**Supplementary Information**

*for*

**Analysis of Airborne Microbial Communities Using 16S ribosomal RNA: Potential Bias due to Air Sampling Stress**

Huajun Zhen, Valdis Krumins, Donna E. Fennell, and Gediminas Mainelis\*

Rutgers University, Department of Environmental Sciences, 14 College Farm Rd., New Brunswick, NJ 08901;

Revised Manuscript

(Ref. No.: STOTEN-D-17-05391)

*Science of the Total Environment*

Oct 2017

\*Corresponding author: Gediminas Mainelis

Phone: 848-932-5712

Fax: 732-932-8644

E-mail: mainelis@envsci.rutgers.edu

Bacterial Culture in Laboratory Experiments

The Gram-negative bacterium *E. coli* (ATCC 15597, Manassas,VA) was precultured in 50 ml Tryptic Soy broth (Becton, Dickinson and Company, Sparks, MD) at 37 °C for ~16 hours prior to each test. The bacteria were harvested and washed once with 1×phosphate buffer solution (1×PBS: 10 g/L NaCl, 0.25 g/L KCl, 1.43 g/L Na2HPO4, 0.25 g/L KH2PO4) by centrifugation at 7000×g (Jouan Inc., Winchester, VA) for 5 min at room temperature (25 °C) and resuspended in 1×PBS solution. The cell numbers in the final bacterial solution were determined by epifluorescence microscopy using the Axioskop 20 (Carl Zeiss Inc. Thornwood, NY) as reported previously1.

Nucleic Acids Extraction

The entire filter from Button sampler and pelleted samples from BioSampler and SpinCon air samplers were suspended in 50 µl buffer A (50 mM glucose, 10 mM EDTA and 25 mM pH=8.0 Tris)2. The solutions were subjected to five freeze/thaw cycles of freezing with liquid nitrogen and thawing in a 55 °C water bath. Then, 200 µl of buffer A, 100 µl of 4 mg/ml lysozyme in buffer A and 50 µl of 500 mM EDTA were added to the liquid. After incubation at room temperature for 10 minutes, a 50 µl 10% sodium dodecyl sulfate (SDS) solution was added followed by extraction with 800 µl phenol-chloroform-isoamyl alcohol mixture (25:24:1, PH=6.7) twice. The aqueous phase was then transferred to a new 1.5 ml microcentrifuge tube, and the total nucleic acids were precipitated by mixing the liquid with 50 µl of 3.0 M sodium acetate, 2 µl of glycogen (20 mg/ml), and 1 ml 100% ice-cold ethanol. The nucleic acids were pelleted by centrifugation at 16,100×g at 4°C for 15 minutes and washed once with 400 µl of cold 70% ethanol solution. The recovered pellets were first dried under a laminar flow hood for 10 minutes and then dissolved in 100 µl diethylpyrocarbonate (DEPC)-treated water for subsequent analysis.

The liquid was extracted twice with 800 µl phenol-chloroform-isoamyl alcohol mixture (25:24:1, PH=6.7) to extract extracellular nucleic acids from liquid suspension after centrifugation, and the aqueous phase was transferred to a separate 1.5 ml microcentrifuge tube after extraction. The total nucleic acids were precipitated by mixing the liquid with 50 µl of 3.0 M sodium acetate, 2 µl of glycogen (20 mg/ml), and 1 ml 100% ice-cold ethanol, pelleted by centrifugation at 16,100×g at 4°C for 15 minutes, and then washed once with 400 µl cold 70% ethanol solutions. The recovered pellets were first dried under a laminar flow hood for 10 minutes and then dissolved in 100 µl diethylpyrocarbonate (DEPC)-treated water for subsequent analysis. The pellet and supernatant extracts were combined into one sample and then subjected to analysis.

DNA Removal and Reverse Transcription

The total nucleic acid extracts from *E. coli* laboratory samples were diluted by a factor of 10, and then 45 µl of the diluted sample was treated with DNase using Ambion TURBO DNA-free DNase kit (Life Technologies, Grand Island, NY) according to the manufacturer’s instructions. The treated samples were then subjected to PCR reaction with primer sets targeting general bacterial 16S rRNA gene sequences to verify the absence of DNA in RNA sample. Then 5 µl RNA sample was reverse transcribed into complementary DNA (cDNA) using SuperScript® VILO cDNA synthesis kit (Life Technologies, Grand Island, NY). The cDNA was diluted by a factor of 10 and then stored at -20°C. The cDNA and the remaining samples with mixed DNA/RNA before DNase treatment were saved for subsequent qPCR and sequencing analysis.

For outdoor air samples, the extracted nucleic acids were processed with the Qiagen AllPrep DNA/RNA Mini kit following the manufacturer’s instructions (Qiagen, Valencia, CA). DNA and RNA were separated from the initial 100 µl mixed solution and eluted in 100 µl TE buffer and 60 µl RNase-free water, respectively. PCR of RNA samples did not produce any amplicons when using primer sets targeting general bacterial 16S rRNA gene sequences, which demonstrated the efficient separation of the DNA from the RNA. A 10 µl of RNA sample was used to make cDNA by using SuperScript® VILO cDNA synthesis kit (Life Technologies, Grand Island, NY). The DNA and cDNA samples were saved at -20°C for subsequent qPCR and sequencing analysis.

Quantitative PCR

In laboratory experiments, a multiplex qPCR method described elsewhere was used to quantify the 16S rRNA gene and reverse transcribe 16S rRNA for both sample (*E. coli)* and reference (*P. fluorescens).* Thus, prior to DNA/RNA co-extraction, *E. coli* samples were spiked with 5 µl of genomic DNA (1.1×106 copies of 16S rRNA gene/µl) and 5 µl 16S rRNA extracts (8.6×109 copies/µl) from *Pseudomonas fluorescens*. The primer-probe sets for *E. coli* 16S rRNA gene included forward primer: GGGAGTAAAGTTAATACCTTTG, reverse primer: CCAGTATCAGATGCAGTTC, and probe: TCACATCTGACTTAACAAACCGCCT-FAM. For *P. fluorescens* 16S rRNA gene, they included forward primer: CCTTGTCCTTAGTTACCAG, reverse primer: CTCTGTACCGACCATTGTA, and probe: CACTCTAAGGAGACTGCCGGTGAC-HEX. Each multiplex qPCR reaction included 10 µl of 2× TaqMan® universal PCR master mix (Life technologies, Grand Island, NY), 650 nM of each forward and reverse primer, 200 nM of each probe, 0.04 mg/ml bovine serum albumin (Sigma, St. Louis, MO) and 4 µl sample DNA or cDNA. The amplification was performed on an iCycler iQ5 RT-PCR detection system (Bio-Rad Laboratories, Hercules, CA) using the following temperature program: 10 min of denaturation at 95°C and 40 cycles of 15 s of denaturation at 95°C and 1 min of annealing/extension at 58°C and 30 s further extension at 72°C. Standard curves were prepared by performing PCR amplification with a 10-fold serial dilution of *E. coli* and *P. fluorescens* genomic DNA mixture with similar 16S rRNA gene copy numbers across five orders of magnitude. The qPCR standards were prepared with the genomic DNA extracted from pure *E. coli* and *P. fluorescens* bacterial cultures. Detailed protocols including extraction, purification, and quantification of DNA standards are reported elsewhere3. PCR amplification for each sample and standard were performed in triplicates.

In order to quantify the 16S rRNA and 16S rRNA gene from bioaerosols collected outdoors, a SYBR-Green-based qPCR assay was performed on iCycler iQ5 RT-PCR detection system (Bio-Rad Laboratories, Hercules, CA). The universal primer sets targeting bacterial 16S rRNA gene included forward primer: 5’-TCCTACGGGAGGCAGCAGT-3’ and reverse primer: 5’-GGACT ACCAGGGTATCTAATCCTGTT-3’ with an amplicon size of 466 bp on the reference *E. coli* genome. Each 20 µl reaction contained 10 µl of 2× SYBR green supermix (Bio-Rad Laboratories, Hercules, CA), 2 µl of each 2.5 µM primer, 5 µl of template DNA, and 1 µl PCR-grade water. The thermocycler was programmed for 10 min of denaturation at 95°C and 40 cycles of 15 s of denaturation at 95°C and 1 min of annealing/extension at 60°C. Upon completion of PCR amplification, a melt curve analysis was performed to check the purity of generated amplicons. A 10-fold serial dilution of *E. coli* genomic DNA was amplified with samples to serve as a standard curve on each reaction plate.

Sequence Analysis

For the air sample collected by BioSampler on Aug. 6th 2014, two sample fractions, the pellet and supernatant lipids were subjected to pyrosequencing analysis separately. In order to generate one sequencing data set for two fractions of the sample, 16S rRNA or 16S rRNA gene sequences from two data sets representing the pellet and supernatant liquid were randomly picked and combined after quality control check. The number of picked sequences from each data set was weighted by the abundance of 16S rRNA gene copy number for each sample from qPCR assay.

General Characteristics of Outdoor Bioaerosols

A range of 27-173 unique bacterial genera were identified with 16S rRNA gene or 16S rRNA sequences in different samples. The number of unique bacterial genera that were detected with 16S rRNA sequences (110±36 on average) was significantly higher (*p*=0.016) than that of 16S rRNA gene sequences (85±31 on average) (Figure S4A in Supplementary Information), while no significant difference was found in the number of detected bacterial genera between the two seasons (*p*=0.207) or sampling devices (*p*=0.427). No significant correlation was observed between the number of bacterial genera detected in paired 16S rRNA gene and 16S rRNA sequences from the same samples (*p*=0.941, Figure S4B in Supplementary Information). For those bacterial genera that were identified with both types of sequences (sixteen paired 16S rRNA and 16S rRNA gene sequences), the sum of their relative abundances were in the range of 22.5-73.1% (58.0±13.0% on average) and 17.8-79.5% (51.2±16.9% on average) for 16S RNA gene and 16S rRNA, respectively (Figure S4C in Supplementary Information). None of the three investigated factors (sequence type, season, and sampling device) had a significant effect on the relative abundance of bacterial genera. There was a positive and statistically significant correlation between the sum of relative abundances of bacterial genera identified with two types of sequences (R2=0.27, *p*=0.039, Figure S4D in Supplementary Information).

**REFERENCES**

1. Zhen, H. J.; Han, T.; Fennell, D. E.; Mainelis, G., Release of free DNA by membrane-impaired bacterial aerosols due to aerosolization and air sampling. *Appl. Environ. Microbiol.* **2013,** *79*, (24), 7780-7789.
2. Krumins, V.; Mainelis, G.; Kerkhof, L. J.; Fennell, D. E., Substrate-dependent rRNA production in an airborne bacterium. *Environ. Sci. Technol. Letters* **2014,** *1*, (9), 376-381.
3. Zhen, H. J.; Krumins, V.; Fennell, D. E.; Mainelis, G., Development of a dual-internal-reference technique to improve accuracy when determining bacterial 16S rRNA:16S rRNA gene ratio with application to *Escherichia coli* liquid and aerosol samples. *J. Microbiol. Methods* **2015,** *117*, 113-121.



Figure S1. Relative abundance of dominant bacterial phyla in outdoor bioaerosol samples collected by Button sampler and BioSampler and analyzed based on 16S rRNA gene and 16S rRNA sequences. Each stacked column shows the average of three samples collected in summer or five samples collected in late winter/early spring.

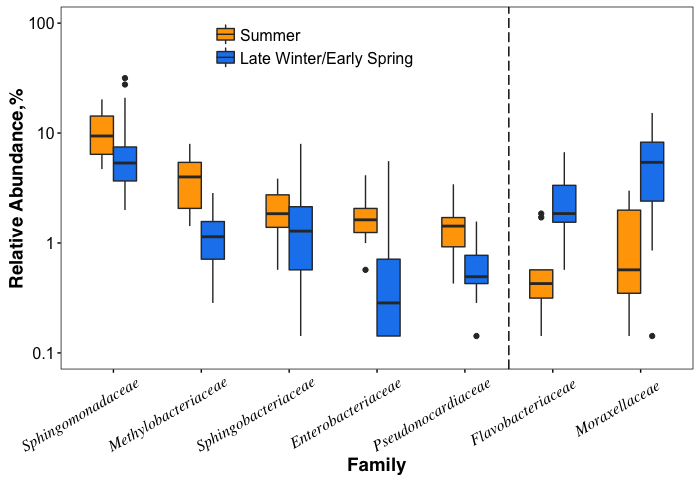
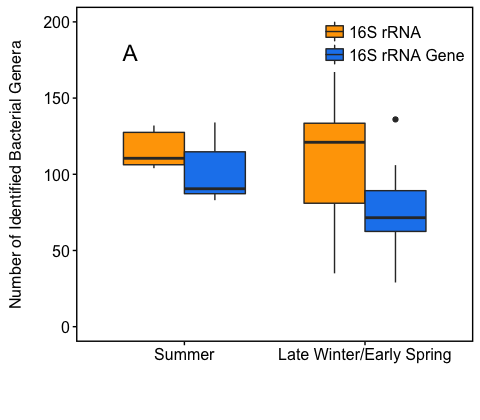
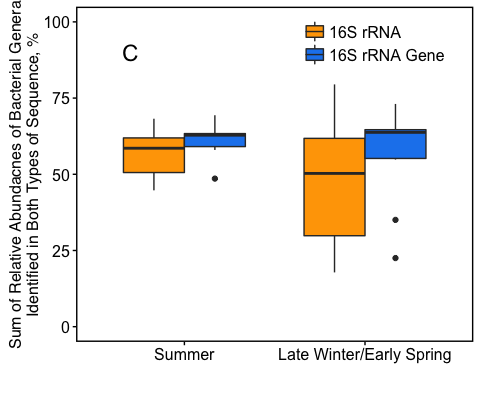


Figure S2. Bacterial taxa that exhibited significantly different abundance (*p*<0.05) in summer (yellow boxplot) and late winter/early spring (blue boxplot) samples collected in an outdoor environment (left of dash line: more abundant in summer than in late winter/early spring; right of dashed line: more abundant in late winter/early spring than in summer). Each boxplot represents the results of 16S rRNA gene and 16S rRNA sequences from six samples collected in summer or ten samples collected in late winter/early spring. There are twelve and twenty data points in each box for summer and later winter/early spring samples, respectively. The line through the middle of each box plot represents the median. The bottom and top of the box are the 25th and 75th percentiles. Outliers are indicated with points beyond the whiskers.



Figure S3. The relationship between the relative abundances of individual identified bacterial genus in paired 16S rRNA and 16S rRNA gene sequences in all collected samples. A total of 2396 points are presented on the plot, and each circle may represent a number of overlaid individual data points.

**Figure S4**. A) The number of identified bacterial genera in 16S rRNA and 16S rRNA gene sequences from all collected outdoor bioaerosol samples separated by seasons. B) No correlation was observed between numbers of bacterial genera identified in paired 16S rRNA gene and 16S rRNA sequences. C) Sum of relative abundances of bacterial genera identified in both 16S rRNA and 16S rRNA gene sequences in each sample. Results for all samples were separated by seasons. D) Correlation relationship between the sum of relative abundances of bacterial genera identified in both types of sequences in each sample.

**Table S1.** The relative abundances of bacterial phyla and subphyla identified in collected outdoor bioaerosol samples. Numbers represent an average percent of each bacterial taxonomic group. Data are grouped by season, sequences types and sampling device.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Bacteria Group** | **Summer** | | | | **Late Winter/early Spring** | | | | **Total** |
|  | 16S rRNA Gene | | 16S rRNA | | 16S rRNA Gene | | 16S rRNA | |
|  | Button Sampler | BioSampler | Button Sampler | BioSampler | Button Sampler | BioSampler | Button Sampler | BioSampler |
| ***Acidobacteria*** | 0.1 | 0.7 | 1.7 | 0.8 | 4.2 | 1.3 | 2.1 | 3.3 | 2.0 |
| ***Actinobacteria*** | 17.1 | 21.5 | 13.6 | 19.1 | 15.1 | 23.2 | 20.3 | 10.9 | 17.6 |
| *Actinomycetales* | 14.2 | 19.8 | 12.5 | 17.8 | 14.4 | 19.6 | 18.7 | 8.7 | 15.6 |
| *Solirubrobacterales* | 2.0 | 1.1 | 0.8 | 1.0 | 0.5 | 1.5 | 0.7 | 1.5 | 1.1 |
| *Actinobacteria Other* | 0.8 | 0.6 | 0.3 | 0.3 | 0.1 | 2.1 | 0.9 | 0.7 | 0.8 |
| ***Bacteroidetes*** | 10.8 | 11.9 | 10.5 | 12.1 | 14.0 | 13.3 | 12.8 | 7.5 | 11.7 |
| *Bacteroidales* | 0.5 | 0.7 | 0.2 | 0.2 | 1.9 | 2.0 | 1.1 | 0.3 | 1.0 |
| *Cytophagales* | 2.4 | 3.5 | 4.8 | 5.5 | 3.1 | 3.3 | 7.0 | 3.0 | 4.1 |
| *Flavobacteriales* | 2.6 | 2.7 | 1.4 | 2.3 | 3.4 | 4.2 | 2.7 | 1.4 | 2.7 |
| *Sphingobacteriales* | 2.0 | 2.5 | 1.3 | 2.5 | 2.8 | 1.8 | 0.7 | 0.7 | 1.7 |
| *Saprospirales* | 2.7 | 2.5 | 2.6 | 1.4 | 2.7 | 1.9 | 1.1 | 1.7 | 2.0 |
| *Bacteroidetes Other* | 0.6 | 0.1 | 0.2 | 0.2 | 0.1 | 0.1 | 0.2 | 0.4 | 0.2 |
| ***Cyanobacteria*** | 16.0 | 9.0 | 9.7 | 9.3 | 2.6 | 7.9 | 7.5 | 10.7 | 8.6 |
| *Streptophyta* | 15.0 | 8.5 | 8.0 | 7.5 | 2.0 | 5.1 | 0.9 | 1.1 | 5.1 |
| *Chroococcales* | 0.0 | 0.0 | 0.0 | 0.4 | 0.3 | 2.4 | 5.5 | 8.1 | 2.6 |
| *Cyanobacteria Other* | 0.9 | 0.5 | 1.8 | 1.4 | 0.3 | 0.5 | 1.1 | 1.5 | 1.0 |
| ***Firmicutes*** | 5.7 | 8.4 | 4.2 | 4.4 | 14.4 | 20.6 | 15.4 | 5.6 | 10.9 |
| *Bacillales* | 2.0 | 5.6 | 2.0 | 2.8 | 7.9 | 8.3 | 5.7 | 1.9 | 4.9 |
| *Lactobacillales* | 0.4 | 0.8 | 0.7 | 0.6 | 2.3 | 2.6 | 3.2 | 2.0 | 1.8 |
| *Clostridiales* | 3.0 | 2.0 | 1.4 | 1.0 | 4.0 | 9.3 | 6.2 | 1.7 | 4.0 |
| *Firmicutes Other* | 0.2 | 0.0 | 0.1 | 0.0 | 0.1 | 0.3 | 0.3 | 0.1 | 0.2 |
| ***Alphaproteobacteria*** | 17.7 | 26.0 | 24.1 | 26.0 | 20.9 | 11.9 | 17.4 | 29.5 | 21.2 |
| *Caulobacterales* | 0.3 | 0.8 | 0.7 | 1.5 | 3.4 | 0.7 | 1.6 | 3.0 | 1.7 |
| *Rhizobiales* | 4.2 | 8.6 | 7.4 | 9.3 | 3.3 | 4.0 | 5.5 | 4.5 | 5.5 |
| *Rhodobacterales* | 0.5 | 1.1 | 1.5 | 0.9 | 0.7 | 0.4 | 1.8 | 1.0 | 1.0 |
| *Rhodospirillales* | 0.1 | 1.9 | 1.9 | 3.6 | 0.3 | 0.7 | 3.9 | 1.9 | 1.8 |
| *Sphingomonadales* | 11.9 | 12.5 | 12.1 | 10.4 | 12.7 | 5.7 | 4.4 | 16.8 | 10.6 |
| *Alphaproteobacteria Other* | 0.7 | 1.0 | 0.6 | 0.4 | 0.5 | 0.4 | 0.3 | 2.3 | 0.8 |
| ***Betaproteobaceria*** | 3.5 | 4.8 | 3.9 | 2.9 | 9.1 | 7.0 | 7.0 | 8.3 | 6.3 |
| *Burkholderiales* | 3.5 | 4.4 | 3.7 | 2.7 | 8.9 | 6.8 | 6.8 | 5.4 | 5.7 |
| *Betaproteobaceria Other* | 0.0 | 0.4 | 0.3 | 0.2 | 0.1 | 0.2 | 0.2 | 2.9 | 0.6 |
| ***Deltaproteobacteria*** | 0.8 | 1.0 | 2.4 | 2.0 | 0.3 | 0.3 | 0.8 | 0.7 | 0.9 |
| ***Gammaproteobacteria*** | 5.7 | 8.6 | 8.0 | 11.5 | 12.0 | 8.2 | 12.6 | 15.4 | 10.7 |
| *Enterobacteriales* | 1.7 | 2.5 | 1.8 | 1.3 | 1.3 | 0.7 | 0.5 | 0.3 | 1.1 |
| *Pseudomonadales* | 2.8 | 4.0 | 2.1 | 2.6 | 9.5 | 6.1 | 9.1 | 11.6 | 6.7 |
| *Xanthomonadales* | 1.0 | 1.4 | 3.7 | 7.3 | 1.2 | 0.9 | 2.5 | 3.2 | 2.5 |
| *Gammaproteobacteria Other* | 0.3 | 0.6 | 0.4 | 0.3 | 0.1 | 0.5 | 0.5 | 0.3 | 0.4 |
| ***Verrucomicrobia*** | 0.0 | 0.2 | 0.1 | 0.1 | 0.2 | 0.1 | 0.1 | 0.0 | 0.1 |
| ***Bacteria Other*** | 22.6 | 7.8 | 21.6 | 11.7 | 7.3 | 6.2 | 4.1 | 8.0 | 10.0 |

**Table S2.** A list of bacterial strains from those genera that exhibited higher abundances of 16S rRNA than 16S rRNA gene relative to all other bacterial genera in bioaerosols collected in summer or late winter/early spring. Information on the complete genome of each strain was based on search results with GenBank database (<http://www.ncbi.nlm.nih.gov/genbank/>, accessed on March 12th, 2016).

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Organism/Name | **Gene Bank No.** | **Assembly Level** | **Chromosome Size (Mb)** | **Copy per Genome** | | | |
| 16S rRNA | 23S rRNA | 5S rRNA | tRNA |
| *Methylobacterium extorquens* AM1 | CP001510.1 | Complete Genome | 5.51 | 5 | 5 | 5 | 63 |
| *Methylobacterium extorquens* PA1 | CP000908.1 | Complete Genome | 5.47 | 5 | 5 | 5 | 58 |
| *Methylobacterium extorquens* CM4 | CP001298.1 | Complete Genome | 5.78 | 5 | 5 | 5 | 62 |
| *Methylobacterium extorquens* DM4 | FP103042.2 | Complete Genome | 5.94 | 5 | 5 | 5 | 58 |
| *Methylobacterium populi* BJ001 | CP001029.1 | Complete Genome | 5.80 | 5 | 5 | 5 | 56 |
| *Methylobacterium nodulans* ORS 2060 | CP001349.1 | Complete Genome | 7.77 | 7 | 7 | 7 | 72 |
| *Methylobacterium oryzae* CBMB20 | CP003811.1 | Complete Genome | 6.29 | 4 | 4 | 4 | 57 |
| *Methylobacterium radiotolerans* JCM 2831 | CP001001.1 | Complete Genome | 6.08 | 6 | 6 | 6 | 59 |
| *Methylobacterium aquaticum* MA-22A | AP014704.1 | Complete Genome | 5.35 | 11 | 11 | 11 | 97 |
| *Methylobacterium sp.* 4-46 | CP000943.1 | Complete Genome | 7.66 | 6 | 6 | 6 | 65 |
| *Methylobacterium sp.* AMS5 | CP006992.1 | Complete Genome | 5.44 | 5 | 5 | 5 | 60 |
| ***Methylobacterium* Average (n=11)** | | | | **5.8** | **5.8** | **5.8** | **64.3** |
| *Hymenobacter swuensis* DY53 | CP007145.1 | Complete Genome | 4.90 | 3 | 3 | 2 | 43 |
| *Hymenobacter sp.* APR13 | CP006587.1 | Complete Genome | 4.87 | 4 | 4 | 4 | 42 |
| *Hymenobacter sp.* DG25B | CP010054.1 | Complete Genome | 3.87 | 2 | 2 | 2 | 43 |
| *Hymenobacter sp.* DG25A | CP012623.1 | Complete Genome | 3.78 | 2 | 2 | 2 | 43 |
| *Hymenobacter sp.* DG5B | CP013909.1 | Complete Genome | 4.87 | 3 | 3 | 3 | 44 |
| *Hymenobacter sp.* PAMC26628 | CP014304.1 | Complete Genome | 5.28 | 3 | 3 | 3 | 41 |
| ***Hymenobacter* Average (n=6)** | | | | **2.8** | **2.8** | **2.8** | **42.7** |
| *Rhodanobacter denitrificans* 2APBS1 | CP003470.1 | Complete Genome | 4.23 | 2 | 2 | 2 | 49 |