

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15

## **Supporting Information**

### **For**

#### **Diversity and functions of bacterial community in drinking water biofilms revealed by high-throughput sequencing**

Yuanqing Chao<sup>1,2</sup>, Yanping Mao<sup>1</sup>, Zhiping Wang<sup>1,3</sup>, Tong Zhang<sup>1,\*</sup>

<sup>1</sup> Environmental Biotechnology Lab, The University of Hong Kong, Pokfulam Road, Hong Kong SAR, China

<sup>2</sup> School of Environmental Science and Engineering, Sun Yat-Sen University, Guangzhou, China

<sup>3</sup> School of Environmental Science and Engineering, Shanghai Jiao Tong University, Shanghai, China

---

\* Corresponding author. Email: zhangt@hku.hk. Phone: (852) 28598551. Fax: (852) 25595337.

16 *Experimental repeatability*

17 To test the repeatability of DNA extraction, PCR and 454 pyrosequencing, the 180 d  
18 biofilm was divided into duplicates. DNA was extracted from the two duplicates, PCR  
19 conducted and then sequenced individually. The results generated indicated that the  
20 analysis was repeatable at the genus level ( $R^2 = 0.976$ , Figure S7).

21 **Table S1** The genus list of several bacterial groups, including potentially pathogenic  
 22 bacteria, iron-oxidizing bacteria, and sulfate-reducing bacteria, according to previous  
 23 studies.

Bacterial groups	Potentially pathogenic bacteria	Iron-oxidizing bacteria	Sulfate-reducing bacteria
Genus list	<i>Aeromonas</i>	<i>Acidimicrobium</i>	<i>Bdellovibrio</i>
	<i>Arcobacter</i>	<i>Acidithiobacillus</i>	<i>Clostridium</i>
	<i>Bordetella</i>	<i>Alicyclobacillus</i>	<i>Desulfobacter</i>
	<i>Borrelia</i>	<i>Aquifex</i>	<i>Desulfobacterium</i>
	<i>Brucella</i>	<i>Chlorobium</i>	<i>Desulfobulbus</i>
	<i>Campylobacter</i>	<i>Ferrimicrobium</i>	<i>Desulfococcus</i>
	<i>Chlamydia</i>	<i>Ferrithrix</i>	<i>Desulfohalobium</i>
	<i>Chlamydomphila</i>	<i>Ferritrophicum</i>	<i>Desulfomicrobium</i>
	<i>Clostridium</i>	<i>Gallionella</i>	<i>Desulfomonas</i>
	<i>Corynebacterium</i>	<i>Leptospirillum</i>	<i>Desulfomonile</i>
	<i>Enterobacter</i>	<i>Leptothrix</i>	<i>Desulfonatronum</i>
	<i>Enterococcus</i>	<i>Marinobacter</i>	<i>Desulfonema</i>
	<i>Escherichia</i>	<i>Mariprofundus</i>	<i>Desulfosarcina</i>
	<i>Francisella</i>	<i>Pseudogulbenkiania</i>	<i>Desulfotomaculum</i>
	<i>Haemophilus</i>	<i>Rhodobacter</i>	<i>Desulfovibrio</i>
	<i>Helicobacter</i>	<i>Rhodopseudomonas</i>	<i>Desulfuromonas</i>
	<i>Klebsiella</i>	<i>Sideroxydans</i>	<i>Desulfuromusa</i>
	<i>Legionella</i>	<i>Sphaerotilus</i>	<i>Geobacter</i>
	<i>Leptospira</i>	<i>Thiobacillus</i>	<i>Myxococcus</i>
	<i>Listeria</i>		<i>Pelobacter</i>

	<i>Mycobacterium</i>		<i>Pseudomonas</i>
	<i>Mycoplasma</i>		<i>Syntrophobacter</i>
	<i>Neisseria</i>		<i>Thermodesulfobacterium</i>
	<i>Pseudomonas</i>		<i>Thermodesulfobivrio</i>
	<i>Salmonella</i>		
	<i>Serratia</i>		
	<i>Shigella</i>		
	<i>Staphylococcus</i>		
	<i>Streptococcus</i>		
	<i>Treponema</i>		
	<i>Vibrio</i>		
	<i>Yersinia</i>		
Total	32	19	24
Reference	Ye and Zhang, 2011	Emerson <i>et al.</i> , 2010	Castro <i>et al.</i> , 2000

24 **Reference:**

- 25 Ye, L. & Zhang, T. Pathogenic bacteria in sewage treatment plants as revealed by 454  
 26 pyrosequencing. *Environ. Sci. Technol.* **45**, 7173-7179 (2011).  
 27 Emerson, D., Fleming, E. J. & McBeth, J. M. Iron-oxidizing bacteria: an environmental and  
 28 genomic perspective. *Annu. Rev. Microbiol.* **64**, 561-583 (2010).  
 29 Castro, H. E., Williams, N. H. & Ogram, A. Phylogeny of sulfate-reducing bacteria. *FEMS*  
 30 *Microbiol. Ecol.* **31**, 1-9 (2000).

31

32 **Table S2** Evaluation of completeness and redundancy of extracted draft genome by core COGs of the reference genomes in order *Rhizobiales*  
 33 and 9 families of this order.

Levels	Name	No. of Reference	Core COGs		Single core COGs		Completeness (%)	Redundancy (%)
			Bin	Reference	Bin	Reference		
Order	<i>Rhizobiales</i>	96	74	75	0	0	99	0
	<i>Bartonellaceae</i>	9	655	725	514	573	90	19
	<i>Beijerinckiaceae</i>	2	1,169	1,487	702	914	79	13
	<i>Bradyrhizobiaceae</i>	16	1,033	1,195	0	1	86	0
	<i>Brucellaceae</i>	21	1,055	1,350	607	808	78	10
Family	<i>Hyphomicrobiaceae</i>	5	928	1,045	481	548	89	7
	<i>Methylobacteriaceae</i>	8	1,190	1,539	479	637	77	5
	<i>Phyllobacteriaceae</i>	6	1,095	1,287	476	561	85	5
	<i>Rhizobiaceae</i>	24	573	629	243	275	91	3
	<i>Xanthobacteraceae</i>	3	1,174	1,460	588	736	80	8

34

35 **Table S3** Phylogenetic analysis of the draft genome by annotating the ORFs.

Level	Closest taxon <sup>a</sup>	Attributed ORFs	Percentage	Criteria <sup>b</sup>	Pass/Fail
Phylum	<i>Proteobacteria</i>	3,178/3,458	92%	>50%	Pass
Class	<i>Alphaproteobacteria</i>	2,641/3,458	76%	>50%	Pass
Order	<i>Rhizobiales</i>	1,568/3,458	45%	>40%	Pass
Family	<i>Bradyrhizobiaceae</i>	618/3,458	18%	>34%	Fail

36 <sup>a</sup> The closest taxon at a specific level was selected according to the number of attributed ORFs;

37 <sup>b</sup> The phylogenetic levels of bins were identified by the following criteria: for phylum and class  
 38 level, > 50% of ORFs in the draft genome could be attributed; for order and family level, > 40%  
 39 and > 34% of ORFs could be attributed, respectively (Ishii et al., 2013) .

40 **Reference:**

41 Ishii, S. et al. A novel metatranscriptomic approach to identify gene expression dynamics during  
 42 extracellular electron transfer. Nat. Commun. **4**, 1601 (2013).

43

44 **Table S4** Surface tension, zeta potential and RMS roughness of the three materials  
45 used in the present study.

Materials	Surface tension (mJ/m <sup>2</sup> ) <sup>a</sup>	Zeta potential (mV) <sup>b</sup>	RMS roughness (nm) <sup>c</sup>
PC	42 ± 1.1	-27 ± 4.6	21 ± 8.5
PE	34 ± 1.6	-29 ± 2.3	21 ± 6.4
SS	41 ± 1.8	40 ± 7.3	83 ± 25

46 <sup>a</sup> Surface tension was evaluated by using contact angle method;

47 <sup>b</sup> Zeta potential was measured in 100 mM PBS solution;

48 <sup>c</sup> Roughness was calculated by the AFM height images with scan sizes of 2 × 2 μm<sup>2</sup>.

49

50 **Table S5** Detailed water quality of the tap water in all 12 time points.

Parameters (Unit)	Time points												Average $\pm$ SD <sup>a</sup>
	1	2	3	4	5	6	7	8	9	10	11	12	
Temperature ( °C)	27	26	26	22	23	25	23	26	25	25	26	26	25 $\pm$ 1.5
pH	7.1	7.5	6.9	7.4	7.7	6.6	7.1	7.4	7.7	7.5	7.0	7.7	7.3 $\pm$ 0.35
Dissolved oxygen (mg/L)	6.9	7.5	7.0	6.8	7.7	7.2	7.1	6.8	7.0	7.6	6.7	7.4	7.1 $\pm$ 0.33
Total organic carbon (mg/L) <sup>b</sup>	1.1	2.6	1.9	2.7	1.7	1.3	1.5	2.3	1.3	2.0	2.1	2.3	1.9 $\pm$ 0.51
	1.1	2.5	1.7	2.5	1.8	1.2	1.5	2.3	1.4	1.9	2.1	2.5	
Inorganic carbon (mg/L) <sup>b</sup>	7.0	6.1	5.5	6.2	5.6	7.6	5.8	6.1	6.8	7.2	7.6	5.5	6.4 $\pm$ 0.79
	7.0	6.1	5.4	6.2	5.7	7.7	5.8	5.9	7.0	7.2	7.5	5.5	
Total nitrogen (mg/L)	2.1	1.4	1.1	2.1	1.0	1.3	0.98	1.2	1.2	1.6	1.9	1.1	1.4 $\pm$ 0.41
Total phosphate (mg/L)	0.021	0.0084	0.0084	0.0084	0.017	0.013	0.013	0.013	0.0084	0	0.0084	0.0042	0.010 $\pm$ 0.0055
Residual chlorine (mg/L)	0.02	0.08	0.1	0.11	0.07	0.09	0.11	0	0.04	0.13	0.02	0.02	0.066 $\pm$ 0.044
Conductivity ( $\mu$ s/cm)	154	144	139	136	148	152	166	170	160	173	170	139	154 $\pm$ 13
Turbidity (NTU <sup>c</sup> )	0.88	0.48	0.26	0.19	0.12	0.12	0.20	0.29	0.17	0.11	0.084	0.097	0.25 $\pm$ 0.23
Total dissolved solid (mg/L)	150	146	88	136	152	108	128	212	100	184	179	102	140 $\pm$ 38

51 <sup>a</sup> Standard deviation; <sup>b</sup> TOC and IC tests were repeated twice for one sample; <sup>c</sup> Nephelometric turbidity units.

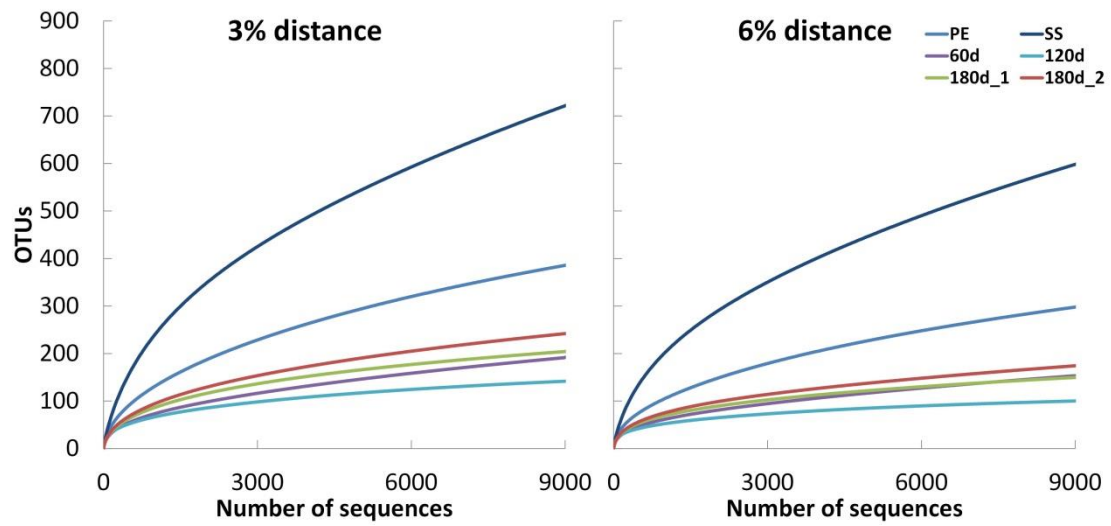


52 **Table S6** The basic information of 30 metagenomic data in three aquatic ecosystems (including ocean, fresh water and drinking water), which  
 53 were applied to principal component analysis in the present study.

Biome	MG-RAST ID	Metagenome size (bp)	Reads No.	Location	Sequence Method
Ocean_1	4441148.3	54,752,102	50,609	Indian Ocean, St. Anne Island, Seychelles	other <sup>a</sup>
Ocean_2	4441139.4	52,667,848	50,096	Indian Ocean, International waters between Madagascar and South Africa	other
Ocean_3	4441150.3	64,230,062	61,020	Indian Ocean	other
Ocean_4	4441155.3	62,752,349	59,813	Indian Ocean	other
Ocean_5	4441608.3	53,607,277	49,597	Indian Ocean	other
Ocean_6	4441135.3	45,710,196	46,052	Indian Ocean, Madagascar Waters, Madagascar	other
Ocean_7	4441149.3	64,223,447	60,932	Outside Seychelles, Indian Ocean, Seychelles	other
Ocean_8	4441147.3	55,638,894	52,118	Indian Ocean	other
Ocean_9	4441156.3	62,072,289	59,080	Indian Ocean	other
Ocean_10	4441134.3	53,607,277	49,597	Indian Ocean	other
Fresh water_1	4534328.3	1,888,723,700	18,887,237	Minnesota, United States	Illumina
Fresh water_2	4534330.3	1,847,964,400	18,479,644	Minnesota, United States	Illumina
Fresh water_3	4534342.3	1,647,353,500	16,473,535	Minnesota, United States	Illumina
Fresh water_4	4534350.3	1,623,038,100	16,230,381	Minnesota, United States	Illumina

Fresh water_5	4534344.3	1,606,029,200	16,060,292	Minnesota, United States	Illumina
Fresh water_6	4534338.3	1,599,106,200	15,991,062	Minnesota, United States	Illumina
Fresh water_7	4534336.3	1,430,592,200	14,305,922	Minnesota, United States	Illumina
Fresh water_8	4534340.3	1,427,643,200	14,276,432	Minnesota, United States	Illumina
Fresh water_9	4534346.3	1,395,406,900	13,954,069	Minnesota, United States	Illumina
Fresh water_10	4534348.3	1,031,183,000	10,311,830	Minnesota, United States	Illumina
DW disinfected_1	4470937.3	86,319,144	155,593	Cincinnati, Ohio, United States	454
DW disinfected_2	4470954.3	197,354,490	362,640	Cincinnati, Ohio, United States	454
DW disinfected_3	4509542.3	1,191,805,800	11,918,058	Hong Kong SAR, China	Illumina
DW disinfected_4	4509336.3	3,455,504,600	34,555,046	Hong Kong SAR, China	Illumina
DW disinfected_5	4615185.3	1,193,976,400	11,939,764	Nanjing, China	Illumina
DW undisinfected_1	4509541.3	1,190,701,600	11,907,016	Hong Kong SAR, China	Illumina
DW undisinfected_2	4509333.3	4,192,434,200	41,924,342	Hong Kong SAR, China	Illumina
DW undisinfected_3	4509334.3	3,607,833,600	36,078,336	Hong Kong SAR, China	Illumina
DW undisinfected_4	4615186.3	1,191,712,400	11,917,124	Nanjing, China	Illumina
DW undisinfected_5	4615323.3	1,078,393,739	10,783,937	Nanjing, China	Illumina

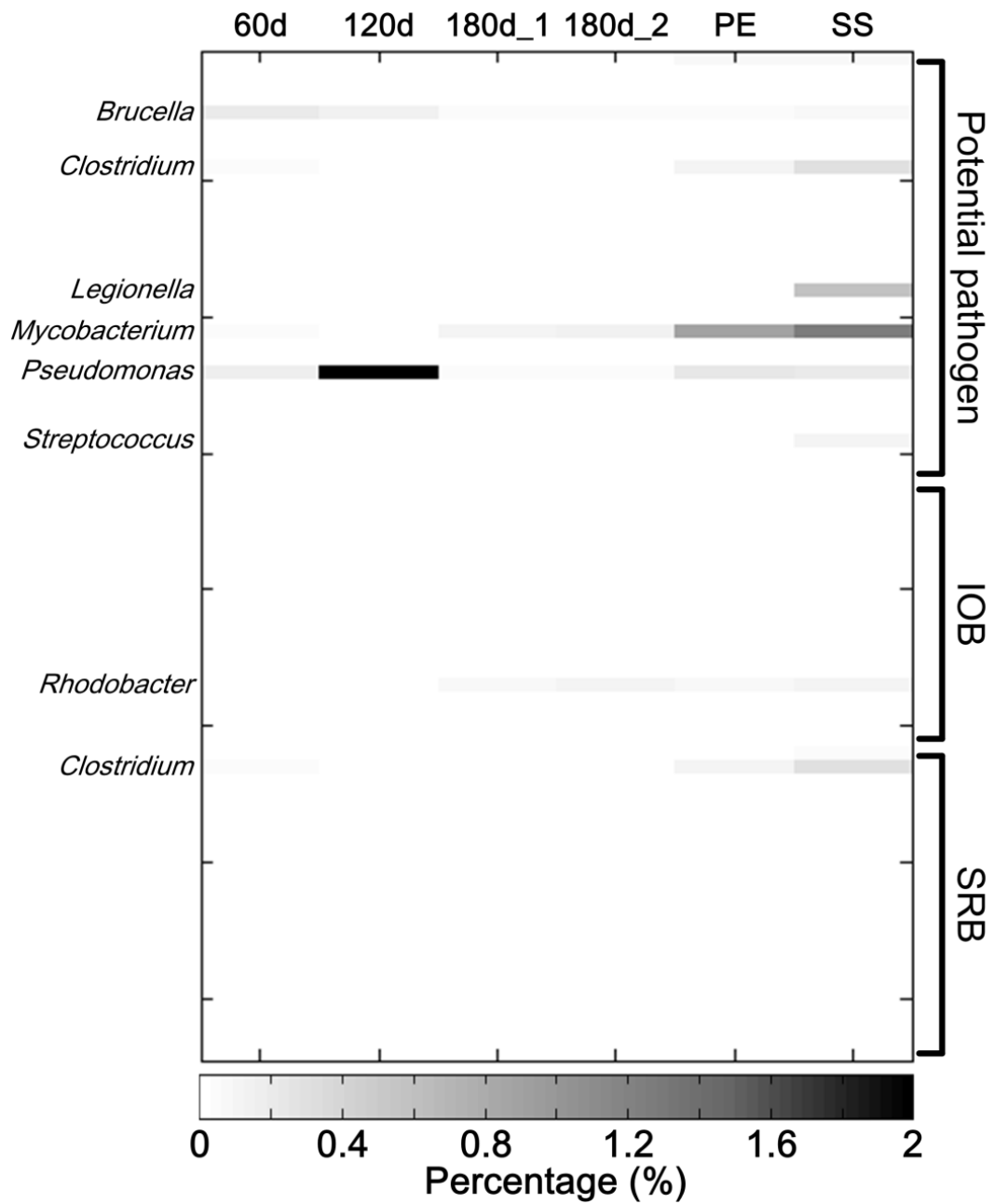
54 <sup>a</sup> indicates the researchers applied other sequencing method, except for sanger, Illumina and 454 methods.



55

56 **Figure S1** Rarefaction curves of OTUs defined by 0.03 and 0.06 cluster distances in  
 57 the DW biofilm samples.

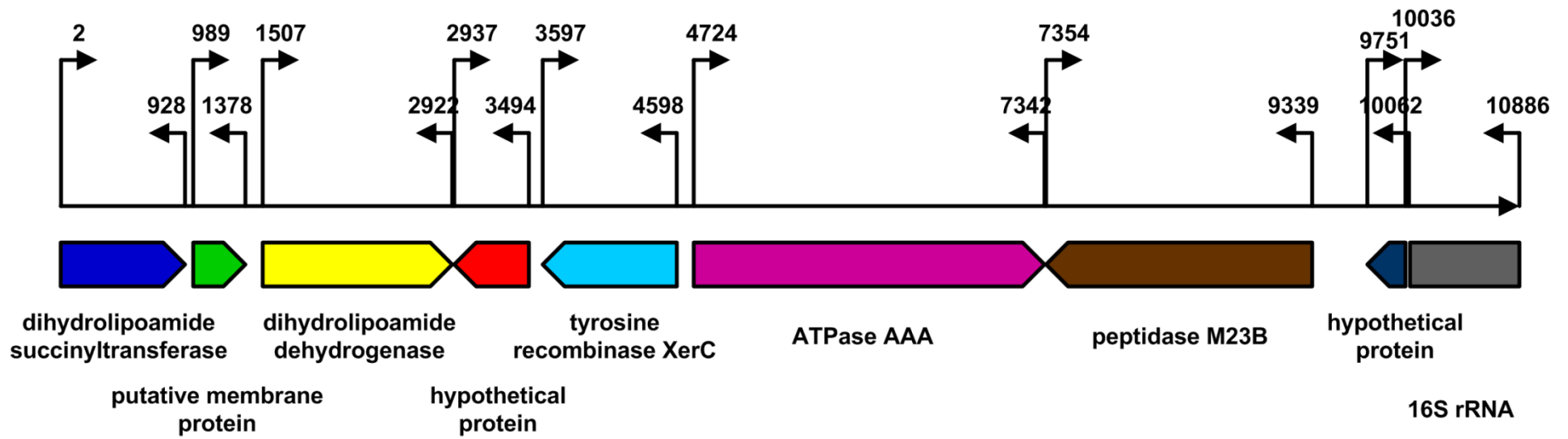
58



59

60 **Figure S2** Relative distribution of 454 reads in several bacterial groups at the genus  
 61 level, including potentially pathogenic bacteria, iron-oxidizing bacteria, and  
 62 sulfate-reducing bacteria ([Table S1](#)).

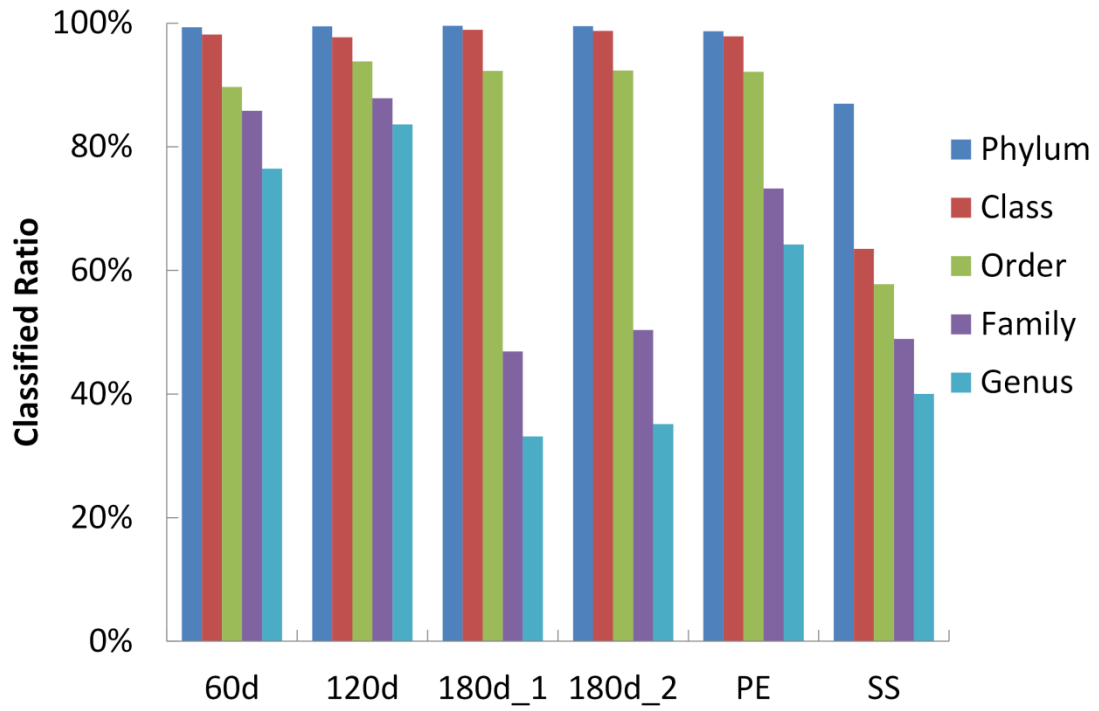
63



64

65 **Figure S3** The contig which carried 16S rRNA in the draft genome.

66

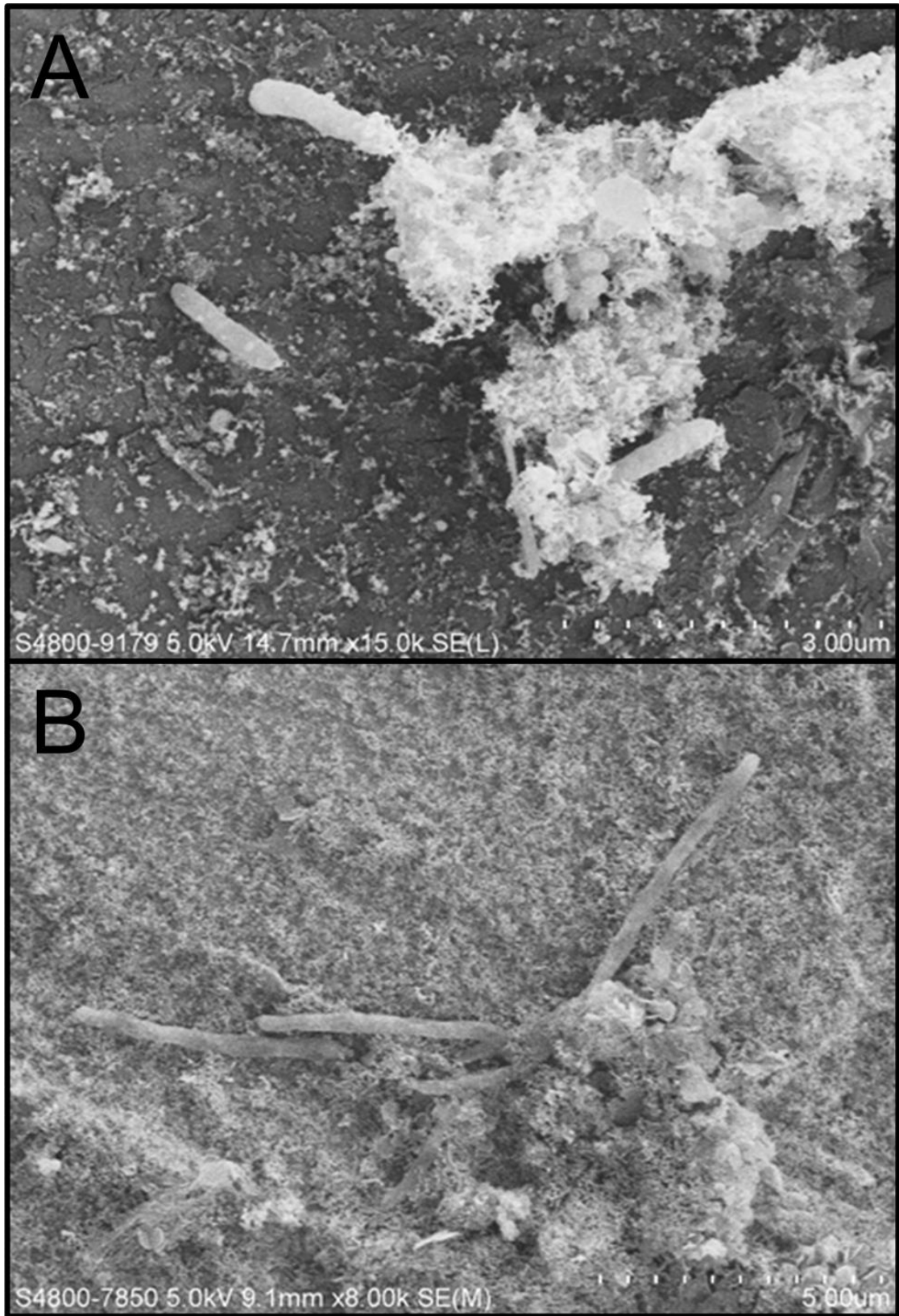


67

68 **Figure S4** Classified ratio of 454 reads at different taxonomic levels, including

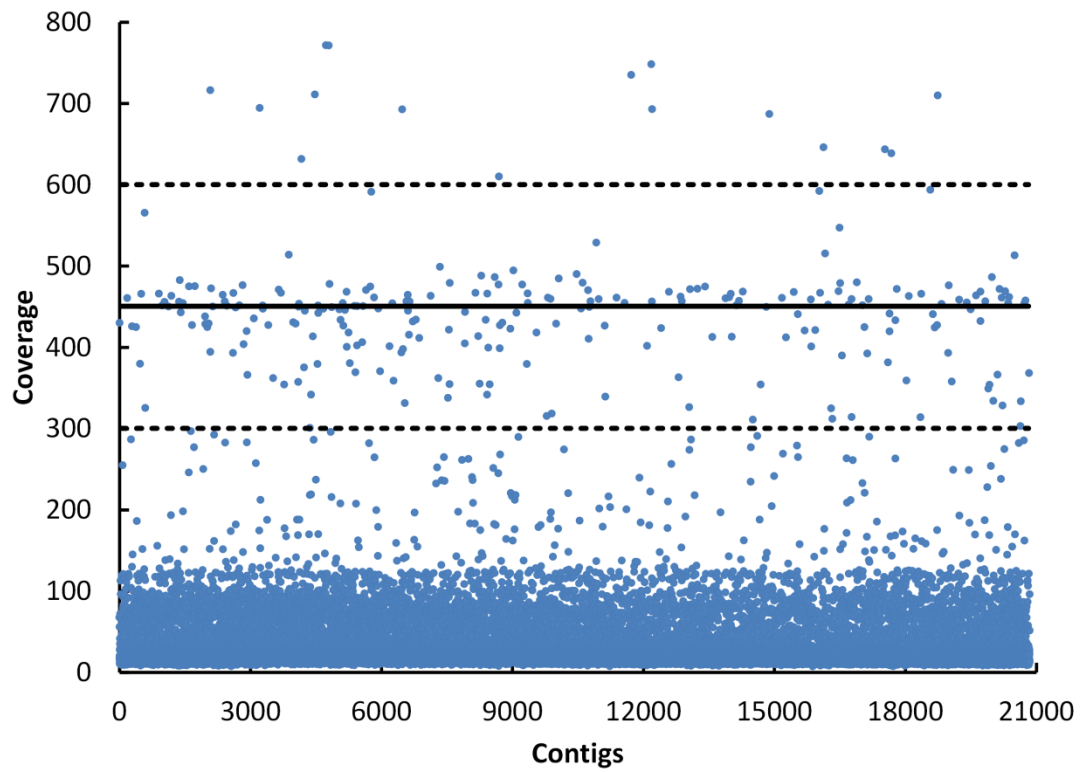
69 phylum, class, order, family and genus, in the 6 biofilms by RDP.

70



71

72 **Figure S5** SEM images of the DW biofilms formed on PE (A) and SS (B).

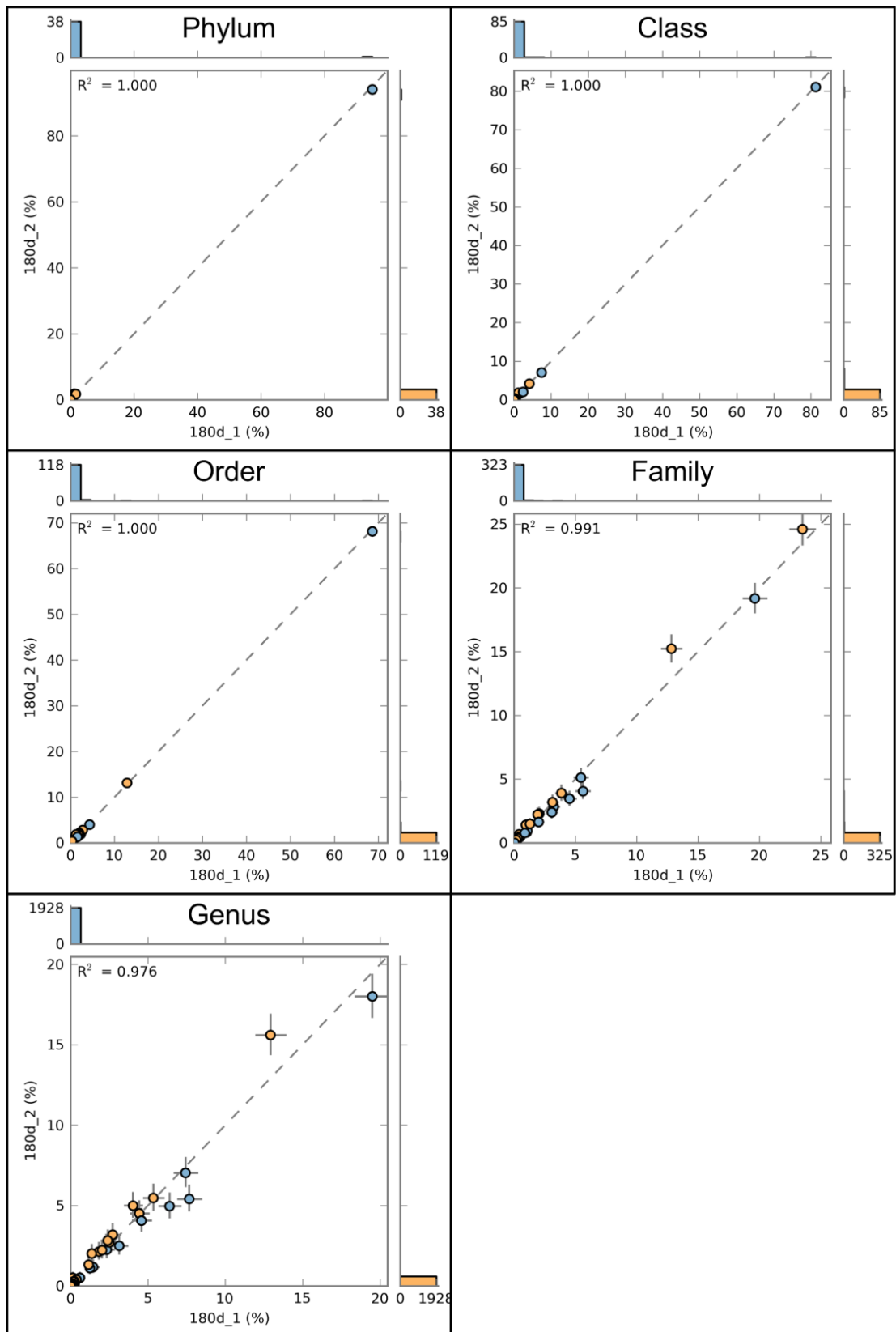


73

74 **Figure S6** Coverage distribution of the assembled contigs in the DW biofilm  
75 metagenome.

76





77

78 **Figure S7** Repeatability evaluation of two duplicates, i.e. 180d\_1 (blue) and 180d\_2

79 (brown) biofilms, by Scatter plots using STAMP software.