



# Computational Analysis of Microbial Flow Cytometry Data

 Peter Rubbens,<sup>a</sup> Ruben Props<sup>b</sup>

<sup>a</sup>Flanders Marine Institute (VLIZ), Ostend, Belgium

<sup>b</sup>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


**Citation** Rubbens P, Props R. 2021. Computational analysis of microbial flow cytometry data. *mSystems* 6:e00895-20. <https://doi.org/10.1128/mSystems.00895-20>.

**Editor** Pieter C. Dorrestein, University of California, San Diego

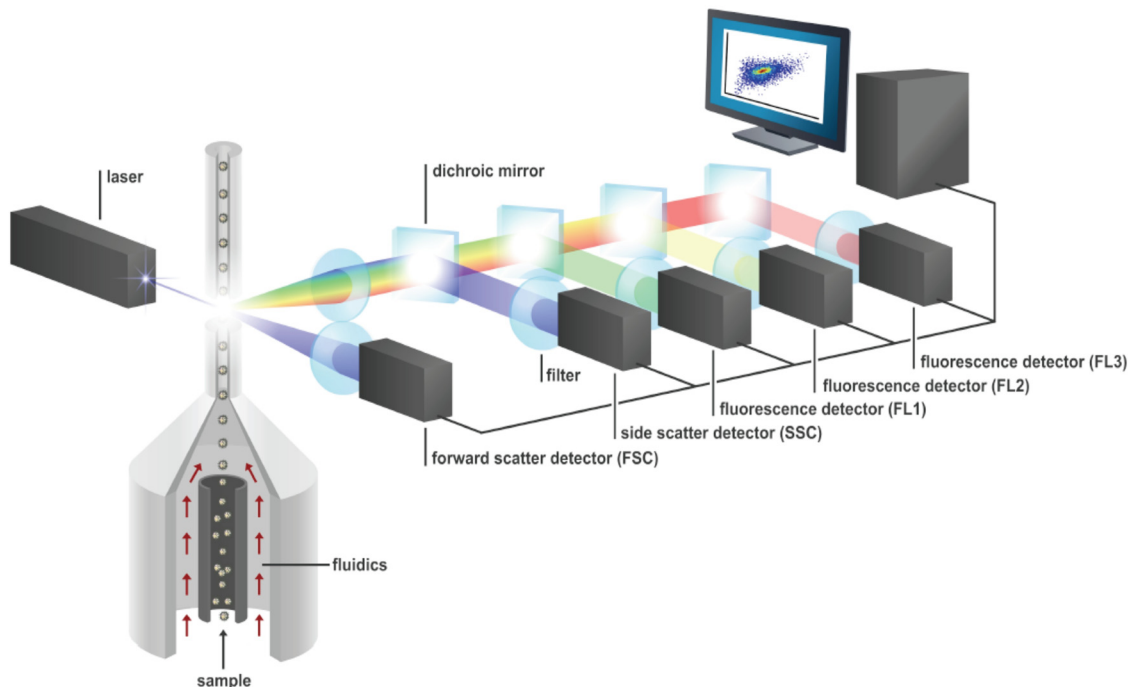
The review history of this article can be read [here](#).

**Copyright** © 2021 Rubbens and Props. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](#).

Address correspondence to Peter Rubbens, [peter.rubbens@vliz.be](mailto:peter.rubbens@vliz.be), or Ruben Props, [ruben.props@ugent.be](mailto:ruben.props@ugent.be).

 Computational Analysis of Microbial Flow Cytometry Data: a step-by-step minireview highlighting the full data analysis pipeline of microbial flow cytometry data.

**Published** 19 January 2021



**FIG 1** Schematic overview of a flow cytometry analysis. Suspended particles are aligned one by one by hydrodynamic focusing. Next, each particle is interrogated by one or more lasers in the flow cell. The resulting scatter (FSC and SSC) and fluorescence signals (denoted “FL”) of each cell are captured by multiple detectors. Fluorescence is measured at multiple wavelength intervals (three in this illustration). The electronic signals originating from these detectors are then finally transformed into digital ones.

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](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 (21), 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 (22). 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  $\mu\text{m}$ . Third, microbial communities comprise high levels of phenotypic and phylogenetic complexity (e.g., 1,000s of taxa) and heterogeneity (16). 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 (23, 24), with the majority of research relying on one or two general markers with phenotypic (e.g., nucleic acids or membrane permeability) (25) 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 (27, 28), issues such as fluorescence instability hamper their widespread use and further development (23). 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

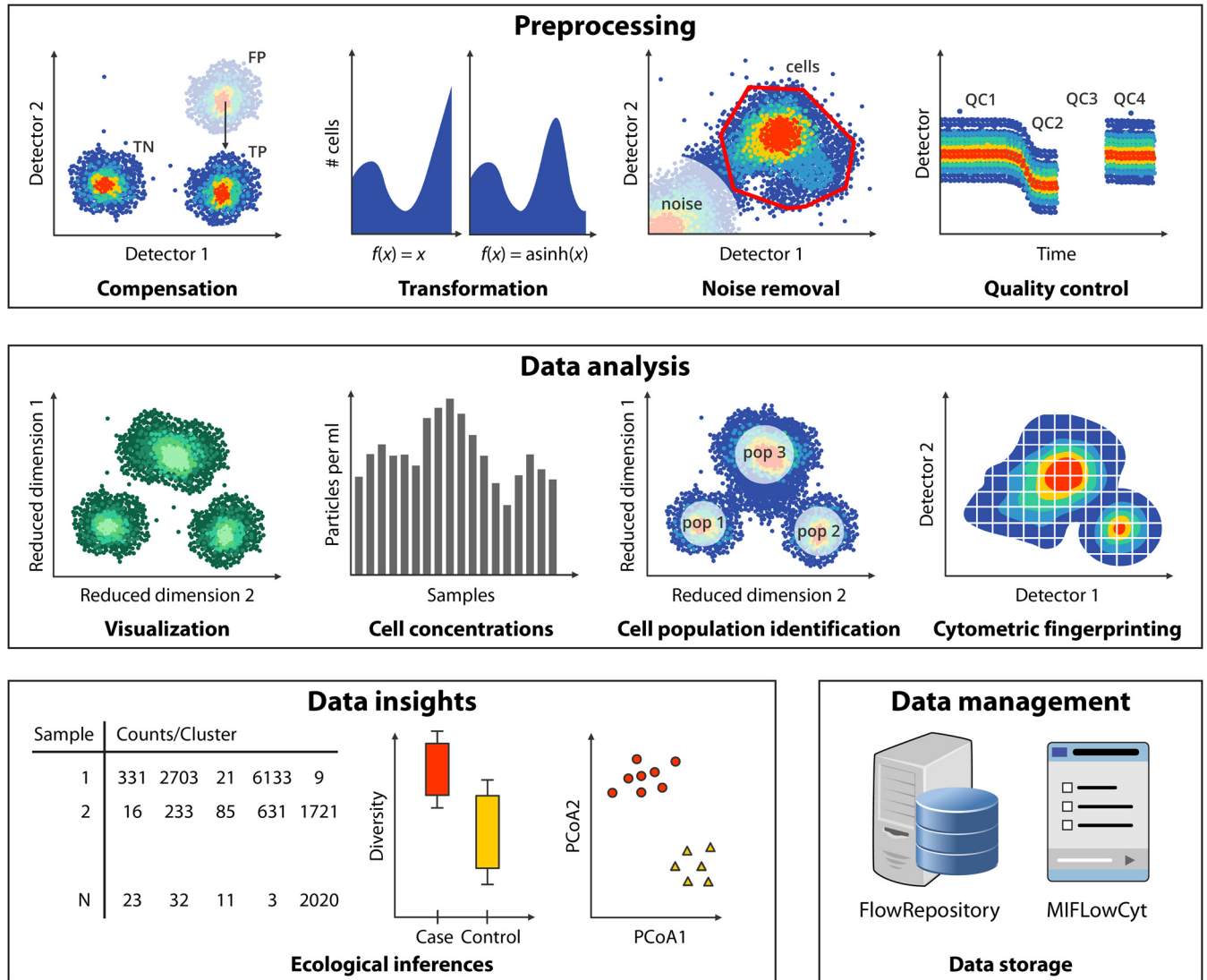
FCM data are stored in flow cytometry standard (FCS) format from commercial software. The most recent version, FCS 3.1, was introduced in 2010 (29). The area (A), the height (H), and sometimes also the width (W) of fluorescence and scatter pulses are recorded for each individual particle. In addition, 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 (30).

**Compensation.** If a detector has an optical filter that captures signals coming from multiple stains, false-positive cells can be detected. Data compensation intends to correct for emission signal spillover from one stain (e.g., Syto59) into the channel designated for another stain (e.g., propidium iodide). Currently, this is only rarely applied due to the limited availability of multicolor FCM protocols to analyze microbial communities, although a few examples can be found in the literature (23, 31–34). Functions to perform compensation are incorporated into the flowCore package.

**Transformation.** Fluorescence and scatter values registered for microbial cells can differ by orders of magnitude and need to be transformed to enable separation of instrument noise and cell signals. These values exhibit linear behavior at small scales,



**FIG 2** Overview of a reproducible microbial FCM data analysis pipeline. First, the data are preprocessed in several steps (compensation, transformation, noise removal, and quality control). Second, the data are analyzed, which can be done in multiple ways: through visualization, determination of cell concentration, cell population identification, and/or cytometric fingerprinting. Once processed, the data can be analyzed at the community level to make ecological inferences. Upon conclusion of the analysis, the raw data are properly annotated and stored in a publicly accessible database. FP, false positives; TN, true negatives; TP, true positives; QC1 to -4, quality control measurements, respectively; pop, population; PCoA 1 and -2, principal coordinate axis 1 and 2, respectively.

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?) <sup>a</sup>
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
Visualization	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

<sup>a</sup>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- $\mu$ m-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 Ca<sup>2+</sup>/Mg<sup>2+</sup>-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 time 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  $L$  by  $L$  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 INFERENCES 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/M Sys\\_minireview](https://github.com/rprops/M Sys_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. Paa AS, Cowles JR, Oro J. 1977. Flow-microfluorometric analysis of *Escherichia coli*, *Rhizobium meliloti*, 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, D'hoel 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>.



8. Gryp T, De Paepe K, Vanholder R, Kerckhof FM, Van Biesen W, Van de Wiele T, Verbeke F, Speckaert M, Joossens M, Couttenye MM, Vaneechoutte M, Glorieux G. 2020. Gut microbiota generation of protein-bound uremic toxins and related metabolites is not altered at different stages of chronic kidney disease. *Kidney Int* 97:1230–1242. <https://doi.org/10.1016/j.kint.2020.01.028>.
9. Waite AJ, Shou W. 2012. Adaptation to a new environment allows cooperators to purge cheaters stochastically. *Proc Natl Acad Sci U S A* 109:19079–19086. <https://doi.org/10.1073/pnas.1210190109>.
10. Bleichrodt RJ, Read ND. 2019. Flow cytometry and FACS applied to filamentous fungi. *Fungal Biol Rev* 33:1–15. <https://doi.org/10.1016/j.fbr.2018.06.001>.
11. Brussaard CP. 2004. Optimization of procedures for counting viruses by flow cytometry. *Appl Environ Microbiol* 70:1506–1513. <https://doi.org/10.1128/AEM.70.3.1506-1513.2004>.
12. Props R, Monsieurs P, Mysara M, Clement L, Boon N. 2016. Measuring the biodiversity of microbial communities by flow cytometry. *Methods Ecol Evol* 7:1376–1385. <https://doi.org/10.1111/2041-210X.12607>.
13. Hammes F, Broger T, Weilenmann HU, Vital M, Helbing J, Bosshart U, Huber P, Peter Odermatt R, Sonnleitner B. 2012. Development and laboratory-scale testing of a fully automated online flow cytometer for drinking water analysis. *Cytometry A* 81A:508–516. <https://doi.org/10.1002/cyto.a.22048>.
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, Buyschaert 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. Hatzenpichler R, Krukenberg V, Spietz RL, Jay ZJ. 2020. Next-generation physiology 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, Lehmann K, Maier R, Hoffmann A, Riedel R, Heiking K, Zimmermann J, Siegmund B, Mashreghi M, Radbruch A, Chang H. 2020. Specific microbiota enhances intestinal IgA levels by inducing TGF- $\beta$  in T follicular helper cells of Peyer's patches in mice. *Eur J Immunol* 50:783–794. <https://doi.org/10.1002/eji.201948474>.
19. O'Neill K, Aghaepour N, Spidlen J, Brinkman R. 2013. Flow cytometry bioinformatics. *PLoS Comput Biol* 9:e1003365. <https://doi.org/10.1371/journal.pcbi.1003365>.
20. Saey Y, Van Gassen S, Lambrecht BN. 2016. Computational flow cytometry: helping to make sense of high-dimensional immunology data. *Nat Rev Immunol* 16:449–462. <https://doi.org/10.1038/nri.2016.56>.
21. Quixabeira VBL, Nabout JC, Rodrigues FM. 2010. Trends in genetic literature with the use of flow cytometry. *Cytometry A* 77A:207–210. <https://doi.org/10.1002/cyto.a.20837>.
22. Mair F, Prlic M. 2018. OMIP 044: 28 color immunophenotyping of the human dendritic cell compartment. *Cytometry A* 93A:402–405. <https://doi.org/10.1002/cyto.a.23331>.
23. Buyschaert B, Byloos B, Leys N, Van Houdt R, Boon N. 2016. Reevaluating multicolor flow cytometry to assess microbial viability. *Appl Microbiol Biotechnol* 100:9037–9051. <https://doi.org/10.1007/s00253-016-7837-5>.
24. Léonard L, Chibane LB, Bouhedda BO, Degraeve P, Oulahl N. 2016. Recent advances on multi-parameter flow cytometry to characterize antimicrobial treatments. *Front Microbiol* 7:1225. <https://doi.org/10.3389/fmicb.2016.01225>.
25. Koch C, Müller S. 2018. Personalized microbiome dynamics—cytometric fingerprints for routine diagnostics. *Mol Aspects Med* 59:123–134. <https://doi.org/10.1016/j.mam.2017.06.005>.
26. Hatzenpichler R, Connon SA, Goudeau D, Malmstrom RR, Woyke T, Orphan VJ. 2016. Visualizing in situ translational activity for identifying and sorting slow-growing archaeal-bacterial consortia. *Proc Natl Acad Sci U S A* 113:E4069–E4078. <https://doi.org/10.1073/pnas.1603757113>.
27. Barbesti S, Citterio S, Labra M, Baroni MD, Neri MG, Sgorbati S. 2000. Two and three-color fluorescence flow cytometric analysis of immunoidentified viable bacteria. *Cytometry* 40:214–218. [https://doi.org/10.1002/1097-0320\(20000701\)40:3<214::AID-CYTO6>3.0.CO;2-M](https://doi.org/10.1002/1097-0320(20000701)40:3<214::AID-CYTO6>3.0.CO;2-M).
28. Duquenoey A, Bellais S, Gasc C, Schwintner C, Dore J, Thomas V. 2020. Assessment of Gram- and viability-staining methods for quantifying bacterial community dynamics using flow cytometry. *Front Microbiol* 11:1469. <https://doi.org/10.3389/fmicb.2020.01469>.
29. Spidlen J, Moore W, Parks D, Goldberg M, Bray C, Bierre P, Gorombey P, Hyun B, Hubbard M, Lange S, Lefebvre R, Leif R, Novo D, Ostruszka L, Treister A, Wood J, Murphy RF, Roederer M, Sudar D, Zigon R, Brinkman RR. 2010. Data file standard for flow cytometry, version FCS 3.1. *Cytometry A* 77A:97–100. <https://doi.org/10.1002/cyto.a.20825>.
30. Hahne F, LeMeur N, Brinkman RR, Ellis B, Haaland P, Sarkar D, Spidlen J, Strain E, Gentleman R. 2009. flowCore: a Bioconductor package for high throughput flow cytometry. *BMC Bioinformatics* 10:106. <https://doi.org/10.1186/1471-2105-10-106>.
31. Arioli S, Guglielmetti S, Amalfitano S, Viti C, Marchi E, Decorosi F, Giovannetti L, Mora D. 2014. Characterization of tetA-like gene encoding for a major facilitator superfamily efflux pump in *Streptococcus thermophilus*. *FEMS Microbiol Lett* 355:61–70. <https://doi.org/10.1111/1574-6968.12449>.
32. Anvarian AHP, Cao Y, Srikumar S, Fanning S, Jordan K. 2016. Flow cytometric and 16S sequencing methodologies for monitoring the physiological status of the microbiome in powdered infant formula production. *Front Microbiol* 7:968. <https://doi.org/10.3389/fmicb.2016.00968>.
33. Conacher CG, Naidoo-Blossoples RK, Roussow D, Bauer FF. 2020. Real-time monitoring of population dynamics and physical interactions in a synthetic yeast ecosystem by use of multicolour flow cytometry. *Appl Microbiol Biotechnol* 104:5547–5562. <https://doi.org/10.1007/s00253-020-10607-x>.
34. Vergine P, Amalfitano S, Salerno C, Berardi G, Pollice A. 2020. Reuse of ultrafiltered effluents for crop irrigation: on-site flow cytometry unveiled microbial removal patterns across a full-scale tertiary treatment. *Sci Total Environ* 718:137298. <https://doi.org/10.1016/j.scitotenv.2020.137298>.
35. Parks DR, Roederer M, Moore W. 2006. A new “Logicle” display method avoids deceptive effects of logarithmic scaling for low signals and compensated data. *Cytometry A* 69A:541–551. <https://doi.org/10.1002/cyto.a.20258>.
36. Herzenberg L, Tung J, Moore W, Herzenberg L, Parks DR. 2006. Interpreting flow cytometry data: a guide for the perplexed. *Nat Immunol* 7:681–685. <https://doi.org/10.1038/ni0706-681>.
37. Finak G, Perez JM, Weng A, Gottardo R. 2010. Optimizing transformations for automated, high throughput analysis of flow cytometry data. *BMC Bioinformatics* 11:546. <https://doi.org/10.1186/1471-2105-11-546>.
38. Sarkar D, Le Meur N, Gentleman R. 2008. Using flowViz to visualize flow cytometry data. *Bioinformatics* 24:878–879. <https://doi.org/10.1093/bioinformatics/btn021>.
39. Van P, Jiang W, Gottardo R, Finak G. 2018. ggCyto: next generation open-source visualization software for cytometry. *Bioinformatics* 34:3951–3953. <https://doi.org/10.1093/bioinformatics/bty441>.
40. Brown MR, Hands CL, Coello-Garcia T, Sani BS, Ott AIG, Smith SJ, Davenport RJ. 2019. A flow cytometry method for bacterial quantification and biomass estimates in activated sludge. *J Microbiol Methods* 160:73–83. <https://doi.org/10.1016/j.mimet.2019.03.022>.
41. Fletez-Brant K, Spidlen J, Brinkman RR, Roederer M, Chattopadhyay PK. 2016. flowClean: automated identification and removal of fluorescence anomalies in flow cytometry data. *Cytometry A* 89A:461–471. <https://doi.org/10.1002/cyto.a.22837>.
42. Monaco G, Chen H, Poidinger M, Chen J, de Magalhães JP, Larbi A. 2016. flowAI: automatic and interactive anomaly discerning tools for flow cytometry data. *Bioinformatics* 32:2473–2480. <https://doi.org/10.1093/bioinformatics/btw191>.
43. Meskas J, Wang S, Brinkman RR. 2020. flowCut—an R package for precise and accurate automated removal of outlier events and flagging of files based on time versus fluorescence analysis. *bioRxiv* <https://doi.org/10.1101/2020.04.23.058545>.
44. Hammes F, Egli T. 2010. Cytometric methods for measuring bacteria in water: advantages, pitfalls and applications. *Anal Bioanal Chem* 397:1083–1095. <https://doi.org/10.1007/s00216-010-3646-3>.
45. Fragoso GM, Poulton AJ, Pratt NJ, Johnsen G, Purdie DA. 2019. Trait-based analysis of subpolar North Atlantic phytoplankton and plastidic ciliate communities using automated flow cytometer. *Limnol Oceanogr* 64:1763–1778. <https://doi.org/10.1002/lno.1189>.
46. Proctor CR, Besmer MD, Langenegger T, Beck K, Walser JC, Ackermann M, Bürgmann H, Hammes F. 2018. Phylogenetic clustering of small low nucleic acid-content bacteria across diverse freshwater ecosystems. *ISME J* 12:1344–1359. <https://doi.org/10.1038/s41396-018-0070-8>.

47. van der Maaten L, Hinton G. 2008. Visualizing data using t-SNE. *J Mach Learn Res* 9:2579–2605.
48. McInnes L, Healy J, Saul N, Großberger L. 2018. UMAP: uniform manifold approximation and projection. *J Open Source Softw* 3:861. <https://doi.org/10.21105/joss.00861>.
49. Aghaeepour N, Finak G, Hoos H, Mosmann TR, Brinkman R, Gottardo R, Scheuermann RH, DREAM Consortium. 2013. Critical assessment of automated flow cytometry data analysis techniques. *Nat Methods* 10:228–238. <https://doi.org/10.1038/nmeth.2365>.
50. Weber LM, Robinson MD. 2016. Comparison of clustering methods for high-dimensional single-cell flow and mass cytometry data. *Cytometry A* 89A:1084–1096. <https://doi.org/10.1002/cyto.a.23030>.
51. Liu X, Song W, Wong BY, Zhang T, Yu S, Lin GN, Ding X. 2019. A comparison framework and guideline of clustering methods for mass cytometry data. *Genome Biol* 20:297. <https://doi.org/10.1186/s13059-019-1917-7>.
52. Van Gassen S, Callebaut B, Van Helden MJ, Lambrecht BN, Demeester P, Dhaene T, Saey Y. 2015. FlowSOM: using self-organizing maps for visualization and interpretation of cytometry data. *Cytometry A* 87A:636–645. <https://doi.org/10.1002/cyto.a.22625>.
53. Levine JH, Simonds EF, Bendall SC, Davis KL, Amir E-a. D, Tadmor MD, Litvin O, Fienberg HG, Jager A, Zunder ER, Finck R, Gedman AL, Radtke I, Downing JR, Pe'er D, Nolan GP. 2015. Data-driven phenotypic dissection of AML reveals progenitor-like cells that correlate with prognosis. *Cell* 162:184–197. <https://doi.org/10.1016/j.cell.2015.05.047>.
54. Ludwig J, zu Siederdisen CH, Liu Z, Stadler PF, Müller S. 2019. flowEMMI: an automated model-based clustering tool for microbial cytometric data. *BMC Bioinformatics* 20:643. <https://doi.org/10.1186/s12859-019-3152-3>.
55. Rajwa B, Venkatapathi M, Ragheb K, Banada PP, Hirleman ED, Lary T, Robinson JP. 2008. Automated classification of bacterial particles in flow by multiangle scatter measurement and support vector machine classifier. *Cytometry A* 73A:369–379. <https://doi.org/10.1002/cyto.a.20515>.
56. Pereira GC, Ebecken NFF. 2011. Combining in situ flow cytometry and artificial neural networks for aquatic systems monitoring. *Expert Syst Appl* 38:9626–9632. <https://doi.org/10.1016/j.eswa.2011.01.140>.
57. Rubbens P, Props R, Boon N, Waegeman W. 2017. Flow cytometric single-cell identification of populations in synthetic bacterial communities. *PLoS One* 12:e0169754. <https://doi.org/10.1371/journal.pone.0169754>.
58. Duygan BDÖ, Hadadi N, Babu AF, Seyfried M, van der Meer JR. 2020. Rapid detection of microbiota cell type diversity using machine-learned classification of flow cytometry data. *Commun Biol* 3:379. <https://doi.org/10.1038/s42003-020-1106-y>.
59. García-Timmermans C, Rubbens P, Heyse J, Kerckhof FM, Props R, Skirtach AG, Waegeman W, Boon N. 2020. Discriminating bacterial phenotypes at the population and single-cell level: a comparison of flow cytometry and Raman spectroscopy fingerprinting. *Cytometry A* 97A:713–714. <https://doi.org/10.1002/cyto.a.23952>.
60. Koch C, Fetzer I, Schmidt T, Harms H, Müller S. 2013. Monitoring functions in managed microbial systems by cytometric bar coding. *Environ Sci Technol* 47:1753–1760. <https://doi.org/10.1021/es3041048>.
61. Koch C, Fetzer I, Harms H, Müller S. 2013. CHIC—an automated approach for the detection of dynamic variations in complex microbial communities. *Cytometry A* 83A:561–567. <https://doi.org/10.1002/cyto.a.22286>.
62. Wanderley BMS, Araújo DS, Quiroga MV, Amado AM, Neto ADD, Sarmiento H, Metz SD, Unrein F. 2019. flowDiv: a new pipeline for analyzing flow cytometric diversity. *BMC Bioinformatics* 20:274. <https://doi.org/10.1186/s12859-019-2787-4>.
63. Roederer M, Moore W, Treister A, Hardy RR, Herzenberg LA. 2001. Probability binning comparison: a metric for quantitating multivariate distribution differences. *Cytometry* 45:47–55. [https://doi.org/10.1002/1097-0320\(20010901\)45:1<47::AID-CYTO1143>3.0.CO;2-A](https://doi.org/10.1002/1097-0320(20010901)45:1<47::AID-CYTO1143>3.0.CO;2-A).
64. Rogers WT, Holyst HA. 2009. FlowFP: a Bioconductor package for fingerprinting flow cytometric data. *Adv Bioinform* 2009:193947. <https://doi.org/10.1155/2009/193947>.
65. Huang TH, Ning X, Wang X, Murthy N, Tzeng YL, Dickson RM. 2015. Rapid cytometric antibiotic susceptibility testing utilizing adaptive multidimensional statistical metrics. *Anal Chem* 87:1941–1949. <https://doi.org/10.1021/ac504241x>.
66. Amalfitano S, Fazi S, Ejarque E, Freixa A, Romani AM, Butturini A. 2018. Deconvolution model to resolve cytometric microbial community patterns in flowing waters. *Cytometry A* 93A:194–200. <https://doi.org/10.1002/cyto.a.23304>.
67. Rubbens P, Props R, Kerckhof FM, Boon N, Waegeman W. PhenoGMM: Gaussian mixture modelling of cytometry data quantifies changes in microbial community structure. *mSphere*, in press.
68. Aghaeepour N, Chattopadhyay P, Chikina M, Dhaene T, Van Gassen S, Kursu M, Lambrecht BN, Malek M, McLachlan GJ, Qian Y, Qiu P, Saey Y, Stanton R, Tong D, Vens C, Walkowiak S, Wang K, Finak G, Gottardo R, Mosmann T, Nolan GP, Scheuermann RH, Brinkman RR. 2016. A benchmark for evaluation of algorithms for identification of cellular correlates of clinical outcomes. *Cytometry A* 89A:16–21. <https://doi.org/10.1002/cyto.a.22732>.
69. Kotecha N, Krutzik PO, Irish JM. 2010. Web-based analysis and publication of flow cytometry experiments. *Curr Protoc Cytom* 53:10.17.1–10.17.24. <https://doi.org/10.1002/0471142956.cy101753>.
70. Spidlen J, Breuer K, Rosenberg C, Kotecha N, Brinkman RR. 2012. Flow-Repository: a resource of annotated flow cytometry datasets associated with peer-reviewed publications. *Cytometry A* 81A:727–731. <https://doi.org/10.1002/cyto.a.22106>.
71. Lee JA, Spidlen J, Boyce K, Cai J, Crosbie N, Dalphin M, Furlong J, Gasparetto M, Goldberg M, Goralczyk EM, Hyun B, Jansen K, Kollmann T, Kong M, Leif R, McWeeney S, Moloshok TD, Moore W, Nolan G, Nolan J, Nikolich-Zugich J, Parrish D, Purcell B, Qian Y, Selvaraj B, Smith C, Tchuvatkina O, Wertheimer A, Wilkinson P, Wilson C, Wood J, Zigon R, Scheuermann RH, Brinkman RR, International Society for Advancement of Cytometry Data Standards Task Force. 2008. MIFlowCyt: the minimum information about a flow cytometry experiment. *Cytometry A* 73A:926–930. <https://doi.org/10.1002/cyto.a.20623>.
72. Chan S, Pullerits K, Riechelmann J, Persson KM, Rådström P, Paul CJ. 2018. Monitoring biofilm function in new and matured full-scale slow sand filters using flow cytometric histogram image comparison (CHIC). *Water Res* 138:27–36. <https://doi.org/10.1016/j.watres.2018.03.032>.
73. Duveau F, Hodgins-Davis A, Metzger BPH, Yang B, Tryban S, Walker EA, Lybrook T, Wittkopp PJ. 2018. Fitness effects of altering gene expression noise in *Saccharomyces cerevisiae*. *Elife* 7:e37272. <https://doi.org/10.7554/eLife.37272>.
74. Karava M, Bracharz F, Kabisch J. 2019. Quantification and isolation of *Bacillus subtilis* spores using cell sorting and automated gating. *PLoS One* 14:e0219892-15. <https://doi.org/10.1371/journal.pone.0219892>.
75. Rubbens P, Props R, Garcia-Timmermans C, Boon N, Waegeman W. 2017. Stripping flow cytometry: How many detectors do we need for bacterial identification? *Cytometry A* 91A:1184–1191. <https://doi.org/10.1002/cyto.a.23284>.
76. Buysschaert B, Kerckhof FM, Vandamme P, De Baets B, Boon N. 2018. Flow cytometric fingerprinting for microbial strain discrimination and physiological characterization. *Cytometry A* 93A:201–212. <https://doi.org/10.1002/cyto.a.23302>.
77. Hahne F, Khodabakhshi AH, Bashashati A, Wong CJ, Gascoyne RD, Weng AP, Seyfert-Margolis V, Bourcier K, Asare A, Lumley T, Gentleman R, Brinkman RR. 2010. Per-channel basis normalization methods for flow cytometry data. *Cytometry A* 77A:121–131. <https://doi.org/10.1002/cyto.a.20823>.
78. Jahn M, Seifert J, Hübschmann T, Von Bergen M, Harms H, Müller S. 2013. Comparison of preservation methods for bacterial cells in cytomics and proteomics. *J Integr OMICS* 3:25–33. <https://doi.org/10.5584/jiomics.v3i1.115>.
79. Vazquez-Albacete D, Cavaleiro AM, Christensen U, Seppälä S, Møller BL, Nørholm MHH. 2017. An expression tag toolbox for microbial production of membrane bound plant cytochromes P450. *Biotechnol Bioeng* 114:751–760. <https://doi.org/10.1002/bit.26203>.
80. Reiss J, Perkins DM, Fussmann KE, Krause S, Canhoto C, Romeijn P, Robertson AL. 2019. Groundwater flooding: ecosystem structure following an extreme recharge event. *Sci Total Environ* 652:1252–1260. <https://doi.org/10.1016/j.scitotenv.2018.10.216>.
81. Amir E-a. D, Davis KL, Tadmor MD, Simonds EF, Levine JH, Bendall SC, Shenfeld DK, Krishnaswamy S, Nolan GP, Pe'er D. 2013. viSNE enables visualization of high dimensional single-cell data and reveals phenotypic heterogeneity of leukemia. *Nat Biotechnol* 31:545–552. <https://doi.org/10.1038/nbt.2594>.
82. Sgier L, Freimann R, Zupanec A, Kroll A. 2016. Flow cytometry combined with viSNE for the analysis of microbial biofilms and detection of microplastics. *Nat Commun* 7:11587. <https://doi.org/10.1038/ncomms11587>.
83. Becht E, McInnes L, Healy J, Dutertre CA, Kwok IWH, Ng LG, Ginhoux F, Newell EW. 2019. Dimensionality reduction for visualizing single-cell data using UMAP. *Nat Biotechnol* 37:38–44. <https://doi.org/10.1038/nbt.4314>.
84. Lo K, Hahne F, Brinkman RR, Gottardo R. 2009. flowClust: a Bioconductor package for automated gating of flow cytometry data. *BMC Bioinformatics* 10:145. <https://doi.org/10.1186/1471-2105-10-145>.

85. García FC, López-Urrutia Á, Morán XAG. 2014. Automated clustering of heterotrophic bacterioplankton in flow cytometry data. *Aquat Microb Ecol* 72:175–185. <https://doi.org/10.3354/ame01691>.
86. Ge Y, Sealfon SC. 2012. flowPeaks: a fast unsupervised clustering for flow cytometry data via K-means and density peak finding. *Bioinformatics* 28:2052–2058. <https://doi.org/10.1093/bioinformatics/bts300>.
87. Bervoets I, Van Brempt 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, Taghiyar 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, Swalwell J, Carlson M, Clayton S, Hennon 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/s41597-019-0292-2>.
91. Koch C, Harnisch F, Schröder U, Müller S. 2014. Cytometric fingerprints: evaluation of new tools for analyzing microbial community dynamics. *Front Microbiol* 5:273. <https://doi.org/10.3389/fmicb.2014.00273>.
92. Menyhart L, Nagy S, Lepossa 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* 134-135:67–79. <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 fast microbial community fingerprinting. *Water Res* 46:907–919. <https://doi.org/10.1016/j.watres.2011.11.076>.
95. Kinet R, Dzaomuh P, Baert J, Taminiua B, Daube G, Nezer C, Brostaux Y, Nguyen F, Dumont G, Thonart P, Delvigne F. 2016. Flow cytometry community fingerprinting and amplicon sequencing for the assessment of landfill leachate cellulolytic bioaugmentation. *Bioresour Technol* 214:450–459. <https://doi.org/10.1016/j.biortech.2016.04.131>.
96. Favere J, Buyschaert B, Boon N, De Gusseme B. 2020. Online microbial fingerprinting for quality management of drinking water: full-scale event detection. *Water Res* 170:115353. <https://doi.org/10.1016/j.watres.2019.115353>.
97. Coggins LX, Larma I, Hinchliffe A, Props R, Ghadouani A. 2020. Flow cytometry for rapid characterisation of microbial community dynamics in waste stabilisation ponds. *Water Res* 169:115243. <https://doi.org/10.1016/j.watres.2019.115243>.
98. De Rudder C, Calatayud Arroyo M, Lebeer S, Van de Wiele T. 2020. Dual and triple epithelial coculture model systems with donor-derived microbiota and THP-1 macrophages to mimic host-microbe interactions in the human sinonasal cavities. *mSphere* 5:e00916-19. <https://doi.org/10.1128/mSphere.00916-19>.
99. Rubbens P, Props R, Kerckhof FM, Boon N, Waegeman W. 2021. Cytometric fingerprints of gut microbiota predict Crohn's disease state. *ISME J* 15:354–358. <https://doi.org/10.1038/s41396-020-00762-4>.
100. Fodelianakis S, Lorz A, Valenzuela-Cuevas A, Barozzi A, Booth JM, Daffonchio D. 2019. Dispersal homogenizes communities via immigration even at low rates in a simplified synthetic bacterial metacommunity. *Nat Commun* 10:1314. <https://doi.org/10.1038/s41467-019-09306-7>.
101. Koch C, Günther S, Desta AF, Hübschmann T, Müller S. 2013. Cytometric fingerprinting for analyzing microbial intracommunity structure variation and identifying subcommunity function. *Nat Protoc* 8:190–202. <https://doi.org/10.1038/nprot.2012.149>.
102. Ramette A. 2007. Multivariate analyses in microbial ecology. *FEMS Microbiol Ecol* 62:142–160. <https://doi.org/10.1111/j.1574-6941.2007.00375.x>.
103. Paliy O, Shankar V. 2016. Application of multivariate statistical techniques in microbial ecology. *Mol Ecol* 25:1032–1057. <https://doi.org/10.1111/mec.13536>.
104. Fontana S, Petchey OL, Pomati F. 2016. Individual level trait diversity concepts and indices to comprehensively describe community change in multidimensional trait space. *Funct Ecol* 30:808–818. <https://doi.org/10.1111/1365-2435.12551>.
105. Liu Z, Cichocki N, Bonk F, Günther S, Schattenberg F, Harms H, Centler F, Müller S. 2018. Ecological stability properties of microbial communities assessed by flow cytometry. *mSphere* 3:e00564-17. <https://doi.org/10.1128/mSphere.00564-17>.
106. Christensen H, Bakken LR, Olsen RA. 1993. Soil bacterial DNA and biovolume profiles measured by flow-cytometry. *FEMS Microbiol Lett* 102:129–140. <https://doi.org/10.1111/j.1574-6968.1993.tb05804.x>.
107. Bouvier T, Troussellier M, Anzil A, Courties C, Servais P. 2001. Using light scatter signal to estimate bacterial biovolume by flow cytometry. *Cytometry* 44:188–194. [https://doi.org/10.1002/1097-0320\(20010701\)44:3<188::AID-CYTO1111>3.0.CO;2-C](https://doi.org/10.1002/1097-0320(20010701)44:3<188::AID-CYTO1111>3.0.CO;2-C).
108. Felip M, Andreatta S, Sommaruga R, Straskrábová V, Catalan J. 2007. Suitability of 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>.
109. 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>.
110. 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>.
111. Bowman JS, Amaral-Zettler LA, J Rich J, M Luria C, Ducklow 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>.
112. Schleich C, Chan S, Pullerits K, Besmer MD, Paul CJ, Rådström 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/w11102137>.
113. Huang TH, Tzeng 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>.
114. Props R, Kerckhof FM, Rubbens P, De Vriese 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>.
115. Props R, Schmidt ML, Heyse J, Vanderploeg 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/10.1111/1462-2920.13953>.
116. García FC, Alonso-Sáez L, Morán XAG, López-Urrutia Á. 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/10.1111/1462-2920.12984>.
117. Jahn M, Seifert J, von Bergen M, Schmid A, Bühler B, Müller S. 2013. Subpopulation-proteomics in prokaryotic populations. *Curr Opin Biotechnol* 24:79–87. <https://doi.org/10.1016/j.copbio.2012.10.017>.
118. Stepanauskas R, Fergusson 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* 8:84. <https://doi.org/10.1038/s41467-017-00128-z>.

**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 Deneff, 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.

