Computational Analysis of Microbial Flow Cytometry Data

Peter Rubbens,* Ruben Props

Flanders Marine Institute (VLIZ), Ostend, Belgium
Center for Microbial Ecology & Technology (CMET), Faculty of Bioscience Engineering, Ghent University, Ghent, Belgium

ABSTRACT
Flow cytometry is an important technology for the study of microbial communities. It grants the ability to rapidly generate phenotypic single-cell data that are both quantitative, multivariate and of high temporal resolution. The complexity and amount of data necessitate an objective and streamlined data processing workflow that extends beyond commercial instrument software. No full overview of the necessary steps regarding the computational analysis of microbial flow cytometry data currently exists. In this review, we provide an overview of the full data analysis pipeline, ranging from measurement to data interpretation, tailored toward studies in microbial ecology. At every step, we highlight computational methods that are potentially useful, for which we provide a short nontechnical description. We place this overview in the context of a number of open challenges to the field and offer further motivation for the use of standardized flow cytometry in microbial ecology research.

KEYWORDS
bioinformatics, cytometry, fingerprinting, data analysis, microbial ecology, single cell, multivariate statistics

Flow cytometry (FCM) is a single-cell technology that provides an optical description of individual particles based on scatter and fluorescence information. Microbial FCM has a long history, and its first applications in the field date back to the late 1970s to investigate the physiological properties of individual cultures (1, 2). The most prevalent application in microbiology remains the quantification of cell population densities in a wide range of matrices, ranging from lab cultures to marine, freshwater, soil, and fecal samples (3–8). FCM has been applied to many types of microorganisms, mostly phytoplankton and bacteria, but other types of microorganisms include single- and multicellular fungi (9, 10) and viruses (11). For many groups of microorganisms, it has proven to be both accurate and reproducible and can generate results faster than existing plate count and marker gene approaches, such as 16S rRNA gene amplicon sequencing (6, 12). The development of online and real-time FCM facilitates the quantification of microbial community dynamics at a very high temporal resolution (13–15).

A large body of research exists on extracting biological information, in addition to cell enumeration measurements, from the multivariate single-cell data acquired by FCM. Phenotypic properties, such as size, shape, morphology, activity, membrane permeability, pigmentation, and nucleic acid content are measured in various degrees, depending on the applied cell-labeling technique (16, 17). The major ongoing wet-lab FCM developments for microbiology research can be broadly classified into (i) development and standardization of novel staining methods (17, 18) and (ii) novel laboratory protocols to efficiently extract cells from complex matrices (4, 8). Much less attention is given to computational methods that can assist in the analysis of microbial cytometry data. As a result, many microbiologists perform manual interventions during their data analysis, including decisions with respect to denoising, quality control, cell population identification and statistical analyses, on a sample-by-sample or batch-by-batch basis. This inevitably results in user biases, such as reduced reproducibility, but it can also obscure meaningful biological information not apparent from the user’s own interpretation. Many...
computational methods have emerged in the biomedical research field over the past few years to address these shortcomings, grouped together under the names “FCM bioinformatics” or “computational FCM” (19, 20). These aim to facilitate and improve the objectivity, speed, and reproducibility of the data analysis. Likewise, microbiologists have the possibility to set up a dedicated data analysis pipeline to benefit from the same advantages as immunologists do.

In this review, we aim to provide a streamlined overview of the data analysis possibilities along a typical computational workflow for microbial FCM data. A demonstration of such a workflow in R can be found online at https://rprops.github.io/MSys_FCMreview/Demo.html. We highlight a number of interesting applications in which FCM is used to perform ecological studies. Additionally, we try to point to a number of challenges in the field of microbial FCM that motivate the use and development of standardized FCM for microbiology research.

MICROBIAL FLOW CYTOMETRY

A basic overview of a flow cytometry analysis is given in Fig. 1. Suspended particles are first aligned on a one-by-one basis by means of hydrodynamic focusing. Each particle is then interrogated by one or more lasers. Optical filters allow one to measure emitted fluorescence at multiple wavelengths, next to forward scatter (FSC) and side scatter (SSC) signals. Photomultiplier tubes are used to convert the fluorescence and scatter signals to an electronic signal. The morphology of the cell is reflected in the FSC (size and shape) and SSC (intracellular complexity). The measured fluorescence is the result of autofluorescent properties (such as pigments) or the interaction with a fluorescent dye. Mostly, generic stains that target properties related to nucleic acid content, membrane integrity and other physiological aspects, such as lipid content, enzyme activity, and translational activity, are used (3, 17). The technology is fast in the sense that it is able to measure more than thousands of particles per second. It is quantitative, because each particle is described by a numeric multivariate measurement that represents a unique optical signature for each particle.
The main applications of FCM are the biological and clinical study of mammalian cells, also known as immunophenotyping FCM. By now, immunophenotyping data routinely represent large antibody panels of up to 28 individual biomarkers, represented by 28 different fluorescence parameters. These applications form the main drivers of instrument development and research. Microbial FCM has a number of different characteristics and challenges compared to immunophenotyping FCM. First, most prokaryotic cells are much smaller in size and volume than human or mammalian cells. Therefore, measurements can lie close to the detection limit of an instrument. Second, while most cells are small, the size range within which microbial cells occur is larger than for mammalian cells, covering a range between 0.2 and 500 μm. Third, microbial communities comprise high levels of phenotypic and phylogenetic complexity (e.g., 1,000s of taxa) and heterogeneity. As such, contrasting results concerning the establishment of multicolor staining panels for microbial communities have been reported. Single- and double-staining methods are routinely used, with the majority of research relying on one or two general markers with phenotypic (e.g., nucleic acids or membrane permeability) or phylogenetic (e.g., see reference 26) specificity. It appears much more difficult to standardize and broadly apply a triple-staining protocol, as the efficiency and stability of cell staining protocols are dependent on the bacterial taxa on which they are applied. Although successful approaches are reported in the literature, issues such as fluorescence instability hamper their widespread use and further development. Therefore, microbial FCM data are characterized by data with fewer dimensions compared to immunophenotyping FCM.

DATA ANALYSIS

A typical FCM data analysis pipeline can be broadly divided into multiple categories, of which an overview is given in Fig. 2. These include preprocessing of the data, visualization, cell enumeration of specific populations or the whole community, cytometric fingerprinting, community-level analysis and data format and storage. While some steps are necessary, others are optional and depend on the research question and experimental setup. We have summarized and ordered the computational methods that we discuss in this minireview (see Table 1). Here, we focus on software packages that are publicly available in the R statistical programming language. Note that a number of packages are also available in other languages, such as Python or Matlab.

DATA FORMAT

FCM data are stored in flow cytometry standard (FCS) format from commercial software. The most recent version, FCS 3.1, was introduced in 2010. The area (A), the height (H), and sometimes also the width (W) of fluorescence and scatter pulses are recorded for each individual particle. In addition, the file format also allows one to store metadata describing the experimental settings.

PREPROCESSING

Before FCM data can be analyzed, a number of preprocessing steps need to be performed. Some of them are optional; others are recommended or necessary. The basic steps are available within the generic FCM software package flowCore.
but as they increase, their values increase exponentially, resulting in values that are orders of magnitude larger. Traditionally, microbial FCM data are transformed using a logarithmic function. However, measurements can be negative as well, making the logarithm unsuitable to transform the data in this case. More advanced transformations are recommended, such as the arcsine hyperbolic function or a generalized extension with one or more adjustable parameters, often referred to as the “biexponential” or “Logicle” transformation (35, 36). These transformations are available within the flowCore package. Forward and side scatter information can also be analyzed on a linear scale for heterogeneous cell populations of larger cells (37), but most microbiological applications require a transformation of the scatter parameters as well.

**Noise removal.** Instrument noise is always present in the data caused by the measurement of (in)organic particles, cell aggregates and electronic noise. Due to the small cell sizes of microbes, live cells can have fluorescence and scatter properties in a manner similar to the instrument noise. Therefore, noise removal is often performed...
TABLE 1 Overview of peer-reviewed computational methods for performing data preprocessing, visualization, cell population identification, cytometric fingerprinting and data storage

| Category                          | Method (reference) | Short description                                                                 | Reference(s) (applied to microbial data?) |
|-----------------------------------|--------------------|------------------------------------------------------------------------------------|------------------------------------------|
| Preprocessing                     | flowCore (30)      | Basic data manipulation, gating, compensation, and transformation                   | 12, 72–74                                |
| Preprocessing                     | flowTrans (37)     | Optimized individual channel transformations                                       |                                          |
| Preprocessing                     | flowAI (42)        | Automated denoising                                                                | 75, 76                                   |
| Preprocessing                     | flowClean (41)     | Automated denoising                                                                |                                          |
| Preprocessing                     | flowStats (77)     | Per-channel normalization                                                           | 78                                       |
| Visualization                     | flowViz (38)       | Customized FCM data visualization                                                  | 9, 79, 80                                |
| Visualization                     | ggcyto (39)        | Customized FCM data visualization with ggplot-like functionality                   | 74                                       |
| Visualization                     | viSNE (81)         | Dimensionality reduction and visualization using t-SNE                               | 59, 82                                   |
| Cell population identification    | UMAP (83)          | Dimensionality reduction using UMAP                                                |                                          |
| Cell population identification    | flowClust (84)     | t-distribution mixture model with Box-Cox transformation                            | 85                                       |
| Cell population identification    | flowEMMI (54)      | Gaussian mixture model with the Bayesian information criterion                      | 54                                       |
| Cell population identification    | flowPeaks (86)     | k-means clustering followed by peak search and merging using a Gaussian mixture model | 87, 88                                   |
| Cell population identification    | flowDensity (89)   | Density-based sequential bivariate gating                                           | 54, 90                                   |
| Cell population identification    | FlowSOM (52)       | Self-organizing map and meta-clustering                                            |                                          |
| Cell population identification    | PhenoGraph (53)    | k-nearest neighbor weighted graph and Louvain method for community detection        | 82                                       |
| Cytometric fingerprinting         | CHIC (61)          | Two-channel histogram image comparison                                             | 61, 72, 91                               |
| Cytometric fingerprinting         | flowCyBar (60)     | Manual annotation of interesting regions                                            | 60, 91–93                                |
| Cytometric fingerprinting         | flowDiv (62)       | Fixed-binning grid over multiple two-channel combinations                          | 62                                       |
| Cytometric fingerprinting         | flowFP (64)        | Distribution-dependent binning in hyper-rectangles                                  | 91, 94–96                                |
| Cytometric fingerprinting         | PhenoFlow (12)     | Fixed-binning grid and kernel density estimation over multiple two-channel combinations | 12, 76, 97, 98                           |
| Cytometric fingerprinting         | PhenoGMM (67)      | Overclustering using a Gaussian mixture model                                       | 67, 99                                   |
| Data storage                      | FlowRepository (70)| Public database to store and annotate FCM data                                     | 34, 75, 100, 101                         |

*This column highlights references in which the method has been applied to microbial FCM data.*

manually, by defining rectangle, quadrant, ellipsoid, or generic polygon regions (i.e., “gates”) containing the cell signals. As stated in the introduction, this can be laborious in time and introduce subjective biases (20). At the individual-sample level, optimal gates can often differ due to intersample variation. It is advised to use, whenever possible, the same gating template within a single experiment, although samples analyzed with different gating templates can still be compared using proper control samples. Negative-control samples are then necessary, which can include heat-killed samples and 0.2-µm-filtered samples, either stained and/or unstained. A gating template can be defined using the flowCore package. Denoising is usually guided by user-dependent iterative visualization of the scatterplots. FCM data visualization is supported by the flowViz and ggcyto packages (38, 39). The number of gating steps are largely dependent on the complexity of the analyzed sample. High degrees of noise may require additional signal filtering from multiple fluorescence (e.g., autofluorescence on violet laser) and/or scatter channels. In most measurements, cell aggregates, such as doublets, triplets, or chains, are measured as well and can be identified through visualizing the area and height parameter of the primary fluorescence or scatter channel. However, for microbiological applications, it remains difficult in practice, and there is no consensus yet on how to best handle cell aggregate signals. We recommend the optimization of sample preparation protocols to reduce the percentage of cell aggregates; these can include the use of filtration, ultrasonication, surfactants (e.g., Tween, Triton X-100), complexing agents (e.g., EDTA, sodium pyrophosphate), and/or Ca²⁺/Mg²⁺-free buffers (4, 40).

**Quality control.** The quality of the data and its acquisition are subject to both instrument and biological variation. The ideal data acquisition consists of the measurement of homogeneous and stable cell signal distribution during sample analysis. However, deviations can occur, for example, due to large particles that clog the system or air bubbles that cause gaps in the data. This results in aberrations in the data, such as spikes, gaps or gradual degradation of the mean fluorescence intensity. These need to be addressed and, depending on the research question, removed. While these
actions can be done manually, algorithms to detect and remove anomalies automatically by inspecting individual cell parameters in the function of the acquisition time have been developed. By applying statistical methods in combination with anomaly detection strategies, deviating segments are annotated and removed. A number of methods exist; see, for example, flowClean, flowAI and flowCut (41–43). Further research is needed to evaluate these algorithms for cell counting, as they can drastically influence the number of cells measured. As such, we currently do not recommend applying these quality control algorithms for cell counting, but we do recommend them for fingerprinting and cell population identification applications.

CELL POPULATION IDENTIFICATION

After the data preprocessing, samples can be further analyzed. The most common analysis in microbial FCM is to characterize the microbial load by enumerating microbial cell densities of the total community, quantified as the number of particles per milliliter or gram (44).

However, the data can contain distinct cell populations caused by differences in cell size, morphology, and autofluorescent properties (e.g., phytoplankton [45]) or due to the use of specific stains (e.g., a nucleic acid stain to detect nucleic acid populations in aquatic environments [46]). While these are routinely gated manually, cell population identification algorithms detect these automatically and therefore reduce the bias and analysis inherent in manual gating procedures by experts (20). Dimensionality reduction algorithms can be used to visualize the multivariate single-cell data at once and to explore whether distinct cell populations are present in the data. These include principal-component analysis (PCA), but more advanced algorithms have demonstrated their advantages for immunophenotyping cytometry, such as t-distributed stochastic neighbor embedding (t-SNE) and uniform manifold approximation and projection (UMAP) (47, 48).

The performance of cell populations identification algorithms has been thoroughly benchmarked in terms of cluster accuracy, stability, rare cell type discovery and computing time using standardized immunophenotyping FCM or mass cytometry data sets (49–51). FlowSOM (52) has been proposed as the least time-intensive algorithm with favorable results for human mass cytometry data (50), with PhenoGraph (53) being an interesting competitor. Recently, flowEMMI (54), a clustering approach based on Gaussian mixture models and the expectation-maximization algorithm, has been proposed and compared to a number of additional algorithms to identify clusters in two-channel bacterial samples. Another option is to perform single-cell classification to identify known bacterial populations (55–58). These can be helpful in case it is known which populations are present in the data and one expects their properties to remain stable throughout the experiment; however, especially the latter is often difficult, due to the phenotypic heterogeneity of bacterial populations (59).

CYTOMETRIC FINGERPRINTING

The second set of algorithms falls under the category of cytometric fingerprinting approaches. In this case, the focus lies on modeling the multivariate distribution of single-cell observations by dividing the parameter space into regions in which cell counts or densities are recorded. The identification of distinct cell populations is, in this case, a secondary objective. Three categories of cytometric fingerprinting approaches can be distinguished, based on how these regions are determined.

- For manual approaches, multiple clusters or gates are manually drawn in regions of interest and applied to all samples (see the FlowCyBar algorithm [60]).
- For fixed-binning approaches, a grid of dimensions $L \times L$ with equally sized bins is placed over one or multiple bivariate channel combinations, and the cell count per bin is registered. Cytometric histogram image comparison (CHIC), Phenoflow, and flowDiv have been specifically developed for microbial cytometry data (12, 61, 62).
For adaptive-binning approaches, a grid or other structure with various region sizes and shapes is placed over a bivariate or multivariate combination of FCM parameters. The size and shape depend on the distribution of the data, with typically small bin sizes for those regions of high density and vice versa. The first adaptive-binning approach, termed probability binning, was already proposed in 2001 and divides the data in hyper-rectangular bins of various sizes (63). The algorithm is publicly available as a software package under the name of flowFP (64). An extension, called PB-sQF, that uses probability binning in combination with the quadratic-form distance statistic to compare two samples has been developed (65). Recent alternatives include the search for local density peaks, after which bins are created using Voronoi tessellation (66), and PhenoGMM, an approach that uses all multivariate information at once by overclustering the data using a predefined large number of Gaussian mixtures (67).

Limited research has been devoted to a comparison of fingerprinting methods. Therefore, it is difficult to provide a clear recommendation on which method(s) a user should use for their data. In terms of time of analysis and objectivity, operator-independent methods are preferred over manual methods. Fixed-binning approaches, such as PhenoFlow and CHIC, model the distribution of the data by using a two-dimensional gridded approach. However, these approaches become less performant when a user wants to incorporate more parameters. In addition, the number of community-describing variables is large. Adaptive-binning approaches require some time to estimate the gating template. However, these are more advantageous to model multivariate data and result in fewer community-describing variables.

Some years ago, the FlowCAP (flow cytometry, critical assessment of population identification methods) initiatives were organized within the immunophenotyping cytometry community (49, 68). In this initiative, a number of highly curated data sets were provided to objectively compare cell population identification algorithms. Currently, these data sets are still the standard to benchmark new computational methods, and their methodology forms the basis for more recent benchmark studies (50, 51). Microbial FCM currently lacks highly curated data sets. These would be, in combination with a set of commonly agreed-upon data analysis objectives, of great value for the development of cytometric fingerprinting methods.

**DATA STORAGE**

It is recommended that the raw data be stored in FCS format in public FCM repositories, such as Cytobank or FlowRepository (69, 70). The corresponding accession identifiers should be added to the data section of every publication and not just the final-count table. FlowRepository is recommended by multiple journals and societies, including Cytometry Part A (the official journal of the International Society for the Advancement of Cytometry [ISAC]), all American Society for Microbiology and PLOS journals, and Springer Nature. Another helpful tool is the minimum information about a flow cytometry experiment (MIFlowCyt) document, which assists in the annotation of the minimum of information that is required to report an FCM experiment (71). A minor caveat is that current guidelines are tailored toward biomedical experiments. MIFlowCyt is incorporated in FlowRepository. An overview of peer-reviewed computational methods for performing data preprocessing, visualization, cell population identification, cytometric fingerprinting, and data storage can be found in Table 1.

**ECOLOGICAL INFERENCE AND APPLICATIONS**

The output of cell population identification and fingerprinting algorithms are contingency tables of counts or densities across the determined multivariate regions, which may be bins, clusters or manually selected regions. With these tables, a variety of traditional multivariate methods (e.g., PCA, canonical-correspondence analysis [CCA], permutational multivariate analysis of variance [PERMANOVA], etc.) can be applied to test for differences among sample groups in the function of experimental conditions.
(e.g., pH, nutrient concentrations, host disease state, etc.) (102, 103). Traditional ecological parameters, such as alpha and beta diversity, and a range of stability and functional diversity metrics have been developed that enable researchers to quantitatively compare changes in community structure (104, 105). The availability of reference bead data with refractive indices matching those of bacterial cells and microscopy-based validation experiments can be used to create predictive models for cell size and biovolumes of individual cells and populations (88, 106–108).

For all the described algorithms and metrics, there exist numerous research applications, most of which are situated in the aquatic research domain, although research in air, soil, sediment, and clinical microbiology is gaining traction (4, 7, 109, 110). A few studies have used cell population identification methods to analyze microbial FCM data. These include the identification of physiological populations (i.e., high and low nucleic acid populations in marine and freshwater systems) (85, 111), phytoplankton populations (88, 90), or different strains of yeast (9). In contrast, fingerprinting methods have been more broadly applied in environmental microbiology, where they have been used to track changes in drinking water, sludge, and soil microbiome structures over time and in functions of environmental conditions (15, 92, 94, 112). In a clinical setting, fingerprinting has been used to infer bactericide treatment effects in saliva microbiomes (93), to train a predictive model for Crohn’s disease in gut microbiomes (99), and to test for antibiotic susceptibility (65, 113). FCM has also proven to be a complementary technique to next-generation sequencing technologies to enable absolute quantification of microbial taxa (114). Even more, the taxonomic community structure based on 16S rRNA gene amplicon sequencing has been associated with cytometric community structures in multiple environments, including freshwater (12, 115), marine (62, 116), and gut communities (99).

CONCLUSIONS
Microbial FCM applications are rapidly evolving, for example through the use of online and real-time FCM (13, 14), the development of polychromatic staining panels dedicated to microbial research (28), and integration with molecular analyses (101, 117, 118). The amount and complexity of the data will continue to increase as the technology is integrated further into clinical, environmental and industrial research. This will necessitate the need for objective and streamlined bioinformatics workflows to achieve a quantitative and reproducible data analysis. The collaboration across cytometry disciplines will be crucial to ensure the adoption of computational methods by a wider user base in the field of microbiology. We hope with this review to have contributed to this end and look forward to new developments that are yet to emerge in the field.

SUPPORTING INFORMATION
A demonstration of a computational workflow can be found at https://github.com/rprops/MSys_minireview.

ACKNOWLEDGMENTS
We thank Jorien Favere, Jo De Vrieze, and Maarten De Rijcke for their valuable feedback on previous versions of the manuscript. We thank Tim Lacoere for his marvelous graphical designs.

REFERENCES
1. Paau AS, Cowles JR, Oro J. 1977. Flow-microfluorometric analysis of Escherichia coli, Rhizobium melloti, and Rhizobium japonicum at different stages of the growth cycle. Can J Microbiol 23:1165–1169. https://doi.org/10.1139/m77-175.
2. Hutter KJ, Eipel HE. 1979. Microbial determinations by flow cytometry. J Gen Microbiol 113:369–375. https://doi.org/10.1099/00221287-113-2-369.
3. Wang Y, Hammes F, De Roy K, Verstraete W, Boon N. 2010. Past, present and future applications of flow cytometry in aquatic microbiology. Trends Biotechnol 28:416–424. https://doi.org/10.1016/j.tibtech.2010.04.006.
4. Frossard A, Hammes F, Gessner MO. 2016. Flow cytometric assessment of bacterial abundance in soils, sediments and sludge. Front Microbiol 7:903. https://doi.org/10.3389/fmicb.2016.00903.
5. Chodkowski JL, Shade A. 2017. A synthetic community system for probing microbial interactions driven by exometabolites. mSystems 2:e00129-17. https://doi.org/10.1128/mSystems.00129-17.
6. Van Nevel S, Koetzsch S, Proctor CR, Besmer MD, Prest EI, Vrouwenvelder JS, Knezev A, Boon N, Hammes F. 2017. Flow cytometric bacterial cell counts challenge conventional heterotrophic plate counts for routine microbiological drinking water monitoring. Water Res 113:191–206. https://doi.org/10.1016/j.watres.2017.01.065.
7. Vandeputte D, Kathagen G, O’hoe K, Vieira-Silva S, Valles-Colomer M, Sabino J, Wang J, Tito RY, De Commer L, Darzi Y, Vermeire S, Falony G, Raes J. 2017. Quantitative microbiome profiling links gut community variation to microbial load. Nature 551:507–511. https://doi.org/10.1038/nature24460.
Minireview

14. Besmer MD, Hammes F. 2016. Short-term microbial dynamics in a drinking water plant treating groundwater with occasional high microbial loads. Water Res 107:11–18. https://doi.org/10.1016/j.watres.2016.10.041.

15. Props R, Rubbens P, Besmer M, Buysschaert B, Sigrist J, Weilenmann H, Waegeman W, Boon N, Hammes F. 2018. Detection of microbial disturbances in a drinking water microbial community through continuous acquisition and advanced analysis of flow cytometry data. Water Res 145:73–82. https://doi.org/10.1016/j.watres.2018.08.013.

16. Müller S, Nebe-von Caron G. 2010. Functional single-cell analyses: flow cytometry and cell sorting of microbial populations and communities. FEMS Microbiol Rev 34:554–587. https://doi.org/10.1111/j.1574-6976.2010.00214.x.

17. Habermayer A, Krukenberg V, Spiezl RL, Jay ZJ. 2020. Next-generation physiological approaches to study microbiome function at single cell level. Nat Rev Microbiol 18:241–256. https://doi.org/10.1038/s41579-020-0323-1.

18. Beller A, Kruglov A, Durek P, Goetze V, Werner K, Heinz GA, Ninnemann J, Huber P, Peter Odenthal R, Sonnleitner B. 2012. Development and laboratory-scale testing of a fully automated online flow cytometer for drinking water analysis. Cytochemistry A 81A:508–516. https://doi.org/10.1016/j.cyto.a.2011.07.001.

19. O'Neill K, Aghaeepour N, Spidlen J, Brinkman RR. 2013. Flow cytometry and cell sorting in clinical research. Cytometry A 81A:516. https://doi.org/10.1002/cyto.a.20825.

20. Waite AJ, Shou W. 2012. Adaptation to a new environment allows cooperators to purge cheaters stochastically. Proc Natl Acad Sci U S A 109:19079–19084. https://doi.org/10.1073/pnas.121091098.

21. Gillam M, Amend JM, Jørgensen BT. 2019. Flow cytometry and sorting slow-growing archaeal-bacterial consortia. Proc Natl Acad Sci U S A 116:4051–4056. https://doi.org/10.1073/pnas.1904234116.

22. Anvarian AHP, Cao Y, Srikumar S, Fanning S, Jordan K. 2016. Flow cytometry avoids deceptive effects of logarithmic scaling for low signals and compensated data. Cytometry A 93A:402–405. https://doi.org/10.1002/cyto.a.2480. https://doi.org/10.1093/cytoba/ctw179.

23. Van P, Jiang W, Gottardo R, Finak G. 2018. ggCyto: next generation open-source visualization software for flow cytometry. Bioinformatics 34:2487–2489. https://doi.org/10.1093/bioinformatics/btx741.

24. Léonard L, Chibane LB, Bouhedda BO, Degraeve P, Oulahal N. 2016. Optimisation of procedures for counting viruses by flow cytometry and FACs applied to microbial treatments. Front Microbiol 7:1225. https://doi.org/10.3389/fmicb.2016.01225.

25. Roger EB, Amal D, Huber P, Peter Odenthal R, Sonnleitner B. 2016. Development and labo-
Minireview

59. García-Timermans C, Rubbens P, Heyse J, Kerckhof FM, Props R, Skirtach
54. Ludwig J, zu Siederdissen CH, Liu Z, Stadler PF, Müller S. 2019.
51. Liu X, Song W, Wong BY, Zhang T, Yu S, Lin GN, Ding X. 2019. A compari-
52. Van Gassen S, Callebaut B, Van Helden MJ, Lambrecht BN, Demeester P,
55. Rajwa B, Venkatapathi M, Ragheb K, Banada PP, Hirleman ED, Lary T,
60. Koch C, Fetzer I, Schmidt T, Harms H, Müller S. 2013. Monitoring func-
48. McInnes L, Healy J, Saul N, Großberger L. 2018. UMAP: uniform manifold
66. Amal
64. Rogers WT, Holyst HA. 2009. FlowFP: a Bioconductor package for
67. Rubbens P, Props R, Kerckhof FM, Boon N, Waegeman W. PhenoGMM:
60. Koch C, Fetzer I, Schmidt T, Harms H, Müller S. 2013. Monitoring func-
55. https://doi.org/10.1002/1097
147. https://doi.org/10.1002/cyto.a.23030.
1106. https://doi.org/10.1038/s42003-020-1106-y.
54. Ludwig J. zu Siederdissen CH, Liu Z, Stadler PF, Müller S. 2019. A compari-
52. Van Gassen S, Callebaut B, Van Helden MJ, Lambrecht BN, Demeester P,
48. McInnes L, Healy J, Saul N, Großberger L. 2018. UMAP: uniform manifold
66. Amal
64. Rogers WT, Holyst HA. 2009. FlowFP: a Bioconductor package for
67. Rubbens P, Props R, Kerckhof FM, Boon N, Waegeman W. PhenoGMM:
60. Koch C, Fetzer I, Schmidt T, Harms H, Müller S. 2013. Monitoring func-
55. https://doi.org/10.1002/1097
147. https://doi.org/10.1002/cyto.a.23030.
1106. https://doi.org/10.1038/s42003-020-1106-y.
54. Ludwig J. zu Siederdissen CH, Liu Z, Stadler PF, Müller S. 2019. A compari-
52. Van Gassen S, Callebaut B, Van Helden MJ, Lambrecht BN, Demeester P,
48. McInnes L, Healy J, Saul N, Großberger L. 2018. UMAP: uniform manifold
66. Amal
64. Rogers WT, Holyst HA. 2009. FlowFP: a Bioconductor package for
67. Rubbens P, Props R, Kerckhof FM, Boon N, Waegeman W. PhenoGMM:
60. Koch C, Fetzer I, Schmidt T, Harms H, Müller S. 2013. Monitoring func-
55. https://doi.org/10.1002/1097
147. https://doi.org/10.1002/cyto.a.23030.
1106. https://doi.org/10.1038/s42003-020-1106-y.
54. Ludwig J. zu Siederdissen CH, Liu Z, Stadler PF, Müller S. 2019. A compari-
52. Van Gassen S, Callebaut B, Van Helden MJ, Lambrecht BN, Demeester P,
48. McInnes L, Healy J, Saul N, Großberger L. 2018. UMAP: uniform manifold
66. Amal
64. Rogers WT, Holyst HA. 2009. FlowFP: a Bioconductor package for
67. Rubbens P, Props R, Kerckhof FM, Boon N, Waegeman W. PhenoGMM:
60. Koch C, Fetzer I, Schmidt T, Harms H, Müller S. 2013. Monitoring func-
55. https://doi.org/10.1002/1097
147. https://doi.org/10.1002/cyto.a.23030.
1106. https://doi.org/10.1038/s42003-020-1106-y.
54. Ludwig J. zu Siederdissen CH, Liu Z, Stadler PF, Müller S. 2019. A compari-
52. Van Gassen S, Callebaut B, Van Helden MJ, Lambrecht BN, Demeester P,
48. McInnes L, Healy J, Saul N, Großberger L. 2018. UMAP: uniform manifold
66. Amal
64. Rogers WT, Holyst HA. 2009. FlowFP: a Bioconductor package for
67. Rubbens P, Props R, Kerckhof FM, Boon N, Waegeman W. PhenoGMM:
60. Koch C, Fetzer I, Schmidt T, Harms H, Müller S. 2013. Monitoring func-
55. https://doi.org/10.1002/1097
147. https://doi.org/10.1002/cyto.a.23030.
1106. https://doi.org/10.1038/s42003-020-1106-y.
54. Ludwig J. zu Siederdissen CH, Liu Z, Stadler PF, Müller S. 2019. A compari-
52. Van Gassen S, Callebaut B, Van Helden MJ, Lambrecht BN, Demeester P,
48. McInnes L, Healy J, Saul N, Großberger L. 2018. UMAP: uniform manifold
66. Amal
64. Rogers WT, Holyst HA. 2009. FlowFP: a Bioconductor package for
67. Rubbens P, Props R, Kerckhof FM, Boon N, Waegeman W. PhenoGMM:
60. Koch C, Fetzer I, Schmidt T, Harms H, Müller S. 2013. Monitoring func-
55. https://doi.org/10.1002/1097
147. https://doi.org/10.1002/cyto.a.23030.
1106. https://doi.org/10.1038/s42003-020-1106-y.
54. Ludwig J. zu Siederdissen CH, Liu Z, Stadler PF, Müller S. 2019. A compari-
52. Van Gassen S, Callebaut B, Van Helden MJ, Lambrecht BN, Demeester P,
85. García FC, López-Urrutia A, Morán XAG, 2014. Automated clustering of heterotrophic bacterioplankton in flow cytometry data. Aquat Microb Ecol 72:175–185. https://doi.org/10.1034/jame01691.

86. Ge Y, Sealfon SC. 2012. 2012. Fodelianakis S, Lorz A, Valenzuela-Cuevas A, Barozzi A, Booth JM, January/February 2021 Volume 6 Issue 1 e00895-20 msystems.asm.org 11

87. Bervoets I, Van Brempot M, Van Nerom K, Van Hove B, Maertens J, De Mey M, Charlier D. 2018. A sigma factor toolbox for orthogonal gene expression in Escherichia coli. Nucleic Acids Res 46:2133–2144. https://doi.org/10.1093/nar/gky010.

88. Thomas MK, Fontana S, Reyes M, Pomati F. 2018. Quantifying cell densities and biovolumes of phytoplankton communities and functional groups using scanning flow cytometry, machine learning and unsupervised clustering. PLoS One 13:e0196225. https://doi.org/10.1371/journal.pone.0196225.

89. Malek M, Taghijiar MJ, Chong L, Finak G, Gottardo R, Brinkman RR. 2015. FlowDensity: reproducing manual gating of flow cytometry data by automated density-based cell population identification. Bioinformatics 31:606–607. https://doi.org/10.1093/bioinformatics/btu677.

90. Ribalet F, Berthiaume C, Hynes A, Swallwell J, Carlson M, Clayton S, Hennen G, Poirier C, Shimabukuro E, White A, Armbrust EV. 2019. SeaFlow data v1, high-resolution abundance, size and biomass of small phytoplankton in the North Pacific. Sci Data 6:277. https://doi.org/10.1038/s41597-019-0229-2.

91. Koch C, Harnisch F, Schröder U, Müller S. 2014. Cytometric fingerprinting: evaluation of new tools for analyzing microbial community dynamics. Front Microbiol 5:273. https://doi.org/10.3389/fmicb.2014.00273.

92. Menyhárt L, Nagy S, Leposka A. 2018. Rapid analysis of photoautotroph microbial communities in soils by flow cytometric barcoding and fingerprinting. Appl Soil Ecol 130:237–240. https://doi.org/10.1016/j.apsoil.2018.06.013.

93. van Gelder S, Röhrig N, Schattenberg F, Cichocki N, Schumann J, Schmalz G, Haak R, Ziebolz D, Müller S. 2018. A cytometric approach to follow variation and dynamics of the salivary microbiota. Methods 135:63–67. https://doi.org/10.1016/j.ymeth.2017.08.009.

94. De Roy K, Clement L, Thas O, Wang Y, Boon N. 2012. Flow cytometry for estimating bacterial biovolume in natural plankton samples: comparison with microscopy data. Appl Environ Microbiol 73:4508–4514. https://doi.org/10.1128/AEM.00733-07.

95. Jang J, Hendriksen NB, Jakobsen HH, Gosewinkel U. 2018. Application of Cytosense flow cytometer for the analysis of airborne bacteria collected by a high volume impingement sampler. J Microbiol Methods 154:63–72. https://doi.org/10.1016/j.mimet.2018.10.012.

96. Singh A, Yelverton CJ, Barnard TG. 2019. Rapid quantification of the total viable bacterial population on human hands using flow cytometry with SYBR Green I. Cytometry B Clin Cytom 96:397–403. https://doi.org/10.1002/cyto.b.21776.

97. Bowman JS, Amaral-Zettler LA, J Rich J, M Luria C, Dukelow HW. 2017. Bacterial community segmentation facilitates the prediction of ecosystem function along the coast of the western Antarctic Peninsula. ISME J 11:1460–1471. https://doi.org/10.1038/ismej.2016.204.

98. Schlechina C, Chen S, Pullerits K, Besmer MD, Paul CJ, Rädrström P, Keucken A. 2019. Mapping dynamics of bacterial communities in a full-scale drinking water distribution system using flow cytometry. Water 11:2137. https://doi.org/10.3390/w111102137.

99. Huang TH, Tseng YL, Dickson RM. 2018. FAST: rapid determinations of antibiotic susceptibility phenotypes using label-free cytometry. Cytometry A 93A:639–648. https://doi.org/10.1002/cyto.a.23370.

100. Props R, Kerckhof FM, Rubbens P, De Vrieze J, Hernandez Sanabria E, Waegeman W, Monsieurs P, Hammes F, Boon N. 2017. Absolute quantification of microbial taxon abundances. ISME J 11:584–587. https://doi.org/10.1038/ismej.2016.117.

101. Props R, Schmidt ML, Heyse J, Vanderplaag HA, Boon N, Denef VJ. 2018. Flow cytometric monitoring of bacterioplankton phenotypic diversity predicts high population-specific feeding rates by invasive dreissenid mussels. Environ Microbiol 20:521–534. https://doi.org/10.1111/1462-2903.13953.

102. García FC, Alonso-Sáez L, Morán XAG, López-Urrutia A, 2015. Seasonality in molecular and cytometric diversity of marine bacterioplankton: the re-shuffling of bacterial taxa by vertical mixing. Environ Microbiol 17:4133–4142. https://doi.org/10.1111/1462-2903.12984.

103. John M, Seifert J, von Bergen M, Schmid A, Bührer B, Müller S. 2013. Sub-population-proteomics in prokaryotic populations. Curr Opin Biotechnol 24:79–87. https://doi.org/10.1016/j.copbio.2012.10.017.

104. Stepanauskas R, Ferguson EA, Brown J, Poulton NJ, Tupper B, Labonté JM, Becraft ED, Brown JM, Pachiadaki MG, Povilaitis T, Thompson BP, Mascena CJ, Bellows WK, Lubys A. 2017. Improved genome recovery and integrated cell-size analyses of individual uncultured microbial cells and viral particles. Nat Commun 8:884. https://doi.org/10.1038/s41467-017-0028-2.
Peter Rubbens is a postdoctoral researcher at the Flanders Marine Institute (VLIZ) in Ostend, Belgium. He obtained his Ph.D. at Ghent University in 2019. During his doctoral study, he investigated and specialized in machine learning approaches for microbial flow cytometry data, in addition to other single-cell technologies. Currently, his research has shifted to a quantitative study of marine microbial communities and their dynamics in response to a changing environment. In addition, he is interested in applications of machine learning for microbial ecology.

Ruben Props is a postdoctoral researcher at the Center for Microbial Ecology and Technology (CMET, Ghent University). He obtained his Ph.D. at Ghent University in 2018. During his doctoral studies, he studied the microbial ecology of both engineered and freshwater ecosystems. He carried out part of his Ph.D. research at the University of Michigan under the supervision of Vincent Denef, during which he studied strain-level variations in abundant bacterioplankton and the effects of invasive mussel species on bacterioplankton composition and function. Currently, he is investigating the phylogenomic basis of phenotypic properties of microbes using single-cell analysis techniques.