Absolute quantification of microbial taxon abundances

Ruben Props[1,2], Frederiek-Maarten Kerckhof[1], Peter Rubbens[3], Jo De Vrieze[1], Emma Hernandez Sanabria[1], Willem Waegeman[3], Pieter Monsieurs[2], Frederik Hammes[4] and Nico Boon[1,*]

[1] Center for Microbial Ecology and Technology (CMET), Ghent University, Coupure Links 653, B-9000 Gent, Belgium

[2] Belgian Nuclear Research Centre (SCK•CEN), Boeretang 200, B-2400 Mol, Belgium

[3] KERMIT, department of Mathematical modelling, Statistics and Bioinformatics, Ghent University, Coupure Links 653, B-9000, Ghent, Belgium

[4] Department of Environmental Microbiology, Eawag - Swiss Federal Institute for Aquatic Science and Technology, Überlandstrasse 133, CH-8600, Dübendorf, Switzerland

* Corresponding author: Nico Boon, Ghent University; Faculty of Bioscience Engineering; Center for Microbial Ecology and Technology (CMET); Coupure Links 653; B-9000 Gent, Belgium; phone: +32 (0)9 264 59 76; fax: +32 (0)9 264 62 48; E-mail: Nico.Boon@UGent.be; Webpage: www.cmet.UGent.be.

Keywords: microbial ecology, flow cytometry, microbial community dynamics, amplicon sequencing, 16S rRNA gene, microbial biodiversity, microbiome

CONFLICTS OF INTEREST

The authors declare that there exist no conflicts of interest.

This document is the accepted manuscript version of the following article:
Props, R., Kerckhof, F. M., Rubbens, P., De Vrieze, J., Hernandez Sanabria, E., Waegeman, W., … Boon, N. (2017). Absolute quantification of microbial taxon abundances. ISME Journal, 11(587), 584-587.
https://doi.org/10.1038/ismej.2016.117

ABSTRACT

High-throughput amplicon sequencing has become a well-established approach for microbial community profiling. Correlating shifts in the relative abundances of bacterial taxa with environmental gradients is the goal of many microbiome surveys. As the abundances generated by this technology are semi-quantitative by definition, the observed dynamics may not accurately reflect those of the actual taxon densities. We combined the sequencing approach (16S rRNA gene) with robust single-cell enumeration technologies (flow cytometry) to quantify the absolute taxon abundances. A detailed longitudinal analysis of the absolute abundances resulted in distinct abundance profiles that were less ambiguous and expressed in units that can be directly compared across studies. We further provide evidence that the enrichment of taxa (increase in relative abundance) does not necessarily relate to the outgrowth of taxa (increase in absolute abundance). Our results highlight that both relative and absolute abundances should be considered for a comprehensive biological interpretation of microbiome surveys.

Recent advancements in high-throughput sequencing of marker genes, such as the 16S rRNA gene, have provided microbial ecologists the tools to accurately infer the relative composition of microbial communities (Franzosa et al 2015). This resulted in a widespread application of the technology in longitudinal studies where shifts in community structure are related to environmental variables and functional outputs (Faust et al 2015, Wilhelm et al 2015). An inherent limitation of the sequencing technology is that the calculated taxon abundances comprise relative values (Widder et al 2016). Hence, caution must be taken with the biological interpretation of these values, since inter-sample differences in cell density are not considered. To our knowledge, there are no descriptive studies that assess the extent to which relative abundances deliver a skewed image of the actual microbial community dynamics. In this study we combined robust cell density measurements from flow cytometry (Prest et al 2013, Van Nevel et al 2013) with the relative abundances derived from 16S rRNA gene amplicon sequencing. We performed two extensive longitudinal surveys on the central water reservoir of a cooling water system. This engineered freshwater ecosystem was subjected to highly controlled operational phases (**Supplementary information (SI) and dataset**). We quantified the absolute taxon abundances and assessed whether additional insights could be attained with the combined approach.

Based on the sample-specified total cell density, the absolute taxon abundances were calculated for each time point. Individual taxon densities ranged from 0.5 to 1 679 cells μl^{-1}. Several inter-taxon differences became apparent by performing ordinary least squares regression analysis for the three most abundant taxa, which originate from different taxonomic clades (**Fig. 1**) (Newton et al 2011). We identified a significant difference between OTU1 (betI-A clade) and both OTU2 and OTU3 (bacI-A clade; $p=0.044$ and $p=0.046$ for OTU2 and OTU3, respectively), however, no significant difference was observed between OTU2 and OTU3 ($p=0.51$). These findings suggest that identical relative abundances of different taxa require taxon-dependent biological interpretation because they do not necessarily reflect the same absolute abundances. To further verify the limitations of relative

abundances, we closely inspected the temporal trajectories of both clades (**Fig. 2**). Throughout the two surveys, the betI-A clade (OTU1) displayed similar variation in relative abundance (coefficient of variation (CV_{rel}) = 54%) and cell density (CV_{dens} = 50%), whereas the bacI-A clade (OTU2 and OTU3) displayed distinct transient behaviour (CV_{rel} = 124%, CV_{dens} = 172%). Overall, both surveys were characterized by dynamic shifts in community composition and density that were interspersed with stable periods (i.e., beginning and end of each survey).

When interpreting the temporal trajectories of the relative and absolute abundances, two primary discrepancies could be detected, potentially leading to misinterpretation if conclusions would have been based solely on the relative abundances. First, during dynamic community growth, such as the start-up and early reactor operation in survey 1, there was a well-defined transition in absolute abundance, showing the systematic outgrowth and decay of the bacI-A clade. In contrast, the relative abundance profiles were more ambiguous. They remained relatively constant during the growth and decay event and only conclusively indicated the beginning and the end. Another striking discrepancy could be observed in the second half of reactor operation during survey 2. The relative abundance profiles displayed an increase from ±40% to >90%, potentially indicating a selective outgrowth of the betI-A clade. If instead the absolute abundances were considered, there was no distinct pattern of active outgrowth visible for this clade; its cell densities never considerably exceeded the maximum density that was observed at the end of the start-up phase (relative abundance = 57%). This suggests that the decay of other taxa is responsible for this enrichment within the community structure, or that environmental constraints limit its effective outgrowth. Overall, the dynamics of the betI-A and bacI-A clades could be more precisely specified with the absolute abundance profiles. Bulk cell density measurements were, in themselves, a poor descriptive parameter of the microbial community dynamics (**Figure SI1**). Only one OTU's relative abundance (OTU2) was strongly correlated to the total cell density (Pearson's correlation: r_p = 0.60, p < 0.01, n = 79), while the mean correlation strength for the entire

88 community composition was -0.08 ± 0.15 (n = 427). This shows that only for OTU2, its
89 outgrowth (increase in absolute abundance) frequently corresponds with its enrichment
90 (increase in relative abundance).

91 From our results we are able to show that absolute quantification of taxon dynamics is
92 essential and can shed additional light on many outstanding questions within microbial
93 ecology. Next to flow cytometry, quantitative PCR (qPCR) and fluorescence in-situ
94 hybridization (FISH) may represent alternative approaches for estimating absolute cell
95 densities. The tandem of qPCR and sequencing may be appealing because qPCR and
96 amplicon sequencing analyses start from the same DNA-extract and thus incorporate similar
97 laboratory-induced bias. However, for environmental samples qPCR is only sensitive enough
98 to separate twofold changes in gene concentration (proxy for cell abundance) (Smith and
99 Osborn 2009). qPCR also suffers from specific limitations such as amplification efficiency
100 and primer specificity, which makes it unadvised to compare the results between studies and
101 even assays on the same device (Brankatschk et al 2012, Smith and Osborn 2009). FISH
102 provides a PCR-independent approach for calculating relative taxa abundances or, in case of
103 a standardized methodological approach, even estimates of absolute abundances (Daims et
104 al 2001). Unfortunately, FISH analyses enumerate only the active fraction of the community
105 since the analysis is based on the hybridization of fluorescent probes with the 16S rRNA
106 (Amann and Fuchs 2008). These analyses are also more laborious, and generally provide
107 limited sample sizes (i.e., hundreds of cells). In contrast, our flow cytometric enumeration
108 approach of absolute cell densities is robust and high-throughput, but it requires supervised
109 denoising strategies to account for instrument and (in)organic noise, as well as cell
110 aggregates (materials and methods section, SI).

111 Methodological limitations become particularly crucial when the central hypothesis pertains to
112 the "*rare microbiome*". This fraction is currently defined at arbitrary thresholds of 0.1 to 0.01%
113 (Lynch and Neufeld 2015). Although the definition of rare taxa will always remain partially
114 based on *ad-hoc* assumptions (Haegeman et al 2013), several bioinformatics-based tools

have shown the substantial impact of varying rarity thresholds on community analyses (Gobet et al 2010). By also taking into account the absolute taxon densities, which are comparable between studies, the development of a more consistent framework may now prove possible. In this study, we did not take into account the number of 16S rRNA gene copies during the calculation of the absolute abundances because there were insufficient closely related reference genomes available (SI and figure SI2). For more characterized environments this additional normalization may improve the resolution of absolute abundance calculations (Langille et al 2013, Stoddard et al 2015). Overall, our results demonstrate that when united, robust cell density measurements and phylogenetic marker gene data are able to project a more comprehensive image of the compositional dynamics occurring in microbial ecosystems.

REFERENCES

Amann R, Fuchs BM (2008). Single-cell identification in microbial communities by improved fluorescence in situ hybridization techniques. *Nature Reviews Microbiology* **6:** 339-348.

Brankatschk R, Bodenhausen N, Zeyer J, Buergmann H (2012). Simple Absolute Quantification Method Correcting for Quantitative PCR Efficiency Variations for Microbial Community Samples. *Applied and Environmental Microbiology* **78:** 4481-4489.

Daims H, Ramsing NB, Schleifer KH, Wagner M (2001). Cultivation-independent, semiautomatic determination of absolute bacterial cell numbers in environmental samples by fluorescence in situ hybridization. *Applied and Environmental Microbiology* **67:** 5810-5818.

Faust K, Lahti L, Gonze D, de Vos WM, Raes J (2015). Metagenomics meets time series analysis: unraveling microbial community dynamics. *Current Opinion in Microbiology* **25:** 56-66.

Franzosa EA, Hsu T, Sirota-Madi A, Shafquat A, Abu-Ali G, Morgan XC *et al* (2015). Sequencing and beyond: integrating molecular 'omics' for microbial community profiling. *Nature Reviews Microbiology* **13:** 360-372.

Gobet A, Quince C, Ramette A (2010). Multivariate Cutoff Level Analysis (MultiCoLA) of large community data sets. *Nucleic Acids Research* **38**.

Haegeman B, Hamelin J, Moriarty J, Neal P, Dushoff J, Weitz JS (2013). Robust estimation of microbial diversity in theory and in practice. *Isme Journal* **7:** 1092-1101.

Langille MGI, Zaneveld J, Caporaso JG, McDonald D, Knights D, Reyes JA *et al* (2013). Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nature Biotechnology* **31:** 814-+.

Lynch MDJ, Neufeld JD (2015). Ecology and exploration of the rare biosphere. *Nature Reviews Microbiology* **13:** 217-229.

Newton RJ, Jones SE, Eiler A, McMahon KD, Bertilsson S (2011). A Guide to the Natural History of Freshwater Lake Bacteria. *Microbiology and Molecular Biology Reviews* **75:** 14-49.

Prest EI, Hammes F, Kotzsch S, van Loosdrecht MCM, Vrouwenvelder JS (2013). Monitoring microbiological changes in drinking water systems using a fast and reproducible flow cytometric method. *Water Research* **47:** 7131-7142.

Smith CJ, Osborn AM (2009). Advantages and limitations of quantitative PCR (Q-PCR)-based approaches in microbial ecology. *FEMS Microbiol Ecol* **67:** 6-20.

Stoddard SF, Smith BJ, Hein R, Roller BRK, Schmidt TM (2015). rrnDB: improved tools for interpreting rRNA gene abundance in bacteria and archaea and a new foundation for future development. *Nucleic Acids Research* **43:** D593-D598.

Van Nevel S, Koetzsch S, Weilenmann H-U, Boon N, Hammes F (2013). Routine bacterial analysis with automated flow cytometry. *Journal of Microbiological Methods* **94:** 73-76.

Widder S, Allen RJ, Pfeiffer T, Curtis TP, Wiuf C, Sloan WT *et al* (2016). Challenges in microbial ecology: building predictive understanding of community function and dynamics. *Isme J*.

Wilhelm L, Besemer K, Fragner L, Peter H, Weckwerth W, Battin TJ (2015). Altitudinal patterns of diversity and functional traits of metabolically active microorganisms in stream biofilms. *Isme Journal* **9:** 2454-2464.

DATA AVAILABILITY

Flow cytometry data (.fcs format) are available on the FlowRepository archive under repository ID FR-FCM-ZZNA and the Dryad Digital Repository (http://dx.doi.org/10.5061/dryad.m1c04). Sequences are available on the NCBI Sequence Read Archive (SRA) under accession number SRP066190.

ACKNOWLEDGEMENTS

This work was supported through the Inter-University Attraction Pole (IUAP) “µ-manager” funded by the Belgian Science Policy (BELSPO, P7/25) and Geconcerteerde Onderzoeksactie (GOA) from Ghent University (BOF15/GOA/006). RP is supported by Ghent University (BOFDOC2015000601) and the Belgian Nuclear Research Centre (SCK•CEN). JDV and EH-S are supported as postdoctoral fellows from the Research Foundation Flanders (FWO- Vlaanderen). F-MK is supported by the Ghent University Geconcerteerde onderzoeksactie (BOF15/GOA/006). PR is supported by Ghent University (BOFSTA2015000501).

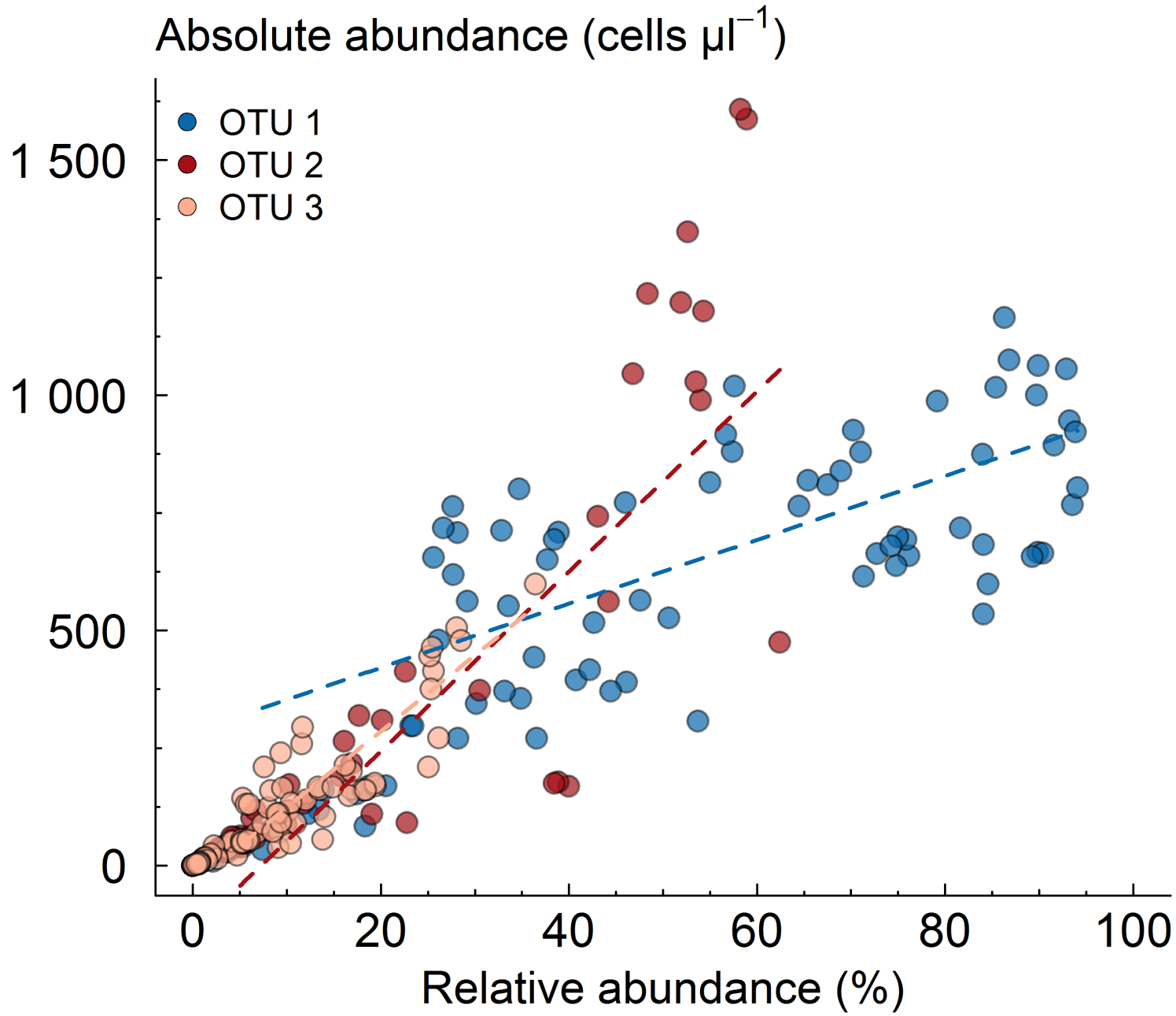
AUTHOR CONTRIBUTIONS

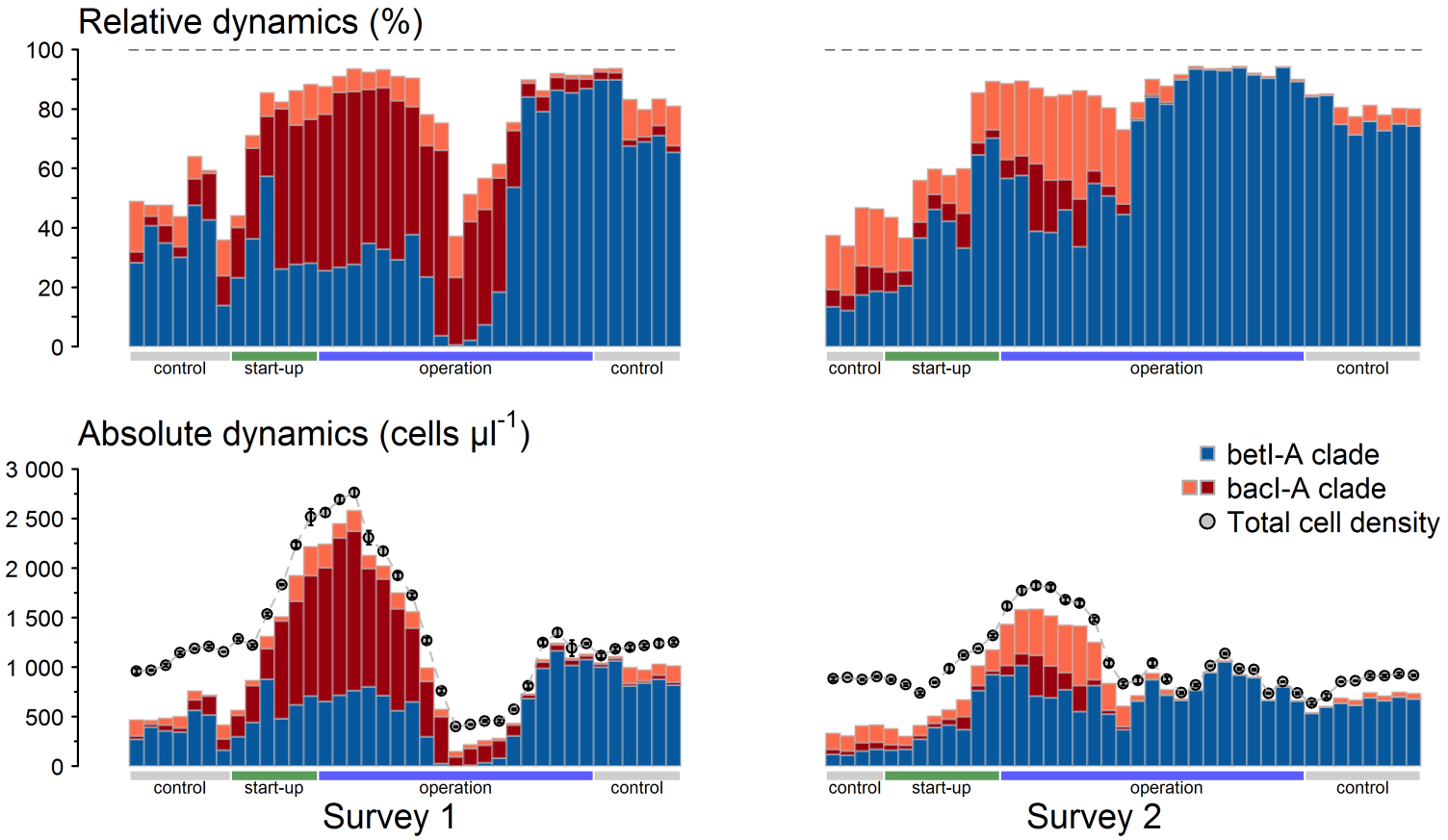
RP performed all the laboratory work and data analysis. PR performed data analysis. All authors interpreted the data and wrote the manuscript.

FIGURE LEGENDS

Figure 1: Scatter plot of the absolute and relative abundance of the three most abundant OTUs registered at 79 time points and throughout two time-separated 40-day surveys of a secondary cooling water circuit that operates on a nuclear test reactor. The variance in the relation between absolute and relative abundances increases at elevated values (Breusch-Pagan test, $p < 0.0001$). OTU1 belongs to the betI-A clade. OTU2 and OTU3 belong to the bacI-A clade. Colored dashed lines depict ordinary least squares regression lines for each OTU. These regressions were used solely for statistical inference and do not necessarily represent the optimal predictive models for these data.

Figure 2: Temporal dynamics for taxa of the two most abundant freshwater clades (i.e., bacI-A (OTU2, red; OTU3, orange) and betI-A (OTU1, blue)) during two time-separated 40-day surveys of a secondary cooling water circuit that operates on a nuclear test reactor. The top panel displays the relative abundances (in %) inferred from the 16S rRNA gene amplicon sequencing data. The bottom panel displays the absolute OTU abundances (in cells μl^{-1}) and the circle labels represent the total cell density of the microbial community (in cells μl^{-1} ± standard deviation). Horizontal stacked bars highlight different phases of the system during surveillance. Gray zones indicate time periods where the cooling water system was not in operation (control phases), green zones indicate the startup and blue zones indicate steady-state operation.


Absolute abundance (cells µl^{-1})
OTU 1
OTU 2
OTU 3
1 500
1 000
500
0
0
20
40
60
80
100
Relative abundance (%)


Relative dynamics (%)
100
80
60
40
20
0
control
start-up
operation
control
Absolute dynamics (cells µl-1)
3 000
2 500
2 000
1 500
1 000
500
0
control
start-up
operation
control
Survey 1
betI-A clade
bacI-A clade
Total cell density
control
start-up
operation
control
Survey 2