established in new environments was determined by both the source communities and new environmental conditions. Moreover, the results provided evidence that biofilm community assembly was based primarily on functional characteristics. Furthermore, metal ion resistance and amino acid metabolism-related functions were identified as major species sorting determinants in the succession of the switched biofilm communities based on evidence from integrated metagenomic and genomic analyses.

Microbial community assembly in new environments and function-based species sorting

Deterministic processes occur in bacterial assemblages across various ecosystems (Stegen *et al.* 2012; Székely *et al.* 2012). Because most of the related studies were performed in communities residing in surface regions, investigation of the species sorting hypothesis in deep sea environments has rarely been performed. The results obtained in the present study revealed the effect of species sorting by changing the incubation conditions for the immigrated biofilm communities. After the immigration of the biofilm to the new environment, the bacterial taxa had to cope with the changes caused by two main selective forces: the change in environmental parameters, and competition with local communities. The PC1 of the taxonomic PCoA plot (Fig. 3A) separated the switched biofilm from the brine and NBW biofilms, which suggested that the taxonomic structure of the switched biofilms was influenced by the new environments. Additionally, the results suggested more colonisation pressure with the NBW than the brine environments. Furthermore, the taxonomic PCoA plot also revealed a close relationship between the switched biofilms and the source communities, suggesting source-dependent community assembly, which could be attributed to the observation that many taxa possess a high degree of functional plasticity.

Interestingly, the PCoA plot and cluster analysis derived from the COG functions (Fig. 3B) revealed a tight relationship between the S-brine and S-NBW biofilms, which were clearly separated from the brine and NBW biofilms. This result led us to propose that only taxa with adaptive functions to both the brine pool and NBW can survive and proliferate. This notion was also consistent with the observation that microbial community assembly is based on functional rather than species composition, as demonstrated in a previous study investigating communities associated with algae (Burke *et al.* 2011). Although a similar phenomenon was observed between the previous and the present study, the underlying mechanisms would be different. Bacteria in communities associated with algae were selected from the same species pool, and the variations in community structure could be explained by the “lottery” model (Sale *et al.* 1976; Burke *et al.* 2011). That is, within a group of species with similar ecological functions, the “lottery” for space is won by whichever species arrives first; space colonization is random from within a group of species with equivalent functions; there should be a core functions consistently present in all communities of a particular habitat. However, for cold seep biofilms, the S-brine and S-NBW biofilms contain different source communities. The species were further selected from two different species pools, which would be the main causes of their different taxonomic structures. Therefore, bacteria inhabiting the S-brine and S-NBW biofilms should share “core ecological functions” that confer the ability of the bacteria to adapt to both environments.

Metal ion resistance and amino acid metabolism as species sorting determinants

The metagenomic comparison revealed an enrichment for metal ion stress resistance and amino acid metabolism-related genes for the switched biofilms, which suggests that these functions were important for bacterial survival in both the brine pool and NBW (Fig. 4 & 5). Consistently, the two representative and dominant bacteria, *Marinobacter* sp. and *Oleispira* sp., showed complete pathways for metal ion transport and resistance as well as amino acid metabolism and protein biosynthesis (Fig. 6). In particular, due to their ability to resist metal ions, *Marinobacter* sp. and *Oleispira* sp. might be versatile generalists that have adapted to the environments of the cold-seep system for a long time. By coordinating the metagenomic analyses and the environmental parameters, we propose that both habitat specialists (such as autotrophs that rely on particular ions as electron donors in the brine pool) and generalists were present in the 3d-NBW biofilms; after immigration of the biofilms from the NBW to the brine pool, only the

generalists in the biofilm with metal ion stress resistance-related functions could survive and proliferate in the S-brine biofilms. Similarly, for the S-NBW biofilms, heterotrophic carbon utilization and amino acid metabolism were important for the community members to be competitive and proliferate after being moved to the NBW (Fig. 7). These results could be further supported by the expression levels of the ion transporter genes and amino acid metabolism-related genes (Fig. S8). Notably, the up-regulation of the catalase of *Marinobacter* sp. in the brine suggested the occurrence of oxidative stress caused by the high metal ion concentration. The importance of amino acids could be explained by the conclusion that amino acids play roles in the bacterial response to numerous stresses and adaptation to new environments (Feehily *et al.* 2013; Horcajo *et al.* 2012; Behrends *et al.* 2013). For example, species in several bacterial genera use the glutamate decarboxylase (GAD) system, which facilitates intracellular pH homeostasis by consuming protons to cope with acidic environmental condition (Feehily *et al.* 2013).

Results of the present study suggest that *Marinobacter* and *Oleispira* are habitat generalists that can survive in both the brine pool and the NBW. Genome analyses of the bacterium *M. aquaeolei* VT8 revealed its metabolic potential to utilize oxygen and nitrate as terminal electron acceptors, iron as an electron donor, and defense mechanisms acquired via horizontal gene transfer that are involved in the perception of environmental fluctuations and antibiotics, phages, toxins, and heavy metal resistance (Singer *et al.* 2011). *O. antarctica* has been reported to be a hydrocarbonoclastic marine bacterium (Kube *et al.* 2013) and possesses an array of amino acid metabolism, alkane degradation, lipid degradation, and siderophore production pathways. However, the functional and genetic features of *Marinobacter* and *Oleispira* remain largely uncharacterized due to limited genomic information. Genome analysis of the *Marinobacter* sp. and *Oleispira* sp. binned in the present study suggest that they adopted a large proportion of genes for amino acid metabolism and resistance to oxidative and metal ion stresses. The ectoine, which may be produced by *Marinobacter* sp., is a compound that is used by many taxa to cope with osmotic stress (Teixidó *et al.* 2005). The presence of the type VI secretion system might also contribute to stress resistance (Zhang *et al.* 2013a; Lin *et al.*, 2015). Moreover, unique features were also observed, such as 2-hydroxy-6-oxonona-2,4-dienedioate hydrolase (K05714) in *Oleispira* sp. This enzyme catalyses a step in the pathway of phenylpropanoid compounds degradation (Lam *et al.* 1997).

Thus, the *Oleispira* sp. and *Marinobacter* sp. have developed specialized adaptation strategies to survive in both the brine pool and the NBW. These results further elucidate the functional bases for the species sorting process by revealing the genomic and metabolic features of two dominant bacteria in the switched biofilms.

Concluding remarks

Through the integration of experimental strategies with metagenomic comparisons, the results of the present study bridge the gap in the species sorting theory by presenting functional bases for the species sorting process. We also provided indications that species sorting may explain the different community structures observed between the brine pool and the NBW. Although six types of substrates were used for biofilm development and core mechanisms were revealed in the metagenomes, additional datasets from larger-scale cold seep environment are required to confirm the general principles governing the species sorting process. Two genome bins were analyzed herein as examples of the generalists inhabiting the Thuwal cold seeps II, which provided some insight regarding bacterial adaptation strategies. Future efforts in the reconstruction of metabolic pathways in other bacteria will facilitate a better understanding of the microbial community assembly in this special habitat.

Acknowledgements

The authors are grateful to Yue Him Wong and Gen Zhang from the Hong Kong University of Science and Technology for sample collection. 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), National Basic Research Program of China (973 Program, No: 2012CB417304), a GRC grant (661611) from the HKSAR government, and the King Abdullah University of Science and Technology (SA-C0040/UK-C0016) to P.Y. Qian.

References

- Albertsen M, Hugenholtz P, Skarszewski A *et al.* (2013) Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. *Nature Biotechnology*, **31**, 533-538.
- Bakker EP, Booth IR, Dinnbier U, Epstein W, Gajewska A (1987) Evidence for multiple K⁺ export systems in *Escherichia coli*. *Journal of Bacteriology*, **169**, 3743-3749.
- Batang ZB, Papathanassiou E, Al-Suwailem A *et al.* (2012) First discovery of a cold seep on the continental margin of the central Red Sea. *Journal of Marine Systems*, **94**, 247-253.
- Behrends V, Ryall B, Zlosnik JEA *et al.* (2013) Metabolic adaptations of *Pseudomonas aeruginosa* during cystic fibrosis chronic lung infections. *Environmental Microbiology*, **15**, 398-408.
- Burke C, Peter S, Doug R *et al.* (2011) Bacterial community assembly based on functional genes rather than species. *Proceedings of the National Academy of Sciences*, **108**, 14288-14293.
- Caporaso JG, Kuczynski J, Stombaugh J *et al.* (2010) QIIME allows analysis of high-throughput community sequencing data. *Nature Methods*, **7**, 335-336.
- Clarke KR (1993) Non-parametric multivariate analyses of changes in community structure. *Austral Ecology*, **18**, 117-143.
- Castresana J (2000) Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular Biology and Evolution*, **17**, 540-552.
- Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, **32**, 1792-1797.
- Feehily C, Karatzas, KAG (2013) Role of glutamate metabolism in bacterial responses towards acid and other stresses. *Journal of Applied Microbiology*, **11**, 11-24.
- Hall BG (2013) Building phylogenetic trees from molecular data with MEGA. *Molecular Biology and Ecology*, **30**, 1229-1235.
- Halm H, Lam P, Ferdelman TG *et al.* (2011) Heterotrophic organisms dominate nitrogen fixation in the South Pacific Gyre. *The ISME Journal*, **6**, 1238-1249.
- Hammer Ø, Harper DAT, Ryan PD (2001) PAST: Paleontological statistics software package for education and data analysis. *Palaeontol Electron*, **4**, 1-9.
- Horcajo P, de Pedro MA, Cava F (2012) Peptidoglycan plasticity in bacteria: stress-induced peptidoglycan editing by noncanonical D-amino acids. *Microbial Drug Resistance*, **18**, 306-313.
- Hyatt D, Chen GL, Locascio PF *et al.* (2010) Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics*, **8**, 119.

Jones SE, McMahon KD (2009) Species-sorting may explain an apparent minimal effect of immigration on freshwater bacterial community dynamics. *Environmental Microbiology*, **11**, 905-913.

Kanehisa M, Susumu G (2000) KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Research*, **28**, 27-30.

Kube M, Chernikova TN, Al-Ramahi Y *et al.* (2013) Genome sequence and functional genomic analysis of the oil-degrading bacterium *Oleispira antarctica*. *Nature Communications*, **4**, 2156.

Lam WW, Bugg TD (1997) Purification, characterization, and stereochemical analysis of a C-C hydrolase: 2-hydroxy-6-keto-nona-2,4-diene-1,9-dioic acid 5,6-hydrolase. *Biochemistry*, **36**, 12242-12251.

Langenheder S, Székely AJ (2011) Species sorting and neutral processes are both important during the initial assembly of bacterial communities. *The ISME Journal*, **5**, 1086-1094.

Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nature Methods*, **9**, 357-359.

Lee OO, Yang J, Bougouffa S *et al.* (2012) Pyrosequencing reveals spatial and species variations in bacterial communities associated with corals from the Red Sea. *Applied and Environmental Microbiology*, AEM-01111.

Li H, Handsaker B, Wysoker A *et al.* (2009) The sequence alignment/map format and SAMtools. *Bioinformatics*, **25**, 2078-2079.

Li SJ, Hua ZS, Huang LN *et al.* (2014) Microbial communities evolve faster in extreme environments. *Scientific Reports*, **4**, 6205.

Lin J, Cheng J, Chen K *et al.* (2015) The icmF3 locus is involved in multiple adaptation-and virulence-related characteristics in *Pseudomonas aeruginosa* PAO1. *Frontiers in Cellular and Infection Microbiology*, **5** doi: 10.3389/fcimb.2015.00070.

Mason OU, Hazen TC, Borglin S *et al.* (2012) Metagenome, metatranscriptome and single-cell sequencing reveal microbial response to Deepwater Horizon oil spill. *The ISME Journal*, **6**, 1715-1727.

McDonagh B, Domínguez-Martín MA, Gómez-Baena G *et al.* (2012) Nitrogen starvation induces extensive changes in the redox proteome of *Prochlorococcus* sp. strain SS120. *Environmental Microbiology Reports*, **4**, 257-267.

Medema MH, Blin K, Cimermanic P *et al.* (2011) antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic Acids Research*, **39**, 339-346.

Meier A, Tsaloglou, NM, Mowlem MC, Keevil CW, Connelly DP (2013) Hyperbaric biofilms on engineering surfaces formed in the deep sea. *Biofouling*, **29**, 1029-1042.

Orsini L, Vanoverbeke J, Swillen I, Mergeay J, Meester L (2013) Drivers of population genetic differentiation in the wild: isolation by dispersal limitation, isolation by adaptation and isolation by colonization. *Molecular Ecology*, **22**, 5983-5999.

Öztürk B, de Jaeger L, Smidt H *et al.* (2013) Culture-dependent and independent approaches for identifying novel halogenases encoded by *Crambe crambe* (marine sponge) microbiota. *Scientific Reports*, **3**, 2780.

Patel RK, Jain M (2012) NGS QC Toolkit: a toolkit for quality control of next generation sequencing data. *PloS one*, **7**, e30619.

Peay KG, Schubert MG, Nguyen NH, Bruns TD (2012) Measuring ectomycorrhizal fungal dispersal: macroecological patterns driven by microscopic propagules. *Molecular Ecology*, **21**, 4122-4136.

Sale PF (1976) Reef fish lottery. *Natural History*, **85**, 60-65.

Scharf B, Clement CC, Zolla V, *et al.* (2014) Molecular analysis of chromium and cobalt-related toxicity. *Scientific Reports*, **4**, 5279.

Singer E, Webb EA, Nelson WC *et al.* (2011) Genomic potential of *Marinobacter aquaeolei*, a biogeochemical “opportunotroph”. *Applied and Environmental Microbiology*, **77**, 2763-2771.

Stegen JC, Lin X, Konopka AE *et al.* (2012) Stochastic and deterministic assembly processes in subsurface microbial communities. *The ISME Journal*, **6**, 1653-1664.

Sullam KE, Essinger SD, Lozupone CA *et al.* (2012) Environmental and ecological factors that shape the gut bacterial communities of fish: a meta-analysis. *Molecular Ecology*, **21**, 3363-3378.

Székely AJ, Berga M, Langenheder S (2012) Mechanisms determining the fate of dispersed bacterial communities in new environments. *The ISME Journal*, **7**, 61-71.

Tamura K, Stecher G, Peterson D, Filipinski A, Kumar S. (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution* **30**, 2725-2729.

Takagi H, Yoshioka K, Awano N, *et al.* (2003) Role of *Saccharomyces cerevisiae* serine O-acetyltransferase in cysteine biosynthesis. *FEMS Microbiology Letters*, **218**, 291-297.

Tatusov RL, Galperin MY, Natale DA *et al.* (2000) The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Research*, **28**, 33-36.

Teixidó N, Cañamás TP, Usall J *et al.* (2005) Accumulation of the compatible solutes, glycine–betaine and ectoine, in osmotic stress adaptation and heat shock cross-protection in the biocontrol agent *Pantoea agglomerans* CPA-2. *Letters in Applied Microbiology*, **41**, 248-252.

Tian RM, Wang Y, Bougouffa S *et al.* (2014) Genomic analysis reveals versatile heterotrophic capacity of a potentially symbiotic sulfur oxidizing bacterium in sponge. *Environmental Microbiology*, **16**, 3548-3561.

Wang Q, Garrity GM, Tiedje JM, Cole JR (2007) Naïve bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Applied and Environmental Microbiolog*, **73**, 5261-5267.

Wu M, Eisen JA (2008) A simple, fast, and accurate method of phylogenomic inference. *Genome Biology*, **9**, R151.

Zhang W, Wang Y, Song Y *et al.* (2013a) A type VI secretion system regulated by OmpR in *Yersinia pseudotuberculosis* functions to maintain intracellular pH homeostasis. *Environmental Microbiology*, **15**, 557-569.

Zhang W, Wang Y, Lee OO *et al.* (2013b) Adaptation of intertidal biofilm communities is driven by metal ion and oxidative stresses. *Scientific Reports*, **3**, 3180.

Zhang W, Wang Y, Tian RM *et al.* (2014) Species sorting during biofilm assembly by artificial substrates deployed in a cold seep system. *Scientific Reports*, **4**, 6647.

Zhang W, Wang Y, Bougouffa S *et al.* (2015) Synchronized dynamics of bacterial niche-specific functions during biofilm development in a cold seep brine pool. *Environmental Microbiology*, doi: 10.1111/1462-2920.12978.

Data accessibility

The raw Illumina sequences, metagenome assembly, metagenome annotation, comparative metagenome data, 16S 454 sequences, aligned 16S sequences, OTU table, SIMPER data, q-RT-PCR data, aligned sequences for tree construction, and genome bins of *Marinobacter* sp. and *Oleispira* sp. are documented in Dryad doi:10.5061/dryad.9ng87.

Figure legends

Figure 1. Schematic model of the experimental design. The biofilms were developed separately in the brine and NBW for three days before being moved to the NBW and brine, respectively, and incubated for an additional three days (“S-brine” for biofilms from NBW to brine and “S-NBW” for biofilms from brine to NBW). The biofilms were developed on six different substrates: aluminum (Al), polyether ether ketone (PEEK), polyvinyl chloride (PVC), polytetrafluoroethylene (PTFE), stainless steel (SS) and titanium (Ti).

Figure 2. Taxonomic classification of qualified bacterial reads retrieved from switched biofilm samples. Operational taxonomic units (OTUs) with an identity of 97% were classified at the genus level using the ribosomal database project (RDP) classifier in the Quantitative Insights into Microbial Ecology (QIIME) pipeline. Genera with a relative abundance >1% in at least one replicate and mean values calculated from 2 replicates are shown. The genera that did not achieve an abundance of greater than 1% in at least one of two replicates were defined as “minor”. A,

S-brine biofilms; B, S-NBW biofilms. Bacterial genera differentiating the switched and source biofilm communities are shown in Figure S3 and S4.

Figure 3. Similarity of the microbial communities in the switched biofilm samples, as illustrated by the taxonomic composition (A) and clusters of orthologous groups (COG) annotation (B). The eigenvalues (percentage of variance for the first two principal components) are indicated in the figure. The colors represent biofilms on different materials: purple for Al, blue for PEEK, bright green for PVC, dark green for PTFE, orange for SS and red for Ti.

Figure 4. SIMPER analysis of the COGs from the S-brine and source biofilms (3d-NBW and 6d-NBW). Comparisons were performed based on the metagenomic annotation. COGs are presented in the order of their contribution to the difference (highest to lowest, top to bottom). Inorganic ion transport and metabolism (P), and amino acid transport and metabolism (E) are indicated in yellow and red, respectively.

Figure 5. SIMPER analysis of the COGs from the S-NBW and source biofilms (3d-brine and 6d-brine). Comparisons were performed based on the metagenomic annotation. COGs are presented in the order of their contribution to the difference (highest to lowest, top to bottom). Inorganic ion transport and metabolism (P), and amino acid transport and metabolism (E) are indicated in yellow and red, respectively.

Figure 6. Genomic construction of *Marinobacter* sp. (A) and *Oleispira* sp. (B). The pathways related to amino acid metabolism, metal ion and oxidative stress resistance and the bacterial secretion system are shown. Complete pathways for type I, II and VI secretion system assembly and ectoine biosynthesis were also identified.

Figure 7. Schematic model of species sorting during the establishment of switched biofilm communities. After the biofilms immigrated to new environments, generalists with the capacity for amino acid metabolism and metal ion stress resistance became more competitive and proliferated.













