# Supplementary Information for *Legionella* relative abundance in shower hose biofilms is associated with specific microbiome members

Alessio Cavallaro[1,2], William J. Rhoads[1], Emile Sylvestre[1], Thierry Marti[1,2], Jean-Claude Walser[2], Frederik Hammes[1]*

[1] Department of Environmental Microbiology, Eawag, Swiss Federal Institute of Aquatic Science and Technology, 8600 Dübendorf, Switzerland

[2] Department of Environmental Systems Science, Institute of Biogeochemistry and Pollutant Dynamics, ETH Zurich, 8092 Zurich, Switzerland

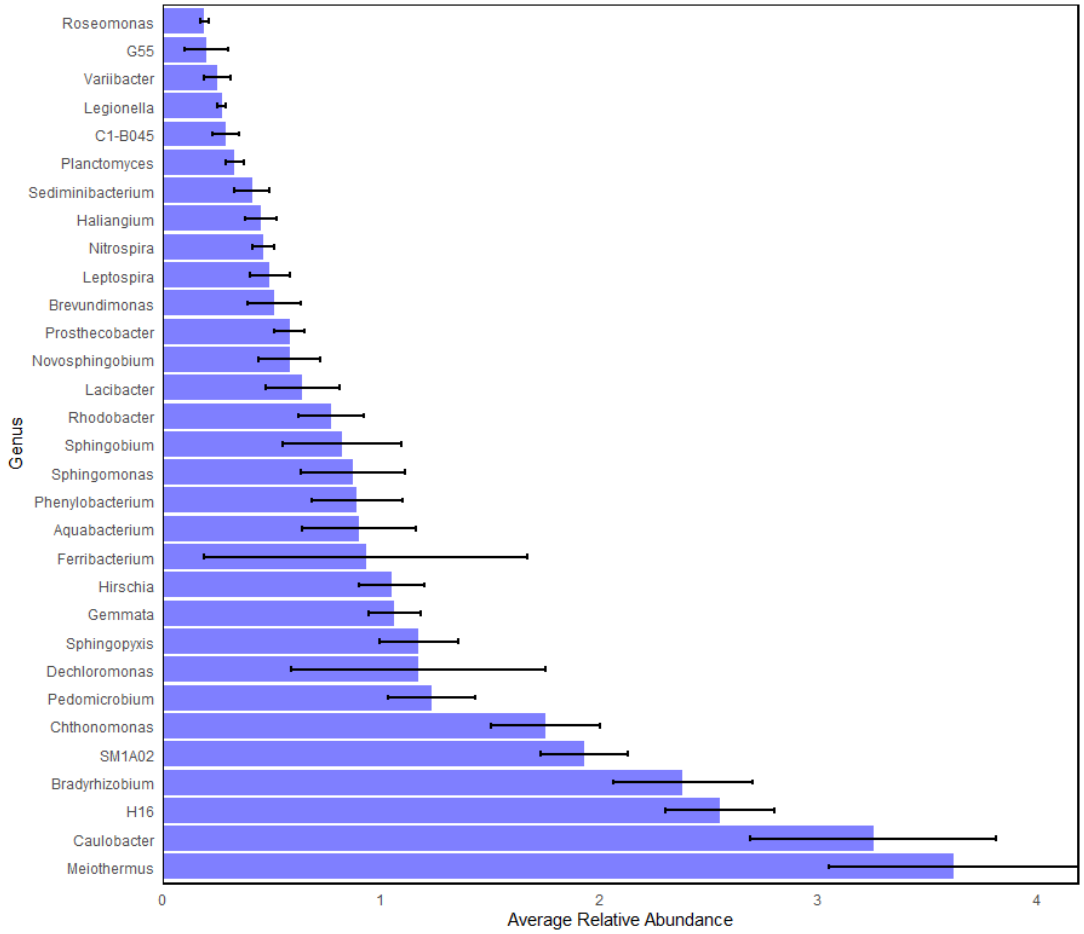
* Corresponding author:

Name: Frederik Hammes

Tel.: +41 58 765 5372

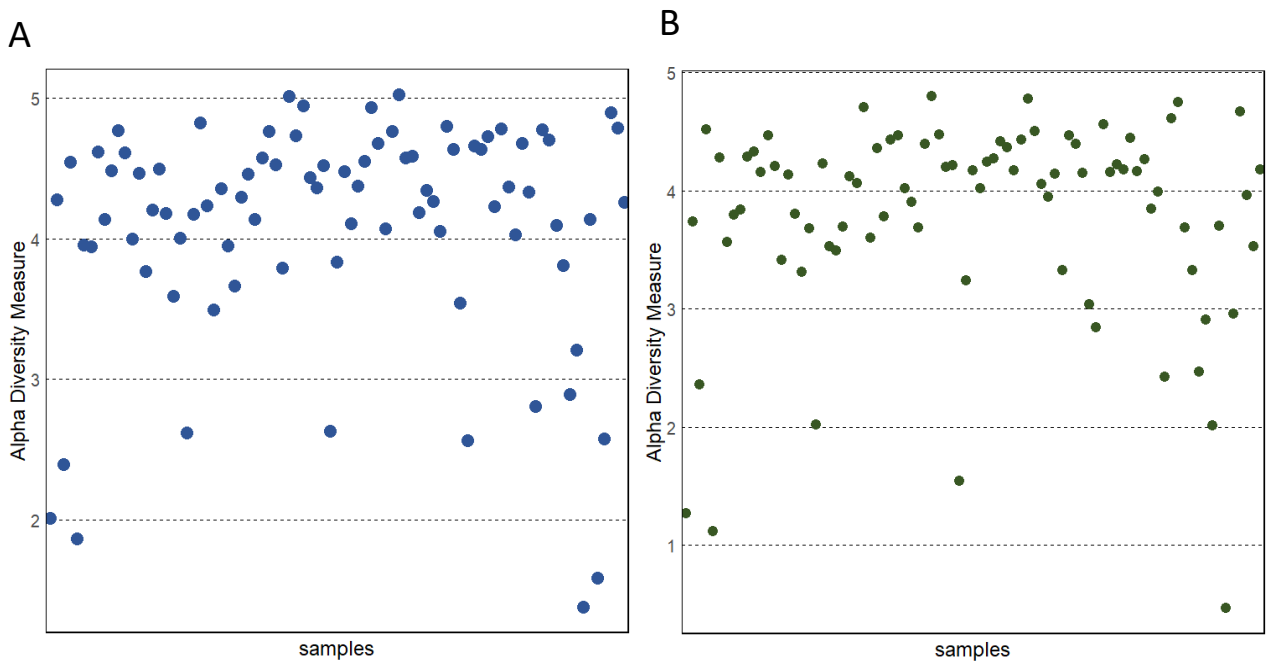
Email: frederik.hammes@eawag.ch

Table of Contents



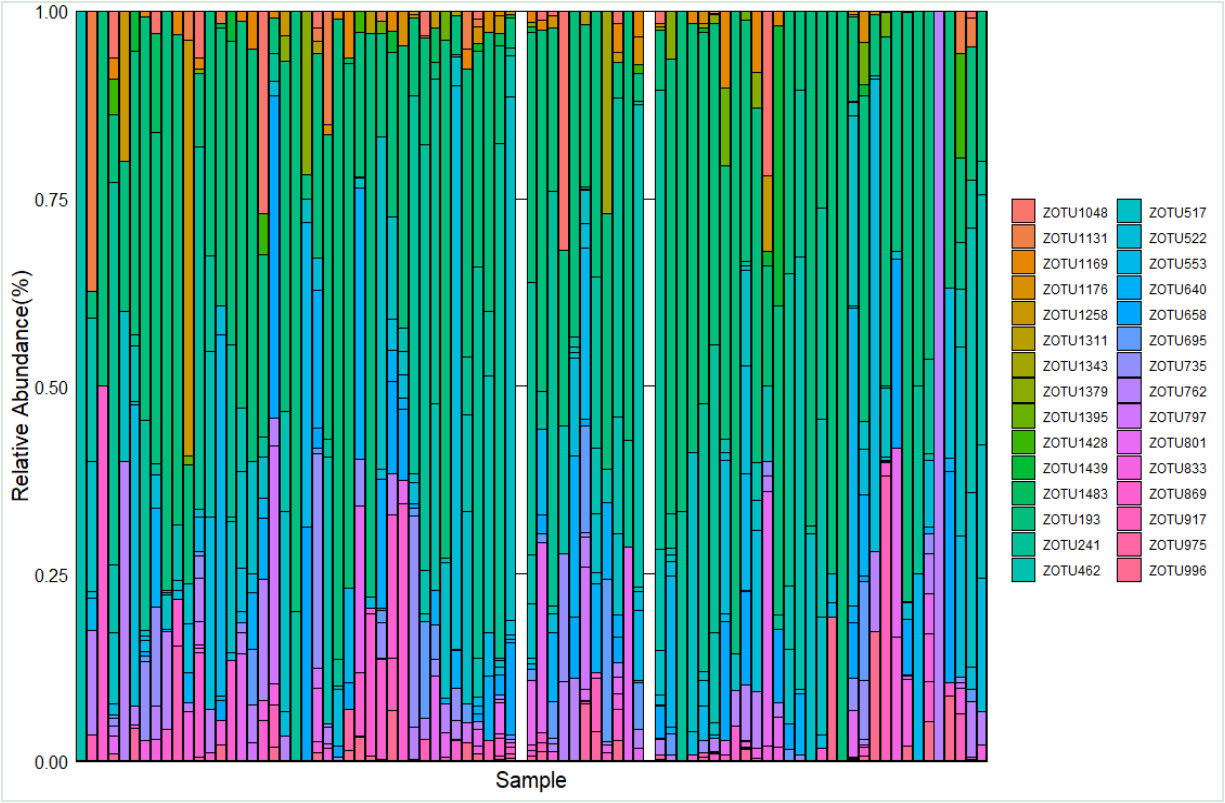

**Supplementary Figure 1. Average Relative Abundance of the 30 top taxa classified at genus level**

The figure shows the average relative abundance across samples of the 30 most abundant taxa classified at genus level. The relative abundance is shown in increasing order from the less abundant to the most abundant, according to the absolute value. A standard error has been included in the plot to show the variations across samples. The relative abundance has been calculated using the ZOTU count table produced with the 16S rRNA gene amplicon sequencing.



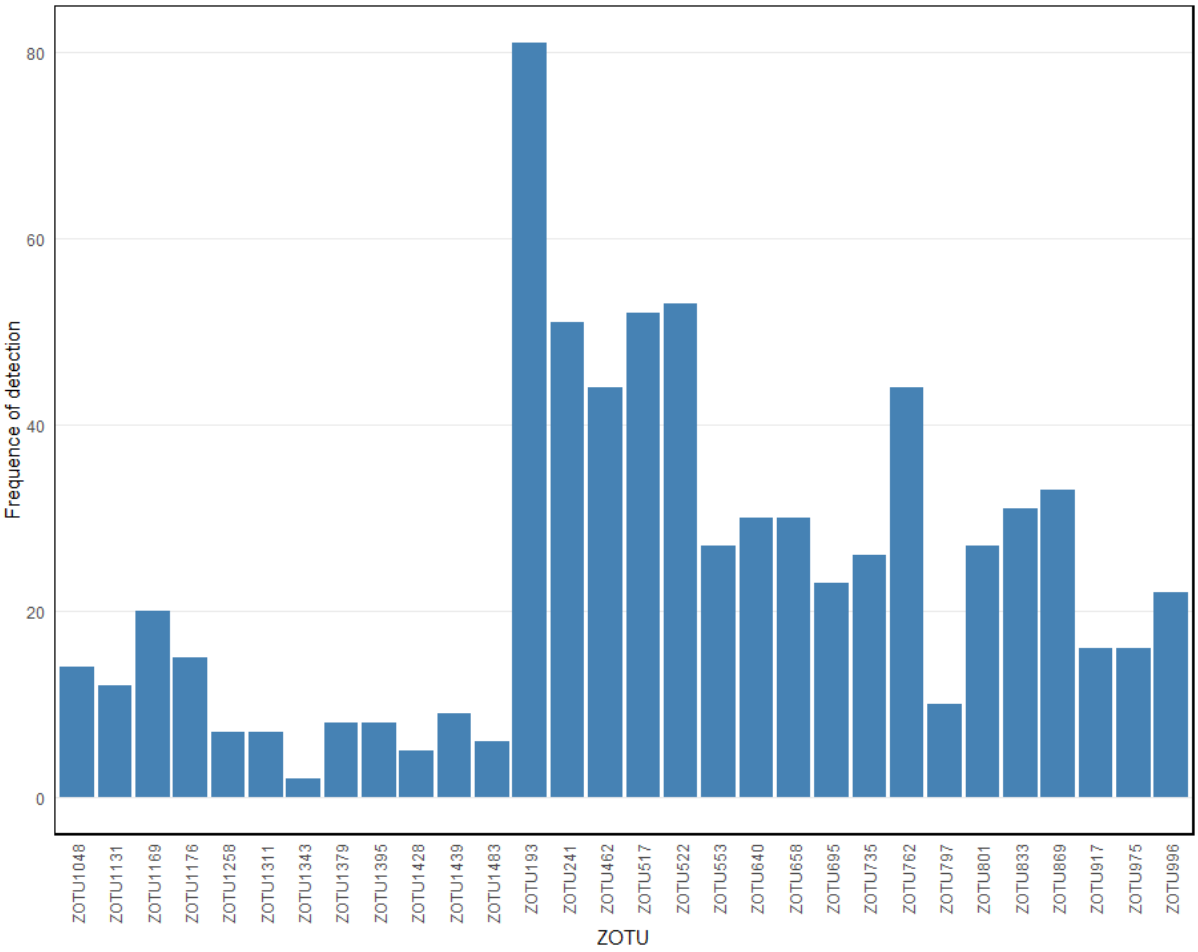

**Supplementary Figure 2. Alpha diversity of the 16S and 18S datasets**

The alpha diversity for the 16S and 18S rRNA gene amplicon sequencing datasets has been calculated using the Shannon index, which takes into account both richness and evenness. The Shannon index (H) is indicated in the y-axis, and the intra-sample diversity is considered higher as the H index increases.



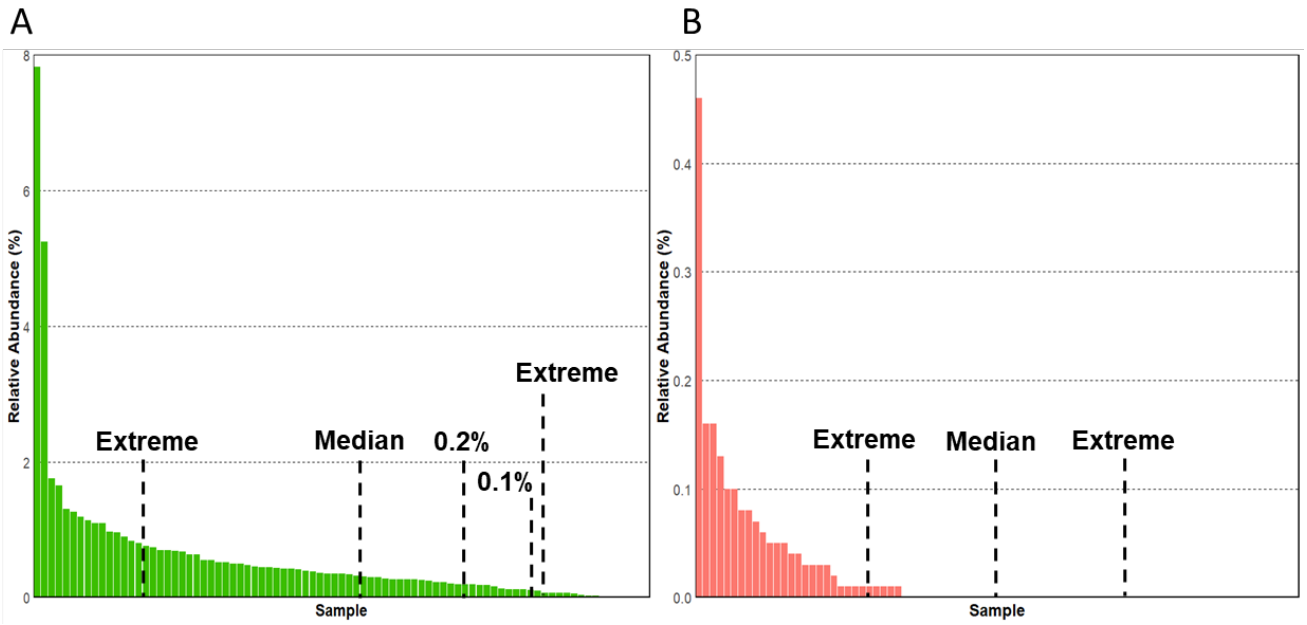

**Supplementary Figure 3. Relative Abundance of the individual *Legionella* ZOTUs across samples**

This stacked bar plot displays the distribution of the 30 ZOTUs assigned to the genus *Legionella* across multiple samples, with each bar representing a single sample. The bar height reflects the total abundance of the genus within each sample, while the segments within the bar indicate the relative abundance of individual ZOTUs. A white bar indicates that no Legionella ZOTUs were detected in a particular sample. This visualization provides insight into the diversity and relative importance of different ZOTUs within the *Legionella* genus across samples.




**Supplementary Figure 4. Frequency of detection of the individual *Legionella* ZOTUs across samples**

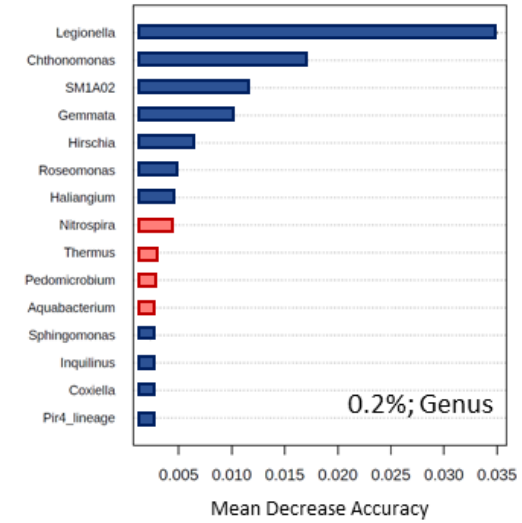
This bar plot displays the frequency of detection of the 30 ZOTUS assigned to the genus Legionella across samples. Each bar corresponds to an individual ZOTU, while the bar height reflects in how many samples the ZOTU was detected (number of samples indicated in the y-axis.

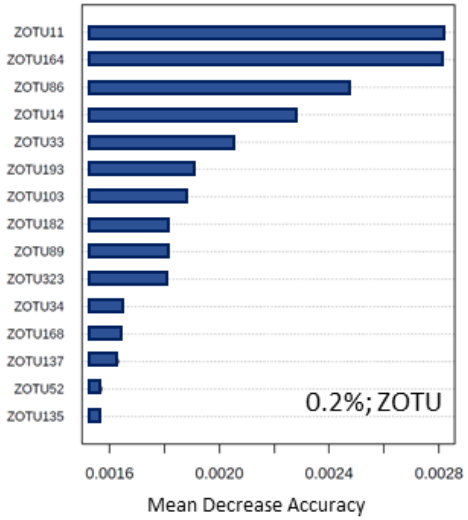


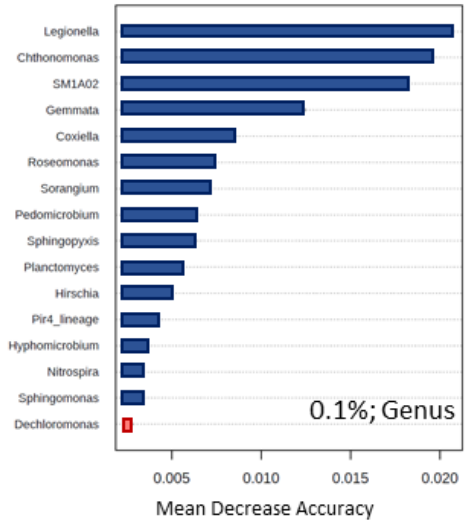

**Supplementary Figure 5. Thresholds used for Random Forest analysis on *Legionella* spp. and *L. pneumophila* relative abundance**

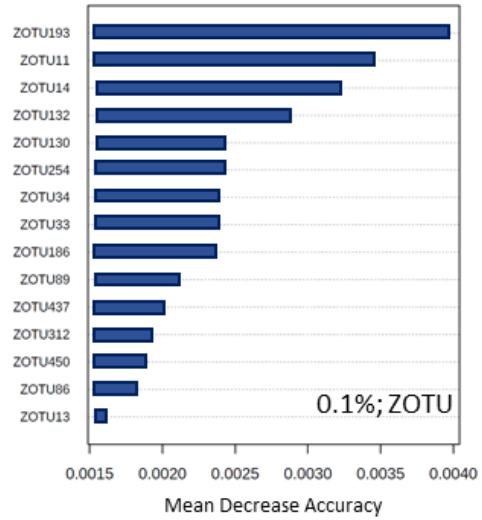
The figure shows how the different thresholds used to determine the variables in the Random Forest analysis were applied to the relative abundance of *Legionella* spp. and *L. pneumophila*. A) The selection of the thresholds for *Legionella* spp. relative abundance. The median value (0.347%) creates two groups of 43 and 42 samples, while only 15 samples per side were choses as extremes based on their relative abundance. Using a threshold of 0.2% results in an uneven sample distribution, with 27 low-*Legionella* samples and 58 high-*Legionella* samples. Further decreasing the threshold to 0.1% creates two groups of 17 and 68 samples. B) Selection of the thresholds for L.pneumophila. The median value (0.002%) creates two groups of 39 and 46 samples (three samples have a relative abundance of 0.002% and have been all included in the same group). 25 samples at the extremes were chosen for this analysis, in order to accommodate all the samples with no *L. pneumophila* in one group.

Prokaryotes

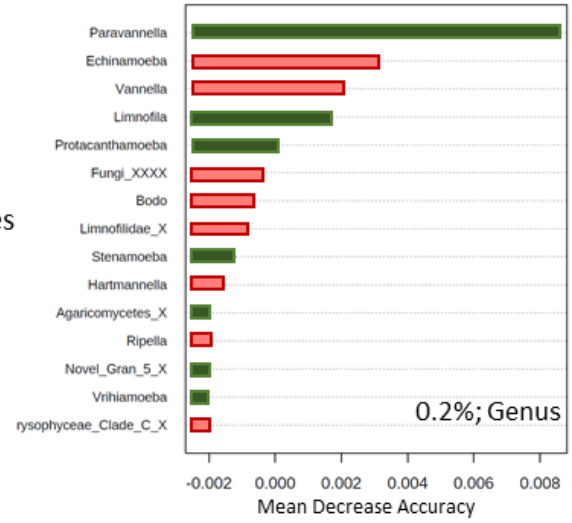


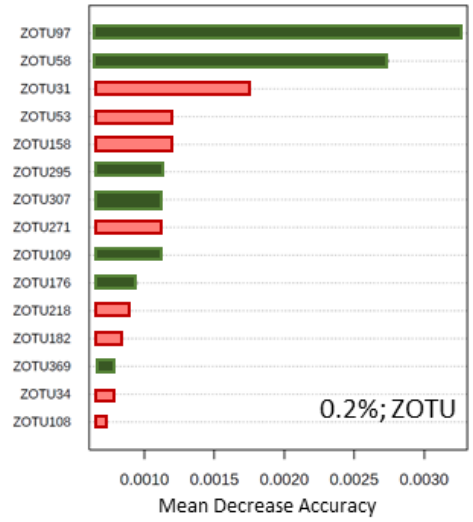



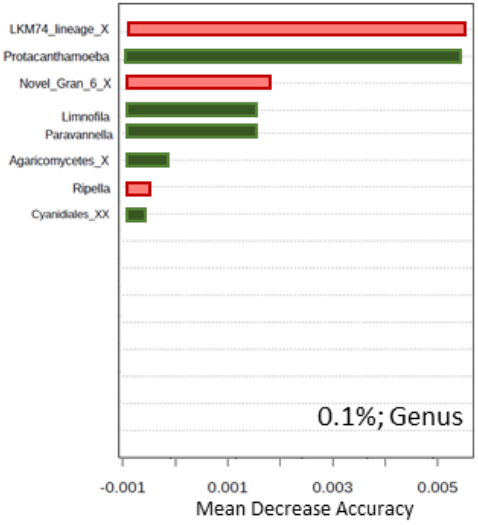



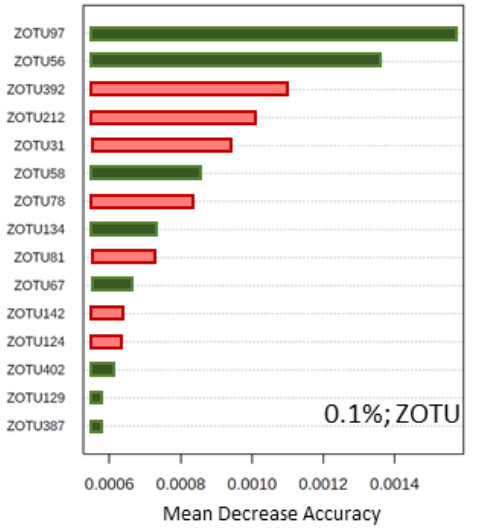




Eukaryotes



**Supplementary Figure 6. Random Forest analysis on *Legionella* spp. relative abundance using additional thresholds**

Random Forest analysis of 16S and 18S datasets using the relative abundance of *Legionella* spp. as variable and additional thresholds (0.2% and 0.1% Legionella spp. relative abundance). A-B: Main prokaryotes (genus level, A; ZOTUs level, B) predicting *Legionella* spp. abundance with a threshold of 0.2% relative abundance; C-D: Main prokaryotes (genus level, C; ZOTU level, D) predicting *Legionella spp.* abundance with a threshold of 0.1% relative abundance. E-F: Main eukaryotes (genus level, E; ZOTU, F) predicting *Legionella* spp. abundance using 0.2% relative abundance as threshold; G-H: Main eukaryotes (ZOTU, G) predicting *Legionella* spp. abundance using a threshold of 0.1% relative abundance.

| Reagent | Final Conc. | Volume (uL) |
|---|---|---|
| Nuclease Free Water | | 10.53 |
| PerfeCTa Multiplex ToughMix 5X | 1X | 5.4 |
| Fluorescein 1uM | 100 nM | 2.7 |
| Forward Primer 1 (20uM) | 0.6 uM | 0.675 |
| Reverse Primer 1 (20uM) | 0.6 uM | 0.675 |
| Probe 1 (20uM) | 0.15 uM | 0.135 |
| Forward Primer 2 (20uM) | 0.4 uM | 0.675 |
| Reverse Primer 2 (20uM) | 0.4 uM | 0.675 |
| Probe 2 (20uM) | 0.15 uM | 0.135 |
| DNA Template | | 5.4 |
| Total | | 27 |

| Primer/Probe | Sequence |
|---|---|
| ssrA forward | GGC GAC CTG GCT TC |
| ssrA reverse | TCA TCG TTT GCA TTT ATA TTT A |
| ssrA probe | HEX-ACG TGG GTT GCA A-BHQ1 |
| mip forward | TTG TCT TAT AGC ATT GGT GCC G |
| mip reverse | CCA ATT GAG CGC CAC TCA TAG |
| mip probe | FAM-CGG AAG CAA TGG CTA AAG GCA TGC A-BHQ1 |

| Step | Settings |
|---|---|
| Partitioning | 12 min / 40°C |
| Enzyme activation | 10 min / 95°C |
| 45 cycles | 15 sec / 95°C; 60 sec / 55°C |
| Depressurization | Ambient Temperature |

**Supplementary Table 1. Primers, probes and ddPCR reagents and conditions for a duplex assay to detect *Legionella* spp. and *Legionella pneumophila***

| Reagent | Final Conc. | Volume (uL) |
|---|---|---|
| Nuclease Free Water | | 14.76 |
| PerfeCTa Multiplex ToughMix 5X | 0.75X | 4.05 |
| Evagreen 20X | 1.5X | 2.025 |
| Forward Primer 1 (10uM) | 0.1 uM | 0.216 |
| Reverse Primer 1 (10uM) | 0.1 uM | 0.27 |
| Alexa 0.1 ug/uL | 0.8 ng/uL | 0.27 |
| DNA Template | | 5.4 |
| Total | | 27 |

| Primer | Sequence |
|---|---|
| 515F | GTG CCA GCM GCC GCG GTA A |
| 805R | GGA CTA CHV GGG TWT CTA AT |

| Step | Settings |
|---|---|
| Partitioning | 12 min / 40°C |
| Enzyme activation | 10 min / 95°C |
| 45 cycles | 15 sec / 95°C; 60 sec / 50°C |
| Depressurization | Ambient Temperature |

**Supplementary Table 2. Primers, probes and ddPCR reagents and conditions for the detection of the total 16S genes**

| Reagent PCR 1 | Final Conc. | Volume (uL) |
|---|---|---|
| Nuclease Free Water | | 6 |
| KAPA HiFi HotStart ReadyMix 2X | 1X | 12.5 |
| Forward Primer (10uM) | 0.3 uM | 0.75 |
| Reverse Primer (10uM) | 0.3 uM | 0.75 |
| DNA Template | | 5 |
| Total | | 25 |
| **Reagent PCR 2** | **Final Conc.** | **Volume (uL)** |
| Nuclease Free Water | | 22.5 |
| KAPA HiFi HotStart ReadyMix 2X | 1X | 12.5 |
| Nextera XT Index Primer 1 | | 5 |
| Nextera XT Index Primer 2 | | 5 |
| DNA Template | | 5 |
| Total | | 50 |

| Primer | Sequence |
|---|---|
| 515F (16S) | GTG CCA GCM GCC GCG GTA A |
| 805R (16S) | GGA CTA CHV GGG TWT CTA AT |
| EUK1391F (18S) | GTAC ACA CCG CCC GTC |
| EUK1510R (18S) | CCT TCY GCA GGT TCA CCT AC |

| Step PCR 1 (16S) | Settings |
|---|---|
| Denaturation | |
| 24 cycles | 20 sec / 98°C; 15 sec / 50°C; 15 sec/72°C |
| Final elongation | 5 min/ 72°C |
| **Step PCR 1 (18S)** | **Settings** |
| Denaturation | |
| 25 cycles | 20 sec / 98°C; 15 sec / 57°C; 15 sec/72°C |
| Final elongation | 5 min/ 72°C |
| **Step PCR 2** | **Settings** |
| Denaturation | 3 min/95°C |
| 10 cycles | 30 sec / 95°C; 30 sec / 55°C; 30 sec/72°C |
| Final elongation | 5 min/ 72°C |

**Supplementary Table 3. Primers, PCR reagents and conditions for the amplicon sequencing library preparation**