



OPEN

Detailed investigation of the microbial community in foaming activated sludge reveals novel foam formers

SUBJECT AREAS:
MICROBIAL ECOLOGY
MOLECULAR ECOLOGY

Feng Guo, Zhi-Ping Wang, Ke Yu & T. Zhang

Received
22 April 2014Accepted
1 December 2014Published
6 January 2015Correspondence and
requests for materials
should be addressed to
T.Z. (zhangt@hku.hk)

Environmental Biotechnology Laboratory, Department of Civil Engineering, The University of Hong Kong, Pokfulam Road, Hong Kong SAR, China.

Foaming of activated sludge (AS) causes adverse impacts on wastewater treatment operation and hygiene. In this study, we investigated the microbial communities of foam, foaming AS and non-foaming AS in a sewage treatment plant via deep-sequencing of the taxonomic marker genes 16S rRNA and mycobacterial *rpoB* and a metagenomic approach. In addition to *Actinobacteria*, many genera (e.g., *Clostridium* XI, *Arcobacter*, *Flavobacterium*) were more abundant in the foam than in the AS. On the other hand, deep-sequencing of *rpoB* did not detect any obligate pathogenic mycobacteria in the foam. We found that unknown factors other than the abundance of *Gordonia* sp. could determine the foaming process, because abundance of the same species was stable before and after a foaming event over six months. More interestingly, although the dominant *Gordonia* foam former was the closest with *G. amarae*, it was identified as an undescribed *Gordonia* species by referring to the 16S rRNA gene, *gyrB* and, most convincingly, the reconstructed draft genome from metagenomic reads. Our results, based on metagenomics and deep sequencing, reveal that foams are derived from diverse taxa, which expands previous understanding and provides new insight into the underlying complications of the foaming phenomenon in AS.

Foaming at water surfaces occurs either chemically or biologically¹. Foaming is a problem for sewage treatment plants (STPs) worldwide and frequently occurs in the aeration tanks, where it is caused by air bubbles, (bio)surfactants and a high abundance of hydrophobic bacteria². Typical foam is a stable, bubbling layer floating on the water surface, with a thickness ranging from centimeters to more than one meter³. It may result in operation failure, deterioration of effluent quality and loss of biomass⁴.

A high abundance of hydrophobic bacterial cells in the water film is essential for the stability of foam². For a long time, before molecular methods were widely adopted^{5,6}, foaming bacteria were mainly identified and quantified by staining and culture-based methods⁷. During the last two decades, 16S rRNA gene-based methods, such as denaturing gradient gel electrophoresis (DGGE)⁸ and fluorescent *in situ* hybridization (FISH)⁹, were the mainstream methods for identifying the foaming species. Previous studies have shown that most foaming bacteria belong to the mycolic acid-producing *Actinomycetales* (mycolata group), whose cell walls usually exhibit high hydrophobicity^{2,3}. Within that group, *Gordonia amarae* is the most well-known foam former, due to its high frequency and abundance in foaming activated sludge (AS) worldwide³. Although a few researchers have reported diverse foam formers in AS other than the mycolata group^{10–12}, the candidates are much less well-known than the mycolata group due to their relatively low abundance and the low resolution of the identification methods used. The foaming threshold for different bacteria might vary, normally ranging from 10⁷ to 10⁹ cells per ml, according to a laboratory study based on isolates². With around 10¹⁰ cells in one milliliter of activated sludge mixed liquor¹³, it could roughly be inferred that only a 0.1% cell abundance of foam formers may be sufficient for foaming. Unfortunately, microbes with such a low abundance are difficult to examine in a community survey by traditional molecular methods, due to the methodological detection limits.

In addition to operating concerns, foaming may also cause hygienic risks³. First, it may release more microorganisms into the environment via the effluent. Second, but more importantly, aerosols from the broken foam may spread into the air if the tanks are not covered. In the mycolata group, the genus *Mycobacterium*, which is usually highly resistant to dehydration¹⁴, contains well-known pathogenic species. Unlike activated sludge and effluent, an aerosol transmission may introduce pathogens onto/into the human body without physical contact, posing a public health risk. Concerning *Mycobacterium*, no study has resolved its population down to the species level for activated sludge samples, due to the low resolution of the 16S rRNA gene^{15–16}. Sequencing of the *rpoB* and



hsp65 is an unambiguous way to obtain mycobacterial species-level assignments^{17,18} and provides a powerful tool for examining mycobacteria in activated sludge and its foam.

Although the foaming of activated sludge is of significance in application and microbial ecophysiology, to the best of our knowledge, there is no high-resolution survey on the microbial community and pathogeny of foaming activated sludge. High-throughput sequencing techniques and advanced data processing are refining our understanding of the microbial world^{19–20}. Deep-sequencing-based metagenomic analysis provides opportunities to obtain comprehensive genetic information. In this study, our AS samples were collected from the Shatin STP in Hong Kong, which we have monitored its microbial community for over five years²¹. Long-term observations indicated that foaming regularly occurred in this STP each year in the early spring and the dominant foam former should be the species in *Gordonia*²¹. However, the detailed microbial community and potential mycobacterial pathogens in foam and foaming AS are unclear. Therefore, high-throughput sequencing was performed to obtain detailed community information from foam samples, particularly by comparing the foam with the corresponding AS, and by comparing the foaming AS with non-foaming AS that was collected before and after the foaming event. With the high-resolution profile of detailed microbial community, our study will expand the understanding of foam formers in activated sludge and unravels the potential microbial resources (e.g., biosurfactants) in the foam.

Results

High-throughput sequencing. The summary of the high-throughput sequencing datasets obtained in this study is listed in Table S1. For the 16S rDNA amplicons, 10,013–19,229 cleaned pyrotags (>300 bp) were obtained for the eight samples (i.e., AS-I, Foam-I, AS-II, Foam-II and AS samples collected in January, February, June and July of 2010). After trimming the primers, 14,427 and 11,554 *rpoB* amplicon cleaned pyrotags were obtained for the Foam-I and Foam-II samples, respectively. For the metagenomic sequencing, approximately 33 million and 31 million paired-end reads and were obtained for both the Foam-II and AS-II samples. Both datasets contained over 12 million Illumina tags (itags) (>130 bp), in which 8,888 and 8,214 were identified as 16S itags for Foam-I and AS-I, respectively.

Bacterial communities in the foam and the foaming AS. *Archaea* only accounted for approximately 0.2% of the total 16S itags in both the Foam-II and AS-II samples, thus we focused only on the bacterial community. As shown in Fig S1, the largest divergence between the foam and the AS at the phylum level was a much higher abundance of *Actinobacteria* present in the foam (1.6 to 2.0 fold higher in foam than in AS for different samples/methods; one-sided Fisher's exact test, corrected $P=0.015$ and 0.024 and for Foam-I & AS-I and Foam-II & AS-II datasets, respectively). In addition to this major difference for *Actinobacteria*, other differences were also observed between the foam and AS samples. For example, *Firmicutes* were also more likely to be present in foam than in AS (1.3 to 2.2 fold higher in foam than in AS), and *Nitrospira*, a major lineage of the nitrite-oxidizing bacteria (NOB) that are vital for biological wastewater treatment, had a quite low abundance in the foam (9.8 fold higher in AS-II pyrotags than in the Foam-II pyrotags). Additionally, among the *Proteobacteria*, which was the most abundant phylum in both the foam and foaming AS, the AS-I and Foam-I were dominated by *Alphaproteobacteria*, while the AS-II and Foam-II had comparable abundances of the *Alpha*, *Beta* and *Gamma* subdivisions, suggesting an inter-annual variation of the foaming bacterial community for the WWTP. Interestingly, there were some differences between the pyrotags of the 16S rRNA gene amplicons and the corresponding metagenomic datasets. The metagenomic data without PCR amplification revealed more phyla, such as *Planctomycetes*, *Verrucomicrobia*,

and others, which may be biased during PCR using the universal primer sets used in the present study²².

The comparison of the bacterial communities in the foam samples and the AS samples, for the top 32 genera that were over 0.5% abundance in at least one sample, is displayed in Fig 1. By RDP Classifier, only approximately 35% of pyrotags from AS and 40–50% pyrotags from foam could be assigned into certain genera. *Gordonia* was dominant in all samples (1.5%–25.7%). Within the mycolata groups other than *Gordonia*, *Mycobacterium* (1.3%–3.5%) and *Williamsia* (0.3%–0.9%) were also present at relatively high abundances, although *Williamsia* was only found in the metagenomic data, which suggests a potential false negative for this genus introduced by PCR amplification.

The distinct distributions of these genera between the foam and the AS is shown on the bottom of Fig 1. *Gordonia*, *Mycobacterium*, *Clostridium* XI, *Simplicispira*, *Arcobacter*, *Flavobacterium* and *Williamsia* were consistently higher in the foam samples than in the AS samples. Approximately 3.0- to 4.5-fold more *Gordonia* were detected in the foam than in the AS (one-sided Fisher's exact test, corrected $P=0.001$ for both Foam-I & AS-I and Foam-II & AS-II datasets), while the difference for *Mycobacterium* was less than *Gordonia* (1.3–2.8 fold). The enrichment ratio of *Gordonia* was quite similar to that in a previous report, which was based on a FISH approach²³. The enrichment of these genera in the foam suggests their hydrophobic feature and roles in foam formation.

Intriguingly, some bacteria were negatively selected in the foam (i.e., less abundant in the foam than in the AS), including *Caldilinea* (54.7% less in Foam-I than in AS-I referring to the pyrosequencing data), *Ilumatobacter* (47.4%), *Phyllobacterium* (38.3%), *Bradyrhizobium* (55.1%) and *Nitrospira* (90.5%). It is especially interesting that many nitrogen-metabolism specialists had a much lower abundance in the foam, including the nitrite oxidizing bacteria (NOB) *Nitrospira* and the ammonia oxidizing bacteria (AOB) *Nitrosomonas* (not listed in Fig 1, due to the low abundance). A shared characteristic of these nitrogen-metabolism specialists is that they usually grow in clusters. Their low presence in the foam might be due to their strong hydrophilicity or their cluster-growth pattern. The microscopic image shown in Fig S2 also suggests that the cluster-growing organisms were rarely observed in the foam.

High abundance of *Gordonia* in non-foaming samples suggested that other factors determined foaming. In addition to the comparison between foam and the corresponding AS, the bacterial communities in the foaming AS and non-foaming AS samples, collected before and after the foaming event in March 2010 were also examined. The variations of 39 OTUs that had over 0.5% abundance in the foaming AS are summarized in Fig 2. First, it was not expected that the abundance of *Gordonia* would be quite stable across the sampling period of about six months (from 3.4% to 6.2%, before and after, and 4.5% during the foaming event), despite the absence of observable stable foaming in the other four samples. Further examination confirmed that the major *Gordonia* species in July 2010 were the same as in March 2010 (over 80% of the *Gordonia* 16S rRNA sequences in March and July could be clustered into one OTU, at a similarity of 98%). Thus, this result indicated that the foaming event was not solely determined by the abundance of *Gordonia* in AS.

As shown in Fig 2, there were ten OTUs with relatively high abundance only in the foaming events, with low abundance in the other four samples (at least two-fold higher in the foaming AS than the highest level of the non-foaming AS). Interestingly, five of them were affiliated with *Flavobacteriaceae*, including one identified as *Winogradskyella* and the other four unclassified at the genus level. These OTUs were nearly undetectable in the non-foaming AS samples. In addition, two OTUs were identified as *Saprospiraceae*, which is a common filamentous bacteria in AS²⁴. Most of these OTUs were

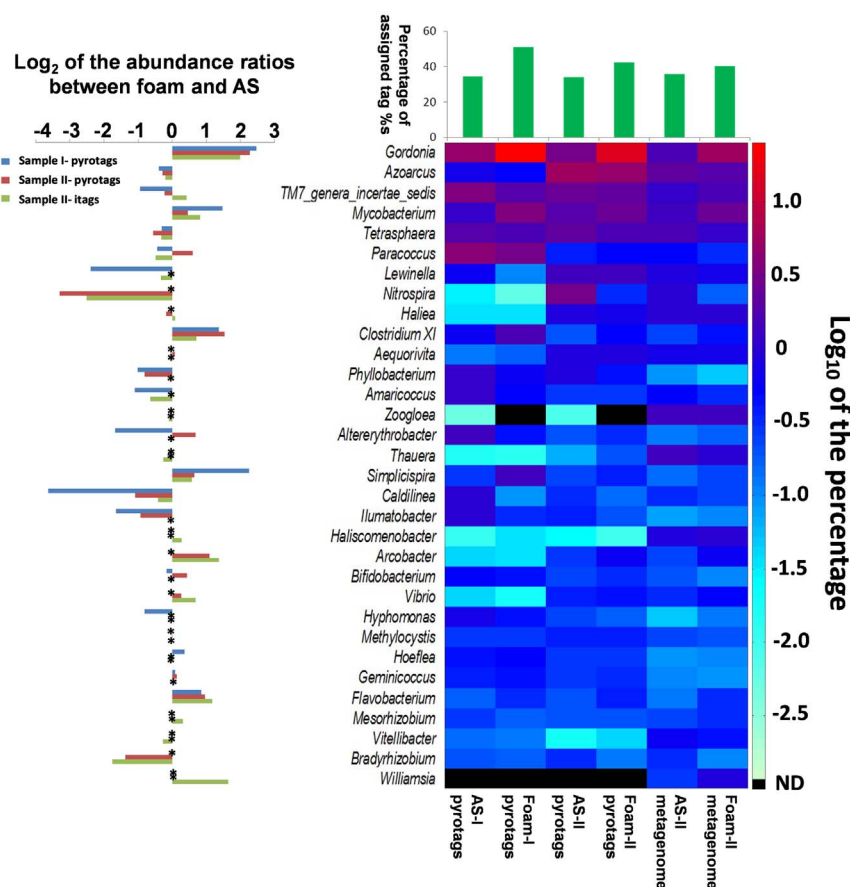


Figure 1 | Abundance and divergence of the major genera in foam samples and corresponding AS samples. The taxonomic classification was performed by RDP Classifier at confidence thresholds of 80% (for 454 pyrotags over 300 bps) and 50% (for 16S rDNA itags, 130–190 bp). The heatmap lists the top 32 genera over 0.5% abundance in at least one sample. The column above the heatmap shows percentages of tags assigned into a genus. The bar chart on the left exhibits the divergence of distribution for each genus between the foam samples and the corresponding foaming AS samples (using the logarithmic ratio between them). Only the ratios of those genera over 0.5% in foam or the corresponding AS are shown; for the others, the chart is marked by an asterisk. ND means “not detected”.

not significantly enriched in foam during the foaming event (data not shown), which suggested that their cells might not be hydrophobic enough. We also characterized the variation of *Gordonia* around the foaming event of 2012. However, the abundances of *Gordonia* in the AS were very low before and after that foaming event, indicating a case-by-case difference and the instability of specific populations.

Mycobacterial population in the foam. To identify and quantify the mycobacteria in Foam-I and Foam-II, a total of 11,554 and 14,427 *rpoB* pyrotags were generated, respectively. They were subjected to BLASTn against a customized database containing the collections of *rpoB* sequences of 131 reference *Mycobacterium* species from the NCBI database. According to a previous survey of 44 species, the inter-species divergence of the targeting *rpoB* segments was larger than 4% for most cases¹⁷. A pairwise comparison between any two species in the database showed that only 7% of all inter-species pairs were higher than 96% similarity (Fig S3). Using a <4% divergence cut-off, the two samples only had 301 (out of 11,554) and 1,177 (out of 14,427) validated hits with close similarities to the reference *rpoB* sequences, respectively. As shown in Fig 3, most of the *rpoB* pyrotags had a lower than 96% similarity to any reference *rpoB* sequences. Considering the possibility of non-targeted amplification of the primer sets used, 300 *rpoB* pyrotags from each sample were randomly picked for BLASTn against the online NCBI (NT) database to estimate how many sequences belonged to *Mycobacterium*. Manually checking results indicated that 28% and 37% sequences from samples Foam-I and -II, respectively, were the most similar to a

mycobacterial *rpoB* reference, suggesting that about one-third to one-fourth of the total hit tags were potentially derived from *Mycobacterium* spp.

Interestingly, no slow-growing mycobacteria that are associated with all the well-known obligate pathogenic species in this genus were detected. This showed that the foam samples in the present study were inhabited by no, or very few, pathogenic mycobacteria, given that the approximate detection limit of the methods used in the present study was approximately 1 in 100,000 bacterial cells (for 3% *Mycobacterium* in the total bacterial population and a 3,000 *rpoB* sequencing depth for this genus). *M. crocinum* related organisms dominated in both Foam-I and Foam-II. *M. vanbaalenii* and *M. aromaticivorans* related organisms were also abundant in Foam-I. All the three reference species have been reported as polycyclic aromatic hydrocarbon (PAH) degraders, and *M. vanbaalenii* has been reported to contain a novel *pht* operon for PAH degradation^{25–26}.

The dominant foam former was an undescribed organism other than ‘*Gordonia amarae*’. To identify the dominant *Gordonia* foam former, we first examined the near full-length 16S rRNA gene of the organism retrieved by constructing a clone library (using *Actinobacteria*-specific 243F: 5'-GGATGAGCCCGCGGCTA-3' and the universal 1492R: 5'-TACCTTGTTACGACTT-3' primer sets during PCR). Unexpectedly, the phylogenetic reconstruction based on the full length rRNA gene (approximately 7% 16S rRNA pyrotags in the Foam-II sample had >99.5% similarity with it) suggested that this organism was distantly related to the reference

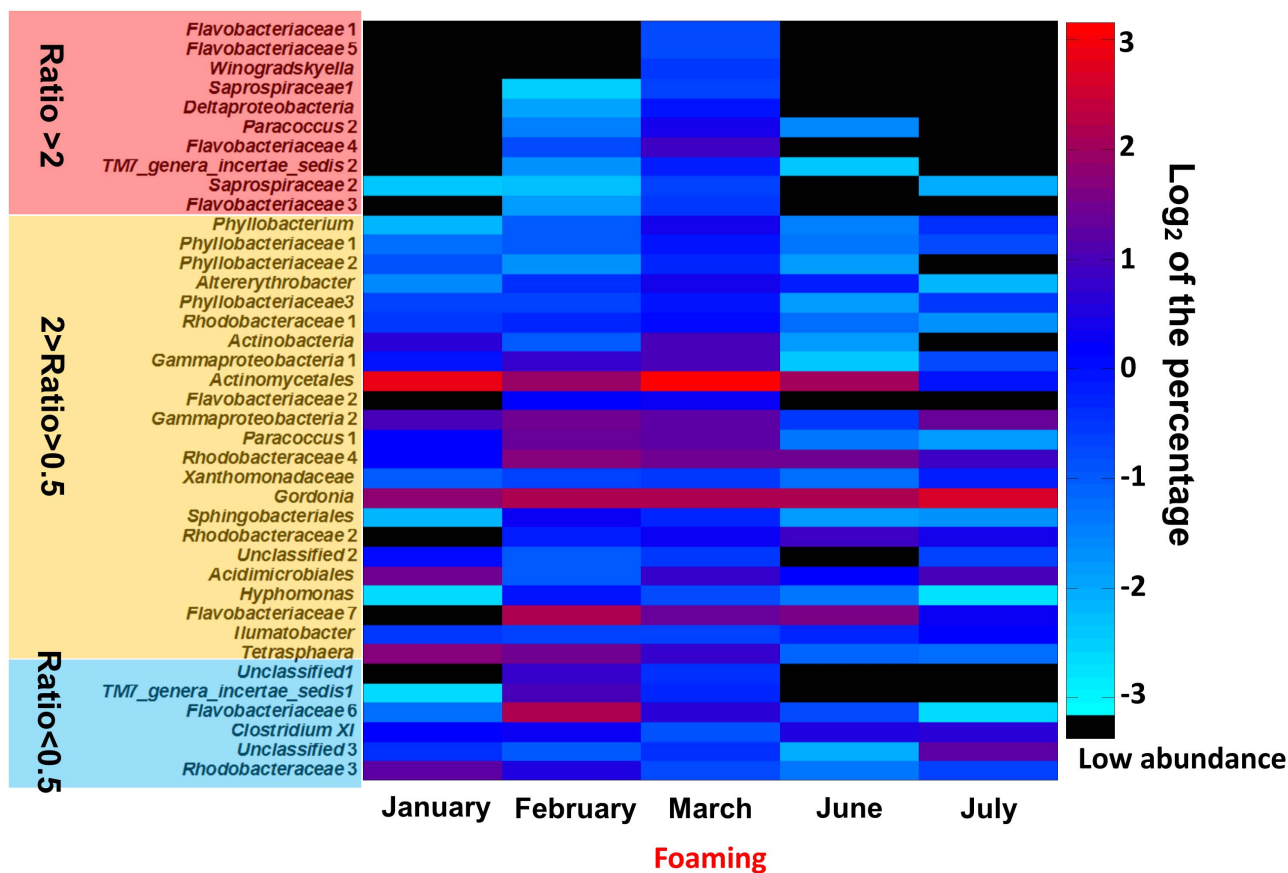


Figure 2 | Variations of the 39 OTUs in AS before, during and after the foaming events in March 2010. These OTUs had more than 0.5% abundance in the foaming AS. A black block in the heatmap means less than 0.1% or not detected. The list of the OTUs (from top to bottom) was determined by the ratio of abundance in the foaming AS and the highest abundance of the non-foaming AS samples. The taxonomic names are the lowest classified level in RDP Classified at a confidence threshold of 80%, by one representative sequence. The number after the taxonomic name indicates that more than one OTU could be assigned to the taxa and the number was sorted by abundance in the foaming AS.

G. amarae strains, including the original strain DSM43392 and other *G. amarae* strains (Fig 4), although it was more divergent from the other *Gordonia* spp. The previous investigation showed that the inter-species divergences of 16S rRNA in *Gordonia* were between

0.2% to 6.0%²⁷. The range of divergence in our analysis ranged from 0 to 4.3%; the difference with the reference can be explained by the fact that a different sequence fragment was used²⁷. We found that the retrieved sequence was the closest to *G. amarae* DSM43392

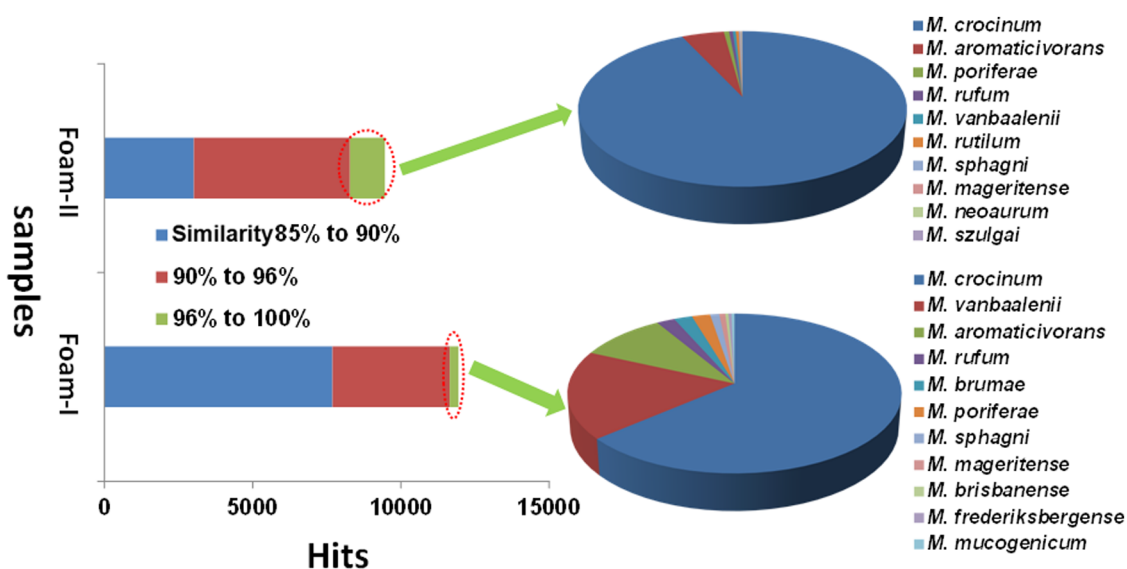


Figure 3 | *Mycobacterial* population in two foam samples by deep-sequencing of the *rpoB* gene amplicons. Only the validated sequences that had more than 96% similarity to a reference were identified as the reference related mycobacteria.



with a distance of 1.9% that was beyond the recently proposed species-level similarity of 98.65%²⁸. It was more divergent from the other *Gordonia* species (2.3% to 3.7%). Moreover, the retrieved sequence does not form a reliable cluster with the *G. amarae* strains (Fig 4). This, therefore, implied that the organism could represent a novel species that is related to the prototypical *G. amarae* DSM43392. Furthermore, the assembled *gyrB* sequence in the draft genome was verified (complete identity) by PCR-clone sequencing of a 1,245 bp *gyrB* amplicon (GenBank accession: KJ021036.1). A phylogenetic tree based on *gyrB* also showed that the organism present was distantly related to the two *G. amarae* references (divergences of 11% to 13%), although the three sequences formed a clade supported by the bootstrap analysis (Fig. S4). The divergence of 11.7% between the assembled sequence and *gyrB* of *G. amarae* DSM 43392 was within the range of inter-species distances (2.4% to 26.4%). Compared with the intra-species divergences of this gene in other *Gordonia* species (<8% in all listed cases; and for most, less than 2%; except for some obviously wrong taxonomic assignments, see Fig S4), the distances also suggested the three sequences (KJ021036.1, AB075554.1 and AB438180.1) could be derived from different species.

The genome bin of the uncultured *Gordonia* sp. enriched in foam was well separated from most other contigs using the two dimensional coverage binning pipeline (Fig 5). The general information for the reconstructed draft genome and other *Gordonia* genomes is summarized in Table S2. The reconstructed draft genome was 3.8 M containing 1,493 contigs (over 1 kb length) and 4,721 protein encoded genes with a GC content of 65%. There were 83 essential single-copy genes (ESCGs) detected in the reconstructed genome,

and three of them were found in two copies. However, further examination strongly suggested that those multi-copies are not contaminants (Table S3). For the species of *Actinobacteria*, the number of ESCGs in a genome was empirically determined to be 105²⁹. Thus, the estimated genome completeness was about 75–80% with little or no contamination.

Based on the similarities between amino acid sequences, the draft genome was the closest to *G. amarae* NBRC15530 (just the same strain as DSM43392, but in a different code). As shown in Fig 6A, 1,537 open reading frames (ORFs, 32.7% of total) in *G. amarae* NBRC15530 had at least a relative with over 90% similarity in the reconstructed genome, while the numbers of such ORFs for the other *Gordonia* species were only 59 to 205. However, a large number of ORFs (26.7%) in the reconstructed genome had less than 50% similarity or lacked homologs to the ORFs from NBRC15530, supporting the conclusion that the organism was only distantly related to the prototypical *G. amarae* and represented a novel species of *Gordonia* (Fig 6B). The average amino acid identity (AAI) between the different strains within *G. terrae* or *G. polyisoprenivorans* was quite higher than that between the reconstructed genome and the prototypical *G. amarae* NBRC15530. Higher similarity was observed between two distinct species, *G. namibiensis* NBRC108229 and *G. rubripertincta* NBRC101908, which were found to be closely related, according to the *gyrB* phylogenetic placement (Fig S4 & Fig 6B). In addition, the average nucleotide identity (ANI) between the draft genome and *G. amarae* NBRC15530 was approximately 87.9%, which was much lower than the proposed species-level cutoff of 95%–96%^{30–31}. By using the Genome-to-Genome Distance Calculator³², the *in silico* DNA-DNA hybridization between the two genomes was about

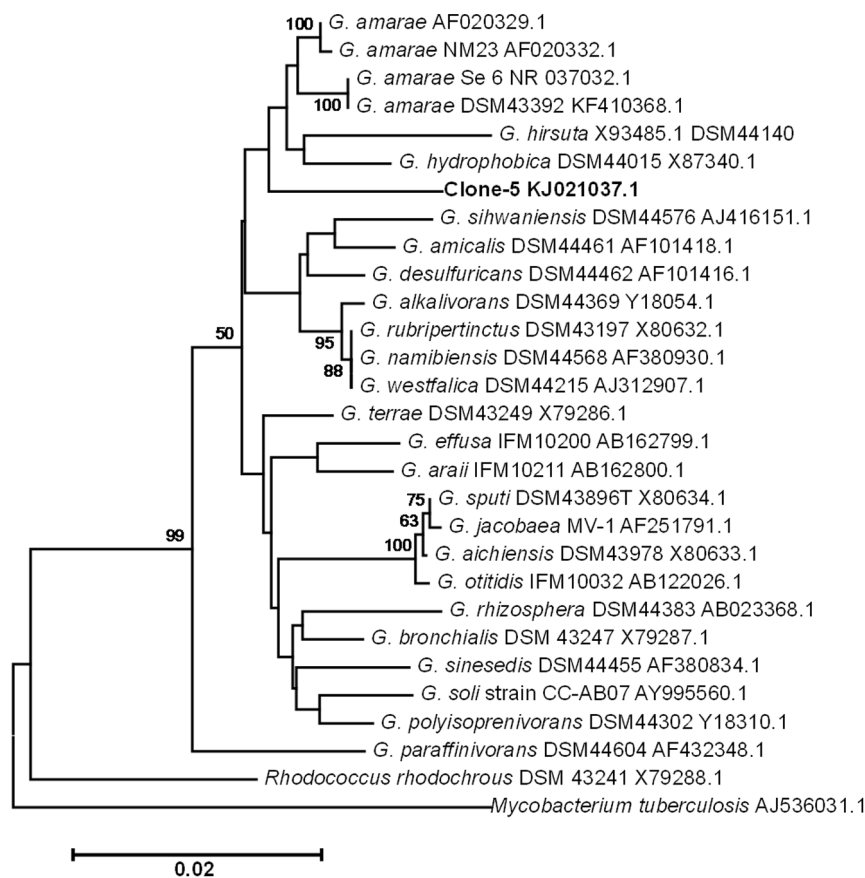


Figure 4 | Deposition of the uncultured *Gordonia* clone from the foam former in a phylogenetic tree of *Gordonia*, based on the near full-length 16S rRNA gene (approximately 1,200 bp). The neighbor-joining method and Jukes-Cantor model were applied with 1,000 bootstrap replications. The sequences from *Mycobacterium tuberculosis* and *Rhodococcus rhodochrous* DSM 43241 were set as out-groups during the tree construction in MEGA 5. Only the bootstrap values of more than 50% were shown. The scale bar represents 2% divergence.

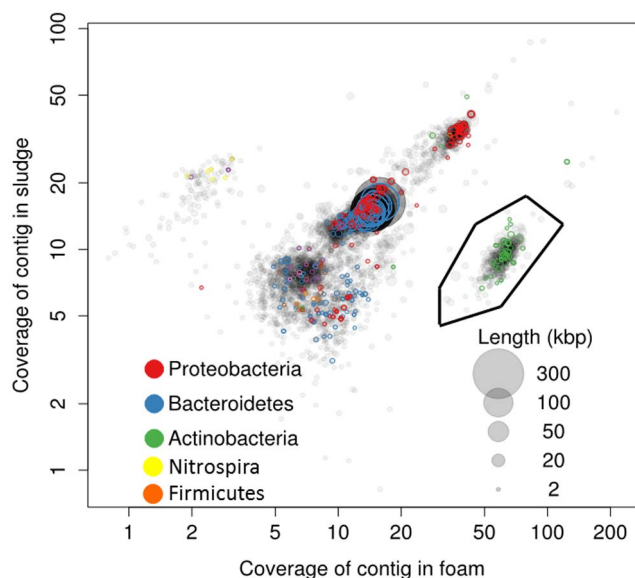


Figure 5 | Genome bin of the uncultured *Gordonia* sp. enriched in foam. The raw reads were assembled into contigs and only contigs over 1 kb were involved in the genomic binning. Only contigs over 2 kb were shown in this figure. Two dimensional coverages were referred to separate the contigs from different genomes. The marked cluster was the genomic bin of the uncultured *Gordonia* sp. with average coverages of 9 and 58 in the activated sludge and the foam datasets, respectively. It contained 83 essential single copy genes as highlighted in colored circles.

29.5% to 30.7%, which also strongly supported that the organism was not *G. amarae*. In summary, consistent with analysis of the 16S rRNA gene and *gyrB*, the genomic comparison further confirmed the conclusion that the organism that was dominant and enriched in the foam of the present study was an undescribed *Gordonia* species, which was closely related to, but distinct from *G. amarae*.

Discussion

As shown by previous studies, the accumulation of certain bacteria in foam relative to AS may not necessarily mean that they are the original foam formers³. Furthermore, the classic Koch's postulates, the gold standard for identification of causative organisms, are unfortunately unsuitable to distinguish the foam formers, due to both their uncertain cultivability and the greatly different conditions between aeration tanks and the lab. Under the modified standards for environmental microbiology^{3,33}, typical foam formers should have one or both of the following two characteristics: I) production of bio-surfactants to form foam; II) a high cell hydrophobicity to form stable foam. Thus, there should be two functional bacteria responsible for foaming in AS. Other than the mycolata group, we found that several genera, *Clostridium* XI, *Arcobacter* and *Flavobacterium*, accumulated more in the foam than in AS. This has not been reported in previous studies about the foaming of AS. Notably, *Clostridium* and *Flavobacterium* have been known as biosurfactant producers^{34–35}, while *Arcobacter* is a frequently detected oil-degrading genus with a potentially high cell hydrophobicity³⁶. It will be worthwhile to isolate these bacteria and find whether they can form foam or produce bio-surfactants in pure cultures.

More interestingly, and for the first time, our results reported different foaming conditions while the same *Gordonia* species maintained a fairly consistent abundance in the same STP. This observation strongly implied that other factors besides the abundance of *Gordonia* spp. determined the foaming process. The underlying reasons could be the different physiological status of the same organism or the involvement of other organisms. One possible determinative factor could be bio-surfactants produced by other organisms. As we

found, *Flavobacteriaceae*, *Saprospiraceae*, and others, could be candidates for further investigation.

In addition to detection of potential foam formers, understanding the particular selectivity of foam also has significance for engineering. The loss of AS biomass caused by foaming can sometimes be up to or more than 10%³⁷, which is not non-selective based on our results. Interestingly, the bacteria selected in the foam are generally hydrophobic and do not include some vital populations, such as the nutrient removal specialists (AOB and NOB), and thus would not cause other performance problems due to the loss of specific functional groups from AS.

Unlike effluent and activated sludge, bacteria in the foam may threaten to spread through the air, where pathogens are very hard to control³. Partially due to an absence of case reports or lack of awareness of foam-induced infections, no study, to our knowledge, has examined this issue. Among the pathogens, mycobacteria pose the greatest concern, due to their universal existence and abundance in global AS^{38–39}, as well as their ability to infect through respiration (*M. tuberculosis*) and skin contact (*M. ulcerans*). In the present study, none of the *Mycobacterium* species detected in the foam belonged to the slow-growing pathogenic *Mycobacterium* species. An investigation of the removal of mycobacterial populations in sewage treatment has been reported, focusing on a fast-growing opportunistic pathogen, *M. chelonae*⁴⁰. This species was not detected in our study. However, because our samples were only two time-point collections from one WWTP that uniquely treats saline sewage, the mycobacteria-related hygiene risks due to foaming AS could not be excluded. Comprehensive investigation of diverse WWTPs is required for a more thorough conclusion.

Our results indicated the dominant foam former in our sample was a novel species that is only distantly related to the prototypical *G. amarae*, according to the 16S rRNA gene, *gyrB* and reconstructed draft genome. Within the metabolically diverse *Gordonia*⁴¹, *G. amarae* seemed to be an independent phylogenetic branch when employing the *gyrB* and *secA1* genes as the markers under a neighbor joining method²⁷. Unfortunately, the 16S rRNA gene does not work well for species identification and the proposal of a novel species within this genus²⁷. For organisms such as *Gordonia* species, a community structure analysis based solely on the 16S rRNA gene (e.g., clone library, FISH, etc.) would obliterate the real inter-genus diversity and divergences across environmental samples. The genome binning method undoubtedly increased the taxonomic resolution, especially for the abundant organisms, whose genomes could be reconstructed in high quality. Moreover, systematic clarification on the taxonomy of *G. amarae*-like foaming organisms should be performed when sufficient isolates, sequences of marker genes and genomes foaming activated sludge become available.

In summary, our study expands the taxonomic width of potential foam formers and provides new insight into the foaming process. Many microbes other than mycolata group could be enriched in foam relative to AS. Furthermore, even within the well-studied genus of *Gordonia*, some undescribed species will cause the foaming of activated sludge and the correlation of their abundances and the foaming events are not consistent. The negative result of mycobacterial pathogen in our study could never be a conclusion to argue the biosafety of foaming activated sludge. Detailed investigations of more foaming samples, as conducted in the present study, will help elucidate the diversity and functional roles of these foam formers and answer the hygienic concerns of foaming events.

Methods

Sample collection. The sampling site was the Shatin WWTP located in Hong Kong, China. It treats saline municipal wastewater (at approximately 1.0% salinity), due to the usage of seawater for toilet flushing. Two foaming samples, with the corresponding AS samples, were collected on March 20, 2010 (named Foam-I and AS-I, respectively) and March 13, 2012 (named Foam-II and AS-II, respectively). Non-foaming AS samples were also collected in January, February, June and July of

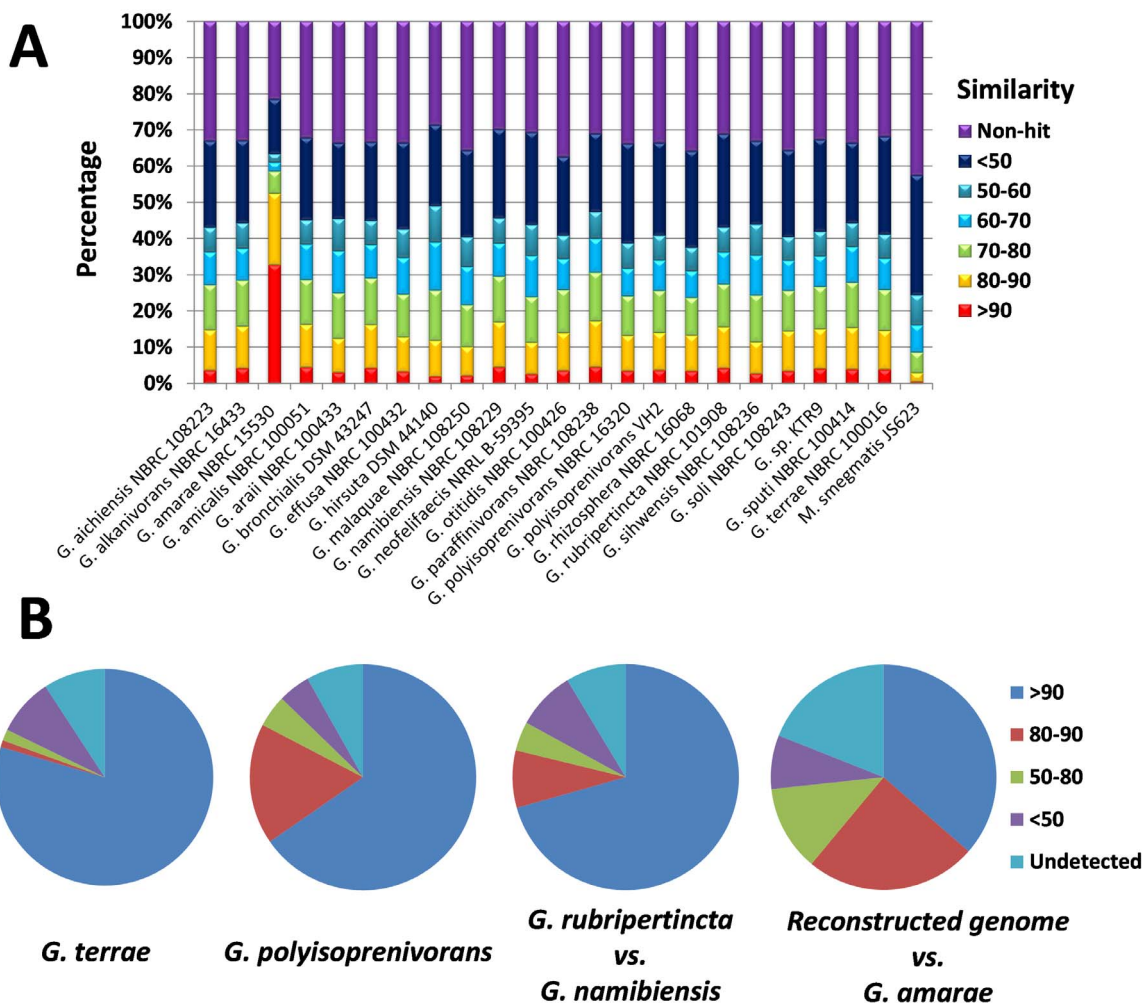


Figure 6 | Amino acid sequences similarities between the reconstructed draft genome and 22 *Gordonia* genomes and an out-group member, *Mycobacterium smegmatis* (A) and intra-species and inter-species *Gordonia* strains (B). The collection of the ORFs of the reconstructed draft genome was used as the database for BLASTx by the genes of all the *Gordonia* genomes in Group A. In Group B, the ORFs genomes used as the databases were *G. terrae* NBRC100016, *G. polyisoprenivorans* VH2, *G. namibiensis* NBRC108229 and *G. amarae* NBRC15530, respectively. The ORFs from *G. terrae* C-6, *G. polyisoprenivorans* NBRC16320, *G. rubripertincta* NBRC 101908 and the reconstructed draft genome were queries to perform BLASTp against the corresponding databases, respectively. Therefore, the denominators here are the total ORFs in the queries. The percentages in the pie charts are the average amino acid identities for the shared ORFs.

2010. The macroscopic and microscopic inspection of the foaming can be seen and compared with normal conditions and the corresponding AS in Fig S2. On the basis of foaming scales, the two foaming samples were classified as level 4–5, which was approximately 5–10 cm in height and very stable⁴². To avoid contamination during sampling, the AS and floating foam were carefully collected separately, and each was mixed with absolute ethanol to a 50% final ethanol solution. Before DNA extraction, the samples were stored at -20°C .

DNA extraction, PCR and high-throughput sequencing. The DNA of the AS and foam samples was extracted in duplicate by the FastDNATM Spin Kit for Soil, following the manufacturer's instruction (MP Biomedicals, USA). We have previously reported its high efficiency for Gram-positive bacteria and high DNA yields for AS samples in comparison with other kits⁴³. After mixing the DNA extraction duplicates, the DNA extract was quantified by both spectrophotometry (NanoDrop ND-1000, USA) and visualization on agarose gels. Then, part of the DNA samples (approximately 6 μg) from Foam-II and AS-II were sent to BGI (Shenzhen, China) for library construction and high-throughput Illumina sequencing. For both samples, libraries of fragments of approximately 170 bp were built and 101 bp paired-end Illumina sequencing was performed. The *fastq* files were uploaded to NCBI Sequence Read Archive (Accession Nos. SRX277355 and SRX277352 for Foam-II and AS-II, respectively).

The PCR primers with adaptors and barcodes for different biomarker genes and PCR conditions in this study are listed in Table S4. For the 16S rRNA gene, the V3 and V4 regions were selected for amplification using barcoded forward primers^{44–45}. The partial sequences of the *rpoB* of were examined for mycobacterial identification¹⁷. Barcoded PCR products of these two genes were purified and pyrosequenced in the

Roche 454 Titanium platform at BGI. For the 16S rRNA gene, all eight samples were examined, while only the Foam-I and -II samples were analyzed for the mycobacterial *rpoB*.

Sequence analysis. For the pyrosequencing data of 16S rDNA amplicons from different samples, the sequence trimming, denoising, chimeric removal and OTU picking were conducted using the Mothur platform following the 454 standard operating procedures^{46–48}. The cleaned pyrotags or the OTU representatives were identified by online RDP Classifier under a confidence threshold of 80%⁴⁹. The mycobacterial *rpoB* reads were filtered by requiring Q-value over 30 in any 20 nt window in Mothur platform.

A database containing the sequences of mycobacterial *rpoB* genes from 131 species (downloaded from Genbank) was built. The double-primer trimmed pyrotags/sequences of *rpoB* segments obtained in the present study were subjected to BLASTn against this database, with an e-value of 10^{-5} , and were applied for species identification based on the Best-Hit approach. Only those hits with over 96% similarity and over 95% aligned length were considered for the taxonomic assignment results.

For Illumina metagenomic data, the raw paired-end reads were pre-filtered by their quality (average $Q > 20$) and merged into itags by a customized Python script that required at least 10 bp of overlap without any mismatch. On the other side, all itags were subjected to BLASTn against the Greengenes 16S rRNA database⁵⁰ (released in May 2011), with the e-value cutoff of 10^{-20} , and the hit itags were extracted as 16S rRNA itags from the metagenomic datasets by a Python script. The extracted itags (130–190 bp) were classified by RDP Classifier at a confidence threshold of 50%⁴⁹. All of the tree visualization for the phylogenetic analysis was performed in MEGA 5⁵¹.



Genome binning of the dominant *Gordonia* species in foam. The paired-end metagenomic data of both the AS and foam samples were used in sequence assembly with the CLC software (kmer-size of 51) and only contigs over 1 kb were kept for subsequent analysis. The genome binning was performed by referring to the two dimensional coverages in the activated sludge and foam datasets following the reported pipeline²⁹. The completeness and redundancy of the reconstructed draft genome was then examined by counting the number of essential single copy genes in *Actinobacteria*²⁹. Genes of all the available reference *Gordonia* genomes and the reconstructed genome were predicted by the Metagenemark program to gain the single-gene or single-protein comprised genomes⁵². The genomic similarity between two organisms was analyzed by referring to the AAI, ANI and genome-to-genome distances (using contigs), according to references^{30,32,53}.

Statistical analysis. The STAMP software was used to compare the abundances of taxa in foam samples and the corresponding activated sludge samples⁵⁴. The one-sided Fisher's exact test with Newcombe-Wilson CI method and Bonferroni multiple test correction was conducted to examine the significance of difference between relative abundances in the foam and the corresponding activated sludge sample⁵⁴. Only dominant phyla (at least over 2% in one sample) and genera (at least over 0.5% in one sample) in the amplicon datasets were analyzed.

- Schilling, K. & Zessner, M. Foam in the aquatic environment. *Water Res.* **45**, 4355–4366 (2011).
- Petrovski, S. *et al.* An examination of the mechanisms for stable foam formation in activated sludge systems. *Water Res.* **45**, 2146–2154 (2011).
- de los Reyes, F. L. Foaming. In: *Microbial Ecology of Activated Sludge*. IWA Publishing, London, UK, pp. 215–258 (2010).
- Soddell, J. A. & Seviour, R. J. Microbiology of foaming in activated sludge plants. *J. Appl. Bacteriol.* **69**, 145–176 (1990).
- Bradford, D. *et al.* 16S rRNA analysis of isolates obtained from gram-negative, filamentous bacteria micromanipulated from activated sludge. *Syst. Appl. Microbiol.* **19**, 334–343 (1996).
- Schuppler, M., Merterns, F., Schon, G. & Gobel, U. B. Molecular characterization of nocardioform actinomycetes in activated-sludge by 16S ribosomal-RNA analysis. *Microbiology* **141**, 513–521 (1995).
- Eikelboom, D. H. Filamentous organisms observed in activated sludge. *Water Res.* **9**, 365–388 (1975).
- Shen, F. T. *et al.* Detection of filamentous genus *Gordonia* in foam samples using genus-specific primers combined with PCR – denaturing gradient gel electrophoresis analysis. *Can. J. Microbiol.* **53**, 768–774 (2007).
- Davenport, R. J., Curtis, T. P., Goodfellow, M., Stainsby, F. M. & Bingley, M. Quantitative use of fluorescent in situ hybridization to examine relationships between mycolic acid-containing actinomycetes and foaming in activated sludge plants. *Appl Environ Microbiol* **66**, 1158–1166 (2000).
- de los Reyes, F. L., Rothauszky, D. & Raskin, L. Microbial community structures in foaming and nonfoaming full-scale wastewater treatment plants. *Water Environ. Res.* **74**, 437–449 (2002).
- Klein, A. N., Frigon, D. & Raskin, L. Populations related to *Alkanindiges*, a novel genus containing obligate alkane degraders, are implicated in biological foaming in activated sludge systems. *Environ. Microbiol.* **9**, 1898–1912 (2007).
- Lemmer, H., Lind, G., Muller, E. & Schade, M. Non-famous scum bacteria: Biological characterization and troubleshooting. *Acta Hydrochimica et Hydrobiologica* **33**, 197–202 (2005).
- Bond, P. L., Hugenholz, P., Keller, J. & Blackall, L. L. Bacterial community structures of phosphate-removing and non-phosphate-removing activated sludges from sequencing batch reactors. *Appl. Environ. Microbiol.* **61**, 1910–1916 (1995).
- Gannon, B. W., Hayes, C. M. & Roe, J. M. Survival rate of airborne *Mycobacterium bovis*. *Res. Vet. Sci.* **82**, 169–172 (2007).
- Fox, G. E., Wisotzkey, J. D. & Jurtschuk, P. How close is close – 16S ribosomal-RNA sequence identity may not be sufficient to guarantee species identity. *Int. J. Syst. Bacteriol.* **42**, 166–170 (1992).
- Ninet, B. *et al.* Two different 16S rRNA genes in a mycobacterial strain. *J. Clin. Microbiol.* **34**, 2531–2536 (1996).
- Kim, B. J. *et al.* Identification of mycobacterial species by comparative sequence analysis of the RNA polymerase gene (*rpoB*). *J. Clin. Microbiol.* **37**, 1714–1720 (1999).
- Ringuet, H. C. *et al.* *hsp65* sequencing for identification of rapidly growing mycobacteria. *J. Clin. Microbiol.* **37**, 852–857 (1999).
- Logares, R. *et al.* Environmental microbiology through the lens of high-throughput DNA sequencing: Synopsis of current platforms and bioinformatics approaches. *J. Microbiol. Meth.* **91**, 106–113 (2012).
- MacLean, D., Jones, J. D. G. & Studholme, D. J. Application of 'next-generation' sequencing technologies to microbial genetics. *Nat. Rev. Microbiol.* **7**, 287–296 (2009).
- Ju, F., Guo, F., Ye, L., Xia, Y. & Zhang, T. Metagenomic analysis on seasonal microbial variations of activated sludge from a full-scale wastewater treatment plant over 4 years. *Environ. Microbiol. Rep.* **6**, 80–89 (2014).
- Daims, H., Bruhl, A., Amann, R., Schleifer, K. H. & Wagner, M. The domain-specific probe EUB338 is insufficient for the detection of all *Bacteria*: Development and evaluation of a more comprehensive probe set. *Syst. Appl. Microbiol.* **22**, 434–444 (1999).
- de los Reyes, M. F., de los Reyes, F. L., Hernandez, M. & Raskin, L. Quantification of *Gordonia amarae* strains in foaming activated sludge and anaerobic digester systems with oligonucleotide hybridization probes. *Appl. Environ. Microbiol.* **64**, 2503–2512 (1998).
- Xia, Y., Kong, Y., Thomsen, T. R. & Nielsen, P. H. Identification and characterization of epiphytic protein-hydrolyzing *Saprosiraceae* (Candidate *Epiflobacter* spp.) in activated sludge. *Appl. Environ. Microbiol.* **74**, 2229–2238 (2008).
- Hennessee, C. T., Seo, J. S., Alvarez, A. M. & Li, Q. X. Polycyclic aromatic hydrocarbon-degrading species isolated from Hawaiian soils: *Mycobacterium crocinum* sp. nov., *Mycobacterium pallens* sp. nov., *Mycobacterium rutilum* sp. nov., *Mycobacterium rufum* sp. nov. and *Mycobacterium aromaticivorans* sp. nov. *Int. J. Syst. Evol. Microbiol.* **59**, 378–387 (2009).
- Stingley, R. L., Brezna, B., Khan, A. A. & Cerniglia, C. E. Novel organization of genes in a phthalate degradation operon of *Mycobacterium vanbaalenii* PYR-1. *Microbiology* **150**, 3749–3761 (2004).
- Kang, Y., Takeda, K., Yazawa, K. & Mikami, Y. Phylogenetic studies of *Gordonia* species based on *gyrB* and *secA1* gene analyses. *Mycopathologia* **167**, 95–105 (2009).
- Kim, M., Oh, H. S., Park, S. C. & Chun, J. Towards a taxonomic coherence between average nucleotide identity and 16S rRNA gene sequence similarity for species demarcation of prokaryotes. *Int. J. Syst. Evol. Microbiol.* **64**, 346–351 (2014).
- Albertsen, M. *et al.* Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. *Nat. Biotechnol.* **31**, 533–538 (2013).
- Goris, J. *et al.* DNA–DNA hybridization values and their relationship to whole-genome sequence similarities. *Int. J. Syst. Evol. Microbiol.* **57**, 81–91 (2007).
- Richter, M. & Rosselló-Móra, R. Shifting the genomic gold standard for the prokaryotic species definition. *Proc. Natl. Acad. Sci.* **106**, 19126–19131 (2009).
- Meier-Kolthoff, J. P., Alexander, A. F., Hans-Peter, K. & Markus, K. Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinform.* **14**, 60 (2013).
- Fredericks, D. N. & Relman, D. A. Sequence-based identification of microbial pathogens: a reconsideration of Koch's postulates. *Clin. Microbiol. Rev.* **9**, 18–33 (1996).
- Richter, M. & Rosselló-Móra, R. Shifting the genomic gold standard for the prokaryotic species definition. *Proc. Natl. Acad. Sci.* **106**, 19126–19131 (2009).
- Cooper, D. G., Zajic, J. E., Gerson, D. F. & Manninen, K. I. Isolation and identification of bio-surfactants produced during anaerobic growth of *Clostridium pasteurianum*. *J. Ferment. Technol.* **58**, 83–86 (1980).
- Hubert, C. R. J. *et al.* Massive dominance of *Epsilonproteobacteria* in formation waters from a Canadian oil sands reservoir containing severely biodegraded oil. *Environ. Microbiol.* **14**, 387–404 (2012).
- Pretorius, W. A. & Laubscher, C. J. P. Control of biological scum in activated-sludge plants by means of selective flotation. *Water Sci. Technol.* **19**, 1003–1011 (1987).
- Guo, F. & Zhang, T. Profiling bulking and foaming bacteria in activated sludge by high throughput sequencing. *Water Res.* **46**, 2772–2782 (2012).
- Zhang, T., Shao, M. F. & Ye, L. 454 Pyrosequencing reveals bacterial diversity of activated sludge from 14 sewage treatment plants. *ISME J.* **6**, 1137–1147 (2012).
- Radomski, N. *et al.* *Mycobacterium* behavior in wastewater treatment plant, a bacterial model distinct from *Escherichia coli* and *Enterococci*. *Environ. Sci. Technol.* **45**, 5380–5386 (2011).
- Arenskötter, M., Broker, D. & Steinbuchel, A. Biology of the metabolically diverse genus *Gordonia*. *Appl. Environ. Microbiol.* **70**, 3195–3204 (2004).
- Blackall, L. L., Harbers, A. E., Greenfield, P. F. & Hayward, A. C. Activated-sludge foams – effects of environmental variables on organism growth and foam formation. *Environ. Technol.* **12**, 241–248 (1991).
- Guo, F. & Zhang, T. Biases during DNA extraction of activated sludge samples revealed by high throughput sequencing. *Appl. Microbiol. Biotechnol.* **97**, 4607–4616 (2013).
- Cai, L., Ye, L., Tong, A. H. Y., Lok, S. & Zhang, T. Biased diversity metrics revealed by bacterial 16S pyrotags derived from different primer sets. *PLoS ONE* **8**, e53649 (2013).
- Bartram, A. K., Lynch, M. D. J., Stearns, J. C., Moreno-Hagelsieb, G. & Neufeld, J. D. Generation of Multimillion-Sequence 16S rRNA Gene Libraries from complex microbial communities by assembling paired-end illumina reads. *Appl. Environ. Microbiol.* **77**, 3846–3852 (2011).
- Haas, B. J. *et al.* Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. *Genome Res.* **21**, 494–504 (2011).
- Huse, S. M., Welch, D. M., Morrison, H. G. & Sogin, M. L. Ironing out the wrinkles in the rare biosphere through improved OTU clustering. *Environ. Microbiol.* **12**, 1889–1898 (2010).
- Schloss, P. D., Gevers, D. & Westcott, S. L. Reducing the effects of PCR amplification and sequencing artifacts on 16S rRNA-based studies. *PLoS ONE* **6**, e27310 (2011).
- Wang, Q., Garrity, G. M., Tiedje, J. M. & Cole, J. R. Naive Bayesian classification for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol.* **73**, 5261–5267 (2007).



50. DeSantis, T. Z. *et al.* Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl. Environ. Microbiol.* **72**, 5069–5072 (2006).
51. Tamura, K. *et al.* MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* **28**, 2731–2739 (2011).
52. Zhu, W., Lomsadze, A. & Borodovsky, M. *Ab initio* gene identification in metagenomic sequences. *Nucleic Acids Res.* **38**, e132 (2010).
53. Konstantinidis, K. T. & Tiedje, J. M. Towards a genome-based taxonomy for prokaryotes. *J. Bacteriol.* **187**, 6258–6264 (2005).
54. Parks, D. H., Tyson, G. W., Hugenholtz, P. & Beiko, R. G. STAMP: Statistical analysis of taxonomic and functional profiles. *Bioinformatics* **30**, 3123–3124 (2014).

Acknowledgments

Dr. Feng Guo and Dr. Zhi-Ping Wang want to thank The University of Hong Kong for their postdoctoral fellowships. Mr. Ke Yu wishes to thank The University of Hong Kong for the postgraduate studentship. Dr. Yu Xia is thanked for her help in genomic binning. We also thank the Research Grants Council of Hong Kong (HKU7201/11E) for financial support of this study. We have confirmed the no conflict of interest statement.

Author contributions

F.G. contributed to sample collection, experimental design and conduct, data analysis and manuscript preparation. Z.P.W. and Y.K. contributed to the data analysis. T.Z. contributed to the experimental design and data analysis. All authors participated in manuscript revisions.

Additional information

Supplementary information accompanies this paper at <http://www.nature.com/scientificreports>

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

How to cite this article: Guo, F., Wang, Z.-P., Yu, K. & Zhang, T. Detailed investigation of the microbial community in foaming activated sludge reveals novel foam formers. *Sci. Rep.* **5**, 7637; DOI:10.1038/srep07637 (2015).



This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder in order to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-sa/4.0/>