Imaging flow cytometry enables label-free cell sorting of morphological variants from unculturable bacterial populations.

Keywords: autofluorescence, imaging flow cytometry, cell sorting, non-model system, *Pasteuria ramosa*

Authors

Daniel Vocelle[1], Lindsey Thompson[2], Matthew P. Bernard[1] & Nina Wale[12,3,4,+]

1. Department of Pharmacology & Toxicology, Michigan State University, East Lansing, MI 48824, USA
2. Department of Microbiology, Genetics & Immunology, Michigan State University, East Lansing, Michigan, USA 48824
3. Department of Integrative Biology, Michigan State University, East Lansing, Michigan, USA 48824
4. Program in Ecology, Evolution & Behavior, Michigan State University, East Lansing, Michigan, USA 48824

+corresponding author

Contributions

DV and NW conceptualized the research.

DV and LT performed the experiments.

DV designed the analysis, with consultation from MB.

DV and NW analyzed the data.

NW wrote the manuscript.

DV, LT, MB edited the manuscript.

Data availability

Description: data will be made available upon article acceptance.

Abstract

Bacterial populations often display remarkable morphological heterogeneity. Flow cytometric cell sorting (often called FACS) is an important tool for understanding this diversity. FACs allows researchers to obtain pure samples of each morphological variant (or morphotype) that is present within a mixed population of cells and thus permits each morphotype to be phenotyped. In FACS, cells are first labeled with fluorescent markers such as antibodies or transgenic constructs, and then separated out based on their possession of these labels. However, since the development of fluorescent labels requires *a priori* knowledge of bacterial biology, it is often impossible to apply FACS to understudied and/or unculturable bacteria. This challenge has limited our capacity to investigate the biology of bacterial size and shape in all but a small, largely culturable subset of bacterial taxa. Here, we present an innovative strategy that permits label-free cell sorting of bacterial morphotypes, using an unculturable, pleiomorphic pathogen (*Pasteuria ramosa*) as a model bacterium. We show that imaging flow cytometry (IFC) can be used to systematically identify light-scattering and autofluorescence “signatures” of bacterial morphotypes, on which basis cell sorting can be conducted. Critically, our IFC-enabled cell sorting strategy yields samples of sufficient purity (>90%) for common downstream analyses e.g., “-omics” analyses. Our work represents an innovative application of IFC and provides an economical, widely applicable solution to a central problem in the study of bacterial diversity.

Importance

Bacteria come in many different shapes and sizes. Why this morphological variation exists is a longstanding question in microbiology, but it remains difficult to answer. To phenotype different morphological variants (morphotypes) within a bacterial population, we need to separate them from one another. This is normally achieved using flow cytometric cell-sorting, whereby morphotypes are labelled with fluorescent antibodies and separated on the basis of their differential fluorescence. Unfortunately, it is difficult to develop fluorescent-labels specific to unculturable or poorly studied bacteria because of the limited availability of appropriate molecular tools. Here, we demonstrate that imaging flow cytometry can be used to design and validate label-free cell sorting strategies. Recently, there has been a resurgence of interest in bacterial morphological diversity and a call to expand its study across the tree of life. Our work will help microbiologists to answer this call.

Introduction

Bacterial populations often contain cells varying in size and shape, and this morphological heterogeneity can have a profound impact on bacterial fitness (1–3). For example, different forms of bacteria (morphotypes) may perform distinct ecological “roles” or exhibit differential susceptibility to environmental stress, permitting population persistence in the face of such stress (4–7). To fully understand the causes and consequences of morphological variation within bacterial populations, we need to extend our investigations beyond well-studied taxa and *in vitro* environments, which can suppress morphological diversity within a bacterial population (2, 8–10). Here, we present imaging flow cytometry (IFC)-enabled cell sorting, a novel procedure that allows for the purification of distinct morphotypes of unculturable bacteria living in complex milieu.

Efforts to understand the mechanistic basis and functional significance of morphological heterogeneity in a bacterial population often begin with isolating distinct bacterial morphotypes from each other and/or their environmental milieu, via flow cytometric cell sorting (or, colloquially, fluorescence activated cell sorting, FACS). However, the application of FACS to unculturable bacteria or those found in complex samples (e.g., soil, host tissues) is a challenge. A chicken-and-egg-scenario exists. In FACS, different cell populations are tagged with distinct fluorescent labels (e.g., fluorescently labelled antibodies) on whose basis the populations are separated. However, the development of morphotype-specific labels requires the characterization of morphotype-specific antigens, which itself requires that morphotypes be separated and studied. In the absence of morphotype specific labels, ‘general’ membrane permeable labels (e.g., CFSE, Mitotracker, Hoechst), which bind to molecules present in most cells (e.g., amines, DNA/RNA, or mitochondria), could be used to discriminate morphotypes in

102 certain contexts (11). However, because general labels indiscriminately permeate all cells within
103 a sample, they are unlikely to be helpful in distinguishing a specific morphotype if it is present in
104 a complex, cell-rich milieu such as host-tissue or soil.
105
106 A potential solution to the problem of sorting bacterial morphotypes in the absence of labels is to
107 exploit intrinsic variation in their light-scattering and -fluorescence properties (i.e., "light
108 signatures"). Cells differing in size and granularity differentially scatter light, and the degree of
109 autofluorescence may also vary among morphotypes (12). For example, viable *Bacillus*
110 *anthracis* spores exhibit a distinctive autofluorescent signature upon excitation with UV light
111 (13). Recent advances in imaging flow cytometry (IFC) simplify the identification of intrinsic light
112 signatures of morphologically distinct cells by generating images of each cell and measures of
113 their light-scattering and autofluorescence properties. Systematic analysis of these data allows
114 for the identification of morphotype-specific light signatures and has been used to identify
115 morphologically distinct subpopulations of mammalian and bacterial cells (14–17).
116
117 Here, using the pleiomorphic bacterial pathogen *Pasteuria ramosa* (hereafter, *Pasteuria*) as a
118 model, we demonstrate that IFC can be used to efficiently design and validate strategies for
119 label-free sorting of bacterial morphotypes. *Pasteuria* is a spore-forming pathogen of
120 zooplankton that exhibits striking morphological variation during its life cycle (Table 1, (18)).
121 Longstanding hypotheses vis-à-vis the "division of labor" among different morphotypes within
122 the *Pasteuria* bacterial population (19) remain untested, largely because it is difficult to separate
123 bacteria from host material for functional -omics and *in vitro* study. By combining state-of-the-art
124 IFC instrumentation, which allows us to rapidly derive thousands of images of bacterial cells,
125 and machine-learning algorithms, we systematically designed and validated autofluorescence-
126 based FACS strategies that yield high-purity samples of each of four distinct morphotypes. Our

work provides proof of principle that IFC-enabled sorting can be used to overcome the first barrier to understanding morphological heterogeneity in bacterial species that cannot be readily labeled.

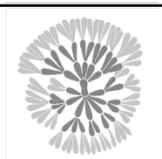
	name (abbreviation)	**Function** (*hypothesized)	**largest dimension** (range, µm)
	Activated (A)	host attachment	-
	`cauliflower' (C)	in-host growth, host manipulation*	4-6.7
	`grape-seed' (G)	sporulation	4-6
	mature spore (S)	transmission, environmental persistence	4.2-5.4

Table 1. Variation in the form and function of *Pasteuria ramosa* morphotypes. Form and function information extracted from (18, 20, 21). Illustrations are schematic representations of morphotypes and are not to scale. *Indicates function is hypothesized.

Materials and methods

Our goal was to design and validate cell sorting strategies suitable for the purification of 4 distinct *Pasteuria* morphotypes from the homogenate of an infected host i.e., a sample containing several *Pasteuria* morphotypes, host cells, food (algae) present in the host gut, and host microbiota. Below we describe our approach in general terms. The final procedure and results of our validation steps are described in the Results.

164

165 **Host & Parasites**

166 Throughout, we used *Daphnia magna* (HuHO-2 genotype) hosts and *Pasteuria* (strain C1),
167 originally collected in Hungary and Russia, respectively (22, 23). Both were generously donated
168 by D. Ebert (U. Basel, Switzerland).

169

170 To generate samples of *Pasteuria*, we experimentally infected *Daphnia magna*. Throughout, we
171 used a standard infection procedure. Briefly, we shook *Pasteuria* spores at 1500 rpm, overnight.
172 On day 0 of the experiment, we filled a vessel (e.g., 24 well-plate or beaker) with media (ADaM
173 or COMBO), added 1-10 *D. magna* <24 hours old and ~37,500 *Pasteuria* spores per host (24).
174 On the following day, we added a second equivalently sized dose of spores. We fed hosts
175 increasing quantities of either *Scenedesmus obliquus* or *Ankistrodesmus falcatus* as they aged,
176 following (25), and maintained them for 30-60 days. To diagnose *Pasteuria* infection, we
177 assessed hosts by eye for signs of infection, including discoloration, lack of offspring and
178 enlargement (26), and then confirmed it using microscopy.

179

180 **Instruments**

181 We used a Thermofisher Attune CytPix flow cytometer (cytometer, hereafter) for IFC. This
182 cytometer is equipped with a high-speed, brightfield camera (20x objective, 0.45 numerical
183 aperture, 0.3 micron/pixel resolution), four lasers and 14 filters (Table S1). Acoustic focusing
184 allows for the alignment of cells and particles (hereafter, events) in a narrow depth of field. As a
185 result, this machine can acquire in-focus images an order of magnitude faster than cytometers
186 with hydrodynamic-focusing (27).

187 For cell sorting, we used a BD Influx equipped with 5 lasers and 12 filters (Table S1). The
188 default measure of fluorescence on the BD Influx is “pulse height” (rather than area). Since our
189 goal was to use the imaging cytometer to identify fluorescence parameters that could be
190 employed in cell sorting, we used pulse height as our measure of fluorescence on both
191 machines.

192

193 **Data acquisition**

194 Immediately following diagnosis, we prepared infected animals for flow cytometric analysis.
195 First, each animal was washed in PBS (no Ca+ or Mg+). We transferred 1-4 animals, depending
196 on the degree of infection and an initial assessment of morphotype composition via microscopy,
197 into a 1.5mL Eppendorf tube containing PBS. We homogenized the animals briefly with an
198 electric pestle (~20 seconds, to avoid damaging cells) to create a slurry. We diluted the sample
199 in 500mL Focusing Buffer (Thermofisher Scientific, catalog no. A24904) or PBS, which have
200 equivalent refractive indices. We ran samples at 12.5mL/min until we acquired 30,000 events.
201 Between runs, we twice flushed the cytometer line to prevent sample carryover.

202

203 **Pre-sort analysis: initial gating & image-processing**

204 We first excluded (“gated out”) algae on the basis of their chlorophyll autofluorescence (high
205 emission in channel R3, see Table S1). We then selected (“gated in”) the events that (i) were
206 successfully photographed and processed and, furthermore, contained only one particle. This
207 yielded a dataset of single events, each of which was associated with an image and a set of
208 autofluorescence measurements.

209

210 To analyze the image data, we employed inbuilt Attune Image-Processing Software (version
211 6.0.1). We used the software’s bead masking model (Beads_Only_Full_Resolution_v1) which is

212 optimized for objects 2-5mm. We chose this algorithm because (i) an alternative model
213 optimized for objects >5mm, excluded ~70% of our target objects and (ii) because we found that
214 the bead model could process a variety of *P. ramosa* morphotypes, including those that
215 exceeded 5mm (e.g., cauliflowers). Image-processing yielded a matrix containing the values of
216 22 imaging parameters including a variety of size and "texture" metrics (Table S2). In total,
217 including the fluorescent parameters (i.e., pulse height, area, and width) each event was
218 associated with 74 parameters. Before conducting further analysis, we rescaled each parameter
219 linearly between 0 and 1.
220
221 We visualized the data using Uniform Manifold Approximation Projection for Dimension
222 Reduction (UMAP) (28), implemented in FCS Express (De Novo Software). Briefly, this
223 algorithm allows for the visualization of high-dimensional data in a 2D plane. By inspecting
224 distinct regions of the UMAP, and the imaging data with which it was associated, we found that
225 different morphotypes clustered in distinct areas of the UMAP. Preliminary analysis of this kind
226 revealed that certain image parameters were particularly informative for distinguishing different
227 morphotypes (Table S2). We employed this subset of parameters in further analyses, described
228 below.
229
230 **Pre-sort analysis: identifying morphotype-specific parameters for cell sorting**
231 For each sample, we used IFC to identify gates that would enable us to sort a specific
232 morphotype ("target morphotype", hereafter) from the remainder of the sample.
233
234 First, we visualized the morphological complexity of the sample using a UMAP analysis of all the
235 scattering and fluorescent parameters, as well as several imaging parameters identified in
236 preliminary analyses as being most-informative (Table S2, S3). We then performed cluster

237 analysis to identify (i) a primary gate, on which basis we could distinguish our target morphotype
238 from other bacterial cells, and (ii) subsequent gate(s), on which basis we could distinguish our
239 target morphotype from debris (see Results). Specifically, we used FlowSOM - a neural
240 network-like algorithm that identifies clusters of events characterized by distinct (combinations
241 of) parameters (29). We set the consensus clustering parameter to between 8-20, depending on
242 the sample, and otherwise used default settings (Table S3).
243
244 The above-described procedure was repeated on a sample-by-sample basis. That is, the
245 parameters used to define the gates varied among samples (particularly in the initial stages of
246 our work, Table S4). Due to differences in the autofluorescence of cells among samples, the
247 exact position of the gates varied from sample to sample. Importantly, however, the gate -
248 identification and -validation strategy - which is our focus here - remained the same across
249 samples.
250
251 **Cell sorting**
252 Prior to cell sorting, we transferred the sample to a 5 mL FACS tube (Corning, catalog no.
253 352063) and filtered it through a 100mm mesh filter. We then used the gating strategy designed
254 using IFC (or a minor variation of the strategy, see Results, Fig. S1), to sort each of the
255 morphotypes in the sample that were of sufficient abundance (see Table S4 for number of
256 events sorted per sample).
257
258 **Post-sort analysis: quantification of sample purity**
259 We verified the purity of each sorted sample using IFC. Briefly, after running the sorted sample
260 on the imaging cytometer, the data was first processed as described above and then “down-
261 sampled” to a subset of ~500 events (with a few exceptions, see Table S4). Using the image
262 data, we manually classified each of the events in the downsampled dataset as an activated-,

263 cauliflower-, grape- or spore- morphotype of *Pasteuria* (Table 1) or as unidentified debris.
264 Finally, we calculated the purity of the sorted sample i.e., the proportion of total events in the
265 down-sampled dataset that were of the target morphotype. This purity measure is conservative
266 given that we include unidentified debris such as salts in our total event count.
267

268 Results

269 **IFC enables the identification of cell sorting strategies that exploit morphotype-specific**
270 **light signatures**

271 We established a robust procedure for identifying light signatures that can be harnessed for the
272 cell sorting of morphologically distinct bacterial morphotypes, in the absence of labels. To
273 demonstrate this approach, we provide a worked example (Fig. 1). In this example, our goal
274 was to sort a spore (S) population from a sample comprising host (*Daphnia*) material, host-
275 microbiota and non-target morphotypes of *Pasteuria* (cauliflower (C), grape (G)). Worked
276 examples of the procedure as applied to activated, cauliflower and grape stages are provided in
277 the supplement (Fig. S1).
278

279 First, we performed a dimensionality reduction (UMAP) analysis to visualize the data in 2D
280 space (Fig. 1A). To identify where in this 2D space each *Pasteuria* morphotype lay, we imposed
281 the results of a cluster analysis onto the UMAP (Fig. 1A, colors/numbers indicate clusters). This
282 cluster analysis was performed using all light-scattering and -fluorescence parameters and a
283 subset of the imaging parameters (Table S2). We manually classified each cluster as either a
284 morphotype-rich or debris-rich cluster using the image data (Fig. 1B). For example, since spores
285 appeared in most of the images associated with cluster 3, we designated cluster 3 the spore-
286 rich cluster (S-cluster; Fig. 1B, purple outline); similarly, clusters 1 and 2 were designated
287 grape- and cauliflower- rich clusters, respectively.

288

289 Next, we used the cluster analysis to identify light-scattering and fluorescence parameters that
290 characterized our target morphotype vs. the other morphotypes (C, G). Analysis of the median
291 parameter values associated with each of the morphotype clusters (Fig. 1Ci), suggested that the
292 different morphotypes could be broadly distinguished by the light scattering parameters FSC
293 and SSC (Fig.1Ci, F,S), consistent with their differences in size and granularity (Table 1). The
294 distribution of G-, C- and S- cluster events in FSC-SSC space confirmed this inference (Fig.
295 1Cii, Ciii). We thus used FSC-SSC to define our primary gate ("S1", Fig. 1Cii).

296

297 Inspection of the images associated with the events in the primary gate revealed that it included
298 debris as well as our target morphotype (Fig. 1Di). We therefore sought to identify a secondary
299 gate that could be used to distinguish the target morphotype from debris. First, we inspected the
300 images associated with the events in the target cluster, S1. This analysis revealed that there
301 was a gradient within the cluster: on the left lay a largely homogenous population of spores (Fig.
302 1Dii, region A, top images); whereas on the right we found a mixed population of spores and
303 other non-target objects (Fig. 1Dii, region B; bottom images). To identify which of the
304 fluorescence parameters shared by the sorter and the cytometer (Table S1) could differentiate
305 events in regions A and B we recolored the events in each cluster by their fluorescence
306 intensity in each of the shared channels (Fig. Diii, 4 parameters shown for readability; all
307 parameters are displayed in Fig. S2). This analysis revealed that V1 could distinguish the spore-
308 rich region A from the mixed region B and could therefore be used in combination with FSC to
309 define a secondary gate. Finally, we established the position of this secondary gate ("S2") by
310 plotting the events in our primary gate in bipartite FSC-V1 space (Fig. 1Div). We inspected the
311 images of the events in this space (Fig. 1Div, left - images of events within dotted rectangle;

312 right - images of events within solid rectangle) and refined the gate so that it included only
313 events of the target morphotype (Fig. 1Div, solid rectangle).
314
315 We then performed cell sorting using the gating strategy identified via IFC. As anticipated, the
316 sample exhibited the same qualitative, but not quantitative, pattern on the sorter as on the
317 imaging-cytometer (Fig. 1, compare. Ei vs. Ciii). We thus positioned our gates in the same
318 position, relative to the distribution of events, on the sorter as on the cytometer. Interestingly, we
319 found that UV1 – a parameter available on the cell sorter but not the cytometer – could better
320 discriminate spores from debris than V1 (Fig. S3). In this worked example, as well as other
321 analyses where spores were the target morphotype, we used UV1 rather than V1 to define our
322 secondary gate. With the gates set, we sorted the sample.
323
324 Finally, to validate our sorting strategy and quantify the purity of the sorted sample, we analyzed
325 a subset of the sorted sample via IFC. First, we confirmed that the sorted events occupied the
326 expected region of FSC-SSC space i.e., that the sorted sample occupied the same position in
327 FSC-SSC space as the S-cluster (compare Fig. 1, Ciii, 1Fi). Second, we confirmed the identity
328 of the sorted events using the image data (Fig. 1Fii).
329
330 **IFC-enabled sorting yields high-purity populations of each morphotype**
331 We used the above-described gate-discovery and validation strategy to conduct label-free cell
332 sorting on 43 occasions. These analyses involved 24 samples varying in initial composition and
333 targeted each of the four different *Pasteuria* morphotypes.
334
335 The sorted samples ranged in purity from 75% to 99% (Fig. 2A). Of the events that were not
336 identifiable as the target morphotype ("non-target events"), the majority were debris (e.g., salt

337 crystals, host material, Fig. 2B). Indeed, on average the sorted sample constituted only 1.8%
338 (range 0-8.4%) of *Pasteuria* morphotypes that were not the target (Fig. 2B).
339

340 # Discussion

341 The first step toward understanding the physiological and ecological significance of
342 morphological variation in a bacterial population is to obtain a pure sample of each morphotype.
343 FACS is commonly employed for cell purification but its application to poorly studied or difficult-
344 to-culture bacteria is limited by a paucity of morphotype-specific fluorescent-labels (antibodies,
345 transgenics). Here, we demonstrate that IFC can be used to design and validate a label-free cell
346 sorting strategy that efficiently separates distinct bacterial morphotypes from each other and
347 from environmental debris. Critically, this procedure yields samples of a purity required for
348 common downstream, "-omics" analyses i.e., >=90%. High-speed image-enabled cell sorting
349 approaches have been innovated (30, 31) and are slowly becoming available, but their expense
350 and limited availability makes them out of reach for most. Our methodology represents an
351 immediately available, relatively economical alternative and could greatly expand our capacity to
352 understand morphological heterogeneity in populations of bacteria and other organisms.
353

354 Recent innovations in the acquisition and analysis of imaging flow cytometry data are critical to
355 the success of our approach. The first innovation is acoustic focusing. Historically, analyzing
356 highly heterogeneous samples using imaging cytometry has been impractical, because of the
357 slow speed of image acquisition. The acoustic focusing capability of the cytometer used herein
358 allowed us to obtain thousands of images per second. As such, even though *Pasteuria* cells
359 were by no means the most common events in our samples, we nonetheless obtained
360 thousands of pictures of them. This dataset, in turn, enabled us to use data-hungry clustering
361 algorithms to identify morphotype-specific parameters. In addition to increasing the speed with

362 which we could acquire data, acoustic focusing also reduces the noise in the imaging and
363 fluorescence parameters by orientating cells in a single, narrow stream (27). It is likely that this
364 further facilitated the discrimination of morphotype-specific light signatures. Finally, acoustic
365 focusing enabled us to rapidly validate our sorting protocol (as compared to e.g., a manual,
366 microscopy approach) by boosting the speed with which we could inspect the morphology of our
367 sorted events.
368
369 The second key innovation critical to our approach are machine-learning image-analysis
370 algorithms, which are increasingly used for the analysis of high-parameter flow cytometry data
371 (reviewed in (32, 33)). The multi-dimensionality of IFC data makes it difficult to analyze with
372 hierarchical, manual approaches alone. Here, we used two approaches to efficiently analyze
373 these data: the dimensionality-reduction algorithm (UMAP) allowed us to visualize the data in 2-
374 D space, while the clustering algorithm (FlowSOM) allowed us to rapidly locate different
375 morphotypes in this high-dimensional space and to extract their features. In future, a supervised
376 clustering approach (e.g., (14, 34, 35)) could be used for feature-detection and may improve the
377 efficiency of IFC-enabled gate identification.
378
379 While semi-automated computational analyses were critical to the design of our label-free cell
380 sorting strategy, its optimization nevertheless required user-input and experience. For example,
381 experimentation was required to elucidate the appropriate meta-clustering parameter for
382 FlowSOM: too few clusters and the target morphotype was binned with debris/other
383 morphotypes; too many and the target morphotype was separated into many different clusters,
384 making it difficult to identify morphotype-specific light signatures. Similarly, user-intervention and
385 observations were important for navigating the differences between the imaging cytometer and
386 the cell sorter. Unlike the imaging cytometer, the cell sorter used herein does not have acoustic
387 focusing. As a result, there was higher variability in the scattering and fluorescence data derived

using the sorter vs. the imaging cytometer (compare Fig. 1Cii, Ei). We could not, therefore, position the gates in the same absolute position on the sorter as on the cytometer. Rather, we needed to manually inspect the distribution of the events and then position the gate in the same *relative* position on the sorter as on the imaging cytometer. Finally, prior knowledge vis-à-vis the UV-autofluorescence of spores (13), enabled us to better discriminate the spores from debris. These examples demonstrate that while label-free sorting is made possible by new innovations in IFC hardware and analyses, it nevertheless requires knowledge of the bacterium's biology and the capabilities of the hardware.

Here, we applied IFC in a novel way - to purify *known* morphotypes in a bacterial population. However, a fortuitous byproduct of the IFC-enabled sorting procedure is that it yields a huge amount of image data that can be used in the more traditional application of IFC: the discovery of hitherto uncharacterized variation within a bacterial population (36). For example, in the course of our analysis, we observed a range of grape morphologies, presumably representing cells at various stages of the spore-development process, as well as events of unknown provenance. This variability is difficult to observe (let alone quantify) using microscopy, particularly when morphotypes are rare. These data may help us to address outstanding questions vis-à-vis the life cycle of this bacterial pathogen e.g., "how does *Pasteuria*'s morphology changes during the first week of infection?".

In principle, IFC-enabled cell sorting is readily generalizable to other study systems and contexts. Nevertheless, we anticipate two hurdles that could narrow its widespread uptake. First, it may not always be possible to distinguish morphotypes on the basis of intrinsic light-scattering or autofluorescence signatures alone, particularly when hardware with limited optical components (lasers/filters) is used. Staining the sample with 'general', membrane permeable probes could increase the dimensionality of the data (and has proved useful in the development

414 of a transcriptomics-based, label-free sorting approach (11)). Second, software limitations could
415 limit the uptake of IFC-enabled cell sorting. To use IFC to identify gates requires that one can
416 simultaneously inspect the output of machine learning analyses (e.g., cluster, UMAP analyses)
417 and the image data. As far as we are aware, there is only one proprietary software platform
418 (FCS express) that permits this type of analysis using data derived with the Attune CytPix. We
419 encourage the development of open-source equivalents, particularly in languages commonly
420 used by microbial ecologists (e.g., R, Python), to overcome this potential barrier.
421
422 Cell sorting is a first step to understanding morphological heterogeneity in bacterial populations
423 but can only be applied to a fraction to species or contexts where that heterogeneity is
424 exhibited. Our work offers a novel, comparatively inexpensive method that will enable
425 microbiologists to purify – and hence phenotype – bacterial morphotypes from across the tree of
426 life and diverse environments.
427
428

429 ## Acknowledgements

430 The Attune CytPix used herein is located in the MSU Flow Cytometry Core and is supported by
431 the Equipment Grants Program, award #2022-70410-38419, from the U.S. Department of
432 Agriculture, National Institute of Food and Agriculture. We are grateful to members of the Wale
433 Lab for technical assistance and to Carrie Cizauskas PhD DVM for the illustrations in Table 1.
434
435
436
437
438
439
440
441

Bibliography

1. Yang DC, Blair KM, Salama NR. 2016. Staying in shape: the impact of cell shape on bacterial survival in diverse environments. Microbiol Mol Biol Rev 80:187–203.

2. Kysela DT, Randich AM, Caccamo PD, Brun YV. 2016. Diversity takes shape: understanding the mechanistic and adaptive basis of bacterial morphology. PLoS Biol 14:e1002565.

3. Young KD. 2006. The selective value of bacterial shape. Microbiology and Molecular Biology Reviews 70:660–703.

4. Horvath DJ, Li B, Casper T, Partida-Sanchez S, Hunstad DA, Hultgren SJ, Justice SS. 2011. Morphological plasticity promotes resistance to phagocyte killing of uropathogenic <i>Escherichia coli<i/>. Microbes Infect 13:426–437.

5. Geerlings NMJ, Karman C, Trashin S, As KS, Kienhuis MVM, Hidalgo-Martinez S, Vasquez-Cardenas D, Boschker HTS, Wael KD, Middelburg JJ, Polerecky L, Meysman FJR. 2020. Division of labor and growth during electrical cooperation in multicellular cable bacteria. Proc Natl Acad Sci 117:5478–5485.

6. Zhang Z, Du C, Barsy F de, Liem M, Liakopoulos A, Wezel GP van, Choi YH, Claessen D, Rozen DE. 2020. Antibiotic production in Streptomyces is organized by a division of labor through terminal genomic differentiation. Sci Adv 6:eaay5781.

7. Prashar A, Bhatia S, Gigliozzi D, Martin T, Duncan C, Guyard C, Terebiznik MR. 2013. Filamentous morphology of bacteria delays the timing of phagosome morphogenesis in macrophages. J Cell Biol 203:1081–1097.

8. Meriläinen L, Herranen A, Schwarzbach A, Gilbert L. 2015. Morphological and biochemical features of Borrelia burgdorferi pleomorphic forms. Microbiology 161:516–527.

9. Dienes L, Weinberger HJ. 1951. The L Forms of Bacteria. Bacteriol Rev 15:245–288.

10. Katz SM, Hashemi S, Brown KR, Habib WA, Hammel JM. 1984. Pleomorphism of <i>Legionella pneumophila<i/>. Ultrastruct Pathol 6:117–129.

11. Baron CS, Barve A, Muraro MJ, Linden R van der, Dharmadhikari G, Lyubimova A, Koning EJP de, Oudenaarden A van. 2019. Cell type purification by single-cell transcriptome-trained sorting. Cell 179:527-542.e19.

12. Shapiro HM. 2003. Practical Flow Cytometry, 4th ed. John Wiley & Sons, Inc, New Jersey, USA.

13. Laflamme C, Verreault D, Lavigne S, Trudel L, Ho J, Duchaine C. 2005. Autofluorescence as a viability marker for detection of bacterial spores. Front Biosci 10:1647.

476 14. Blasi T, Hennig H, Summers HD, Theis FJ, Cerveira J, Patterson JO, Davies D, Filby A,
477 Carpenter AE, Rees P. 2016. Label-free cell cycle analysis for high-throughput imaging flow
478 cytometry. Nat Commun 7:10256.

479 15. Kiepś J, Juzwa W, Dembczyński R. 2023. Imaging flow cytometry demonstrates
480 physiological and morphological diversity within treated probiotic bacteria groups. Int J Mol Sci
481 24:6841.

482 16. Pinto RN, Sebastian JA, Parsons MJ, Chang TC, Turner TR, Acker JP, Kolios MC. 2019.
483 Label-Free Analysis of Red Blood Cell Storage Lesions Using Imaging Flow Cytometry. Cytom
484 Part A 95:976–984.

485 17. Dey R, Rieger AM, Stephens C, Ashbolt NJ. 2019. Interactions of Pseudomonas aeruginosa
486 with Acanthamoeba polyphaga Observed by Imaging Flow Cytometry. Cytom Part A 95:555–
487 564.

488 18. Ebert D, Rainey P, Embley TM, Scholz D. 1996. Development, life cycle, ultrastructure and
489 phylogenetic position of *Pasteuria ramosa* Metchnikoff 1888: Rediscovery of an obligate
490 endoparasite of *Daphnia magna* Straus. Phil Trans R Soc B 351:1689–1701.

491 19. Ebert D, Duneau D, Hall MD, Luijckx P, Andras JP, Pasquier LD, Ben-Ami F. 2016. A
492 population biology perspective on the stepwise infection process of the bacterial pathogen
493 *Pasteuria ramosa* in *Daphnia*. Adv Parasit 91:265–310.

494 20. Duneau D, Luijckx P, Ben-Ami F, Laforsch C, Ebert D. 2011. Resolving the infection process
495 reveals striking differences in the contribution of environment, genetics and phylogeny to host-
496 parasite interactions. BMC biology 9:11.

497 21. Clerc M, Ebert D, Hall MD. 2015. Expression of parasite genetic variation changes over the
498 course of infection: implications of within-host dynamics for the evolution of virulence. Proc
499 Royal Soc B Biological Sci 282:20142820.

500 22. Fields PD, Reisser C, Dukić M, Haag CR, Ebert D. 2015. Genes mirror geography in
501 Daphnia magna. Mol Ecol 24:4521–4536.

502 23. Luijckx P, Ben-Ami F, Mouton L, Pasquier LD, Ebert D. 2010. Cloning of the unculturable
503 parasite Pasteuria ramosa and its Daphnia host reveals extreme genotype-genotype
504 interactions. Ecol Lett 14:125–31.

505 24. Klüttgen B, Dülmer U, Engels M, Ratte HT. 1994. ADaM, an artificial freshwater for the
506 culture of zooplankton. Water Res 28:743–746.

507 25. Ben-Ami F, Ebert D, Regoes RR. 2010. Pathogen Dose Infectivity Curves as a Method to
508 Analyze the Distribution of Host Susceptibility: A Quantitative Assessment of Maternal
509 Effects after Food Stress and Pathogen Exposure. Am Nat 175:106–115.

510 26. Ebert D, Zschokke-Rohringer CD, Carius HJ. 2000. Dose effects and density-dependent
511 regulation of two microparasites of Daphnia magna. Oecologia.

27. Ward MD, Kaduchak G. 2018. Fundamentals of Acoustic Cytometry. Curr Protoc Cytom 84:e36.

28. McInnes L, Healy J, Saul N, Großberger L. 2018. UMAP: Uniform Manifold Approximation and Projection. J Open Source Softw 3:861.

29. Gassen SV, Callebaut B, Helden MJV, Lambrecht BN, Demeester P, Dhaene T, Saeys Y. 2015. FlowSOM: Using self-organizing maps for visualization and interpretation of cytometry data. Cytom Part A 87:636–645.

30. Nitta N, Sugimura T, Isozaki A, Mikami H, Hiraki K, Sakuma S, Iino T, Arai F, Endo T, Fujiwaki Y, Fukuzawa H, Hase M, Hayakawa T, Hiramatsu K, Hoshino Y, Inaba M, Ito T, Karakawa H, Kasai Y, Koizumi K, Lee S, Lei C, Li M, Maeno T, Matsusaka S, Murakami D, Nakagawa A, Oguchi Y, Oikawa M, Ota T, Shiba K, Shintaku H, Shirasaki Y, Suga K, Suzuki Y, Suzuki N, Tanaka Y, Tezuka H, Toyokawa C, Yalikun Y, Yamada M, Yamagishi M, Yamano T, Yasumoto A, Yatomi Y, Yazawa M, Carlo DD, Hosokawa Y, Uemura S, Ozeki Y, Goda K. 2018. Intelligent Image-Activated Cell Sorting. Cell 175:266-276.e13.

31. Schraivogel D, Kuhn TM, Rauscher B, Rodríguez-Martínez M, Paulsen M, Owsley K, Middlebrook A, Tischer C, Ramasz B, Ordoñez-Rueda D, Dees M, Cuylen-Haering S, Diebold E, Steinmetz LM. 2022. High-speed fluorescence image–enabled cell sorting. Science 375:315–320.

32. Stolarek I, Samelak-Czajka A, Figlerowicz M, Jackowiak P. 2022. Dimensionality reduction by UMAP for visualizing and aiding in classification of imaging flow cytometry data. iScience 25:105142.

33. Marsh-Wakefield FM, Mitchell AJ, Norton SE, Ashhurst TM, Leman JK, Roberts JM, Harte JE, McGuire HM, Kemp RA. 2021. Making the most of high-dimensional cytometry data. Immunol Cell Biol 99:680–696.

34. Hennig H, Rees P, Blasi T, Kamentsky L, Hung J, Dao D, Carpenter AE, Filby A. 2017. An open-source solution for advanced imaging flow cytometry data analysis using machine learning. Methods 112:201–210.

35. Haridas V, Ranjbar S, Vorobjev IA, Goldfeld AE, Barteneva NS. 2017. Imaging flow cytometry analysis of intracellular pathogens. Methods 112:91–104.

36. Power AL, Barber DG, Groenhof SRM, Wagley S, Liu P, Parker DA, Love J. 2021. The application of imaging flow cytometry for characterisation and quantification of bacterial phenotypes. Front Cell Infect Microbiol 11:716592.

37. 2024. Attune Cytometric Software User Guide. MAN0026553. Thermo Fisher Scientific Ltd., Singapore.

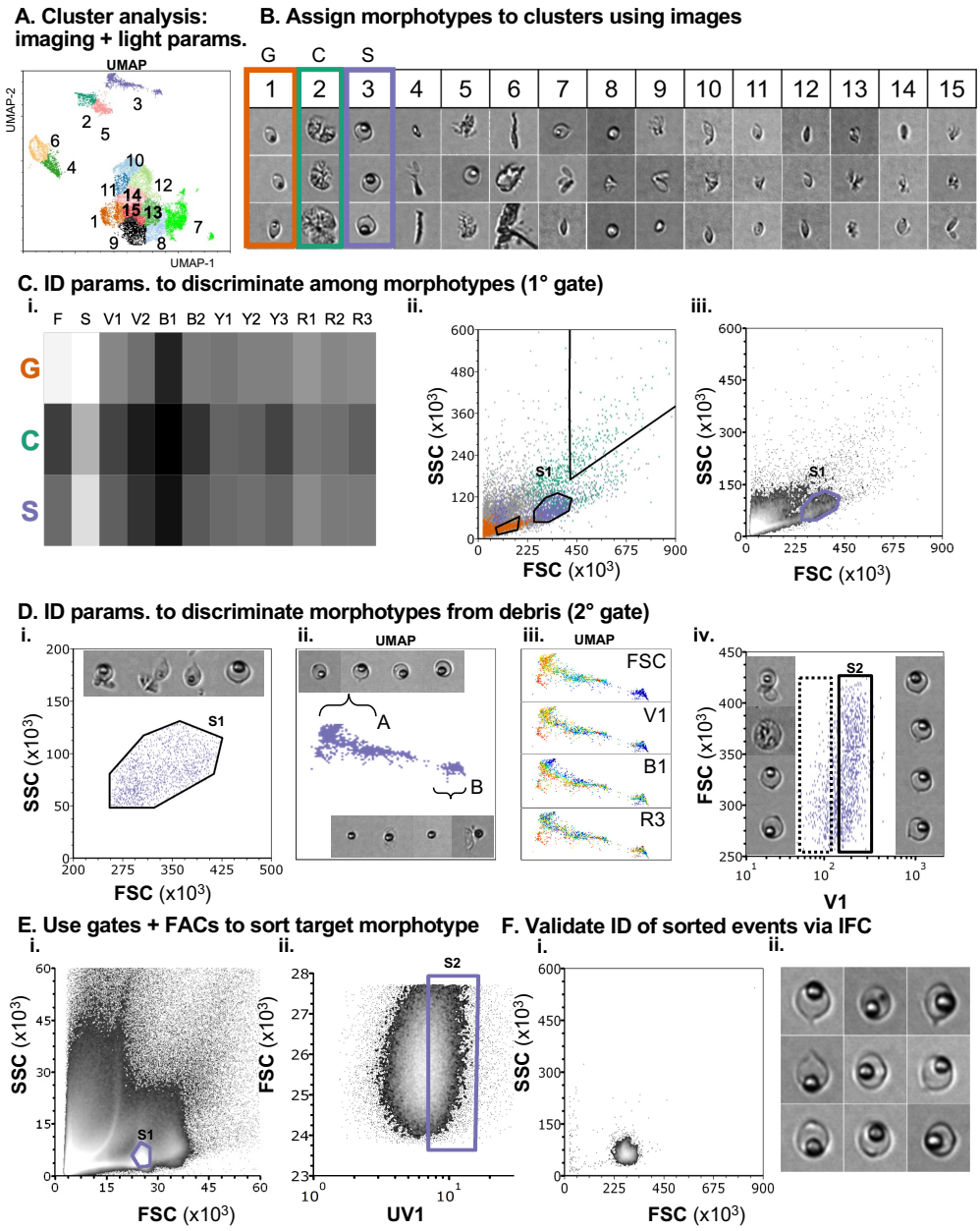


Figure 1: Design & validation of a label-free cell sorting strategy using imaging flow cytometry (IFC). Worked example of IFC-enabled cell-sorting, for the purification of *Pasteuria* spores (the "target morphotype") from a sample of host material, microbiota and (non-target) *Pasteuria* morphotypes (cauliflower, C; grapes, G). **A.** Cluster analysis of imaging and light parameters (scattering, fluorescence) is performed and imposed onto a UMAP. **B.** Each cluster is assigned to a morphotype or debris using the image data. Here, clusters 1, 2, & 3 were denoted the G (orange), C (green) and S (lilac) clusters, respectively. **C.** A primary (1°) gate (S1) that distinguishes the target morphotype from non-target morphotypes is identified using (i) a "parameter heatmap", which displays the median value of each parameter in each cluster (i; only those parameters shared among the sorter and cytometer displayed), and (ii, iii) visualization of the parameters in bipartite space (ii, events in S cluster, lilac, G cluster, orange, C cluster, green; iii, shaded from black to white to indicate increasing event density). **D.** We first inspect the images associated with the events in the 1° gate. To identify a secondary (2°) gate for separating target morphotypes from debris, regions of the target cluster that contain a pure population of the target morphotype (a) vs. a mixed population (b) are identified using the image data. Fluorescence parameters that distinguish the regions are identified by coloring the cluster by each parameter (iii, warmer colors indicate higher fluorescence intensity; note that regions A, B are differentially colored by FSC, V1 indicating the suitability of these parameters for 2° gating). Finally, events in the primary gate are plotted in bipartite space defined by the secondary parameters. The location of the secondary gate (S2) is set with respect to the image data (iv; right images, events within the gate (solid box), left images, events outside of the gate (dashed box)). **E.** The sample is analyzed using the cell sorter. It is verified that the events take approximately the same distribution on the sorter and imaging cytometer (Ei, Ciii respectively; note differences in scale); the gates are then set to the same, relative position with respect to the distribution. Here, analysis of the sample using the sorter revealed that the sorter's UV1 channel had more discriminatory power than V1 (Fig. S3) and so we used this parameter for sorting in place of V1(Eii) **F.** To verify that sorted events are of the target morphotype and to quantify the purity of the sorted sample, the sorted sample is analyzed via IFC. First, it is confirmed that sorted events lie in expected region of bipartite space (compare Fi, Cii), then the image data is inspected and the proportion of target morphotypes quantified (Fii, representative images from sorted sample)

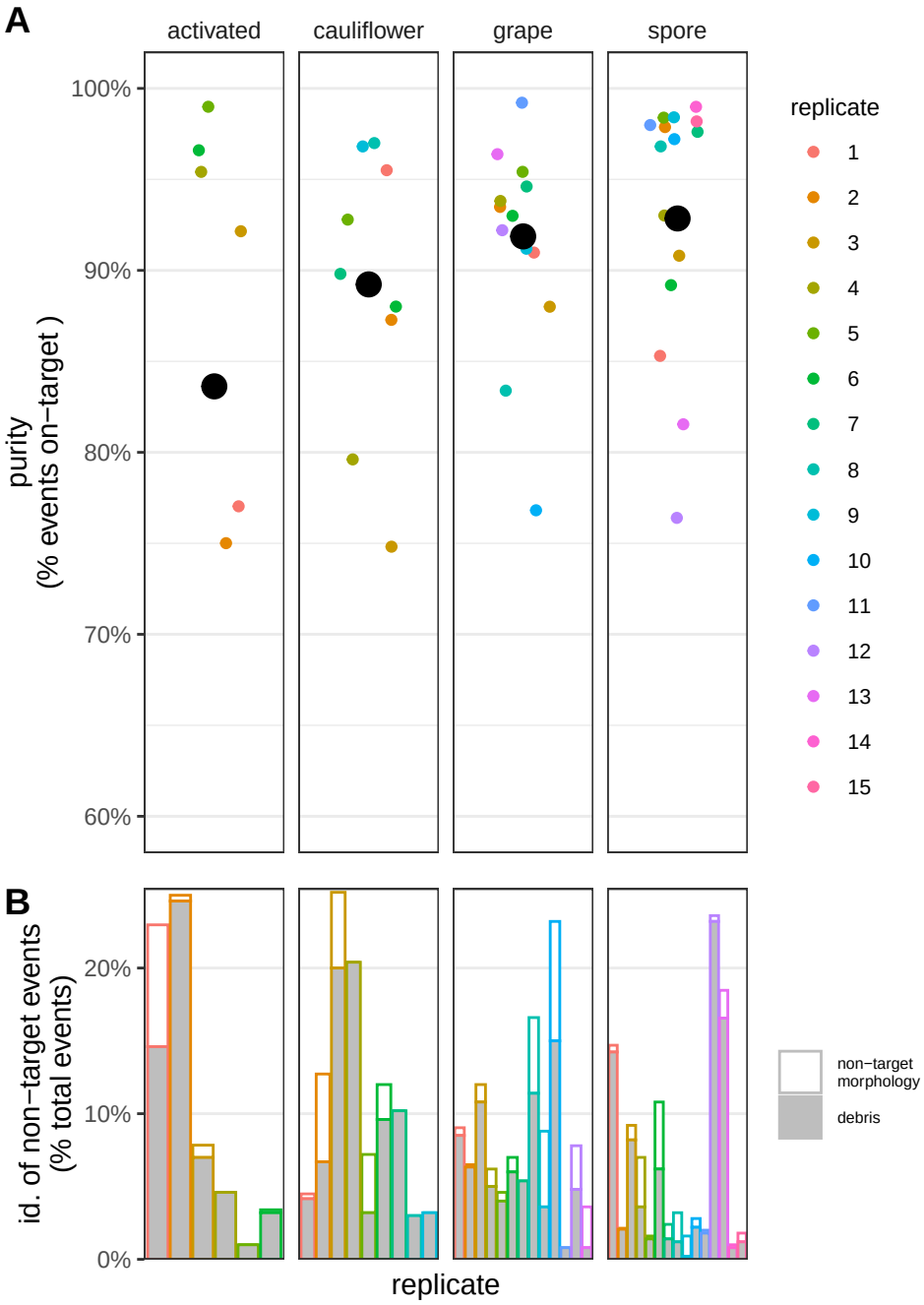


Figure 2: IFC-enabled cell sorting yields high purity samples of four distinct *Pasteuria* morphotypes. **A.** Purity of sorted samples, determined using image data acquired via IFC (mean purity across sorts, black, large points; purity of each replicate sort, colored, small points). Note that a replicate refers to an attempt to sort a specific morphotype i.e., replicate 1 under the "activated " and "cauliflower" columns may not have involved the same sample or been performed at the same time. **B.** Composition of the non-target fraction of the sorted sample in each replicate (colored as in **A**).