

ARTICLE

Received 2 Oct 2015 | Accepted 11 Apr 2016 | Published 18 May 2016

DOI: 10.1038/ncomms11587

OPEN

Flow cytometry combined with viSNE for the analysis of microbial biofilms and detection of microplastics

Linn Sgier¹, Remo Freimann², Anze Zupanic^{1,*} & Alexandra Kroll^{1,*}

Biofilms serve essential ecosystem functions and are used in different technical applications. Studies from stream ecology and waste-water treatment have shown that biofilm functionality depends to a great extent on community structure. Here we present a fast and easy-to-use method for individual cell-based analysis of stream biofilms, based on stain-free flow cytometry and visualization of the high-dimensional data by viSNE. The method allows the combined assessment of community structure, decay of phototrophic organisms and presence of abiotic particles. In laboratory experiments, it allows quantification of cellular decay and detection of survival of larger cells after temperature stress, while in the field it enables detection of community structure changes that correlate with known environmental drivers (flow conditions, dissolved organic carbon, calcium) and detection of microplastic contamination. The method can potentially be applied to other biofilm types, for example, for inferring community structure for environmental and industrial research and monitoring.

¹Department of Environmental Toxicology, Eawag—Swiss Federal Institute for Aquatic Science and Technology, 8600 Dübendorf, Switzerland.

²ETHZ—Department of Biology, Institute of Molecular Health Sciences, 8093 Zürich, Switzerland. * These authors contributed equally to this work. Correspondence and requests for materials should be addressed to A.Z. (email: anze.zupanic@eawag.ch) or to A.K. (email: alexandra.kroll@eawag.ch).

Surfaces at solid–liquid interfaces are often colonized by microbial communities organized in biofilms, which serve essential ecosystem functions¹ and are also used in different technical applications, such as waste-water treatment and biofuel production². Their environmental and technical functions are, among others, determined by species composition and phenotypic variability (further referred to as community structure), as has been shown for the efficiency of biological waste-water treatment³, the beneficial or harmful effects of epibiotic biofilms⁴ and the role of stream biofilms in nutrient cycling⁵.

Depending on species and phenotype, individual microbes vary in size, internal structure and chemical composition. Since these parameters are reflected in optical properties, such as light scattering and autofluorescence, they can be measured in single cells by stain-free flow cytometry (FC). While this concept has been made use of in phytoplankton research^{6–11}, we propose that autofluorescence-based FC is also a useful tool for biofilms. One of the main challenges of using autofluorescence-based polychromatic FC is the extraction of relevant information from the high-dimensional data and clustering single cells into phenotype-based groups^{12,13}. This general issue has been approached by different clustering methods, however, it remains difficult to combine good clustering performance with meaningful visualization¹⁴, making biological interpretation difficult and time consuming.

Here we present a method to characterize biofilms at the single-cell level, on the basis of the FC measurements of optical scatter and autofluorescence, combined with single-cell visualization by viSNE (visual stochastic network embedding)¹⁵. viSNE is a tool for nonlinear dimension reduction and visualization of high-dimensional data, developed for the analysis of human blood cells. It was originally used to visualize mass cytometry data from healthy and leukaemic blood samples, to qualitatively distinguish between blood cell types and to detect aberrant phenotypic shifts in the blood cell community¹⁵. In the present study, we optimize the published viSNE procedures for use with data from heterogeneous biofilms to distinguish between species and different phenotypes present in the biofilms, and introduce a procedure for quantification of changes in biofilm community structure. As model we selected stream biofilms, which are communities of predominantly algae, cyanobacteria and heterotrophic bacteria, for their complexity¹⁶, significance to stream ecosystems¹ and lack of reports on characterization by FC. We demonstrate the robustness and biological relevance of the method by tracking phenotypic shifts in mixtures of species, quantifying effects of temperature stress on the biofilms and identifying correlations between environmental parameters and biofilm community structure. In addition, due to the

sensitivity of the method to rare events, we can detect microplastic contamination in stream biofilms.

Results

viSNE protocol adapted to biofilm analysis. Data analysis by viSNE and quantification of community structure are the basis for all results presented in this manuscript, therefore we first present a short overview of the workflow and corresponding inputs and outputs. The input into viSNE is normalized high-dimensional FC data. The output of viSNE analysis is a two-dimensional scatter plot (viSNE map) in which cells are positioned according to similarity in the high-dimensional space and grouped into visually separable clusters (viSNE clusters, for example, in Fig. 1). viSNE allows for direct cross-sample comparison, by running the SNE algorithm on the data points from all samples together to create a single viSNE map, and then visually comparing the subsets of the map that present the single samples (for example, Fig. 1a is the viSNE map of 30 single species, while in Fig. 1b,c the submaps belonging to individual species are presented).

In our study, we assigned viSNE clusters to subpopulations based on expert knowledge of optical scattering and autofluorescence properties of the biofilm species and *a posteriori* projection of reference data sets on the viSNE maps (that is, single species, pigment-bleached reference samples) (for example, Fig. 2). On the basis of the categorization, a quantitative comparison between the samples can be done if the number of cells belonging to each subpopulation in each sample is counted (for example, Fig. 3). For reference, analysis of 18 samples (150,000 cells in total) by viSNE took us ca. 1 h 30 min on a desktop computer (Supplementary Table 1).

Species separation. The phenotypic variability within stream biofilms is much greater than the variability of a human blood sample. Therefore, we first tested whether viSNE can be used to separate single species as successfully as the blood cell types were separated in the original study¹⁵. For this, we used samples from continuous cultures of 30 selected reference species, covering four main microbial groups constituting stream biofilms: cyanobacteria, diatoms, green algae and red algae (Supplementary Table 2). We recorded light absorbance and fluorescence spectra of the species and adjusted the flow cytometer hardware settings to cover light ranges that would reveal differences in fluorescence intensities (Supplementary Fig. 1, see the ‘Methods’ section). The FC acquisition protocol was adapted to include at least 99% of all cells of all species grown at identical conditions. The resulting light scattering and fluorescence profiles of all species show that there are no optical markers that clearly separate between groups of organisms (cyanobacteria, green algae, diatoms, red algae;

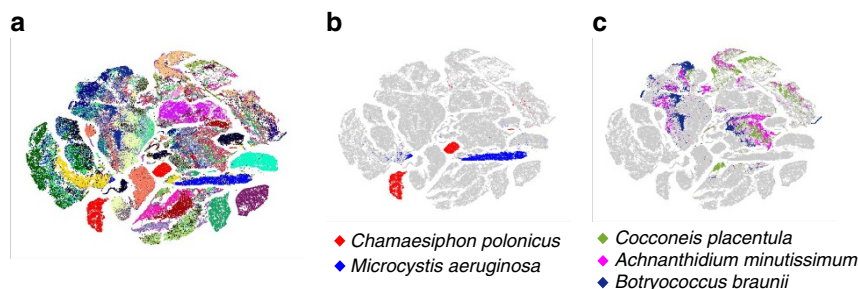


Figure 1 | viSNE map and submaps of reference species. (a) A single viSNE map was created using single-cell non-gated FC data on 30 reference species cultured and measured individually (Supplementary Table 1). Each species is represented by a single colour. (b) Submaps of cyanobacteria *Chamaesiphon polonicus* and *Microcystis aeruginosa* show very little overlap, while (c) submaps of diatoms *Achnantheidum minutissimum* and *Cocconeis placentula* and green algae *Botryococcus braunii* show high overlap and are therefore impossible to separate.

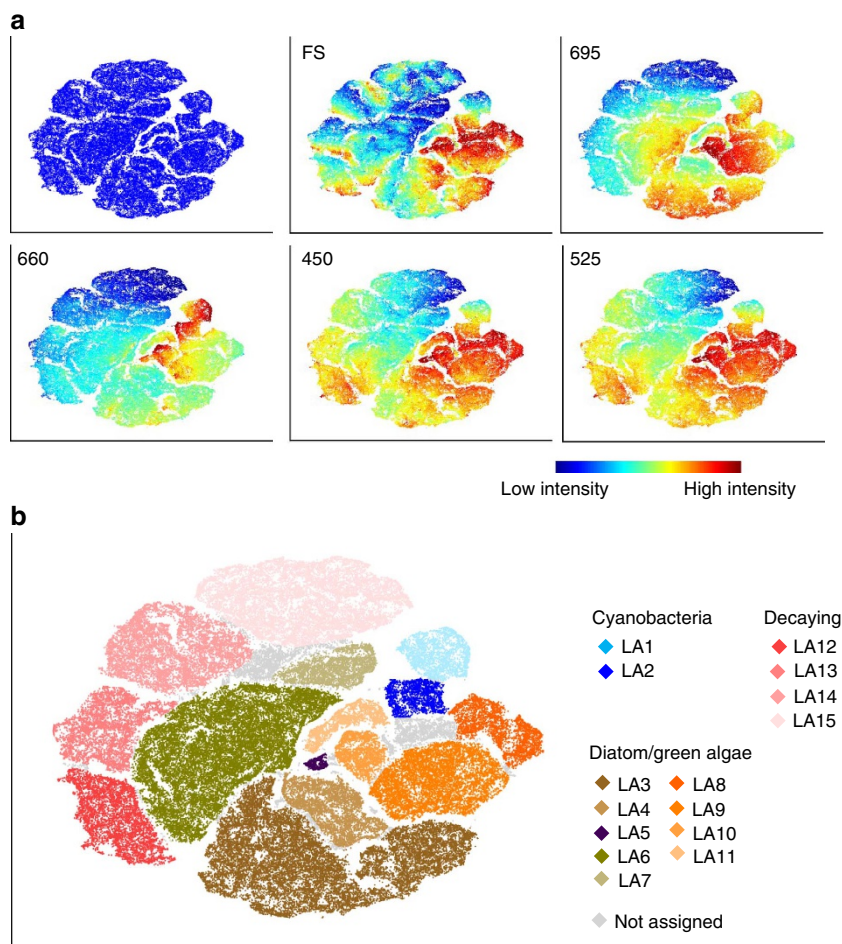


Figure 2 | Categorizing subpopulations in temperature-stressed stream biofilms. (a) Stream biofilms were assessed by FC directly after transfer to higher temperature, after 1, 2 and 3 weeks and the acquired data was altogether mapped by viSNE. viSNE maps are shown in single colour, with each point in the viSNE map representing a single cell from the biofilms, or coloured according to FS and fluorescence intensity at specific wavelengths (nm) measured by FC (full set of wavelengths displayed in Supplementary Fig. 6). (b) Subpopulations (LA1-LA15) categorized based on the viSNE map, optical scatter and fluorescence intensities (a) (Supplementary Fig. 6). A fraction of the particles (4.5–5.8%) was not categorized due to lack of distinct properties. Comparison of subpopulation properties with data acquired from reference species and pigment-bleached reference samples allowed for assigning subpopulations to types of organisms and potentially decaying cells (Supplementary Fig. 8).

Supplementary Fig. 2). For example, of the cyanobacteria used in this study, high fluorescence around 620 nm would only have distinguished *Chaemesiphon polonicus* from most green algae, diatoms and red algae, although previously, fluorescence in the orange range has been used to identify phycoerythrin and/or phycocyanin containing cells⁸. High fluorescence at 660 nm was a slightly better indicator separating two in six cyanobacteria, however, several green algae and diatoms show equally high fluorescence in this range. When analysed by viSNE, some species produced clearly defined clusters, while most were spread across different parts of the viSNE map (Fig. 1a). For example, cyanobacteria *Chamaesiphon polonicus* and *Microcystis aeruginosa* were associated with two clusters each, and can thus be clearly separated (Fig. 1b). Diatom species *Achnanthes minutissimum* and *Cocconeis placentula* and green algae species *Botryococcus braunii*, although morphologically different, shared clusters and thus could not be separated (Fig. 1c). Some species, such as *C. polonicus*, were composed of more than one distinct cluster, indicating the presence of several phenotypes (Fig. 1b). A similarity analysis of the reference species viSNE submaps using maximum mean discrepancy (MMD; see the ‘Methods’ section) showed that only three species from our set were distinctly different from all others (Supplementary Fig. 3). We hypothesized

that the capacity of viSNE for species separation in mixed samples strongly depends on species composition.

To assess the separation capacity in simple mixtures of known species, we tracked temporal phenotypic shifts in simple artificial communities. We mixed two (*Achnanthes minutissimum* and *Nitzschia palea*, (A + N)) or four (A, N, *Botryococcus braunii* and *Chamaesiphon polonicus*, (A + N + B + C)) reference species, followed by FC and light microscopy (LM). Single-species cultures were used as control. After a week, the phenotypes of single species and mixtures moderately changed in the planktonic fractions and strongly changed in the bottom fractions (Supplementary Fig. 4), especially *C. polonicus* (Supplementary Fig. 5). In all single species, bottom fractions contained a subpopulation of cells of low forward scatter (FS) and fluorescence, which can be explained by sedimentation of decaying or decayed cells. LM confirmed the presence of debris in these samples. In the A + N mixtures, A became dominant in both planktonic and bottom fractions, while in the A + N + B + C mixtures, B and C dominated the planktonic fraction, while A and N dominated the bottom fraction, as confirmed by LM. Phenotypic shifts in simple mixtures were successfully tracked, but to associate observed phenotypes with species, single-species controls were needed.

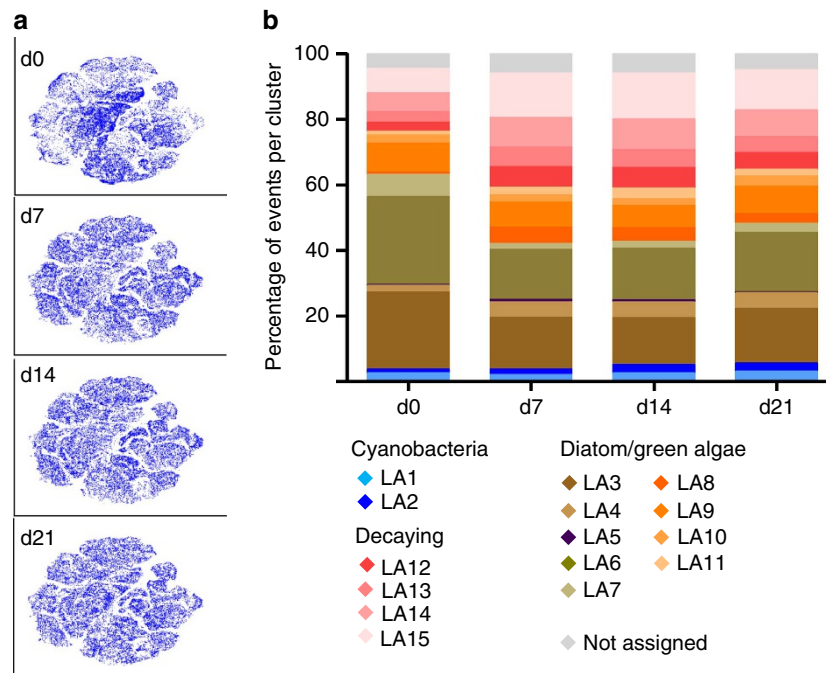


Figure 3 | Tracking and quantification of biofilm community structure changes after temperature increase. (a) viSNE submaps (Fig. 2a) belonging to days 0, 7, 14 and 21 post temperature increase. More colour-intense regions of the submaps depict regions of higher cell density. Similarity analysis of technical replicates (that is, three per sample), biological replicates (that is, five independent microcosms) and time points indicated that the detected changes in community structure were governed by time point and not biological or technical noise (Supplementary Fig. 10). (b) Quantification of subpopulations defined in Fig. 2b, pooled from five biological replicates for each time point after temperature increase (all biological replicates are depicted in Supplementary Fig. 11). Statistical analysis of subpopulation sizes is available in Supplementary Fig. 9.

Quantifying structural shifts in complex biofilms. We then tested the performance of the method on complex stream biofilms of unknown composition. Biofilms were grown from stream water in five independent microcosms in the laboratory and exposed to a temperature increase (median 13.4–17.75 °C). They were assessed by LM and FC at four time points and the FC data were analysed by viSNE. Using the viSNE map, optical scatter and fluorescence properties (Fig. 2a, Supplementary Fig. 6), we categorized the measured cells into 15 subpopulations (Fig. 2b, Supplementary Fig. 7). According to *a posteriori* projection of reference data sets, subpopulations LA1 and LA2 mostly overlapped with cyanobacteria species, LA3–11 with green algae and diatom species and LA12–15 with pigment-bleached reference samples (Supplementary Fig. 8). LA12–15 were characterized by decreasing red fluorescence, which corresponds to reported consecutive decay of chlorophylls¹⁷.

The categorization into subpopulations forms the basis for a quantitative analysis of changes in biofilm community structure. Here, the viSNE submaps for each time point and corresponding subpopulation sizes show that the community structure changed significantly within the first week, while it remained relatively stable for the following 2 weeks (Fig. 3, statistical analysis in Supplementary Fig. 9). A prominent change was an increase in the fraction of decaying cells (LA12–15) by day 7, however some subpopulations have moved towards the original state by day 21 (Supplementary Fig. 9). The increase of decaying cells can be explained by a corresponding decrease in the most dominant healthy diatom and green algae subpopulations (LA3,6,7). Generally, subpopulations that increased contained cells with high FS indicating a general increase in larger cells as FS is a proxy for cell size (LA2,4,8,11). This trend was also confirmed by LM. Subpopulations LA8 and LA11, which showed the greatest relative increase, overlapped strongly with red alga *Bangia*

atropurpurea (Supplementary Fig. 8). Since the presence of *B. atropurpurea* was also confirmed by LM (Supplementary Table 3), this indicates that *B. atropurpurea* tolerates temperature increase better than most other species in our biofilm community, possibly owing to its temperature-dependent life stages¹⁸.

Field assessment of biofilm community structure. Adding another level of complexity, we tested the method in the field setting, for analysis of stream biofilms originating from variable physico-chemical conditions. We selected six sites along the stream Mönchaltorfer Aa (Supplementary Fig. 12), based on expected differences in water chemistry and anthropogenic contamination (Supplementary Tables 4–6). Biofilm samples were taken from the sites, analysed by LM and FC and the FC data analysed by viSNE.

On the basis of the viSNE map, optical scatter and fluorescence intensities (Fig. 4a, Supplementary Fig. 13), the cells were categorized into 11 subpopulations (Fig. 4b, Supplementary Fig. 14). Compared with the laboratory-based biofilms, a phenotypically distinct subpopulation (MA10) appeared in the field, with high sideward scatter (SS) and the lowest fluorescence across all wavelengths, probably consisting of non-organic material. Its absence from the laboratory-based biofilms can be explained by the use of a sediment trap in the laboratory setup. Otherwise, subpopulations with similar characteristics to the laboratory-based subpopulations were derived (Supplementary Fig. 14). MA1–2 correspond to cyanobacterial LA1–2, MA7–9 correspond to decaying cells LA12–15, while MA3–6 correspond to diatom and green algae like LA3–7. We found no subpopulations that would correspond to LA8–11, that is, large and strongly fluorescent diatom and green algae cells.

The viSNE submaps (Fig. 5a) and subpopulation quantification across the sites (Fig. 5b) show significant differences in biofilm

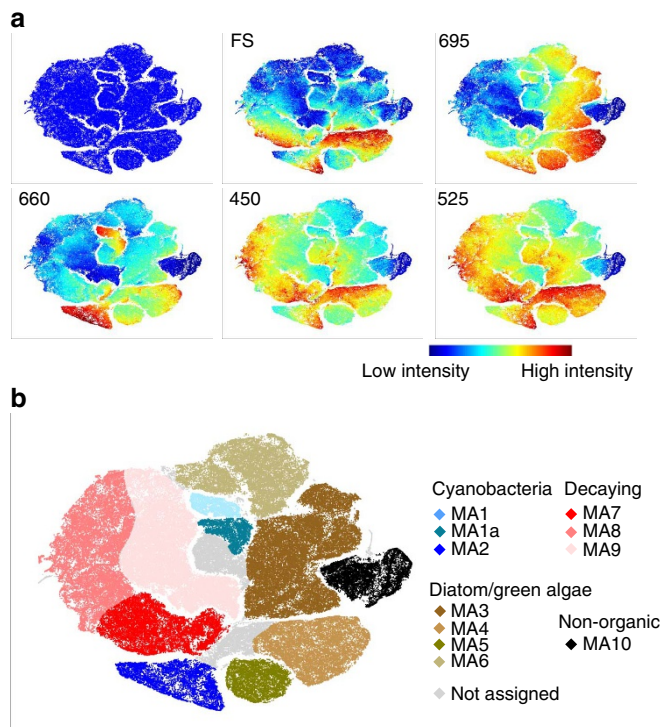


Figure 4 | Categorizing subpopulations in stream biofilms sampled in the field. (a) Stream biofilms were assessed by FC after sampling at six sites along the stream Mönchaltorfer Aa and altogether mapped by viSNE. viSNE maps are shown in single colour, with each point in the viSNE map representing a single cell or particle from the biofilms or coloured according to FS and fluorescence intensity at specific wavelengths (nm) measured by FC (full set displayed in Supplementary Fig. 13). (b) Subpopulations (MA 1–10) categorized based on the viSNE map and optical scatter and fluorescence intensities (a) (Supplementary Fig. 13). Some cells (range 2.7–11.2 %) were not categorized due to lack of distinct properties. Comparison of subpopulation properties with data acquired from reference species and pigment-bleached reference samples (Supplementary Fig. 15) allowed for assigning subpopulations to types of organisms and potentially decaying cells.

community structure (statistical analysis: Supplementary Fig. 16). The non-organic fraction features the strongest at sites E and F, which have a naturally high sediment load. The fraction of decaying cells was lowest at site C, followed by site A (close to the spring), while at other sites it was close to 50% of the particle population. Accordingly, the green algae/diatom fraction was strongest at site C. Large cyanobacteria (MA2) featured most strongly at sites A and B. These observations were confirmed independently by LM (Supplementary Table 7). Our approach proved to be sensitive to detect significant differences in biofilms sampled from sites with different physico-chemical conditions.

Evaluation of biological relevance. To assess the biological relevance of our approach, we tested the correlation between the measured physico-chemical parameters (Supplementary Tables 4 and 6) and the measured community structure (Fig. 5b) by selecting a linear model based on redundancy analysis (RDA). Sampling sites were well-separated by the subpopulation structure. Flow, dissolved organic carbon (DOC) and calcium concentration (Ca) explained 71.3 % of the variation in subpopulation sizes (Fig. 5c). While the linear contribution of chloride, potassium, sodium, silicate, conductivity and pH was

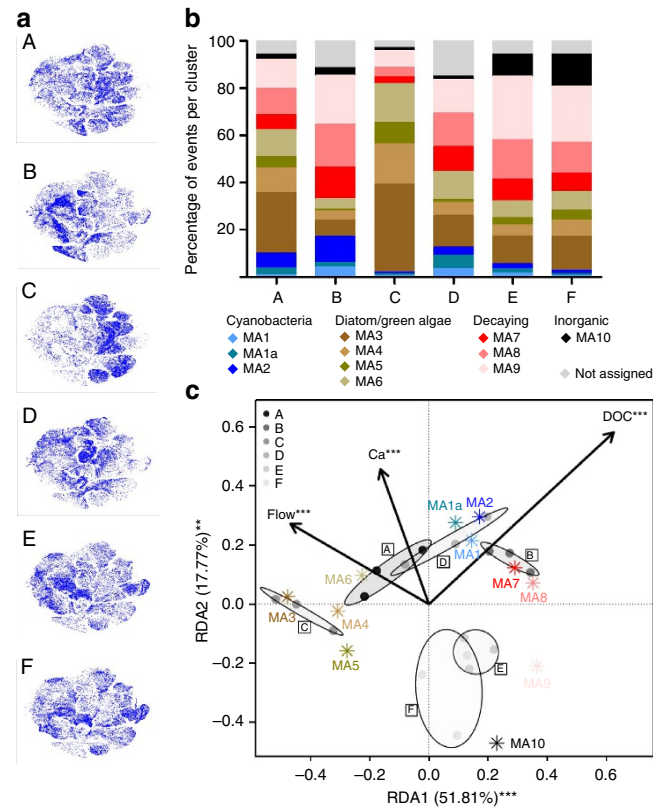


Figure 5 | Tracking biofilm community structure changes along a stream.

(a) viSNE submaps belonging to six sites (A–F) along the stream Mönchaltorfer Aa. More colour-intense regions of the submaps depict regions of higher cell or particle density. Similarity analysis of technical replicates (that is, three per sample), biological replicates (that is, 3 stones per site) and sites indicated that the detected changes in community structure were governed by site and not biological or technical noise (Supplementary Fig. 17). The sites are characterized in Supplementary Tables 4–6. Site A is at the spring of the stream in the forest, site B is in an unshaded stretch, site C is shaded but in the straightened section of the stream-like sites D–F, which are additionally influenced by waste-water treatment plant effluents, site D being situated immediately downstream a treatment plant. (b) Subpopulations are defined in Fig. 4b, pooled from three biological replicates taken from each site (all biological replicates are depicted in Supplementary Fig. 18). Statistical analysis of subpopulation sizes is available in Supplementary Fig. 16. (c) Biplot of the redundancy analysis (RDA) based on the fraction of cells/particles in the subpopulations in Fig. 4b constrained by forward selected field physico-chemical parameters (Supplementary Tables 4,6). Dots/grey tones: specific sampling sites. Dispersion of standard error of the weighted scores of sampling sites are shown as ellipses in the respective grey tone (confidence limit = 0.95). Centroids of the subpopulations (MA1a–MA10) are given. Significantly tested model variables are depicted ($***P < 0.001$). Explained variation for the first two constraint axes is given. Tested significance of axes are depicted ($**P < 0.01$, $***P < 0.001$).

not sufficient to drive model selection, they showed a significant fit to the RDA results (Supplementary Fig. 19), indicating that they may be of biological relevance to the distribution of subpopulations. Flow conditions correlated positively with subpopulations MA3–6 associated with green algae and diatoms. MA1–2 (associated with cyanobacteria) and MA7–8 (associated with decaying material) correlated positively with DOC content. Our community structure was thus largely explained by the same environmental parameters as reported

previously for biofilm diatom composition, biomass and chlorophyll-a content^{19,20} showing the biological relevance of the categorized subpopulations. The sites with high sediment load (E and F) correlated with MA10, which exhibits properties of non-organic particles. Notably, it was shown before that deposition of fine sediment can strongly influence the community structure²¹.

Detecting microplastics in field samples. Because it was shown before that viSNE facilitates detection of rare events, we paid special attention to all small viSNE clusters. We discovered that samples from site D, immediately downstream of a waste-water treatment plant, contained a small cluster (Fig. 6a) with fluorescence properties strikingly similar to those of polystyrene beads that we use for FC calibration (Supplementary Fig. 20).

Using a fluorescence-activated cell sorter with the fluorescence gates defined by the small cluster, we isolated ca. 1% of particles from site D, and <0.2% from all other sites. For control, we spiked reference biofilm samples with 13.6 μm polystyrene beads. LM revealed that the isolated fraction from site D

contained mostly non-identifiable particles 1–20 μm in diameter (Supplementary Fig. 21b), while the isolated fractions of the other sites contained mostly broken diatom shells. The fraction isolated from the control contained the polystyrene beads (Supplementary Fig. 21a). The remaining fractions contained the biofilm community (Supplementary Fig. 21c,d). LM data was confirmed by scanning electron microscopy (SEM) (Supplementary Fig. 22a: particles isolated from site D; Supplementary Fig. 22b: control polystyrene beads). The energy spectra of one isolated and one control particle (Fig. 6b,c) showed a high carbon signal and the same oxygen background (energy-dispersive X-ray spectroscopy (EDS; Fig. 6d,e) corresponding to polystyrene EDS-spectra reported previously²². While this is not enough to fully chemically characterize this particular isolated particle, it indicates strongly that it is a microplastic particle.

The small cluster was not detectable 2.5 km downstream (sites E and F) indicating that the contamination of stream biofilms may still be limited. In contrast, in another study all sediment samples taken from rivers Rhine and Main in Germany were found to contain microplastic particles (75% consisting of polystyrene, polyethylene and polypropylene)²³, illustrating

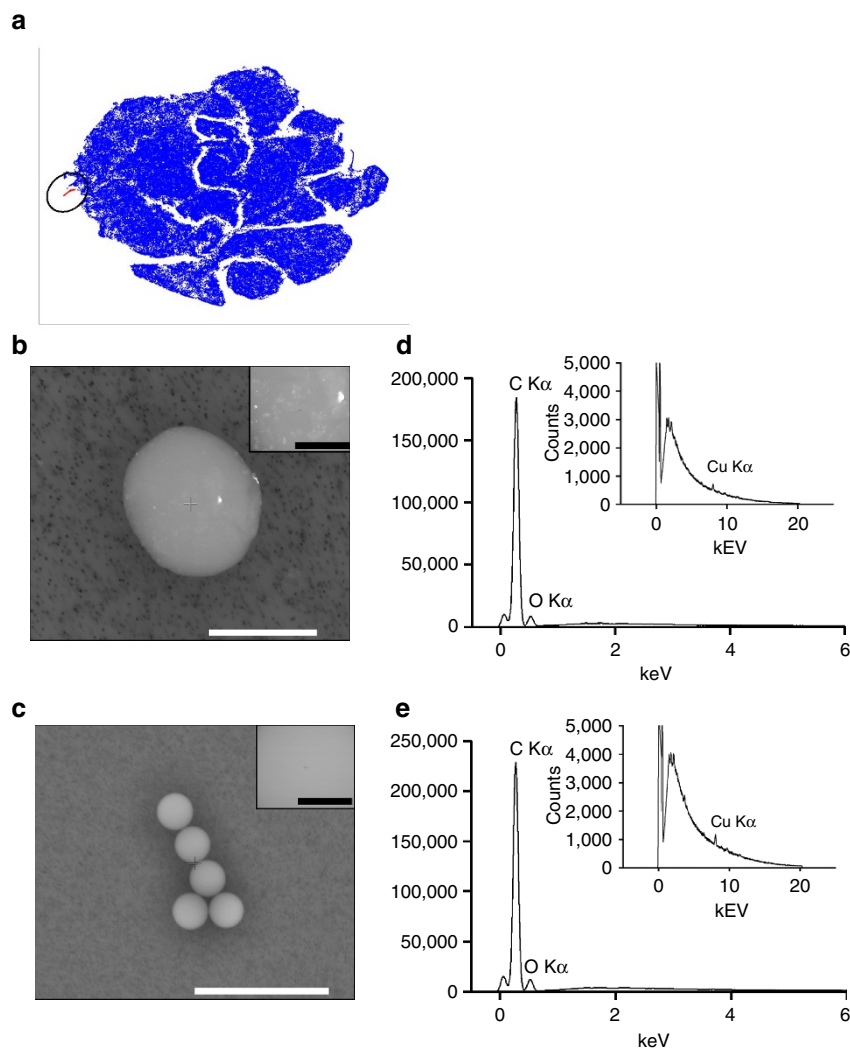


Figure 6 | Detection of microplastics in stream biofilms. (a) viSNE map shown in Fig. 4a with circled hypothetical microplastic cluster. (b) SEM image of a potential microplastic particle isolated from site D. Scale bar, 10 μm . (c) SEM image of polystyrene beads isolated from a spiked sample. Scale bar, 50 μm . The squares in the top right corner of the SEM images (b,c) with separate scale bars, 1 μm , depict the region scanned for EDS analysis. (d) EDS spectrum of a potential microplastic particle isolated from site D. (e) EDS spectrum of polystyrene beads isolated from a spiked sample.

the widespread plastic penetration of the environment and the timeliness of our method.

Discussion

We present a method for the analysis of biofilm community structure, based on stain-free FC and visualization of the acquired high-dimensional single-particle data by viSNE. We built our approach on published viSNE procedures, established for detection of leukaemic cells¹⁵, but optimized viSNE parameters for biofilm analysis. By introducing categorization of subpopulations based on expert knowledge and projection of reference data sets on viSNE submaps, and quantification of the categorized subpopulations, we were able to characterize temporal and spatial differences in biofilm community structures and provide biologically relevant interpretations. Altogether with use of biological and non-biological reference data sets, viSNE enabled detection and interpretation of rare events—in our case microplastic particles. By including all cells and particles irrespective of fluorescence intensity, the method is unique in the combined assessment of community structure, decay of phototrophic organisms and presence of abiotic particles, for instance sediment load and contaminants. This provides insight into the physiological state of biofilms, but also indicates that analysis of low-fluorescent particles such as heterotrophic bacteria and protozoa is feasible. Indeed, in preliminary experiments we were able to characterize three out of six common model protozoa (Supplementary Table 8)²⁴. We determined that the method is robust to subsampling, meaning that our analysis produced the same results if different subsets of cells from the samples were used (Supplementary Fig. 23). The reproducibility of explanatory variables confirms the biological relevance of the categorization strategy.

We tested the redundancy of information provided in different fluorescent channels and determined that 510 nm and 542 nm, 575 nm and 620 nm, 725 nm and 755 nm form redundant pairs, respectively (Supplementary Fig. 24), but that all three lasers were necessary for subpopulation identification. In particular, the blue laser (488 nm) was crucial for identification of the diatom/green algae subpopulation and the decaying cells, the red laser (638 nm) for the identification of the cyanobacterial subpopulation, while in combination with the violet laser (405 nm), it was necessary for the identification of decaying cells and the microplastic cluster. Our approach also proved to be independent of the FC equipment used, as we were able to transfer the FC protocol to a different instrument for isolation of the microplastic particles. However, the method is limited by the size range of particles that can be analysed by FC (here ca. 0.2–200 μm) and the number of particles that can be treated simultaneously by viSNE, which due to crowding is around 150,000 (refs 25,26).

We used the method in settings of increasing complexity, with different target applications. We showed that the method cannot separate between 30 reference species studied, but we were able to detect intraspecies variability and phenotypic shifts in simple communities. We found that larger cells in biofilms are favoured under temperature-stress conditions, indicating the presence of more resilient phenotypes which may be relevant to food-web changes in climate change scenarios. It has been shown before that the interaction of warming with sediment and nutrient load differentially favours smaller or larger diatom growth forms²⁷. We also found that biofilm community structure described by the method robustly correlates with known environmental drivers. We thus propose that prediction of site characteristics and community dynamics based on FC data is possible. Finally, combining viSNE with the newly developed tools enabled the first report of microplastic-like particles being detected directly in a

complex environmental sample and of their presence in stream biofilms. The microplastic accounted for ca 1% of the particles at site D (0.17 % of all analysed particles). The microplastic was robustly detected down to total samples size of 30,000 particles, which corresponds to 5,000 cells per sample (Supplementary Fig. 25).

To evaluate the performance of viSNE, we compared the subpopulations identified with viSNE to clusters returned by the PhenoGraph clustering method, both applied to the same subsamples of the field data. While PhenoGraph detected more clusters, the variability in number of clusters was similar and the groups the clusters/subpopulations belonged to were generally the same (Supplementary Figs 23,26,27). The microplastic cluster was detected in all subsamples by both methods. However, RDA of the viSNE subpopulations was more robust than that of the PhenoGraph clusters in always identifying the same significant drivers of community structure. Combined with the visualization advantage and possibility to project reference data sets onto viSNE maps, we consider our viSNE-based approach to be the best current option to analyse the type of complex flow cytometry data presented.

While this is the first time FC was used for characterization of stream biofilms, it has been used before for phytoplankton analysis (reviewed in ref. 28). Most of these studies were interested in group or species identification^{6,7,9,10}, however, a few have used clustering approaches to sub-group phytoplankton communities at the level of phenotypes (optical scatter and fluorescence intensities), which are similar to our subpopulations⁸. An advantage of the visualization approach enabled by viSNE is that evaluation of the clustering performance and biological interpretation based on expert knowledge and reference data sets can be done simultaneously, making the procedure more user friendly and less time consuming. A related advantage is that evaluation of rare events/small clusters of potential biological significance is easier. In cases where inorganic particle contamination has hindered researchers from quantifying target organisms (for example, see ref. 29), our method can easily separate the respective groups.

As the importance of biofilm community structure to environmental and technical functions has been recognized, there is a need for a fast, reliable and easy-to-use quantification of community structural changes. Here we show that FC combined with our extended viSNE analysis meets these requirements for simple microbial communities and complex natural biofilms. We envisage that the method is easily transferable to other types of biofilms and microbial communities and should therefore be useful in various applications aimed at inferring community structure for environmental and industrial research and monitoring.

Methods

Chemicals. Chemicals were obtained from Sigma-Aldrich and Merck. Polystyrene particle (PS) standards were obtained from BS-Partikel GmbH (13.6 μm LS1400-05, and 19.5 μm LS2000-05).

Algal and cyanobacteria strains. Algal and cyanobacterial strains were obtained from Experimental Phycology and Culture Collection of Algae at the University of Goettingen (EPSAG), Culture Collection of Algae at the University of Cologne (CCAC), Thonon Culture Collection (TCC) and University of Texas Culture Collection of Algae (UTEX). These cultures naturally express benthic and/or planktic, uni- and/or multicellular, some filamentous growth forms, depending on the species. The strains are listed in Supplementary Table 2. All strains were cultured in LA medium, which was specifically developed to provide suitable conditions for all single species and to be close in composition to local surface waters regarding nutrients, P:N ratio, salts and carbonate buffer (Supplementary Table 9). Cultures were exposed to the natural light–dark cycle at $\sim 60 \mu\text{E m}^{-2} \text{s}^{-1}$ and room temperature (20–23 °C) and were not agitated. For long-term culturing, algae were streaked on 15% LA-Agar plates.

Artificial microbial communities. Two different artificial microbial communities were created by mixing *A. minutissimum* (A), *N. palea* (N), *B. braunii* (B) and *C. polonicus* (C) (Supplementary Table 2) to yield A + N or A + N + B + C. Species were mixed in equal numbers ($2 \times 10^5 \text{ ml}^{-1}$ A and N) or equal volumes (1 ml of B and C culture). Cell numbers of A and N were determined by a Casy Cell Counter TTC (Roche Innovatis AG), while B and C were not counted as they are colonial organisms. Single species were cultured as controls. All cultures, in triplicate, were kept in sterilized circular glass microcosms (8 cm diameter, covered by a glass lid) in a final volume of 20 ml LA medium (Supplementary Table 9) at $\sim 60 \mu\text{E m}^{-2} \text{ s}^{-1}$ light and 20–23 °C. For flow cytometric analysis, 1 ml was sampled from the planktonic (suspended) fraction immediately after transfer to the microcosms and after 1 week, while the bottom fraction was removed with a sterile tissue scraper and resuspended in 1 ml of LA medium after one week.

Culturing and sampling of natural biofilms. Biofilms were colonized on glass microscope slides ($\sim 38 \times 26 \text{ mm}$, Thermo Scientific) placed horizontally in Plexiglas channels in a flow-through system fed by water from river Chriesbach (on campus, Dübendorf, Switzerland)^{30,31}. A sediment trap ($0.51 \times 0.7 \times 2.6 \text{ m}$, residence time $\sim 20 \text{ min}$) was used to remove part of the dispersed particles. The flow rate at the inflow of the channels was maintained at $\sim 1 \text{ cm s}^{-1}$ corresponding to a volume flow of 31 min^{-1} . Illumination was provided in 12:12 h light:dark cycles by BioSun fluorescent tubes with a radiation similar to the sunlight spectrum (Radium Lampenwerk GmbH, Germany, ML-T8 36W/965/G13B) at about $80\text{--}100 \mu\text{E m}^{-2} \text{ s}^{-1}$. Temperature and photosynthetically active radiation were monitored with HOBO Pendant Temperature/Light Data Loggers (UA-002-64; median water temperature: 13.4 °C (11.7–16.1 °C), median light intensity in the light period: 1011.8 Lux).

After 3 weeks of colonization (15 October to 5 November 2014), glass slides were randomly placed in five smaller microcosms (Plexiglas ($10 \times 25 \text{ cm}$), 32 glass slides ($3.7 \times 2.6 \text{ cm}$) per microcosm), each connected to a 101 carboy containing 5 l of LA medium (Supplementary Table 9). Room temperature was maintained between 14.9–16.1 °C resulting in median water temperature of 17.75 °C (17.7–17.8 °C). Light conditions and flow rate at the inflow were the same as during colonization (median light intensity in the light period: 947.2 Lux). Oxygen saturation and temperature were monitored in one microcosm (No. 5) with a Presens Microx TX3 system and an NTH-PS1-L2.5-TS-NS40/0.8-YOP-EXT1 oxygen microsensors over 3 weeks (minimum (O_2)/dark period: 7 mg l^{-1} , maximum (O_2)/light period: 7.9 mg l^{-1}) and measured after 21 days in all microcosms. Conductivity, pH and water chemistry were determined in parallel to sampling as described above (Supplementary Table 10). Immediately after transfer to the indoor microcosms, and after 7, 14 and 21 days, 8 slides were taken from each microcosm and biofilms were pooled into 40 ml of LA medium with a plastic scraper. Samples were then prepared for analysis by FC.

Field sampling of stream biofilms. Samples were taken on 18 March 2015 from Mönchaltorfer Aa at six sites (A–F, Supplementary Table 4, Supplementary Fig. 9) already used in previous research³². Mönchaltorfer Aa is affected by different sources of nutrients and pollutants from agricultural and urban sources (Supplementary Table 5). The catchment is situated in the Greifensee basin in the Swiss Plateau and covers an area of 46 km^2 with altitudes ranging from 853 to 445 m a.s.l. Soil types are mostly cambisoles (hillsides) and gley soils (flat areas).

Three stones of similar size from similar microenvironments in the river bed were selected at each sampling site (Supplementary Fig. 28). Biofilms were brushed off the stones with tooth brushes into 15 ml of stream water, which was sampled at the site and filtered through two layers of paper towel. The stones were washed with another 15 ml of filtered stream water. Filtered water was used as control for FC.

Creating pigment-bleached reference samples. Aliquots of 1 ml of biofilm field samples were centrifuged at 23 °C and $8,000g$ for 10 min in a table top centrifuge. Bleaching was performed by resuspending the resulting pellet in 1 ml 90% ethanol (v/v in H_2O). The suspension was vortexed briefly and stored at 4 °C overnight. The procedure was repeated once on the following day.

Sample preparation and FC. Suspensions of reference species and biofilms grown indoors or sampled in the field were sonicated (45 kHz 60 W, VWR Ultrasonic Cleaner) for 1 min to break up colonies, filtered through $50 \mu\text{m}$ filters (Partec) and immediately fixed (0.01% paraformaldehyde and 0.1% glutaraldehyde (w/v, stock in tap water)) at 4 °C overnight.

On the basis of the light absorption and fluorescence spectra of the cultured strains and the sampled biofilms measured with a plate reader, dichroic splitters and filters for the Beckman Coulter Gallios flow cytometer (using 405, 488 and 638 nm lasers) were selected to cover fluorescence emission from 425 to $> 755 \text{ nm}$ (Supplementary Table 11). Voltages of the photomultipliers were adjusted to include at least 99% of all events of the samples taken from the reference species. In total, 36 parameters were measured (FS and SS, 10 fluorescence ranges) as signal area, height and width. Signal area was used for further analysis as height and width did not improve resolution in this case.

Three technical replicates per sample were measured by FC. A total of 3,000 events per strain and 10,000 events per sample were acquired and gated to

exclude all events that were above the signal saturation limit in any of the 12 parameters. The area signal intensity of all parameters was exported to a csv-file for data analysis.

Taxonomic analysis by LM. Fixed samples were sub-sampled three times and were diluted 1:10 in tap water. One ml was transferred to an Uthermol's chamber for microscopic analysis. An inverted microscope (Zeiss Axiovert 135) was used to identify and count (Zeiss EC Plan-Neofluar 40x/0.75 objective, Zeiss EC Plan-Neofluar 100×1.3 Oil objective, if necessary) phototrophic organisms at the level of genera and, if possible, species within three fields of vision. For samples obtained from laboratory experiments, semi-quantitative abundance was described as dominant ('5', > 30 of the cell number counted), frequent ('4', 10–30%), regular ('3', 3–10%), scarce ('2', 1–3%) and sporadic ('1', $< 1\%$) based on two taxonomic references and the recommendations by the Swiss Federal Office for the Environment^{33–35}. For field samples, presence-absence data were documented.

Water analytics. At each field sampling site, spot measures of physical parameters were taken roughly at the same distance above ground as the surfaces of the stones sampled from sites A–F. Water temperature was measured with a DIEHL rigoton thermometer, flow velocity with a Schiltknecht MiniAir2 Micro anemometer (flow accuracy 1.0% fs, 3.0% rdg). Water chemistry was determined in grab samples (500 ml) from each sampling site: Na^+ , K^+ , Ca^{2+} , Mg^{2+} , NO_3^- , SO_4^{2-} and Cl^- content was quantified by ion-chromatography (Metrohm 761 Compact IC, with chemical suppression for NO_3^- , SO_4^{2-} and Cl^-), with a 8 mM HNO_3 /1.197 mM dipicolinic acid solution as mobile phase and a Metrosep C 6–250/4.0 separation column (Metrohm, Na^+ , K^+ , Ca^{2+} , Mg^{2+}) or a Metrosep A Supp 5 100/4 mm column (Metrohm, NO_3^- , SO_4^{2-} and Cl^-) as stationary phase. PO_4^{3-} was quantified colorimetrically (Varian Cary 50 Bio Spectrophotometer) based on the formation of molybdenum blue. Silica was determined colorimetrically based on the reduction of silicomolybdate to silicomolybdous acid in the presence of ascorbic acid using the Autoanalyzer AA3, Bran + Luebbe (Contrec). TOC and DOC were measured with a Shimadzu TOC-L CSH system. LOQ were: $2.5 \text{ mg l}^{-1} \text{ Na}^+$, $1 \text{ mg l}^{-1} \text{ K}^+$, $5 \text{ mg l}^{-1} \text{ Ca}^{2+}$, $2.5 \text{ mg l}^{-1} \text{ Mg}^{2+}$, $0.5 \text{ mg l}^{-1} \text{ NO}_3^-$, $5 \text{ mg l}^{-1} \text{ SO}_4^{2-}$, $0.5 \text{ mg l}^{-1} \text{ Cl}^-$, $1 \text{ mg l}^{-1} \text{ H}_4\text{SiO}_4$, $0.5 \text{ mg l}^{-1} \text{ OC}$.

Fluorescence-activated cell sorting. A MoFlo Astrios cell sorter (Beckman Coulter) was used to sort potentially detected microplastic particles, PS particles and the remaining fraction of the environmental samples for further analysis. The MoFlo Astrios was configured with 4 lasers (355 nm, 488 nm, 561 nm and 640 nm) which were run at 100% power. Ultraviolet-laser (355 nm) was attenuated by the native neutral density filter type 'E'. Specific filter configurations and electronic settings can be seen in Supplementary Table 12. The software settings were calibrated using PS particles, measured with both, the Gallios flow cytometer and the MoFlo Astrios cell sorter. The gating strategy for potential microplastic particles was based on the corresponding viSNE cluster fluorescence intensities as assessed by the Gallios flow cytometer (Beckman Coulter, see above) and the fluorescence characteristics of PS particles. In summary, we used the fluorescence characteristics at 448 nm, 576 nm, 620 nm, 695 nm and 722 nm to sort the potential microplastic particles. Samples from site D and control communities spiked with standard $13.6 \mu\text{m}$ PS particles (at roughly 150:1) were sorted. We used an $80 \mu\text{m}$ tip and ran the fluidics at 60 psi with an average of 470 events per second. The purify mode with a drop envelope of 1–2 was used for sorting. Samples were then analysed by LM and SEM.

SEM and X-ray spectroscopy. The gated fractions obtained from FACS of control field samples spiked with $13.6 \mu\text{m}$ PS particles and of site D samples were analysed by SEM and EDS. Liquid samples were vacuum-filtered onto gold-coated polycarbonate Whatman Nuclepore Track-Etched Membranes (pore size $0.2 \mu\text{m}$, WHA112106 Aldrich). Samples were imaged with a FEI NovaNanoSEM 230 at 15 kV using a gaseous analytical detector. EDS was done at 30 kV using an Oxford instruments X-Max silicon drift detector.

Stochastic network embedding. Stochastic network embedding (SNE) is a nonlinear dimension reduction algorithm that maps data points from a high- to a low-dimensional spaces by minimizing the difference in similarities between points in the high- and low-dimensional spaces. SNE first calculates the pairwise similarity between each multivariate data point in the high-dimensional space (for details, see refs 15,25,36). It then sets a random starting position for each data point in the low-dimensional space and computes the pairwise similarity between the points. The algorithm then, based on gradient descent, iteratively updates the position of points in low-dimensional space until the Kullback–Leibler divergence between the high- and low-dimensional presentations reaches a minimum.

viSNE analysis. In this study, we used the bh-SNE version of SNE³⁶, implemented as viSNE in the cyt software¹⁵ (<http://www.c2b2.columbia.edu/danapeerlab/html/cyt-download.html>, downloaded in January 2015). After importing flow cytometric data into cyt, the data were transformed using hyperbolic arcsin with a cofactor of

150. For each sample or collection of samples, we ran viSNE in cyt (as described in ref. 15), and visualized the results in a 2D scatter plot (viSNE map). For viSNE map creation, all technical and biological replicates were merged into a single sample.

At high number of data points, SNE suffers from the ‘crowding problem’^{15,36}—distant points collapse onto nearby areas of the map, creating a large single dense region with very little separation between points. To avoid this problem, we kept our sample sizes below 150,000 data points²⁶.

Subpopulation identification and quantitative analysis. For comparative analysis of different field and laboratory samples, viSNE maps were created from a mixture of equal-sized samples, for example, the viSNE map presented in the paper (Fig. 4a) was constructed from 25,000 data points taken from each of the six fields sites (150,000 data points in total). From the viSNE maps, we identified subpopulations of interest based on the visual separation between regions of the map and the distribution of the flow cytometry markers (expert knowledge). For quantification, we counted the numbers of particles/cells that belonged to each subpopulation in each of our samples.

Quantifying similarity between viSNE maps. We quantified the difference between viSNE maps with the MMD³⁷. MMD is defined as the largest difference in expectation over functions in the unit ball of a reproducing kernel Hilbert space (RKHS). It is defined between two distributions as:

$$MMD[\mathcal{F}, X, Y] = \left(\frac{1}{m^2} \sum_{i,j=1}^m k(x_i, x_j) - \frac{1}{nm} \sum_{i,j=1}^{m,n} k(x_i, y_j) + \frac{1}{n^2} \sum_{i,j=1}^n k(y_i, y_j) \right)^{1/2}$$

where \mathcal{F} is a unit ball in a universal RKHS, with an associated continuous kernel k , while m and n are the number of samples from each distribution. For more details on the MMD algorithm, please see ref. 37.

To enable better comparison between similarities between different viSNE map pairs, all MMD calculations were made on viSNE maps containing 3,000 data points.

Projecting data points onto a viSNE map. We projected data points, which were not used for viSNE map creation, onto viSNE maps based on similarity in the 12-dimensional space (FS, SS, 10 fluorescences). For each projected point, we determined the viSNE data point closest to it by Euclidean distance and then drew the projected point on the viSNE map at corresponding coordinates.

Clustering with PhenoGraph. As an alternative to viSNE-based subpopulation analysis, we used the PhenoGraph clustering method³⁸ on the same samples as used for viSNE. The number of nearest neighbours was set 45, and the algorithm was run 50-times for each sample. The runs with the highest modularity³⁸ was used for comparison with viSNE. To visualize and compare the PhenoGraph clusters with the viSNE clusters, we projected the PhenoGraph clusters to the viSNE map.

Redundancy analysis. Statistical analysis was carried out using the vegan and multcomp packages in R^{39–41}. Identified subpopulations of the field sampling and the temperature increase experiment were analysed using RDA based on selected physico-chemical variables⁴². In particular, we performed model selection by forward, backward and marginal selection of the variables. Variables that were selected at least twice were tested for their variation inflation factors to further identify redundant constraints. The final RDA model, the constraint axes and the physico-chemical constraints were then tested by permutation tests (999 permutations)⁴³. *A priori* vector fitting and generalized additive models for all physico-chemical variables was performed on the first two RDA axes to test their relation (r^2) to the subpopulations structure as depicted in the biplot⁴⁴. RDA models are based on Hellinger transformed subpopulation data and logarithmic transformed physico-chemical data (except pH)⁴⁵. *A priori* variable fitting was performed with raw physico-chemical data.

Statistics. Differences between subpopulations of the field sites and the different sampling dates of the laboratory experiment were assessed by analysis of variance followed by Tukey’s test with Holmes corrected P values for multiple comparisons.

Data availability. The raw flow cytometry data (in fcs and csv format) are available from Dryad with the identifier doi:10.5061/dryad.62fv1 (ref. 46). Scripts for comparison and post-processing of viSNE maps, RDA and statistical analysis are available from GitHub (https://github.com/anzezupanic/FC_analysis, https://github.com/RemoFreimann/SubPop_analysis.git). The authors declare that all other data supporting the findings of this study are available within the article and its Supplementary Information files.

References

1. Allan, J. D. & Castillo, M. M. in *Stream Ecology: Structure and function of running waters* (Springer, 2007).

- Brenner, K., You, L. & Arnold, F. H. Engineering microbial consortia: a new frontier in synthetic biology. *Trends. Biotechnol.* **26**, 483–489 (2008).
- Wagner, M. & Loy, A. Bacterial community composition and function in sewage treatment systems. *Curr. Opin. Biotechnol.* **13**, 218–227 (2002).
- Wahl, M., Goecke, F., Labes, A., Dobretsov, S. & Weinberger, F. The second skin: ecological role of epibiotic biofilms on marine organisms. *Front. Microbiol.* **3**, 292 (2012).
- Battin, T. J. *et al.* Biophysical controls on organic carbon fluxes in fluvial networks. *Nat. Geosci.* **1**, 95–100 (2008).
- Becker, A., Meister, A. & Wilhelm, C. Flow cytometric discrimination of various phycobilin-containing phytoplankton groups in a hypertrophic reservoir. *Cytometry* **48**, 45–57 (2002).
- Dubelaar, G. B. J., Geerders, P. J. F. & Jonker, R. R. High frequency monitoring reveals phytoplankton dynamics. *J. Environ. Monit.* **6**, 946–952 (2004).
- Pomati, F., Jokela, J., Simona, M., Veronesi, M. & Ibelings, B. W. An automated platform for phytoplankton ecology and aquatic ecosystem monitoring. *Environ. Sci. Technol.* **45**, 9658–9665 (2011).
- Read, D. S., Bows, M. J., Newbold, L. K. & Whiteley, A. S. Weekly flow cytometric analysis of riverine phytoplankton to determine seasonal bloom dynamics. *Environ. Sci. Process. Impacts* **16**, 594–603 (2014).
- Wilkins, M. F., Boddy, L., Morris, C. W. & Jonker, R. R. Identification of phytoplankton from flow cytometry data by using radial basis function neural networks. *Appl. Environ. Microbiol.* **65**, 4404–4410 (1999).
- Estrada, M., Henriksen, P., Gasol, J. M., Casamayor, E. O. & Pedrós-Alió, C. Diversity of planktonic photoautotrophic microorganisms along a salinity gradient as depicted by microscopy, flow cytometry, pigment analysis and DNA-based methods. *FEMS. Microbiol. Ecol.* **49**, 281–293 (2004).
- Aghaeepour, N. *et al.* Critical assessment of automated flow cytometry data analysis techniques. *Nat. Methods* **10**, 228–238 (2013).
- Lugli, E., Roederer, M. & Cossarizza, A. Data analysis in flow cytometry: the future just started. *Cytometry Part A* **77A**, 705–713 (2010).
- Glüge, S., Pomati, F., Albert, C., Kauf, P. & Ott, T. in *Nonlinear Dynamics of Electronic Systems*. (eds Mladenov, V. M. & Ivanov, P. C.) (Springer, 2014).
- Amir E.-a.D. *et al.* viSNE enables visualization of high dimensional single-cell data and reveals phenotypic heterogeneity of leukemia. *Nat. Biotech.* **31**, 545–552 (2013).
- Azim, M. E., Verdegem, M. C. J., van Dam, A. A. & Beveridge, M. C. M. *Periphyton: Ecology, Exploitation and Management* (CABI, 2006).
- Satoh, Y. & Hama, T. Stepwise alteration from fluorescent to non-fluorescent chlorophyll derivatives during early diagenesis of phytoplankton in aquatic environments. *J. Exp. Mar. Bio. Ecol.* **449**, 36–44 (2013).
- Gargiulo, G. M., Culosio, F., Genovesi, G. & De Masi, F. Karyology and effects of temperature and photoperiod on the life- history of *Bangia atropurpurea* (Roth) C. Ag. (Bangiales, Rhodophyta) from the Mediterranean Sea. *Cryptogamie: Algologie* **17**, 45–55 (1996).
- Leonard, G. H., Levine, J. M., Schmidt, P. R. & Bertness, M. D. Flow-driven variation in intertidal community structure in a Maine estuary. *Ecology* **79**, 1395–1411 (1998).
- Schönfelder, I., Gelbrecht, J., Schönfelder, J. & Steinberg, C. E. W. Relationships between littoral diatoms and their chemical environment in northeastern German lakes and rivers. *J. Phycol.* **38**, 66–89 (2002).
- Moss, B. Adaptations of epipelic and epipsammic freshwater algae. *Oecologia* **28**, 103–108 (1977).
- Lee, H. J., Cho, W. & Oh, M. Advanced fabrication of metal-organic frameworks: template-directed formation of polystyrene@ZIF-8 core-shell and hollow ZIF-8 microspheres. *Chem. Commun.* **48**, 221–223 (2012).
- Klein, S., Worch, E. & Knepper, T. P. Occurrence and spatial distribution of microplastics in river shore sediments of the Rhine-Main area in Germany. *Environ. Sci. Technol.* **49**, 6070–6076 (2015).
- Altermatt, F. *et al.* Big answers from small worlds: a user’s guide for protist microcosms as a model system in ecology and evolution. *Methods Ecol. Evol.* **6**, 218–231 (2015).
- van der Maaten, L. & Hinton, G. Visualizing data using t-SNE. *J. Mach. Learn. Res.* **9**, 2579–2605 (2008).
- DiGiuseppe, J. A., Tadmor, M. D. & Pe’er, D. Detection of minimal residual disease in B lymphoblastic leukemia using viSNE. *Cytometry Part B: Clin. Cytometry* **88**, 294–304 (2015).
- Piggott, J. J. *et al.* Climate warming and agricultural stressors interact to determine stream periphyton community composition. *Global Change Biol.* **21**, 206–222 (2015).
- Czechowska, K., Johnson, D. R. & van der Meer, J. R. Use of flow cytometric methods for single-cell analysis in environmental microbiology. *Curr. Opin. Microbiol.* **11**, 205–212 (2008).

29. Schiaffino, M., Gasol, J., Izaguirre, I. & Unrein, F. Picoplankton abundance and cytometric group diversity along a trophic and latitudinal lake gradient. *Aquat. Microbial Ecol.* **68**, 231–250 (2013).
30. Kroll, A., Behra, R., Kaegi, R. & Sigg, L. Extracellular polymeric substances (EPS) of freshwater biofilms stabilize and modify CeO₂ and Ag nanoparticles. *PLoS ONE* **9**, e110709 (2014).
31. Navarro, E., Robinson, C. T., Wagner, B. & Behra, R. Influence of ultraviolet radiation on UVR-absorbing compounds in freshwater algal biofilms and *Scenedesmus vacuolatus* cultures. *J. Toxicol. Environ. Health-Part A: Curr. Issues* **70**, 760–767 (2007).
32. Robinson, C. T., Schuwirth, N., Baumgartner, S. & Stamm, C. Spatial relationships between land-use, habitat, water quality and lotic macroinvertebrates in two Swiss catchments. *Aquat. Sci.* **76**, 375–392 (2014).
33. Hürlimann, J. *Methoden zur Untersuchung und Beurteilung der Fließgewässer. Kieselalgen Stufe F (flächendeckend)* (Bundesamt für Umwelt, 2007).
34. Hustedt, F. *Bacillariophyta (Diatomeae)* 1930 edn. (Fischer, 1976).
35. Pascher, A., Büdel, B. & Ettl, H. in *Süßwasserflora von Mitteleuropa* (Fischer, 1987–1997).
36. van der Maaten, L. Accelerating t-SNE using tree-based algorithms. *J. Mach. Learn. Res.* **15**, 3221–3245 (2014).
37. Gretton, A., Borgwardt, K. M., Rasch, M. J., Schölkopf, B. & Smola, A. A kernel two-sample test. *J. Mach. Learn. Res.* **13**, 723–773 (2012).
38. Levine, Jacob H *et al.* Data-driven phenotypic dissection of AML reveals progenitor-like cells that correlate with prognosis. *Cell* **162**, 184–197 (2015).
39. Oksanen, J. *et al.* *vegan: Community Ecology Package*. R package version 2.2-1. <http://CRAN.R-project.org/package=vegan> (2015).
40. Hothorn, T., Bretz, F. & Westfall, P. Simultaneous inference in general parametric models. *Biomet. J.* **50**, 346–363 (2008).
41. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. <http://www.R-project.org/> (2013).
42. Blanchet, F. G., Legendre, P. & Borcard, D. Forward selection of explanatory variables. *Ecology* **89**, 2623–2632 (2008).
43. Legendre, P., Oksanen, J. & ter Braak, C. J. F. Testing the significance of canonical axes in redundancy analysis. *Methods Ecol. Evol.* **2**, 269–277 (2011).
44. Wood, S. N. Thin plate regression splines. *J. R. Stat. Soc. B Stat. Methodol.* **65**, 95–114 (2003).
45. Legendre, P. & Gallagher, E. Ecologically meaningful transformations for ordination of species data. *Oecologia* **129**, 271–280 (2001).
46. Sgier, L., Freimann, R., Zupanic, A. & Kroll, A. Data from: flow cytometry combined with viSNE for the analysis of microbial biofilms and detection of microplastics. *Dryad Digital Repository* <http://dx.doi.org/10.5061/dryad.62fv1> (2016).

Acknowledgements

This study was supported by a SNF Ambizione Fellowship (PZ00P2_142533) to A.K. We thank Michelle Tadmor from the cyt support team for advice on using their software, Renaud Delmelle and Andreas Borgschulte (EMPA) for performing the SEM-EDS analysis, Christopher Robinson (Eawag) for advice on site selection and discussion of the field results, Rosi Siber (Eawag) for providing land use data for Mönchaltorfer Aa catchment, Esther Keller-Koch and Florian Altermatt (Eawag) for providing protozoa culture and Kristin Schirmer, Renata Behra, Anita Narwani and Francesco Pomati (Eawag) for helpful discussions and critical reading of the manuscript.

Author contributions

A.K., L.S. and A.Z. planned the study, A.K. and L.S. planned the experiments, R.F., A.K. and L.S. performed the experiments, all authors performed parts of the data analysis and wrote the paper.

Additional information

Supplementary Information accompanies this paper at <http://www.nature.com/naturecommunications>

Competing financial interests: The authors declare no competing financial interests.

Reprints and permission information is available online at <http://npng.nature.com/reprintsandpermissions/>

How to cite this article: Sgier, L. *et al.* Flow cytometry combined with viSNE for the analysis of microbial biofilms and detection of microplastics. *Nat. Commun.* **7**:11587 doi: 10.1038/ncomms11587 (2016).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>