

1 **SUPPORTING INFORMATION**

2  
3 **New drinking water genome catalog identifies a globally distributed bacterial genus**

4 **adapted to disinfected drinking water systems**

5 Ashwin S Sudarshan<sup>1</sup>, Zihan Dai<sup>1</sup>, Marco Gabrielli<sup>2</sup>, Solize Oosthuizen-Vosloo<sup>3</sup>, Konstantinos

6 T. Konstantinidis<sup>1</sup>, Ameet J Pinto<sup>1,4\*</sup>

7  
8 <sup>1</sup> School of Civil and Environmental Engineering, Georgia Institute of Technology, Atlanta,

9 USA

10 <sup>2</sup> Department of Environmental Microbiology, Eawag, Swiss Federal Institute of Aquatic

11 Science and Technology, Dübendorf, Switzerland,

12 <sup>3</sup> Institute for Cellular and Molecular Medicine, Department of Immunology, Faculty of Health

13 Sciences, University of Pretoria, Pretoria, South Africa

14 <sup>4</sup> School of Earth and Atmospheric Sciences, Georgia Institute of Technology, Atlanta, USA

15  
16 \*Corresponding authors: [ameet.pinto@ce.gatech.edu](mailto:ameet.pinto@ce.gatech.edu)

17  
18 **Supporting information:** Four pages, eight tables, and two figures.

19

20

21

22

23

24

25

26

27

28

29

30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75

**Supplementary Table S1:** (A) Details of metagenomes used to construct the DWGC. (B) Accession numbers for isolate genomes obtained from NCBI and their isolation sources (see Excel spreadsheet).

**Supplementary Table S2:** SILVA v138.1 Phreatobacter sequences used for comparative analysis (see Excel spreadsheet).

**Supplementary Table S3:** Summary of metagenome assembly and bins across different distribution systems (see Excel spreadsheet).

**Supplementary Table S4:** Detailed taxonomy and genomic information for 1141 MAGs within the DWGC. Taxonomy was estimated using GTRDB (release 207\_v2), completeness/redundancy were estimated using CheckM2, while presence/absence of tRNAs/rRNA genes was determined using Prokka v1.14.6 (see Excel spreadsheet).

**Supplementary Table S5:** Relative Abundance of Species across the 80 distribution systems used to determine the core microbiome (see Excel spreadsheet).

**Supplementary Table S6:** Detection frequency and average relative abundance of genera in the DWGC (see Excel spreadsheet).

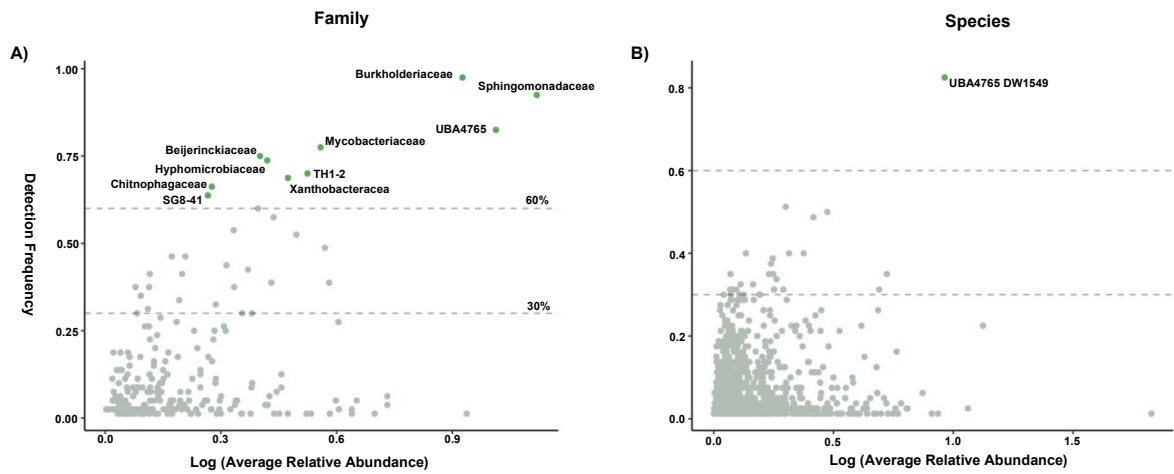
**Supplementary Table S7:** Metabolism annotation results for Lineage 1 (L1) and Lineage 2 (L2) MAGs (see Excel spreadsheet).

76  
77

78 **Supplementary Table S8:** Complete list of names proposed in the current register SeqCode list.

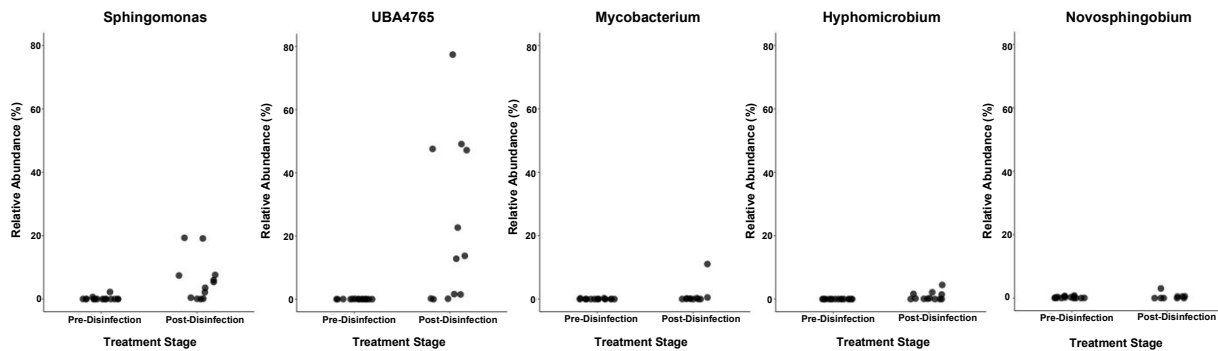
Proposed Taxon	Etymology	Description	Parent Taxon	Type	Registry URL
Genus Raskinella	[Ras.ki.ne'lla] <b>N.L. fem. n.</b> Raskin, Derived from Dr. Lutgarde Raskin; <b>N.L. fem. dim. n.</b> Raskinella, Named after Dr. Lutgarde Raskin	Raskinella is a novel genus under the family Phreatobacteraceae and order Rhizobiales. Organisms within this genus were identified in disinfected drinking water distribution systems.	<i>Phreatobacteraceae</i>	Raskinella chloraquaticus <sup>Ts</sup>	seqco.de/i:46 679
Species Raskinella chloraquaticus <sup>Ts</sup>	[chlor.a.qua'ti.cus] <b>N.L. neut. n.</b> chlorum, chlorine; <b>L. masc. adj.</b> aquaticus, of or from water; <b>N.L. masc. n.</b> chloraquaticus , relating to chlorinated water	Raskinella Chlorumaquaticus is a globally prevalent and abundant species that has been identified within disinfected drinking water systems. It is part of a unclassified genus "Raskinella" under the family Phreatobacteraceae.	<i>Raskinella</i>	NCBI Assembly: GCA_00236517 5.1 <sup>Ts</sup>	seqco.de/i:46 680

79  
80  
81  
82  
83  
84  
85



87  
88  
89  
90  
91  
92

**Supplementary Figure S1:** A) Detection Frequency Vs Log of Average Relative abundance of the DWGC genomes at Family level. B) Detection Frequency Vs Log of Average Relative abundance of the DWGC genomes at Species level.



93  
94  
95  
96

**Supplementary Figure 2:** Relative abundance of MAGs from five genera observed in more than 60% of the systems Pre- and Post-disinfection in studies where Pre-disinfection metagenomes were available.