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

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

Authors

Daniel Vocelle[1], Lindsey Thompson[2], Matthew 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.

DV, MB acquired funding for the instrumentation.

Data availability

Description: data will be made available upon article acceptance.

Abstract

Bacterial populations often display remarkable morphological heterogeneity. Fluorescence activated cell sorting (FACS) is an important tool for understanding the biological significance of this morphological diversity, as it permits the separation and study of distinct morphological variants (morphotypes) from each other and their environmental milieu. In FACS, cells are first labeled with fluorescent markers such as antibodies or transgenic constructs, and then sorted based on their possession of such 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 role 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.

55 # Introduction

56 Bacterial populations often contain cells varying in size and shape, and this morphological
57 heterogeneity can have a profound impact on bacterial fitness (1–3). For example, different
58 forms of bacteria (morphotypes) may perform distinct ecological "roles" or exhibit differential
59 susceptibility to environmental stress, permitting population persistence in the face of
60 environmental stress (4–7). To fully understand the causes and consequences of morphological
61 variation within bacterial populations, we need to extend our investigations beyond well-studied
62 taxa and *in vitro* environments, which can suppress morphological diversity within a bacterial
63 population (2, 8–10). Here, we present imaging flow cytometry (IFC)-enabled cell sorting, a
64 novel procedure that allows for the purification of distinct morphotypes from unculturable,
65 pleiomorphic bacteria living in complex milieu.
66
67 Efforts to understand the significance of morphological heterogeneity in a bacterial population
68 often begin with isolating distinct bacterial morphotypes from each other and/or their
69 environmental milieu, via fluorescence activated cell sorting (FACS). However, the application of
70 FACS to unculturable bacteria or those found in complex samples (e.g., soil, host tissues) is a
71 challenge. A chicken-and-egg-scenario exists. In FACS, different cell populations are tagged
72 with distinct fluorescent labels (e.g., fluorescently labelled antibodies) on whose basis the
73 populations are separated. However, the development of morphotype-specific labels requires
74 the characterization of morphotype-specific antigens, which itself requires that morphotypes be
75 separated and studied. In the absence of morphotype specific labels, "general' membrane
76 permeable labels (e.g., CFSE, Mitotracker), which bind to molecules present in most cells (e.g.,
77 amines, RNA or mitochondria), could be used to discriminate morphotypes in certain contexts
78 (11). However, because general labels indiscriminately permeate all cells within a sample, they

are unlikely to be helpful in distinguishing a specific morphotype if it is present in a complex, cell-rich milieu such has host-tissue or soil.

A potential solution to the problem of sorting bacterial morphotypes in the absence of labels is to exploit intrinsic variation in their light-scattering and -fluorescence properties (i.e., “light signatures”). Cells differing in size and granularity differentially scatter light, and the degree of autofluorescence may also vary among morphotypes (12). For example, viable *Bacillus anthracis* spores exhibit a distinctive autofluorescent signature upon excitation with UV light (13). Recent advances in imaging flow cytometry (IFC) simplify the identification of intrinsic light signatures of morphologically distinct cells, by generating images of each cell and measures of their light-scattering and autofluorescence properties. Systematic analysis of these data allows for the identification of morphotype-specific light signatures and has been used to identify morphologically distinct subpopulations of mammalian and bacterial cells (14–17).

Here, using the pleiomorphic bacterial pathogen *Pasteuria ramosa* (hereafter, *Pasteuria*) as a model, we demonstrate that IFC can be used to efficiently design and validate strategies for label-free sorting of distinct bacterial morphotypes. *Pasteuria* is a spore-forming pathogen of zooplankton that exhibits striking morphological variation during its life cycle (Table 1, (18)). Longstanding hypothesis vis-à-vis the “division of labor” among different morphotypes within the *Pasteuria* bacterial population (19) remain untested, largely because it is difficult to separate bacteria from host material for functional -omics and *in vitro* study. By combining state-of-the-art IFC instrumentation, which allows us to rapidly derive thousands of images of bacterial cells, with machine learning algorithms we systematically designed and validated autofluorescence-based FACS strategies that yield high-purity samples of each of four distinct morphotypes. Our work thus provides proof of principle that IFC-enabled sorting can be used to overcome the first

104 barrier to understanding morphological heterogeneity in bacterial species that cannot be readily

105 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 (`grape', G)	sporulation	4-6
	Mature spore (`spore', S)	transmission, environmental persistence	4.2-5.4

128

129 **Table 1. Variation in the form and function of *Pasteuria ramosa* morphotypes. Form and**
130 function information extracted from (18, 20, 21). Schematic representations of morphotypes not
131 to scale.

132

133 # Materials and methods

134 Our goal was to design and validate cell sorting strategies suitable for the purification 4 distinct

135 *Pasteuria* morphotypes from the homogenate of an infected host i.e., a sample containing

136 several *Pasteuria* morphotypes, host-cells, host food (algae) present in the gut, and host

137 microbiota. Below we describe our approach in general terms. The final procedure and results

138 of our validation steps are described in the results.

139

140 **Host & Parasites**

141 Throughout, we used *Daphnia magna* (HuHO-2 genotype) hosts and *Pasteuria ramosa* (strain
142 C1), originally collected in Hungary and Russia, respectively (22, 23). Both were generously
143 donated by D. Ebert (U. Basel, Switzerland).
144
145 To generate samples of *Pasteuria ramosa*, we experimentally infected *Daphnia magna*. The
146 infection procedure took the same, general format, with minor variation in husbandry details.
147 Briefly, *P. ramosa* spores were shaken at 1500 rpm overnight. On day 0 of the experiment, we
148 filled a vessel (e.g., 24 well plate or beaker) with media (ADaM or COMBO), added 1-10 *D.*
149 *magna* <24 hours old (24). A second equivalently sized dose of spores was added on the
150 following day. Animals were fed an increasing quantity of either *Scenedesmus obliquus* or
151 *Ankistrodesmus falcatus* as they aged, up to ~30-60 days. To diagnose *Pasteuria* infection,
152 animals were assessed for visual signs of infection, including discoloration, lack of offspring and
153 enlargement (25). Infection was then confirmed by inspection using microscopy.
154
155 **Instruments**
156 We used a Thermofisher Attune CytPix flow cytometer (the cytometer, hereafter) for IFC. This
157 cytometer is equipped with a high-speed, brightfield camera (20x objective, 0.45 numerical
158 aperture, 0.3 micron/pixel resolution), four lasers and 14 filters (Table S1). Acoustic focusing
159 allows for the alignment of cells and particles (hereafter, events) in a narrow depth of field. This
160 property (along with the high-speed camera) allows for the acquisition of in-focus images at a
161 rate an order of magnitude faster than cytometers with hydrodynamic-focusing (26).

162 For cell sorting, we used a BD Influx equipped with 5 lasers and 12 filters (Table S1). The
163 default measure of fluorescence on the BD Influx is “pulse height” (rather than area). Since our
164 goal was to use the imaging cytometer to identify fluorescence parameters that could be

165 employed in cell sorting, we used pulse height as our measure of fluorescence on both
166 machines.
167
168 **Data acquisition**
169 Immediately following diagnosis, we prepared infected animals for flow cytometric analysis.
170 First, each animal was washed in PBS (no Ca+ or Mg+). We transferred 1-4 animals, depending
171 on the degree of infection and an initial assessment of morphotype composition via microscopy,
172 into a 1.5mL Eppendorf tube containing PBS. We homogenized the animals briefly with an
173 electric pestle (~20 seconds, to avoid damaging cells) to create a slurry. We diluted the sample
174 for analysis in 500mL Focusing Buffer (Thermofisher Scientific, catalog no. A24904) or PBS,
175 which have equivalent refractive indices. We ran samples at 12.5mL/min until we acquired
176 30,000 events. Between runs, we twice flushed the cytometer line to prevent sample carry over.
177
178 **Pre-gating**
179 We first excluded ("gated out") algae on the basis of their chlorophyll autofluorescence (high
180 emission in channel R3, cf. Table S1). We then selected ("gated in") the events that (i) were
181 successfully photographed and processed ("Is Processed" positive events) and, furthermore,
182 contained only one particle (Event Particle Count = 1). This yielded a dataset of single particles,
183 each of which had had both imaging and light-parameter data associated with it.
184
185 To analyze the image data, we employed inbuilt Attune Image-Processing Software (version
186 6.0.1). We used the software's bead masking model (Beads_Only_Full_Resolution_v1) which is
187 optimized for objects 2-5μm. We chose this algorithm because (i) an alternative model
188 optimized for objects >5μm, excluded ~70% of our target objects and (ii) because we found that
189 the bead model could process a variety of *P. ramosa* morphotypes, including those that

190 exceeded 5μm (e.g., cauliflowers). Image-processing yielded a matrix containing the values of
191 22 imaging parameters including a variety of size and "texture" metrics (Table S2). In total,
192 including the fluorescent parameters (and pulse parameters height, area, and width) each event
193 was associated with 74 parameters. Before conducting further analysis, we rescaled each
194 parameter linearly between 0 and 1.
195
196 We visualized the data using Uniform Manifold Approximation Projection for Dimension
197 Reduction (UMAP) (27), implemented in FCS Express (De Novo Software). Briefly, this
198 algorithm allows for the visualization of high-dimensional data in a 2D plane. By inspecting
199 distinct regions of the UMAP, and the imaging data with which it was associated, we found that
200 different morphotypes clustered in distinct areas of the UMAP. Preliminary analysis of this kind
201 revealed that certain image parameters were particularly informative for distinguishing different
202 morphotypes (Table S2). We employed this subset of parameters in further analyses, described
203 below.
204
205 **Pre-sort analysis: identifying morphotype-specific parameters for cell sorting**
206 For each sample, we used IFC to identify gates that would enable us to sort a specific
207 morphotype ("target morphotype", hereafter) from the remainder of the sample.
208
209 First, we visualized the morphological complexity of the sample using a UMAP analysis of all
210 scattering and fluorescent parameters, as well as several imaging parameters identified in
211 preliminary analyses as being most-informative (Table S2, S3). We then performed cluster
212 analysis to identify (i) a primary gate, on which basis we could distinguish our target morphotype
213 from other bacterial cells, and (ii) secondary gate(s), suitable for distinguishing our target
214 morphotype from debris within the primary gate (see Results). Specifically, we used FlowSOM -

215 a neural network-like algorithm that identifies clusters of events characterized by distinct
216 (combinations of) parameters (28). We set the consensus clustering parameter to between 8-
217 20, depending on the sample, and otherwise used default settings (Table S3).
218
219 The above-described procedure was repeated on a sample-by-sample basis. Although, over the
220 course of this project, we characterized distinct "signatures" of each morphotype that held
221 among samples, the fluorescence parameters used to define the gates varied among samples,
222 particularly in the initial stages of our work. Due to differences among samples in the
223 autofluorescence of the cells, moreover the exact position of the gates varied from sample to
224 sample. Importantly, however, the gate -identification and -validation strategy - which is our
225 focus here - remained the same for all samples.
226
227 **Cell sorting**
228 Prior to analysis via cell sorting, we transferred the sample to a 5 mL FACS tube (Corning,
229 catalog no. 352063) and filtered it through a 100μm mesh filter. We then used the gating
230 strategy designed via IFC ("pre-sort" analysis), or a minor variation of it (see Results, Fig. S1) to
231 sort each of the morphotypes in the sample that were of sufficient abundance.
232
233 **Post-sort analysis: quantification of sample purity**
234 We verified the purity of each sorted sample using IFC. We analyzed up to 30000 events (mean
235 events analyzed = 6993, range 526-30000); the exact number varied according to the relative
236 abundance of the morphotype present in the pre-sorted sample. We analyzed the IFC data as
237 before (i.e., same image processing model, and initial gating on single-processed events) and
238 then "down-sampled" it to 500 events (images). Using the image data, we manually classified
239 each of the 500 events as an identifiable bacterial morphotype (activated, cauliflower, grape or
240 spore) or unidentified debris. Finally, we calculated the purity of the sorted sample i.e., the

241 proportion of total events in the down-sampled dataset that were of the target morphotype. This
242 purity measure is conservative given that we include unidentified debris such as salts in our total
243 event count.
244

245 Results

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

248 We established a robust procedure for identifying light signatures that can be harnessed for the
249 cell sorting of morphologically distinct bacterial morphotypes, in the absence of labels. To
250 demonstrate this approach, we provide a worked example (Fig. 1). In this example, our goal
251 was to sort a spore (S) population from a homogenate comprising host (*Daphnia*) material, host-
252 microbiota and non-target morphotypes of *Pasteuria* (cauliflower (C), grape (G)). Worked
253 examples of the procedure as applied to activated, cauliflower and grape stages are provided in
254 the supplement (Fig. S1).
255

256 First, we performed a dimensionality reduction (UMAP) analysis to visualize the data in 2D
257 space (Fig. 1A). To identify where in this 2D space each *Pasteuria* morphotype lay, we imposed
258 upon this UMAP the results of a cluster analysis (Fig. 1A, colors/numbers indicate clusters).
259 This cluster analysis was performed using all light-scattering and -fluorescence parameters and
260 a subset of the imaging parameters (Table S2). We manually classified each cluster as either a
261 morphotype-rich or debris-rich cluster using the image data (Fig. 1B). For example, since spores
262 appeared in most of the images associated with cluster 3, we designated cluster 3 the spore-
263 rich cluster (S-cluster; Fig. 1B, purple outline); similarly, clusters 1 and 2 were designated
264 grape- and cauliflower- rich clusters, respectively.
265

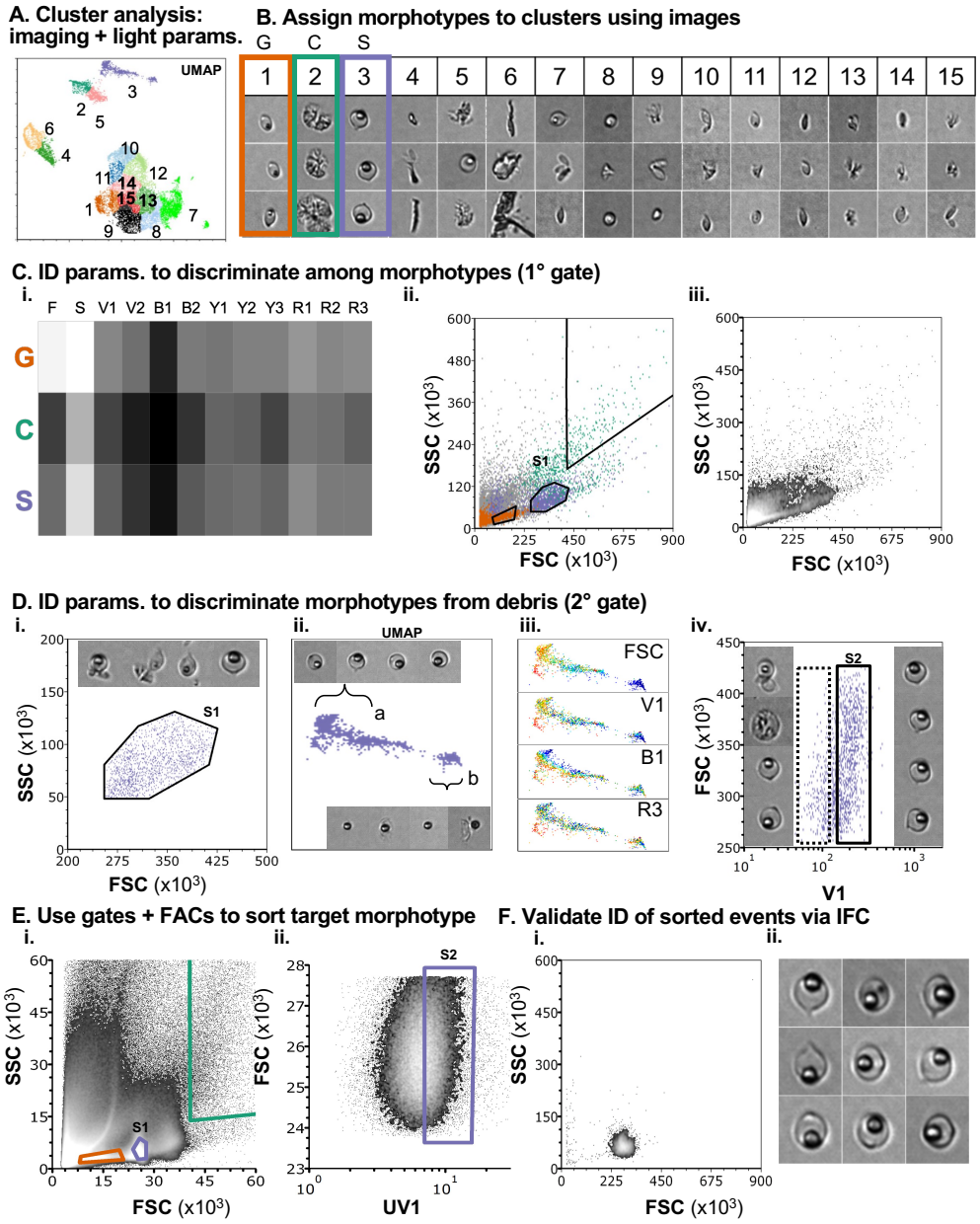

Figure 1: The design & validation of a label-free cell sorting strategy using imaging flow cytometry (IFC). Worked example wherein we used IFC-enabled cell-sorting to purify *Pasteuria* spores (the "target morphotype") from a sample of host material, microbiota and (non-target) *Pasteuria* morphotypes (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 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). The primary gate (S1) is thus set. **D.** We first inspect the images associated with the events in the primary 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)

266 Next, we used the cluster analysis to identify light-scattering and fluorescence parameters that
267 characterized our target morphotype vs. the other morphotypes (C, G). Analysis of the median
268 parameter values associated with each of the morphotype clusters (Fig. 1Ci), suggested that the
269 different morphotypes could be broadly distinguished by the light scattering parameters FSC
270 and SSC (Fig.1Ci, F,S), consistent with their differences in size and granularity (Table 1). The
271 distribution of G-, C- and S- cluster events in FSC-SSC space confirmed this inference (Fig.
272 1Cii, Ciii). We thus used FSC-SSC to define our primary gate (S1, Fig. 1Cii).
273
274 Inspection of the images associated with the events in this primary gate revealed that it included
275 debris as well as our target morphotype (Fig. 1Di). We thus sought to identify a secondary gate
276 that could be used to distinguish the target morphotype from debris. First, we inspected the
277 images associated with the events in our target cluster. This analysis revealed that there was a
278 gradient within the cluster: on the left lay a largely homogenous population of spores (Fig. 1Dii,
279 region a); whereas on the right we found a mixed population of spores and other non-target
280 objects (Fig. 1Dii, region b). To identify which of the fluorescence parameters shared by the
281 sorter and the cytometer (Table S1) could differentiate events in regions a and b, we recolored
282 the events in each cluster by their fluorescence intensity in each of the shared channels (Fig.
283 Diii, 4 parameters shown for readability; all parameters are displayed in Fig. S2). This analysis
284 revealed that V1 could distinguish the spore-rich region (a) from the mixed region (b) and could
285 therefore be used in combination with FSC to define a secondary gate. Finally, we established
286 the position of this secondary gate ("S2") by plotting the events in our primary gate in bipartite
287 FSC-V1 space (Fig. 1Div). We inspected the image data (Fig. 1Div, left images, events within
288 dotted rectangle; right images, images from solid rectangle) and refined the gate so that it
289 included only events of the target morphotype (Fig. 1Div, solid rectangle).
290

291 With a gating strategy in hand, we then performed cell sorting. As anticipated, the sample
292 exhibited the same qualitative, but not quantitative, pattern on the sorter as on the imaging-
293 cytometer (compare Fig. 1Cii, Ciii vs. Ei). We thus positioned our gates in the same position,
294 relative to the distribution of events, on the sorter as on the cytometer. Interestingly, we found
295 that UV1 – a parameter available on the cell sorter but not the cytometer - could better
296 discriminate spores from debris than V1 (Fig. S3). In this worked example, as well as other
297 analyses wherein spores were the target morphotype, we thus used UV1 rather than V1 to
298 define our secondary gate. With the gates set, we sorted the sample.
299
300 Finally, to validate our sorting strategy and quantify the purity of the sorted sample, we analyzed
301 a subset of the sorted sample via IFC. First, we confirmed that the sorted events occupied the
302 expected region of FSC-SSC space i.e., that the sorted sample occupied the same position in
303 FSC-SSC space as the S-cluster (compare Figs. Ciii, 1Fi). Second, we confirmed the identity of
304 the sorted events using the image data (Fig. 1Fii, Fig. 2).
305
306 **IFC-enabled sorting yields high-purity populations of each morphotype**
307 We used the above-described gate-discovery and validation strategy to conduct label-free cell
308 sorting on 43 occasions. These analyses involved 24 samples varying in initial composition and
309 targeted each of the four different *Pasteuria* morphotypes.
310
311 The sorted samples ranged in purity from 75% to 99% (Fig. 2A). Of the events that were not
312 identifiable as the target morphotype ("non-target events"), the majority were debris (e.g., salt
313 crystals, host material, Fig. 2B). Indeed, on average the sorted sample constituted only 1.8%
314 (range 0-8.4%) of *Pasteuria* morphotypes that were not the target (Fig. 2B).
315

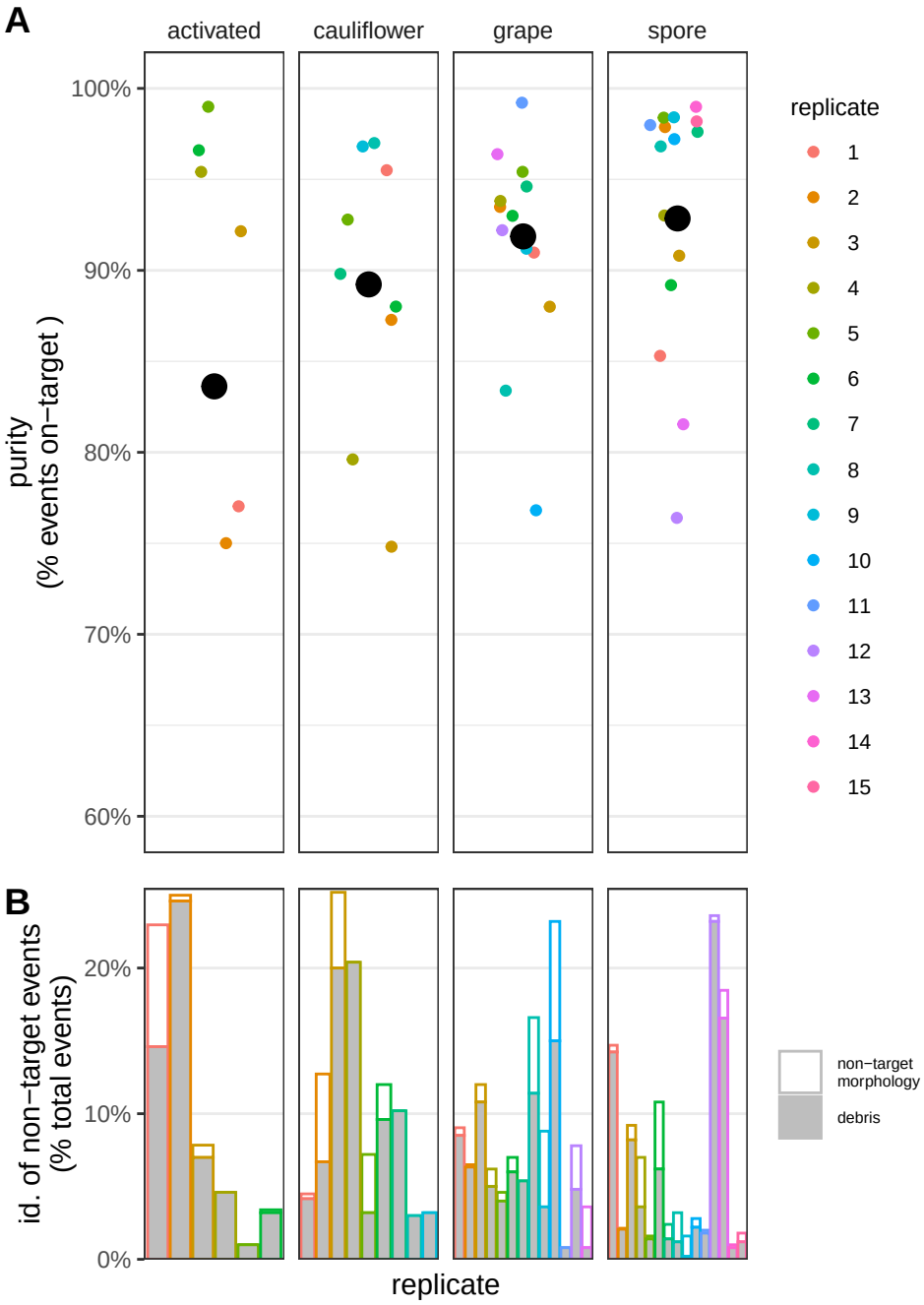

Figure 2: IFC-enabled cell sorting yields high purity samples of four distinct *Pasteuria* morphotypes. **A.** Purity of sorted samples, determined using image data acquired via IFC (mean purity across sorts, black, large points; purity of each replicate sort, colored, small points). Note that a replicate 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**).

316 # Discussion

317 The first step toward understanding the physiological and ecological significance of
318 morphological variation in a cellular population is to purify distinct morphotypes. FACS is
319 commonly employed for cell purification but its application to poorly studied or difficult-to-culture
320 bacteria is limited by a paucity of fluorescent-labels (antibodies, transgenics). Here, we
321 demonstrate that IFC can be used to design and validate a label-free cell sorting strategy that
322 efficiently separates distinct bacterial morphotypes from each other and from environmental
323 debris. Critically, this procedure yields samples of a purity required for common downstream, "-
324 omics" analyses i.e., >=90%. High-speed image-enabled cell sorting approaches have been
325 innovated (29, 30) and are slowly becoming available, but their expense and limited availability
326 makes them out of reach for most. Our methodology represents an immediately available,
327 relatively economical alternative and could greatly expand our capacity to understand
328 heterogeneity in cellular populations of bacteria, and indeed, other organisms or tissues.
329

330 Recent innovations in the acquisition and analysis of imaging flow cytometry data are critical to
331 the success of our approach. The first innovation is acoustic focusing. Historically, analyzing
332 highly heterogeneous samples using imaging cytometry has been impractical, because of the
333 slow speed of image acquisition. The acoustic focusing capability of the cytometer used herein
334 allowed us to obtain thousands of images per second. As such, even though *Pasteuria* cells
335 were by no means the most common particles in our samples, we nonetheless obtained
336 thousands of pictures of these cells. This dataset, in turn, enabled us to use data-hungry
337 clustering algorithms to identify morphotype-specific parameters. In addition to increasing the
338 speed with which we could acquire data, acoustic focusing also reduces the noise in the
339 imaging and fluorescence parameters by orientating cells in a single, narrow stream (26). It is
340 likely that this further facilitated the discrimination of morphotype-specific light signatures.

341 Finally, acoustic focusing enabled us to rapidly validate our sorting protocol (as compared to
342 e.g., a manual, microscopic approach) by boosting the speed with which we could inspect the
343 morphology of our sorted events.
344
345 The second key innovation critical to our approach are machine-learning analysis algorithms,
346 which are increasingly used for the analysis of high-parameter flow cytometry data (reviewed in
347 (31, 32)). The multi-dimensionality of IFC data makes it difficult to analyze with hierarchical,
348 manual approaches alone. Here, we used two approaches to efficiently analyze these data: the
349 dimensionality-reduction algorithm (UMAP) allowed us to visualize the data in 2-D space; while
350 the clustering algorithm (FlowSOM) allowed us to rapidly locate different morphotypes in this
351 high-dimensional space and to extract their features. In future, a supervised clustering approach
352 (e.g., (14, 33, 34)) could be used for feature-detection and may improve the efficiency of our
353 approach.
354
355 While semi-automated computational analyses were critical to the design of our label-free cell
356 sorting strategy, its optimization nevertheless required user-input and experience. For example,
357 experimentation was required to elucidate the appropriate meta-clustering parameter for
358 FlowSOM: too few clusters and the morphotypes were binned with debris/other morphotypes;
359 too many and our target morphotypes were separated into many different clusters, making it
360 difficult to identify morphotype-specific light signatures. Similarly, user-intervention and
361 observations were important for navigating the differences among the imaging-cytometer and
362 the cell sorter. Our cell sorter does not have acoustic focusing like the imaging-cytometer. As a
363 result, there is higher variability in the scattering and fluorescence parameters on the sorter than
364 on the cytometer (compare Fig. 1Cii, Ei), and we could not position the sorting gates in the
365 same absolute position on the sorter as on the cytometer. Instead, we positioned the gates in
366 the same *relative* position on each machine and, at least in our initial runs, confirmed that the

367 gates were correctly positioned by pausing sorting after a few thousand events and verifying the
368 composition of the sorted sample using IFC. Finally, alert to previous findings vis-à-vis the UV-
369 autofluorescence of spores (13), we discovered that the UV1 channel – which was available on
370 the sorter but not the imaging cytometer – allowed us to better discriminate the spores from
371 debris. Thus, while IFC-enabled sorting is arguably made possible by new innovations in IFC
372 hardware and analyses, it nevertheless requires user knowledge of the focal bacterium's biology
373 and the capabilities of the hardware.
374
375 Here, we applied IFC in a novel way - to purify *known* morphotypes in a bacterial population.
376 However, a fortuitous byproduct of the IFC-enabled sorting procedure is that it yields a huge
377 amount of image data that can be used in the more traditional application of IFC: the discovery
378 of hitherto uncharacterized variation within a bacterial population (35). For example, in the
379 course of our analysis, we observed a full range of "grape" morphologies, from those without an
380 "eye" to almost-spore-like stages; we also observed events of unknown provenance. Such
381 variability is difficult to observe (let alone quantify) using microscopy, particularly when
382 morphotypes are rare and may enable us to address outstanding questions in vis-à-vis the life
383 cycle of this pathogen e.g., what morphological stage mediates the first ~week of infection.
384
385 In principle, IFC-enabled cell sorting is readily generalizable to other study systems and
386 contexts. Nevertheless, we anticipate two hurdles that could narrow its widespread uptake.
387 First, it may not always be possible to distinguish morphotypes on the basis of intrinsic light-
388 scattering or autofluorescence signatures alone, particularly when hardware with limited optical
389 components (lasers/filters) is being used. Staining the sample with membrane permeable
390 probes could increase the dimensionality of the data (and proved useful in the development of a
391 transcriptomics based, label-free sorting approach (11)). Second, software suitable for these
392 analyses could limit the uptake of IFC-enabled cell sorting. This analysis requires the capacity to

simultaneously inspect the output of cluster and UMAP analyses and image data. As far as we are aware, there is only one proprietary software platform (FCS express) that permits this type of analysis using the imaging cytometer used herein (Attune Cytpix). We encourage the development of open-source equivalents, particularly in languages commonly used by microbial ecologist (e.g., R, Python).

Cell sorting is a first step to understanding morphological heterogeneity in bacterial populations but can only be applied to a fraction to species or contexts where that heterogeneity is exhibited. Our work offers a novel, comparatively inexpensive method that will enable microbiologists to purify – and hence phenotype – bacterial morphotypes from across the tree of life and from diverse environments.

Acknowledgements

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

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 Escherichia coli. 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 Legionella Pneumophila. 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.

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

458 15. Kiepś J, Juzwa W, Dembczyński R. 2023. Imaging Flow Cytometry Demonstrates
459 Physiological and Morphological Diversity within Treated Probiotic Bacteria Groups. Int J Mol
460 Sci 24:6841.

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

464 17. Dey R, Rieger AM, Stephens C, Ashbolt NJ. 2019. Interactions of Pseudomonas aeruginosa
465 with Acanthamoeba polyphaga Observed by Imaging Flow Cytometry. Cytom Part A 95:555–
466 564.

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

470 19. Ebert D, Duneau D, Hall MD, Luijckx P, Andras JP, Pasquier LD, Ben-Ami F. 2016. A
471 Population Biology Perspective on the Stepwise Infection Process of the Bacterial Pathogen
472 Pasteuria ramosa in Daphnia. Adv Parasit 91:265–310.

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

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

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

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

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

486 25. Ebert D, Zschokke-Rohringer CD, Carius HJ. 2000. Dose effects and density-dependent
487 regulation of two microparasites of Daphnia magna. Oecologia.

488 26. Ward MD, Kaduchak G. 2018. Fundamentals of Acoustic Cytometry. Curr Protoc Cytom
489 84:e36.

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

28. 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.

29. 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.

30. 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.

31. 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.

32. 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.

33. 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.

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

35. 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.