An automated Raman-based platform for the sorting of live cells by functional properties

Citation for published version:
Lee, KS, Palatinszky, M, Pereira, FC, Nguyen, J, Fernandez, V, Mueller, A, Menolascina, F, Daims, H, Berry, D, Wagner, M & Stocker, R 2019, 'An automated Raman-based platform for the sorting of live cells by functional properties', *Nature Microbiology*, vol. 4, pp. 1035–1048. https://doi.org/10.1038/s41564-019-0394-9

Digital Object Identifier (DOI):
10.1038/s41564-019-0394-9

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published In:
Nature Microbiology

An automated Raman-based platform for sorting of live cells by functional properties

Kang Soo Lee[1,2], Márton Palatinszky[3], Fátima C. Pereira[3], Jen Nguyen[1,2], Vicente I. Fernandez[1,2], Anna J. Mueller[3], Filippo Menolascina[4], Holger Daims[3,5], David Berry[3], Michael Wagner[3,5] and Roman Stocker[1,2,*]

[1] Ralph M. Parsons Laboratory, Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA

[2] Institute for Environmental Engineering, Department of Civil, Environmental and Geomatic Engineering, ETH Zurich, Zurich, Switzerland

[3] Department of Microbiology and Ecosystem Science, Division of Microbial Ecology, Research Network Chemistry Meets Microbiology, University of Vienna, Vienna, Austria

[4] School of Engineering, Institute for Bioengineering, The University of Edinburgh, Edinburgh, EG9 3DW, UK

[5] The Comammox Research Platform, University of Vienna, Austria

*Correspondence should be addressed to R.S. (romanstocker@ethz.ch)

January, 2019

Abstract

Stable isotope probing (SIP) is widely used to study the function of microbial taxa in their natural environment, but sorting of isotopically labeled microbial cells from complex samples for subsequent genomic analysis or cultivation is still in its early infancy. Here we introduce an optofluidic platform for automated sorting of SIP-labeled microbial cells, combining microfluidics, optical tweezing, and Raman microspectroscopy, which yields live cells suitable for subsequent single-cell genomics, mini-metagenomics, or cultivation. We describe the design and optimization of this Raman-activated cell sorting (RACS) approach, illustrate its operation with four model bacteria (two intestinal, one soil, and one marine), and demonstrate its high sorting accuracy (98.3 ± 1.7%), throughput (200–500 cells/hr; 3.3–8.3 cells/min) and compatibility with cultivation. Application of this sorting approach for the metagenomic characterization of bacteria involved in mucin degradation in the mouse colon revealed a diverse consortium of bacteria, including several members of the underexplored family Muribaculaceae, highlighting both the complexity of this niche and the potential of RACS for identifying key players in targeted processes.

Environmental and host-associated microbiome research aims to understand the composition and function of complex microbial communities. Single-cell studies have become important in this field as they provide unique insights into phylogeny and genomic (micro)diversity[1-4], phenotypic[5-7] and microenvironmental heterogeneity[8], and microbe-eukaryote[9] and microbe-virus interactions[10]. Physiological *in situ* analyses of microbes are also performed at the single cell level[11,12], by introducing an isotopically labeled compound into a microbial community. After a short incubation, the microbial cells that have consumed the compound and have incorporated the isotope into their biomass are detected by microautoradiography (for radioactive isotopes)[13], NanoSIMS[14-16] or Raman microspectroscopy (for stable isotopes)[17-20]. For simultaneous identification of the substrate-consuming cells, all three approaches can be combined with fluorescence *in situ* hybridization (FISH)[21].

Ideally, microbiologists would be able to combine techniques for functional analyses of individual members of complex microbial communities with single-cell genomics to directly retrieve the genomes of those cells that perform a function of interest. However, in most single-cell genomic studies to date[22-25], individual cells are chosen at random or based on genetic criteria using PCR or FISH. Two recent approaches enable the selection of metabolically active microbial cells for whole genome amplification (WGA) and sequencing. One approach relies upon fluorescence labeling of translationally active microbes via bioorthogonal noncanonical amino acid tagging, followed by fluorescence-activated cell sorting (FACS)[26]. The other approach is based on labeling with the stable isotope deuterium (D) during brief incubation with heavy water (D_2O). Labeling using deuterium has several advantages: it is cheap, generally applicable, and can be performed without modifying the natural substrate pool or introducing toxicity effects[27]. After incubation, deuterated cells can be identified via Raman microspectroscopy by inspecting the C-D fingerprint region of the

Raman spectrum (2040-2300 cm^{-1})[27]. This approach, which is non-destructive and requires no fixation of the cells, will detect all metabolically active cells in a sample, for example, those cells active when incubated in the presence of a specific substrate (which does not itself need to be labeled)[27]. This method can be combined with WGA by performing the Raman measurements in a microcapillary and manually sorting the labeled cells to a sterile end of the capillary with optical tweezers[27,28]. However, although Raman microspectroscopy is fast, manual sorting in this approach is very labor-intensive and offers only very low throughput: generally not exceeding 1–2 sorted cells per hour (in samples with a relative abundance of 20–50% deuterated cells).

Here we introduce a fully automated optofluidic platform for Raman-activated microbial cell sorting (RACS) that can analyze up to 200–500 cells per hour (3.3–8.3 cells/min) and is thus appropriate for both function-based cell culturing, mini-metagenomics, or as a front end to standard single-cell genomics platforms. Unlike previous microfluidic RACS approaches[29,30], this method is not limited to cells containing compounds that enhance measurement sensitivity by virtue of Raman resonance (e.g., carotenoids, chlorophyll), and is thus applicable to a wide range of bacterial and archaeal cells, and potentially to eukaryotic cells. After extensive optimization of the platform, it was applied to metagenomically analyze microbes capable of mucin degradation in the mouse colon.

The secreted mucus layer that covers the mammalian intestine acts both as a physical barrier to prevent bacteria from penetrating the epithelium as well as a nutrient source that supports the growth of a subset of gut microbes[31]. The equilibrium between mucus production and degradation can easily be disrupted upon expansion of bacteria that degrade mucin, the primary component in mucus, which can lead to mucus thinning and barrier defects that make the host susceptible to pathogen invasion, inflammation or cancer[32-35]. Despite their importance for health, a comprehensive catalog of bacteria degrading mucin *in vivo* is still lacking[36]. Using RACS we show that a phylogenetically diverse consortium of bacteria is involved in mucin degradation and that members of the underexplored family Muribaculaceae are highly active mucin-degraders in the mouse gut.

Results

Microfluidic sorter, 3D focusing, and system configuration

The RACS platform uses a microfluidic device to capture and move microbial cells in preparation for the Raman measurement and to sort cells after evaluation of their Raman spectra. The device relies on a three-dimensional flow focusing technique to control the position of the sample stream in the polydimethylsiloxane (PDMS) microfluidic sorter. Vertical focusing is first used to focus the sample stream close to the glass coverslip (**Fig. 1a**, **Supplementary Fig. 1** and **Supplementary Videos 1** and **2**). This ensures a high capture rate of cells by the optical tweezers and prevents the complex Raman spectrum of the PDMS lower surface from interfering with the cell's spectrum[37]. Glass rather than quartz coverslips[38] were used after confirming that this results in no interference with microbial spectra in the spectral region of interest. Downstream of the vertical focusing region, the sample stream is focused horizontally by a second sheath flow, which confines it to one side of the channel so that cells flow by default into the waste outlet (**Fig. 1a**, **Supplementary Fig. 2a** and **Supplementary Video 3**).

In the analysis region of the microfluidic sorter (**Fig. 1a**), individual cells are randomly captured by the optical tweezers and their Raman spectrum is measured (**Fig. 1b,c**). A commercial confocal Raman microscope (**Fig. 1d**; see **Methods**) was used, in which two laser sources are focused at the same position through a high numerical aperture, 60× objective. The two lasers can be translocated simultaneously, to allow the capture, movement and Raman measurement of individual cells. To visually monitor cell trapping in the optical tweezers during Raman measurement, which is paramount for confirming the reliability of cell trapping and release in the automated execution of the RACS, a second, imaging microscope was fitted underneath the stage. After focusing the two (Raman and optical tweezing) lasers on the surface of the glass coverslip using CCD camera #1, the beam splitter is removed and the Raman microscope is operated in 'Raman measurement mode' for the duration of the RACS operation, while the cell trapping, transport and release processes are monitored with the imaging microscope (CCD camera #2). The microscope stage is then moved so that the focus lies 10 μm below the glass coverslip within the flow region to begin sorting.

Sorting begins after acquisition of a reference spectrum for the fluid (see below). Each step in the sorting procedure was designed to allow full automation, with no human intervention. The detailed workflow is described in **Fig. 2** and its caption (see also **Supplementary Video 4**). The automation criteria for the different stages are described below.

Cell detection and analysis of deuterium-labeling status

The automation of (i) the recognition of a cell in the optical tweezers and (ii) the decision to release a cell into the collection vs. the waste outlet based on its deuterium-labeling status required the determination of the acquisition times necessary to obtain reliable Raman spectra and criteria to detect cellular and isotopic signals from the Raman spectra. This was achieved and tested using three Gram-negative (two intestinal and one marine) and one Gram-positive (soil) bacterial species: *Escherichia coli*, *Salmonella typhimurium*, *Marinobacter adhaerens*, and *Bacillus subtilis*.

To determine whether a cell is present in the optical tweezers, the magnitude of the Raman signal in the portion of the spectrum between the wave numbers of 1620 cm^{-1} and 1670 cm^{-1}, with integrated intensity $I_{1620\text{-}1670}$ (**Fig. 3a**), was considered. This region was chosen because it is largely unaffected by the Raman spectrum of PDMS and glass (whereas they contribute a detectable signal in other regions, e.g., <1500 cm^{-1} and >2500 cm^{-1}) and is thus insensitive to small variations in the vertical position of the measurement location. Specifically, the 'cell index' P_C was determined

$$P_\text{C} = \frac{I_{1620\text{-}1670}}{I_{\text{Fluid},1620\text{-}1670}}, \qquad (1)$$

where $I_{\text{Fluid},1620\text{-}1670}$ is computed for the fluid reference spectrum. Visual observation using the imaging microscope (CCD camera #2) showed that the presence of a cell in the optical tweezers always results in $P_\text{C} > 1.0$, for each of the bacterial species tested and regardless of their deuterium-labeling status (**Fig. 3c**). Specifically, the following P_C values were measured: 2.05 ± 0.44 (mean ± s.d.; $n = 31$) for deuterium-labeled *E. coli*; 2.86 ± 0.72 ($n = 19$) for unlabeled *E. coli*; 2.94 ± 1.08 ($n = 24$) for deuterium-labeled *B. subtilis;* 4.84 ± 1.73 ($n = 24$) for unlabeled *B. subtilis*; 2.33 ± 0.56 ($n = 31$) for deuterium-labeled *S. typhimurium*; 3.17 ± 0.82 ($n = 28$) for unlabeled *S. typhimurium*; 1.34 ± 0.11 ($n = 22$) for deuterium-labeled *M. adhaerens*; 1.35 ± 0.10 ($n = 16$) for unlabeled *M. adhaerens*. Labeled cells exhibited somewhat lower P_C values than unlabeled cells, possibly due to lower absorption of the laser power at 532 nm by incorporated deuterium (D) than by hydrogen (H)[39]. Considering that the P_C value of the tweezers when empty varied over time, $P_\text{C} > 1.7$ ($P_\text{C} > 1.2$ for *M. adhaerens*) was selected as the criterion for cell capture in the automated algorithm. In a particular application, the user can set the threshold value based on initial trials using CCD camera #2 to monitor capture by the optical tweezers.

Upon capture of a cell, the system moves the optical tweezers containing the trapped cell to the evaluation location (**Fig. 2a**). Here a second Raman spectrum with longer (5 s) acquisition time is measured, to robustly determine whether the cell is labeled with deuterium. Deuterium affects the intensity of the Raman spectrum in the region between the wave numbers 2040 and 2300 cm^{-1} (the 'C-D peak'; integrated intensity: $I_{2040\text{-}2300}$)[27]. Specifically, for each captured cell the 'labeling index' P_{L} was computed

$$P_{\mathrm{L}} = \frac{I_{2040\text{-}2300}}{I_{1850\text{-}1900}}, \qquad (2)$$

to determine whether a cell is labeled, based on comparison with the region between 1850 and 1900 cm^{-1} (integrated intensity: $I_{1850\text{-}1900}$), chosen as a reference because its intensity is low and insensitive to the Raman emission of the device materials (PDMS and glass). Measurements conducted with deuterium-labeled cells and separately with unlabeled cells showed that $P_{\mathrm{L}} > 6.69$ represents a reliable criterion for the identification of labeled cells. For unlabeled cells, all cells measured had $P_{\mathrm{L}} < 6.69$ (n = 19, 24, 28, 16 for *E. coli*, *B. subtilis*, *S. typhimurium*, and *M. adhaerens*, respectively; **Fig. 3d**). Of the labeled cells measured, on the other hand, 96.8% (30 out of 31), 91.7% (22 out of 24), 48.4% (15 out of 31), and 95.5% (21 out of 22) had $P_{\mathrm{L}} > 6.69$ for *E. coli*, *B. subtilis*, *S. typhimurium*, and *M. adhaerens*, respectively. Thus, for *E. coli*, *B. subtilis*, and *M. adhaerens* almost every labeled cell is successfully recognized as such with this criterion, whereas for *S. typhimurium* only one cell out of two was identified as labeled (the criterion was chosen to be conservative, so that unlabeled cells would not be misidentified as labeled and hence collected).

The Raman acquisition times given above were chosen on the basis of an analysis of the effect of the acquisition time on the quality of the Raman spectrum using ten individual *E. coli* cells (**Fig. 3b**). Upon capture, a cell was moved to the evaluation location and its Raman spectrum repeatedly measured with increasingly long acquisition times, from 0.1 to 5.0 s in 0.1 s increments. For each spectrum, P_C and P_L were computed. The cell index P_C saturated for an acquisition time of ~2 s, whereas the labeling index P_L did so for an acquisition time of ~5 s. These acquisition times were thus used in the automated program to detect cell capture (P_C) and labeling (P_L). Using these parameters, we achieved the analysis of 200 cells/h (3.3 cells/min; see below).

Sorting, recovery efficiency, accuracy, and throughput

To automate sorting, we designed a customizable, user-friendly program to repeatedly capture an individual cell from the capture location, move it to the evaluation location, and accept or reject it based on the analysis of its Raman spectrum (**Supplementary Video 4**). The program controls all hardware elements, including the optical tweezers laser, the Raman laser, the optical shutter, CCD camera #1, and the microscope stage; it acquires and analyzes the Raman spectra; and it applies the decision criteria, P_C and P_L, to sort cells.

Based on the thresholds for the two criteria (P_C and P_L), we tested the RACS process with a mixture of deuterium-labeled and unlabeled cells of the same species (1:1 ratio). We examined 185 *E. coli* cells (100 captured and selected; 85 captured and rejected) (**Fig. 3e**), 153 *B. subtilis* cells (84 captured and selected; 69 captured and rejected), and 203 *S. typhimurium* cells (110 captured and selected; 93 captured and rejected) (**Supplementary Fig. 3**). Among 'captured and selected' cells, 99.0% of *E. coli*, 85.7% of *B. subtilis*, and

96.4% of *S. typhimurium* showed a distinct C-D peak. The remaining events registered as selection did not show a typical Raman cell spectrum because cell loss from the optical tweezers had occurred while translocating the cell from the capture location to the evaluation location. P_{L} measured in the absence of a cell in the optical tweezers is higher than 6.69 (i.e., the program wrongly recognizes the fluid in the second Raman measurement as a deuterium-labeled cell as there is a weak signal in the C-D peak region (**Fig. 3a**; inset for P_{L} calculation). We compared $I_{1620\text{-}1670}$ (the region for P_{C} calculation) of the second recorded Raman spectrum of each 'captured and selected' cell with the reference spectrum of the fluid and this confirmed that there was no longer a cell in the optical tweezers during the second measurement in these cases. Dividing the flow stream in the analysis region into collection and waste streams, we identified six possible scenarios resulting from cell loss and theoretically predicted the potential error (collection of unlabeled cells; **Supplementary Fig. 4**). Based solely on the measured spectra during RACS (**Fig. 3e** and **Supplementary Fig. 3**), we estimated that cell loss from the optical tweezers could reduce sorting accuracy by 0.3% (*E. coli*), 5.1% (*B. subtilis*), and 1.2% (*S. typhimurium*) (see **Methods**).

To quantify the sorting accuracy experimentally, we first measured the recovery efficiency (**Supplementary Fig. 5**), defined as the fraction of cells that the program scores as 'collected' that actually enter the collection reservoir. To this end, we ran the RACS system for 1 h with deuterium-labeled *E. coli* cells stained by DAPI. Cells in the collection outlet were then placed on a black filter and the DAPI-stained cells counted by epifluorescence microscopy. Dividing this value by the number of cells recorded as 'collected' by the RACS program yielded a recovery efficiency of 82.1 ± 2.7% (over 3 repetitions). To determine the sorting accuracy, *E. coli* cells not labeled with deuterium were stained with DAPI and mixed 1:1 with deuterium-labeled *E. coli* cells that were not DAPI stained. This mixture was flown in the

RACS device for 1 h and the number of DAPI-stained (i.e., unlabeled) cells in the collection outlet was determined by counting on a filter, as above. This number, representing incorrectly collected cells, was used, after accounting for the recovery efficiency, to estimate the sorting accuracy, which we found to be 98.3 ± 1.7% (over 3 repetitions).

The RACS platform can be customized to sort cells based on parameters other than deuterium incorporation, such as storage compounds and pigments, as long as they are represented by sufficiently large Raman peaks (**Supplementary Fig. 6**). Furthermore, the RACS protocol is also suitable for sorting isotopically labeled living cells for subsequent cultivation (**Supplementary Fig. 7**).

Sorting and sequencing of mucin utilizers from the mouse gut microbiota

To demonstrate the applicability of the RACS platform to complex microbial communities, we first analyzed microbes from mouse colon samples to determine their P_C values. As for the pure cultures (**Fig. 3c**), the P_C criterion (in this case $P_C > 1.1$) successfully identified cells when captured in the optical tweezers, as monitored visually using CCD camera #2 (**Supplementary Fig. 8a** and **Supplementary Video 4**). The mouse gut microbiota consists of hundreds of bacterial species of diverse sizes and morphologies[40]. To investigate if RACS could capture this diversity or was biased by its laser trapping mechanism against certain cell sizes or morphologies, cells were randomly sorted (i.e., using only the P_C criterion) from the colon samples and subjected to 16S rRNA gene amplicon sequencing. From the sorted fraction we recovered 13 out of 20 of the most abundant operational taxonomic units (OTUs; frequencies >1%) present in the sample, which included members of the major phyla in the mouse colon (Bacteroidetes, Firmicutes, Proteobacteria, Verrucomicrobia) (**Supplementary**

Table 1; $n = 68$ cells sorted), demonstrating that RACS can indeed capture a wide variety of cell types occurring within a complex community. Despite this, we could not recover the most abundant OTU in the colon samples, which corresponds to a Gram-positive non-motile coccobacillus, *Blautia marasmi*. However, the bias may not lie in the RACS procedure – discrepancies in the taxonomic composition of amplified genomes compared with data provided by 16S rRNA gene amplicon sequencing have been reported by others[41] and could be due to biases introduced by the cell lysis method employed during WGA. Control cells from the mouse colon microbiota that had been amended with glucose and incubated in the absence of deuterated water always had $P_L < 6.14$ ($n = 130$ cells). Using this threshold value ($P_L = 6.14$) to sort cells that had been amended with glucose but incubated in the presence of deuterated water, 37% (25 out of 68) of the captured cells were recognized as deuterium-labeled and sorted (**Supplementary Fig. 8b**). The sensitivity of automated sorting is comparable to that of the much slower manual sorting in a glass capillary setup[27] (**Supplementary Fig. 9**). Unlike in pure cultures (**Fig. 3d**), the degree of deuterium incorporation can vary considerably among the cells of a complex community due to variability in the response of different species to the added substrate[16].

To investigate the mucin-foraging capability of the mouse microbiota, we sorted 180 deuterium-labeled cells from three different mucin-amended mouse colon samples and metagenomically sequenced the sorted fractions (mini-metagenomics) (**Fig. 4a** and **Supplementary Table 2**). Comparison of the 16S rRNA gene sequences from the sorted fractions with those present in the initial samples revealed that mucin-stimulated organisms constitute 27.4 ± 6.8% of the initial community (under the assumption that no physiological heterogeneity within an OTU exists) (**Supplementary Table 3**). These results are in good agreement with the percentage of captured cells that were directly identified as labeled by the

RACS platform (23.7 ± 9.9%, **Supplementary Table 2**). Mucin-stimulated bacteria were phylogenetically diverse, belonging to four phyla: Bacteroidetes, Firmicutes, Proteobacteria and *Candidatus* Saccharibacteria (**Supplementary Table 3** and **Supplementary Fig. 10**). Phylogenomic analysis of 25 distinct low-contaminated metagenome-assembled genomes (MAGs) recovered from the shotgun-sequenced sorted fractions revealed that mucin-stimulated bacteria were mainly *Bacteroides* spp., most of which were related to organisms belonging to the uncultured family Muribaculaceae (**Supplementary Fig. 10** and **Supplementary Table 4**). Six of the 25 MAGs were closely related to sequenced genomes (as defined by an average nucleotide identity [ANIm] > 96%), but most MAGs had no sequenced close relatives (**Supplementary Table 5**). To confirm that the identified organisms were indeed mucin degraders, we searched the recovered MAGs for the presence of enzymes involved in mucin degradation (**Fig. 4b**). This revealed that most (84%) of these genomes encode at least one of the enzymes necessary to break down mucin *O*-glycans (**Fig. 4c**). None of these enzymes could be identified in the genomes of the recovered members of the Proteobacteria and *Candidatus* Saccharibacteria.

Discussion

For microbiologists, a major limitation to fully capitalizing on the power of single-cell genomics has been the scarcity of generally applicable techniques to sort individual cells from complex communities according to their functional properties. To fill this gap, by directly connecting stable isotope probing with single-cell genomics, we have demonstrated a fully automated tunable optofluidic platform for sorting of physiologically-active microbial cells based on their Raman signal, which allows functional cell sorting with high throughput and very high accuracy. Furthermore, the developed RACS platform is also suitable for sorting live cells by their metabolic activity for subsequent cultivation (**Supplementary Fig. 7**). The sorting throughput was measured in a sample in which 50% of the cells were deuterated, resulting in the collection of 100 cells/h (1.7 cells/min). In additional experiments, it was observed that the system could analyze up to 500 cells/h (8.3 cells/min) when cells do not show significant photophoretic damage upon exposure to a more powerful, 100 mW laser for the Raman measurements: in this case, simultaneous calculation of P_C and P_L values with only a 0.3 s exposure time is possible, which enables an enhanced throughput (e.g., the *M. adhaerens* in **Fig. 3c,d** were measured with these settings; see also **Methods**). While most of the experiments performed in this study focused on detection and sorting of cells that had incorporated deuterium from heavy water, the user-friendly program controlling RACS can be easily adjusted to sort microbial cells according to other characteristics of their Raman spectra, as demonstrated by the sorting of microbial cells showing a cytochrome *c* signature from a marine enrichment culture (**Supplementary Fig. 6**).

We applied the RACS platform to investigate mucin degradation by members of the mouse colon microbiota. We found that diverse bacteria are involved in mucin degradation, with many Bacteroidetes being primary degraders (**Fig. 4c**). There were indications for specialization within the mucin-stimulated community, as only a small subset of identified species, including members of the under-studied Muribaculaceae, encoded enzymes required for cleaving the terminal sialic acid and sulfate residues from the mucin *O*-glycans (see also **Methods**). Applications of the RACS platform like this will open avenues for environmental and host microbiome research and promise to bridge the knowledge gap between the phenotypic and genotypic properties of cells at the individual level. We anticipate that, with this approach, key questions in contemporary microbial ecology, like the contributions of members of the various branches of the tree of life to biogeochemically important processes, can be tackled in an insightful manner.

Methods

Microfluidic device fabrication

The microfluidic sorter (**Fig. 1a**) was fabricated by conventional soft-lithography out of polydimethylsiloxane (PDMS)[42] and fixed to a glass coverslip (60 mm × 20 mm × 0.15 mm). Although CaF_2 and quartz coverslips have been used previously to avoid complicated Raman signals from glass coverslips[38,43], we did not observe significant issues with a regular glass coverslip. The coverslip-side of the device was fixed to the microscope stage facing upward — to prevent interference from the Raman spectrum of PDMS — by a support fabricated out of laser cut acrylic (**Supplementary Fig. 11**). To allow visual monitoring of the RACS process from below by CCD camera #2 (**Fig. 1d**), the layer of PDMS is thin (~1 mm). The depth of the channel is 44 μm and the 'analysis region' measures 300 μm × 350 μm (**Fig. 1a**).

Three-dimensional hydrodynamic focusing

Among the many possible particle focusing techniques (electric, acoustic, hydrodynamic and geometry-induced)[44-49], we chose hydrodynamic focusing for its simplicity and versatility. In our design, a vertical sheath flow hydrodynamically pushes the cells upwards, focusing the sample stream in a region close to the glass coverslip. This ensures a high capture rate of cells by the optical tweezers, which are focused at a point 10 μm below the coverslip and have a tweezing force that attenuates rapidly with distance[39], and it also prevents the complex Raman spectrum of PDMS from interfering with the spectrum of the cells[37], since Raman measurements are acquired >30 μm away from the PDMS lower wall (**Fig. 1a**). The vertically-focused sample stream is then focused horizontally to one side by a second sheath flow. Thus, cells that are not pulled out of the sample stream flow into the waste outlet by default (**Supplementary Videos 3** and **5**).

The final design and flow rates were chosen to optimize vertical focusing, based on simulations performed with COMSOL Multiphysics (**Supplementary Video 1**) and experiments. To experimentally determine the effectiveness of vertical focusing, we seeded the sample stream with 1-μm polystyrene latex microbeads (Polysciences, Inc., USA) and scanned the depth of the microfluidic device — top to bottom — in 2-μm steps using a phase-contrast microscope with 20× objective (Nikon, Japan; **Supplementary Video 2**). We determined vertical focusing effectiveness using:

$$\eta(\%) = \frac{D\text{-}d}{D}\times 100, \qquad (3)$$

where D denotes the depth of the microfluidic device in the analysis region (44 μm) and d the depth of the sample stream as visualized by the latex beads (**Supplementary Fig. 1a**). As the flow rate ratio $Q_{\text{Sheath}} / Q_{\text{Sample}}$ between the vertical sheath flow and the sample flow increases, vertical focusing effectiveness increases up to a maximum of 61% (for $Q_{\text{Sheath}}/Q_{\text{Sample}} = 8.0$). This ratio was used for all RACS experiments.

The final flow rates of the sample fluid, vertical sheath fluid, and horizontal sheath fluid are 0.01 μL/min, 0.08 μL/min and 0.5 μL/min, respectively. The withdrawal rate from the collection outlet is 0.15 μL/min, while the waste outlet empties into a MilliQ-filled Eppendorf tube (see below; Experimental setup).

We noted that the precise location of the hole punched in the PDMS to deliver the vertical focusing flow is important in determining the focusing effectiveness: a misaligned hole causes flow recirculation that disrupts focusing (**Supplementary Fig. 1c**), and too large a hole reduces the focusing effectiveness due to the reduced fluidic momentum (**Supplementary Fig. 1d**).

Experimental setup

We based the system on a commercial Raman microspectroscope (LabRAM HR800, Horiba Scientific, France), which is a modified upright microscope (BX-41, Olympus, Germany). Because Raman signals are intrinsically weak, imaging and Raman measurement must occur in isolation. Thus, the original system had a single optical path and used a removable beam splitter to alternate between two mutually exclusive modes: Raman measurement and microscopic imaging. To enable simultaneous measurement and visualization, the system was modified to create a 'double-microscope setup' (**Fig. 1d**).

The first optical path in the setup uses two continuous-wave neodymium-doped yttrium aluminum garnet (CW Nd:YAG) lasers: one for Raman measurements (532 nm) and one acting as optical tweezers (1064 nm). Both are focused on the same location within the microfluidic device by a 60× water-immersion objective (UPLSAPO 60XW, NA = 1.20, Olympus, Japan). The 1064 nm laser beam fills the back aperture of the objective, ensuring maximal optical tweezing efficiency for a given laser power[50]. After the two lasers have been focused on the glass coverslip (confirmed via inspection with CCD camera #1), a removable beam splitter is withdrawn and scattered light is directed instead to the spectrometer for Raman measurements (**Fig. 1d**). The two lasers are then focused 10 μm into the fluid below the coverslip (i.e., >30 μm above the PDMS surface) for RACS. Two notch filters (cutoff wavelengths: 532 nm and 1064 nm) eliminate laser transmittance into the spectrometer, preventing the laser wavelengths from drowning the Raman signal. A grating divides the Raman signal into discrete wavelengths.

The second microscope yields dark-field images, in which cells appear as bright spots on a dark background. To visualize single cells (~2 µm) captured by the optical tweezers, we mounted a low-angle blue ring illuminator (30°; 470 nm; bandwidth: 450–475 nm; MLRL-CB25, Moritex Corporation, Japan) surrounding a 10× objective (MPlan N, NA = 0.25, Olympus Inc.), an optical array (3.5× magnification; Navitar Inc., USA) and a CCD camera ('CCD camera #2'; PCO1600, PCO-TECH Inc.) below the microscope stage. A shortpass filter with a cutoff wavelength of 500 nm (Thorlabs Inc., USA) is placed after the objective to avoid laser-induced photo-damage to CCD camera #2.

The ability to visualize tweezed cells allowed us to examine single cell capture, translocation, and release events to simulate the automated RACS during the choice of parameters (**Supplementary Videos 6** and **7**). We were also able to detect when captured cells were lost from the optical tweezers. This loss rate, a function of the flow velocity (drag) and the laser power for a given objective, was minimized by maintaining a flow velocity of ~600 µm/s. However, the low flow speed made it difficult to build a stable pressure balance between the collection and waste outlets. To address this issue, a constant flow rate (0.15 µL/min) was withdrawn from the collection outlet. Additionally, to prevent sample fluid from invading the collection outlet during initial flow stabilization before the start of RACS, fluid was first introduced from the collection outlet before reversing the direction of flow. The microfluidic tubing at the waste outlet was connected to an Eppendorf tube filled with working fluid (reservoir).

To prevent any extraneous biological matter from being collected, a syringe filter (0.1-µm pore size, polyethersulfone, Sartorius Stedim Biotech) was equipped at the glass syringe used to generate the horizontal sheath flow, and 0.5% tween 20 (Sigma-Aldrich) in MilliQ water was used in the collection syringe. Tween 20 in MilliQ water was not used at the sample and sheath inlets, because it interfered with Raman measurements.

To allow the stable injection of cells into the microfluidic device and prevent changes of concentration over time due to cells settling in the syringe, we designed an in-house magnetic stirrer (**Supplementary Video 8**) to mix continuously the sample fluid inside the syringe. Magnetic rods were attached to a compact 12V DC motor shaft placed next to the syringe, and a parylene-encapsulated samarium-cobalt magnet disc (bio-compatible; 3 mm in diameter and 0.7 mm in thickness; VP 782S-3, V&P Scientific Inc.) stirred the cell-suspended sample fluid within the syringe. The DC motor (maximum 300 rpm) was tuned using a current controller to give 100–150 rpm.

Four main system parameters affect the Raman measurements: (i) objective: we used a 60×, high NA (1.2) water-immersion objective. (ii) Laser power: maximal laser power is desirable for a strong Raman signal, but high intensity causes photophoretic cell damage (e.g., 750 mW; **Supplementary Video 9**). We thus used a laser power of 15 mW for the 532 nm laser (this was chosen based on *S. typhimurium* because it was most strongly affected by photo-bleaching among the four pure cultures tested). This is equivalent to an optical energy density of 65.5 MW/m^2, given the objective NA of 1.2 and the beam spot size of 0.54 μm. Photo-induced damage of cells exposed to the 400 mW optical tweezers laser (1064 nm) was not observed, which is consistent with a previous report[39]. (iii) Grating: for faster RACS, a grating with 300 lines/mm was chosen, which can sample 8 times faster than a 1800 lines/mm grating for the spectral region of interest (400–3300 cm^{-1}) (**Supplementary Fig. 12**). (iv) Pinhole (located before the grating): after testing Raman measurements with 100, 200, 300, 400, and 500 μm pinhole sizes, the 300-μm pinhole was selected as it produced the highest ratio of cellular signal to background under the conditions applied.

Raman intensity of the optically tweezed cell and background

To compare the Raman intensity coming from the cell and background, we used a geometric optics approach. A rod-shaped bacterium aligned along the direction of beam propagation in the optical tweezers[51,52] and 532-nm laser has Gaussian intensity distribution[53]:

$$I(r, z) = \frac{2P}{\pi\omega(z)^2} \exp\left[-\frac{2r^2}{\omega(z)^2}\right], \qquad (4)$$

where P is the laser power, r is the radial position from the beam center axis and $\omega(z)$ is the radius of the laser beam waist ($z = 0$ refers to the center of laser focus). A diffraction-limited laser beam spot size (ω_o) is calculated from the Rayleigh criterion:

$$\omega_o = 1.22\lambda/NA, \qquad (5)$$

where λ and NA denote the laser wavelength and the objective numerical aperture. The laser beam waist along the z-direction is then expressed as

$$\omega(z) = [\omega_o^2 + z^2(\tan\alpha)^2]^{1/2}, \qquad (6)$$

where $\alpha = \sin^{-1}(NA/n)$ and n is the refractive index of the background fluid. The confocal length (depth of focus) is expressed by[54]

$$CL = s_o \cot(\alpha), \qquad (7)$$

where $s_o = s/M$, M is the objective magnification and s is the pinhole radius. To account for the collection rate of scattered Raman signal at the spectrometer from the local point in the confocal volume, a radial point source collection efficiency is used[54]:

$$\kappa(r, z) \cong \left[1+(1/2)\left(\frac{z}{CL}\right)^2\right]^{-1} \exp\left\{-\left(\frac{r}{s_o}\right)^2 \left[1+(1/2)\left(\frac{z}{CL}\right)^2\right]^{-1}\right\}. \qquad (8)$$

Because Raman signal intensity is proportional to the intensity of the incident photons and inversely proportional to the fourth order of the laser wavelength ($P_s \propto I_o/\lambda^4$), the Raman intensities from the cell and the background fluid in the confocal volume are respectively:

$$I_{\text{cell}} = \int_{-L_o/2}^{L_o/2} \int_{o}^{r_{\text{cell}}} \int_{o}^{2\pi} r\kappa(r, z) I(r, z) \zeta_{\text{cell}} \mathrm{d}\theta \mathrm{d}r \mathrm{d}z, \tag{9-A}$$

$$I_{\text{fluid}} = \int_{-CL/2}^{CL/2} \int_{o}^{\omega(z)} \int_{o}^{2\pi} r\kappa(r, z) I(r, z) \zeta_{\text{fluid}} \mathrm{d}\theta \mathrm{d}r \mathrm{d}z$$

$$- \int_{-L_o/2}^{L_o/2} \int_{o}^{r_{\text{cell}}} \int_{o}^{2\pi} r\kappa(r, z) I(r, z) \zeta_{\text{fluid}} \mathrm{d}\theta \mathrm{d}r \mathrm{d}z, \tag{9-B}$$

where L_o is the cell length. ζ_{cell} and ζ_{fluid} refer to the absorption coefficients of the cell and the fluid, for which we used 0.93 and 0.16, respectively, to simulate *E. coli* suspended in an aqueous medium[55]. In Eqns. (9-A and B), we assumed that the cell has a rod-shape. From these calculations, the *NA* and magnification of the objective and the pinhole size are the main determinants of detection volume (confocal volume) of the confocal Raman microspectroscopy (**Supplementary Fig. 13**). We found that the Raman intensity from a cell is substantially higher than that from the background fluid, allowing the RACS system to sort based on cell signals. With a 2.0 μm × 0.5 μm rod-shaped *E. coli* or a 0.75 μm spherical-shaped coccus cell, for example, approximately 85% of the Raman signal originates from the cell and only 15% from the surrounding medium ($I_{\text{cell}}/I_{\text{fluid}} = 6.22$ and 6.13 for *E. coli* and coccus cell, respectively) when the 60× 1.2 NA objective and 300-μm pinhole size are used. Spectrum acquisition can be optimized for different cell types by changing the optical pinhole.

Sample preparation

Generally, we followed the procedure described by Berry *et al.* for sample preparation[27]. Four bacterial model species were used: *Escherichia coli* (non-motile mutant; NCM3722 Δ*motA*), *Bacillus subtilis* (wild type; OI1085), *Salmonella typhimurium* (wild type; LT2 TSS495), and

CFP-labeled *Marinobacter adhaerens* (wild type; HP15 eCFP). For *E. coli*, *B. subtilis*, and *S. typhimurium*, a single colony was picked from a Luria broth (LB) agar plate, inoculated into 3 mL of LB medium, and cultured overnight in a shaking incubator (300 rpm and 30°C). The next day, cells from the culture were diluted 1:200 into fresh media made with either 0% (i.e., 100% MilliQ) or 50% D_2O LB medium. Cells were then cultured overnight. For *M. adhaerens*, 2216 medium (BD Difco^{TM}) containing 50 μg/ml ampicillin was used. The media were either filter sterilized (0.2-μm pore size) or autoclaved before use. For experiments, cells were harvested from stationary phase, rinsed three times in MilliQ water by centrifugation (3 min at 7,000 rpm), and re-suspended in MilliQ water. For *M. adhaerens*, artificial seawater was used to rinse and re-suspend.

For the mouse colon microbial community sample, colon contents of 6-8 week old C57BL/6J mice (*n* = 3) were harvested in an anaerobic tent. Contents from both male and female mice were collected and pooled. Sample size was determined based on the amount of colon contents required to perform each amendment and all necessary controls in three replicates. Colon contents were suspended in either non-D_2O-containing phosphate buffered saline (PBS) or 50% D_2O-containing PBS, homogenized by vortexing, and 2.5 mg/mL glucose or 2.5 mg/mL mucin from porcine stomach (both from Sigma-Aldrich) added. After incubation for 6 h at 37°C under anaerobic conditions, glycerol was added (to achieve a final 20% (v/v) of glycerol in the microcosms) and the vials were crimp-sealed with rubber stoppers and stored at -80ºC until further processing. The 6-hour incubation led to only minor shifts in the microbiota composition (**Supplementary Fig. 14**). For experiments, these glycerol-cryopreserved cells were rinsed in PBS by centrifugation (3 min at 7,000 rpm), and re-suspended in 0.2 M glycerol (MilliQ balanced). Sorting of cells from colon microbiota incubated for 6 h in 50% D_2O-containing PBS in the absence of supplement (neither glucose nor mucin) shows that only 0.7% of cells (1 out of 134) display residual metabolic activity in the absence of an added substrate (**Supplementary Fig. 15**). For the mucin amendments, sorted fractions (a total of 7 fractions collected) were obtained as replicates from 3 biologically distinct microcosms: microcosms MucA, MucB and MucC (**Supplementary Table 2**). Animal experiments were approved by the Institutional Ethics Committee of the University of Veterinary Medicine, Vienna, and conducted in accordance with Austrian laws (BMWF-66.006/0002-II/10b/2010).

For the marine enrichment culture, a marine sediment surface sample was collected from the Pacific Ocean off the coast of Vancouver, Canada. A small spatula of the collected sediment was transferred to marine mineral salt medium, incubated in the dark at 28°C without agitation, and repeatedly provided with 1 mM $NaNO_2$ upon depletion to enrich for nitrite oxidizing cells. The marine mineral salt medium was modified from that described by Lücker *et al.*[56] by replacing the natural seawater-water mix with 1 L MilliQ containing 33.4 g red sea salts (Red Sea Aquatics Ltd.). The medium was supplemented with 4.2 µL vitamin solution per L medium (0.02 g/L biotin, 0.02 g/L folic acid, 0.10 g/L pyridoxine HCl, 0.05 g/L thiamine HCl, 0.05 g/L riboflavin, 0.05 g/L nicotinic acid, 0.05 g/L DL pantothenic acid, 0.05 g/L P-aminobenzoic acid, 2.00 g/L choline chloride, 0.01 g/L vitamin B_{12}). Enrichment was performed by transferring 10% of the culture into fresh medium four times before the RACS experiment. Cells were harvested by centrifugation (4500g for 20 min at 28°C, swinging rotors) and re-suspended in 1 mL of artificial seawater for the RACS experiment.

Interface for automated RACS

We employed MATLAB’s (Mathworks, USA) graphical user interface development environment (GUIDE) and ActiveX to bridge MATLAB and LabSpec6, the manufacturer software of the Raman microspectroscope (Horiba Scientific, France). As shown in **Supplementary Fig. 16a**, the program window of the platform allows the user to: (1 & 2) switch CCD camera #1 on and off, (3) start the calibration process, (4) start the automated RACS process, (5) set the threshold of P_C, (6) set the threshold of P_L, (7) view the measured P_C value in real time, (8) view the measured P_L value in real time, (9) track the number of analyzed cells, (10) track the number of captured and selected cells (identified as of interest according to the P_L criterion), (11) track the number of captured and rejected cells (identified as not of interest according to the P_L criterion), (12) view the proportion of selected cells among the analyzed cells, (13) determine the stage position in the *y*-direction (perpendicular to the flow direction), (14) stop the RACS process, (15) close the program window, (16) view the real time spectrum from the LabSpec6 monitor, (17) view the calibration spectrum, and (18) view the measured spectrum upon cell capture in the optical tweezers.

When the user clicks button (1), CCD camera #1 is activated and the Raman (532 nm) and optical tweezers (1064 nm) lasers are visualized. Once the laser spots are focused at the bottom surface of the glass coverslip, CCD camera #1 can be disengaged by clicking button (2) and the system can be switched into Raman measurement mode as the user removes the beam splitter (**Supplementary Fig. 16b**; see **Fig. 1d**). Upon clicking button (3), the program starts the calibration process: the optical tweezers move to the sample-free stream (270 µm away from the initial position) and measure 20 Raman spectra of fluid (background) with a 2-s acquisition time. The measured spectra are then averaged, displayed at window (17), and used as a control to detect single cells captured by the optical tweezers. Upon clicking button (4), the program starts the RACS process: a Raman spectrum with 2-s acquisition time is taken every 2 s and the calculated P_C value is displayed at window (7) in real time. When the program recognizes an optically captured cell (on the basis of the P_C criterion), the spectrum is shown at window (18), and the optical tweezers move to the evaluation location (their current lateral position is displayed at window (13)). After calculating P_L using a new Raman spectrum (5-s acquisition time), the program displays the P_L value at window (8) and makes a decision according to the threshold set by the user at window (6). RACS is designed to be adjustable for diverse samples and molecules of interest, so that the acquisition times and thresholds must be set on the basis of experiments with control and test samples. Window (5) allows the user to adjust P_C to specific strains and experimental conditions (*e.g.,* objective specification and detection volume). If the cell is selected according to the criterion (here, identified as deuterium-labeled based on $P_L > 6.69$ for the four test species), the program closes the optical shutter — blocking the tweezers laser — and the flow carries the released cell into the collection outlet. The program then returns the optical tweezers back to their initial position, and opens the optical shutter, starting the RACS process over. Otherwise, if the cell is to be rejected, the program will move the optical tweezers back to their initial position and then close the optical shutter, so that the flow carries the released cell into the waste outlet. The program then opens the optical shutter and starts over. All spectra are automatically stored, allowing for population characterization by subsequent statistical analysis. The program continues automated collection until the user clicks button (14). Button (15) closes the connection between MATLAB and LabSpec6.

Recovery efficiency, sorting accuracy, and live cell sorting

We evaluated recovery efficiency (the percentage of sorted cells as counted by the program that can be retrieved from the collection outlet for subsequent experiments) and sorting accuracy (the percentage of sorted cells that are actually labeled) with separate experiments (**Supplementary Fig. 5**). To estimate recovery efficiency, overnight cultured deuterium-labeled *E. coli* were stained with DAPI (1:100 concentration; 358/461 nm for excitation/emission; this does not interfere with Raman measurements with the 532-nm laser; Thermo scientific, USA) for 10 min at room temperature to allow detection by epifluorescence and used as a sample. After running the RACS procedure for 1 h, the collected cells were deposited on a black filter (0.2-μm GTBP Isopore$^{\text{TM}}$ membrane filter, Merck Millipore Ltd., Ireland). Deposition onto a small area (~1 mm diameter) was achieved by directly contacting a needle (connected to the tubing) and the black filter, over a vacuum-assisted glass filtration apparatus. We then scanned the area with epifluorescence microscopy (LMD 7000, Leica), stitched the images together, and counted the number of cells visible. We calculated the recovery efficiency:

$$\gamma_{\text{L}} = R_{\text{L}}/N_{\text{L}} \,, \qquad (10)$$

where R_{L} and N_{L} are the number of cells counted on the black filter (i.e., successfully recovered) and identified as deuterium-labeled by the software, respectively.

To measure sorting accuracy, unlabeled *E. coli* cells were stained with DAPI (1:100 concentration) for 10 min at room temperature and mixed 1:1 with deuterium-labeled cells. The procedure described above for the recovery efficiency was then conducted. In this case, stained cells counted by epifluorescence on the black filter represented false positives, for which the cells had likely fallen from the optical tweezers before being moved back to the waste stream. Factoring in the measured recovery efficiency, sorting accuracy was calculated by:

$$\epsilon = \frac{N_{\text{L}}\gamma_{\text{L}} - R_{\text{U}}}{N_{\text{L}}\gamma_{\text{L}}} \,, \qquad (11\text{-A})$$

where R_{U} represents the number of unlabeled (DAPI-stained) cells recovered on the black filter. Hence, a 0% sorting accuracy means that cells observed on the black filter are all deuterium-unlabeled. To estimate the sorting accuracy while accounting for the mixing ratio between deuterium-labeled and -unlabeled cells, we used the parameter, $\gamma_{\text{U}} = R_{\text{U}}/N_{\text{U}}$ (where

N_U is the number of 'captured and rejected' cells displayed in the software window) to convert Eqn. (11-A) into a population ratio-dependent form:

$$\epsilon = \frac{N_L\gamma_L - R_U}{N_L\gamma_L} = 1 - \frac{\gamma_U}{\gamma_L}\frac{N_U}{N_L} = 1 - \frac{\gamma_U}{\gamma_L}\frac{C_U}{C_L}\ , \qquad (11\text{-B})$$

where C_L and C_U are population densities of the deuterium-labeled and -unlabeled cells in the sample flow.

To demonstrate that the RACS protocol is also suitable for sorting isotopically labeled living cells for subsequent cultivation (without significant photophoretic damage by the 532-nm Raman laser exposure), we performed additional experiments and evaluated the recovery efficiency by cell culture (**Supplementary Fig. 7**). We ran the RACS system for 1 h with deuterium-labeled, CFP-labeled *M. adhaerens* cells (CFP has excitation and emission wavelengths of 434 nm and 479 nm, respectively, and thus does not interfere with Raman measurements with the 532-nm laser), instead of DAPI staining (that is typically used for staining fixed cells). Cells recovered from the collection outlet were spread on a 2216 agar plate and colonies grown overnight at 37°C were counted by epifluorescence microscopy. Dividing counts of growing cells by the number of cells recorded as 'collected' by the RACS program (Eqn. (10)) yielded a recovery efficiency of viable cells of 81.8 ± 5.9% (over 3 replicates).

Error analysis

To determine the effects of cell loss from the optical tweezers on sorting accuracy, we visualized the flow pattern in the microfluidic device by injecting a cell-suspended fluid simultaneously into all three inlets (the sample inlet and both sheath inlets) (**Supplementary Fig. 2b** and **Supplementary Video 5**). We found that the fate of cells lost from the tweezers depends on where during the 270 μm translocation (between capture and evaluation;

Supplementary Fig. 4a) the cells are lost. Flow carries cells lost within the first 190 μm into the waste outlet, while those lost within the final 80 μm are carried into the collection outlet. Given this, we classify cell loss into six cases (**Supplementary Fig. 4)**. In two cases (4 and 5), it is possible for cells lost in the collection stream to end up collected as false positives. The probability of these two cases are expressed as $N(1\text{-}\alpha)f_L(1\text{-}f_D)$ (the same for each case), where N is the number of 'captured and evaluated' cells, α is the proportion of the travel distance where flow ends in the waste outlet (i.e., waste stream), f_L is the cell loss frequency, and f_D is the frequency of deuterium-labeled cells among the population, N. From our estimation of cell loss from the optical tweezers, f_L and f_D can be expressed as $\left(N_a+\frac{N_c(1\text{-}\Omega)N_a}{N}\right)/N$ and $(N_b+\Omega N_a)/N$, respectively, where N_a is the number of cells lost in cases 1–4 (i.e., cells lost before they could be evaluated), N_b is N_a subtracted from the number of 'captured and selected' cells, N_c is the number of 'captured and rejected' cells, and Ω is the proportion of deuterium-labeled cells within population N (i.e., $N_b/(N_b + N_c)$). Based solely on the measured spectra during RACS (**Fig. 3e** and **Supplementary Fig. 3**), we computed that cell loss from the optical tweezers could reduce the sorting accuracy by up to 0.3% (185 *E. coli* cells; $N_a = 1$, $N_b = 99$, $N_c = 85$), 5.1% (154 *B. subtilis* cells; $N_a = 12$, $N_b = 72$, $N_c = 69$), and 1.2% (203 *S. typhimurium* cells; $N_a = 4$, $N_b = 106$, $N_c = 93$) under the assumption that N_b contains only deuterium-labeled cells. These predictions assume a constant probability of cell loss from the optical tweezers along their travel path, whereas during measurement we found that the probability is in fact higher at the onset of the tweezer movement – in this region, lost cells pass into the waste outlet, thus decreasing the actual frequency of wrongly sorted cells.

Evaluation of sorting throughput

The sorting throughput depends on the waiting time (the time needed to capture a cell in the optical tweezers, which in turn depends on the cell concentration in the sample), the analysis time (the time for the quantification of P_C and P_L, and for the movement and return of the stage between the capture and evaluation locations), and the probability of a successful Raman measurement for the evaluation of P_C at the capture location without loss of the cell from the tweezers due to collision with another cell.

To evaluate throughput, we first modeled the probability of cell capture and successful P_C measurement as a function of the rate at which cells enter the optical tweezers and the measurement time for each Raman measurement. The rate at which cells enter the optical tweezers (cells/s) is $\delta = C\,Q\,a\,/\,A$, where C is the cell concentration in the sample (cells/mL), Q is the sample flow rate (here, 5.4×10^{-7} mL/s for the sample flow velocity in our experiments, $U = 600$ µm/s), a is the cross-sectional area of the optical tweezers (here, 2 µm × 2 µm, obtained by using the laser beam waist, $2\omega_o = 2$ µm), and A is the cross-sectional area of the sample flow (here, 53 µm × 17 µm; see **Supplementary Figs. 1** and **2**). Assuming that every cell entering the optical tweezers is captured, the probability P of successful Raman measurement of the captured cell to determine whether a cell is present in the optical tweezers (i.e., the evaluation of P_C) depends only on the timing of the cell's arrival with respect to the measurement and whether another cell collides with the captured cell during the measurement. Assuming that the cell arrivals are independent and that the process is stationary, we use a Poisson process model to calculate the probability P of a cell capture and successful measurement of its P_C value, which yields $P = e^{-2\delta TC} \sinh(\delta T_C) / [1 - e^{-\delta TC}\{\cosh(\delta T_C) - \sinh(\delta T_C) - 1\}]$, where T_C is the measurement time used to quantify P_C ($T_C = 2$ s in our experiments). The probability increases with cell concentration up to a maximum of 0.155 (regardless of T_C) as the rate δ of cell entry into the optical tweezers increases, but then decreases for further increases in δ due to the higher rate of collision between the cell in the tweezers and other cells in the sample flow (**Supplementary Fig. 17a**).

The overall sorting throughput of the RACS system (number of cells collected per hour) is $R = [1 / (T_C / P + T_S)]\kappa \times 3{,}600$, where κ is the fraction of cells of interest (in this case, deuterium-labeled) within the sample and T_S is the sorting time, which comprises the time to compute P_L (5 s in our experiments) and the time to move the stage back and forth between the capture and evaluation locations (2 s in total in our experiments). For the values used in our experiments ($T_C = 2$ s, $T_S = 7$ s), the sorting throughput is predicted to be $R = 4$, 33, and 90 cells/h for cell concentrations of 10^6, 10^7, and 10^8 cells/mL, respectively, when κ for example is 50% (**Supplementary Fig. 17b**). These values decrease to $R = 0.8$, 7, and 18 cells/h for the same cell concentrations, when κ is 10%.

The rate-limiting step in the RACS sorting approach depends on the cell concentration. At low cell concentrations, the rate-limiting step is the entry of a new cell into the optical tweezers. For example, at 10^6 cells/mL a new cell enters the tweezers only every 6.9 min. In this case a higher flow rate would increase the sorting throughput, for example from $R = 4$ to 19 cells/h when the sample flow velocity is increased from $U = 600$ μm/s to 3000 μm/s and κ is 50% (**Supplementary Fig. 17a**). We note, however, that stronger flows also increase the probability of cell loss from the optical tweezers due to the increased drag on the captured cell (**Supplementary Fig. 4**). Pre-concentrating cells increases throughput, but only up to a certain point, because the rate of cell collisions also increases: for example, for 5×10^8 cells/mL, the predicted throughput drops to $R = 33$ cells/h for $\kappa = 50\%$ (compared to 90 cells/h for 10^8 cells/mL; **Supplementary Fig. 17b**).

In contrast, at high cell concentrations the rate-limiting step is the analysis time. At 10^8 cells/mL, a concentration of the same order as that used in our experiments, a cell enters the optical tweezers every 3.8 s and P is at its maximum value (**Supplementary Fig. 17a**), which means the system achieves its maximum sorting throughput for the given values of T_C and T_S. In this case, shorter measurement times T_C and T_S would yield a higher sorting throughput R (**Supplementary Fig. 17c**). This could be achieved by improvements to the instrument, for example by increasing the power of the laser or the sensitivity of the Raman signal detector (see below). If the system could be improved to reduce T_C to 0.1 s (retaining $T_S = 7$ s), the throughput could increase from $R = 90$ to 157 cells/h for $\kappa = 50\%$ at a concentration of 10^8 cells/mL. Moreover, if the system were to allow the simultaneous evaluation of P_C and P_L values at the capture location (e.g., $T_S = 2$ s, which comprises only the movement of the stage), the predicted throughput becomes $R = [1 / (T_C / P + T_S / \kappa)] \times 3{,}600$. In this case, κ affects T_S exclusively because only the cells of interest are moved to the evaluation location, which becomes simply a release location for the collection of sorted cells, and stage movement for non-labeled cells is avoided. With this approach, R increases considerably, to 425 cells/h for $\kappa = 50\%$ (with $C = 10^8$ cells/mL and $T_C = 0.1$ s). More importantly still, with this approach, sorting becomes feasible also for samples with very low proportions of cells of interest: for example, when only $\kappa = 1\%$ of the cells are of interest (for $C = 10^8$ cells/mL), R increases from 2 cells/h (with the current approach, i.e., $T_C = 2$ s and $T_S = 7$ s) to 18 cells/h (for $T_C = 0.1$ s and $T_S = 2$ s) (**Supplementary Fig. 17c**).

Flexibility of the RACS platform for the operation with other parameters

The RACS platform can easily be customized to sort cells based on parameters other than deuterium incorporation such as storage compounds, pigments and other compounds, as long as they are represented by sufficiently large peaks in the cellular Raman spectra. This capability was demonstrated by sorting cells from a marine enrichment culture (see above for sample preparation) that included cells with a high cytochrome *c* content as indicated by the presence of the typical four strong Raman peaks[57] of the electron transfer proteins at 750, 1127, 1314, and 1585 cm^{-1}. Initially, the RACS platform was tuned to detect single cells from this enrichment that were randomly captured in the optical tweezers (in this case, using $P_{\mathrm{C}} > 1.1$). In the next step the RACS program was customized to identify and sort cells containing cytochrome *c* from within the complex enrichment community by using the spectral region 750–760 cm^{-1} as marker, as it showed the strongest signal intensity (**Supplementary Fig. 6a**). For all captured cells, $P_{\mathrm{cytochrome}}$ was calculated:

$$P_{\mathrm{cytochrome}} = \frac{I_{750\text{-}760}}{I_{1850\text{-}1900}}. \qquad (12)$$

As for the P_{L} determination, integrated intensity $I_{1850\text{-}1900}$ was used as reference region. We first classified cells by hand as containing cytochrome *c* or not on the basis of their full Raman spectra (**Supplementary Fig. 6a**) and used this manual binning to determine a threshold value of $P_{\mathrm{cytochrome}} > 0.6$ (**Supplementary Fig. 6b**). The platform correctly identified and sorted 75.8% (55 out of 66 cells) of cells that contained cytochrome *c*, while it rejected all cells that lacked cytochrome *c* (i.e., zero false-positives).

Further sorting criteria can include signals from cells containing storage compounds like elemental sulphur[58], compounds that induce Raman resonance with the 532 nm laser like carotenoids[59], and ^{13}C-isotope labeled derivatives[19] after stable isotope probing experiments. Moreover, although not experimentally demonstrated, multiplexing of different Raman signals as sorting criterion is also possible – e.g., to collect only cells that display both a cytochrome *c* and a deuterium signal.

Preparation of 16S rRNA gene amplicon libraries and 16S rRNA gene sequence analyses

DNA extracted from the mouse colon microcosms using a phenol-chloroform bead-beating protocol[60] or obtained by whole genome amplification (WGA) of sorted cells was used as a template for PCR. PCR amplification was performed with a two-step barcoding approach[61]. In the first-step PCR, the following oligonucleotide primers targeting the 16S rRNA gene of most bacteria were used as forward and reverse primers, respectively: S-D-bact-0341-b-S-17 (5-CCTACGGGNGGCWGCAG-3′) and S-D-bact-0785-a-A-21 (5-GACTACHVGGGTATCTAATCC-3′). These first step oligonucleotide primers contained head adaptors (5′-GCTATGCGCGAGCTGC-3′) in order to be barcoded in a second step PCR. Barcode primers consisted of the 16 bp head sequence and a sample-specific 8 bp barcode from a previously published list at the 5′ end[62]. The barcoded amplicons were purified with the ZR-96 DNA Clean-up Kit (Zymo Research, USA) and quantified using the Quant-iT PicoGreen dsDNA Assay (Invitrogen, USA). An equimolar library was constructed by pooling samples, and the resulting library was sent for sequencing on an Illumina MiSeq platform at Microsynth AG (Balgach, Switzerland). Sequences were quality-filtered, paired end reads were concatenated and reads were then clustered into species-level operational taxonomic units (OTUs) of 97% sequence identity as described in Herbold *et al.*[61]. Sequencing libraries were rarefied using the vegan package 2.4-3 of the software R (https://www.r-project.org/).

Mini-metagenome sequencing and analyses

Labeled RACS cells were collected into PCR tubes, lysed and subjected to WGA using the Repli-g Single Cell Kit (QIAGEN), according to the manufacturer's instructions. Shotgun libraries generated using the amplified DNA from WGA reactions as a template and Nextera XT (Illumina) reagents were sequenced with a HiSeq 3000 (Illumina) in 2 × 150 bp mode at the Biomedical Sequencing Facility, Medical University of Vienna, Austria. The sequence reads were quality-trimmed and filtered using AdapterRemoval v2.1.7[63]. The remaining reads were assembled *de novo* using SPAdes 3.11.1[64] in single-cell mode (k-mer sizes: 21, 35, 55).

Binning of the assembled reads into metagenome-assembled genomes (MAGs) was performed with MetaBAT 2 (v2.12.1)[65] using the following parameters: minContig 2000, minCV 1.0, minCVSum 1.0, maxP 95%, minS 60, and maxEdges 200. The quality and contamination of all MAGs were checked with CheckM 1.0.6[66] (**Supplementary Table 4**). MAGs >200 kb obtained from all samples were compared and de-replicated using dRep 1.4.3[67]. Automatic genome annotation of contigs >2 kb within each de-replicated MAG was performed with RAST 2.0[68]. In order to identify which OTUs from the initial mucin-amended microcosms were present in the sorted fractions, a BLAST custom-made database composed of the OTU representative sequences detected in the initial microcosms was generated. Total assemblies from sorted samples were then queried against the custom database with the following parameters: E-value < 1e-98 (to exclude sequence alignments shorter than ~200 base pairs) and percent identity of 97% (**Supplementary Table 3**).

Phylogenomic analyses

An initial reference tree was generated using the maximum likelihood algorithm of FastTree 2.1.10[69] using an alignment of 43 concatenated single-copy marker protein sequences from >7,000 available reference genomes. MAGs were placed onto the reference tree using the evolutionary placement algorithm of the RAxML8.2.11 package[70]. Based on the distribution of the MAGs within this tree, 135 reference genomes were selected for a simplified tree. Query MAGs were again placed onto this simplified tree, as described above. Phylogenomic trees were visualized and formatted using iTOL v4 (https://itol.embl.de/). In order to identify the closest relative for each MAG, the query MAG and close reference genomes (based on the generated phylogenomic tree) were compared using dRep[67]. Compared genomes with a whole-genome based average nucleotide identity (ANIm[71]) >99% were considered to be the same organism. Only closely related genomes with ANIm >96% were considered for **Fig. 4c**.

An alignment coverage cutoff of 0.1 was applied for all genome comparisons.

Identification of mucin utilizers

The RACS platform is highly versatile in its potential applications. It enables direct access to the genomes of microbes from complex microbial communities that respond metabolically to substrate additions or changes in other environmental conditions of interest. In this study the RACS platform was applied to investigate mucin degradation by members of the mouse colon microbiota. Mucin is a highly-glycosylated protein secreted by the mammalian colon that acts as an important barrier between the intestinal tissue and the microbiota, but which is also used as a nutrient source by gut bacteria[31]. Mucin degradation, mediated by the secretion of enzymes such as neuraminidases, α- and β-galactosidases, N-acetylglucosaminidases, L-fucosidades and β-N-hexosaminidases, is thought to be confined to a small group of specialized organisms[36]. Though genes involved in the partial or complete degradation of mucin are present in the genomes of a large fraction of gut inhabitants[72], only a small fraction of genes in the microbiota are transcribed *in vivo*[73], and it therefore remains unclear how broad the mucin degradation guild is *in situ*. Our results suggest that mucin degradation is a process mediated by a phylogenetically diverse group of bacteria, though primary degraders belong largely to the Bacteroidetes (**Fig. 4c**). Interestingly, there appears to be specialization within the mucin-stimulated community, with only a small subset of species encoding enzymes necessary to cleave the terminal sialic acid and sulfate residues from the mucin *O*-glycans (**Fig. 4c**; enzymes EC 3.1.6.14 and EC 3.2.1.18). Sialylated and sulfated glycans are abundant in mouse colon mucins[74], and therefore sialidases and mucin-desulfating sulfatases provide bacteria access to a fraction of mucin glycans inaccessible to the majority of potential mucin degraders. Organisms belonging to the family Muribaculaceae (**Fig. 4c**; Muribaculaceae bacterium MAG RACS_021 and Muribaculaceae bacterium MAG

RACS_016) provide examples with this capacity. Members of the family Muribaculaceae are found in the guts of homoeothermic animals where they are thought to play an important role in modulating the host's health[75,76]. Previously, insights about the role and niche occupied by these abundant components of the mouse gut microbiota were based solely on genomic analyses or *in vitro* phenotypic characterization of a single isolate[40,77]. Here we have shown that members of this important family are stimulated by mucin, with at least two being part of a bacterial "elite" that can degrade nearly all of the *O*-glycans that constitute the mucus layer. Bacteria not able to cleave glycans were also identified, suggesting cross-feeding of sugars liberated by saccharolytic bacteria or utilization of their metabolic products. Together, these results suggest a model in which mucin degradation is mediated by a diverse group of primary degraders that cooperatively degrade this complex glycoprotein, and concomitantly support a community of cross-feeding heterotrophs as well as organisms such as sulfate-reducing bacteria that use the end products of mucin fermentation.

Potential for the improvement of sorting throughput

The throughput of the microfluidic RACS device is far greater than that of manual Raman sorting[27,28]. Throughput is key to the value of the device, and the throughput required for practical applications simply cannot be achieved through manual means. To compare the throughput and capability of manual vs. automated sorting, the following example provides valuable insights. If we consider the abundance distribution of OTUs detected as labeled in the mucin incubations (**Supplementary Table 1**), we can ask how long it would take to detect all the labeled OTUs (OTUs with over 1% relative abundance, and assuming 75% of the cells from each OTU are labeled). Under these conditions, our automated platform sorting at 200–500 cells/h (3.3–8.3 cells/min) could retrieve all of the labeled taxa with 95% probability within 2–5 hours, while manual sorting at 1–2 cells/h would require 265–530 hours (or 33–66 eight-hour days of nonstop sorting) (**Supplementary Fig. 18**), which is certainly a prohibitive level of effort for nearly all projects. Throughput could be further increased in the future via system optimization. Theoretical calculations (see above) suggest that the collection of labeled cells could increase considerably, to 425 cells/h (for a sample cell concentration $C = 10^8$ cells/mL and 50% deuteration) if the system were to perform simultaneous evaluation of P_C and P_L with 0.1 s exposure time. Cells collected by RACS can directly be cultivated or injected into a FACS device for single-cell isolation into multiwell plates for subsequent molecular analysis. As there is no chemical treatment before or during RACS, the cells are compatible with whole genome amplification (WGA) and PCR[27].

Sample preparation for other microbiomes

Similar to the requirements of conventional FACS in the standard single-cell genomics pipeline, the microfluidic RACS system requires the presence of individual cells in solution. For aquatic samples, after heavy water incubation, microbial cells should be harvested by centrifugation and washed in order to remove D_2O, then re-suspended in water or an isotonic solution (that does not interfere with Raman measurements) before injection into the RACS device. For soil samples, suitable protocols for heavy water labeling and extraction of microbial cells have been developed[78]. For other samples such as sediments, fecal samples, biofilms and sludge, additional sample preparation to disaggregate the cells and remove non-cellular particles is required[22]. For environmental studies, the sorting criteria will need to be optimized on a case-by-case basis by measuring control samples. There is naturally a tradeoff between the sensitivity in detecting less-highly-labeled cells and the minimization of false positive sorts, so the sorting criteria chosen will depend on the specific requirements of each study.

Advantages of the platform in minimizing contamination risk

Avoiding contamination with cells not originating from the analyzed sample or with extracellular DNA from the sample is a general challenge for single-cell genomics. Our microfluidic RACS protocol minimizes the risk of contamination. Firstly, only deuterated cells are sorted, and due to the high sorting accuracy, this largely excludes contamination by other cells. Secondly, the liquid used for flow focusing can be autoclaved and treated with DNAses before use, and passed through polyethersulfone syringe filters (0.1-μm pore size) when leaving the syringe pumps to remove any contaminating cells. Thirdly, cell spectra are measured outside of the sample stream in flow and thus all sorted cells undergo a rigorous washing while held individually in the optical tweezers, minimizing the risk of carrying extracellular DNA on their surface. Finally, production of our microfluidic RACS device is cheap and fast, and thus a new device can be used for each experiment to avoid cross-contamination between experiments.

Technical discussion and potential for improvement

In this study, we identified several sources of potential error that should be considered when extending or adapting this technique. (i) There is a discrepancy between the optical tweezers (1064 nm) and Raman (532 nm) laser foci (of ~1.0 μm), as our objective only corrects for chromatic aberrations in the visible wavelengths. This discrepancy reduces the Raman intensity collected from the captured cell. This could be overcome by using an objective precisely corrected over visible-IR wavelengths. (ii) The MilliQ water used to suspend the cells during sorting produces a weak Raman signal that overlaps with the CD peak (**Fig. 3a**; inset for P_L calculation), possibly reducing the sensitivity if weakly deuterium-labeled cells are targeted. (iii) In applications with natural microbial communities, different cell populations may require different threshold values for P_C and P_L, and the sample may contain non-cellular particles. From our observations, such particles often generate large P_C values. These can be screened out by adding an upper bound — for example, $1.7 < P_C < 7.0$ — permitting the analysis of captured cells while particles are automatically released without further measurement. (iv) The program could be modified so that the Raman measurement for the P_L calculation took place within the part of the sample-free region where a cell accidently lost from the optical tweezers would be carried by flow into the waste outlet, and only a cell recognized as deuterium-labeled would be moved to a releasing location where the flow would carry it into the collection outlet. The error which stems from cell loss in the optical tweezers during the translocation between capture location and evaluation location could then be eliminated. (v) For the mouse colon microbial community sample (**Supplementary Fig. 8**), we found that the P_L value of cells in the control case increased over time (**Supplementary Fig. 19**), possibly as a result of osmotic stress. When a slightly more strict (i.e., higher) threshold value of P_L was used ($P_L = 6.23$ in place of 6.14), 28% (19 out of 68) of the deuterium-labeled cells were recognized as such (**Supplementary Fig. 8b**), while unlabeled cells were moved to the waste outlet with an accuracy of 99.7% (385 out of 386; **Supplementary Fig. 19**). We anticipate that ultimately this could be addressed by gradually changing the cell-suspended fluid from 50% D_2O-containing PBS to 0.2 M glycerol, or by using an optimized, perfectly isotonic glycerol or glucose solution. (vi) In this study, we could reliably measure the P_C value of cells of size down to ~1 μm. Generally, optical tweezing efficiency is the key parameter for reliable P_C measurement in the flow. It depends on the cell size and morphology (which influence the fluid drag), and the contrast in refractive index between the cell and the surrounding medium. It can be calculated using a geometric optics approach (by tracing the rays) when the cell size is larger than the laser wavelength ($d_p > \lambda$)[79], and using an electromagnetic approach (by solving Maxwell's equations) when the cell size is comparable with the laser wavelength ($d_p \sim \lambda$)[80]. In general, the larger the cell, the greater the optical tweezing efficiency. Due to the equilibrium of the trapping force, the cell locates at the laser beam focus regardless of its shape (e.g., the center of a spherical cell locates at the laser beam focus; the center of a rod-shaped cell locates at the laser beam focus and it aligns along the laser beam propagation direction as shown in **Fig. 1b**). Given that the Gaussian-shaped laser beam width is minimal at the laser beam focus, the Raman signal is proportional to the intensity of the incident photons, and the Raman system is a confocal system, the majority of the Raman signal comes from the laser beam focus, which means that the majority of the signal comes from the cell (**Fig. 1c**; see also **Methods**). Given these considerations, if the optical tweezers reliably capture the cell in the flow, P_C values can be measured in our platform. Similarly, P_L calculations and the cell loss from the optical tweezers are also influenced by the optical tweezing efficiency. This will lead to some variation in performance depending on cell morphology. For example, we believe that the decreased discrimination of labeled cells for *S. typhimurium* (**Fig. 3d**) and the variation in the rate of cell loss, which appears to be higher in *B. subtilis* (**Supplementary Fig. 3a**), were mostly probably a result of differences in cell morphology and cell envelope structure, which lead to different optical tweezing efficiencies. (vii) Stable-isotope probing has become a core tool in microbial ecology and methods that directly link this with single cell sorting for subsequent cultivation or genomic analysis thus have a very wide field of potential applications. Therefore, our platform that enables sorting isotope-labeled cells represents a major step forward. Beyond this, our platform has already demonstrated the capability to sort based on cytochrome *c* content, and thus similarly can sort unlabeled cells on the basis of natural storage compounds that are ecologically or biotechnologically important (e.g., polyhydroxybutyrate[81], sulphur[58,82], and polyphosphate[83]). This sorting of cells that contain compounds that induce a Raman resonance signal (like cytochromes as described in our manuscript, carotenoids[84], chlorophylls[85], vitamin B_{12}[86], heme[87], and rhodopsin[88]) is possible without labeling. Even the oxidation state of cytochromes in microbes can be measured by the Raman microspectroscopy[89] and potentially be used as sorting criterion. (viii) While considerable effort was devoted to optimizing the RACS process in order to increase the sensitivity of the Raman analysis to allow for shorter measurement time and thus increased throughput, there is still room for further improvement. In particular, using a higher-power laser (on the order of 10^3 mW) would reduce the Raman spectrum acquisition time, while lower magnification and NA objective would be required to prevent photophoretic damage of the cell – this depends on the energy density, not the laser power itself. This would allow for rapid and simultaneous determination of the cell index P_C and the labeling index P_L from a single spectrum, acquired directly within a more rapid single-file sample stream, without optical tweezing of cells. Thus, only labeled cells would need to be sorted into the collection stream downstream by applying an external force (e.g., an electric force). We envisage that this could ultimately increase throughput by up to a hundred-fold (up to 36,000 cells/h when P_C and P_L are measured simultaneously in 0.1 s), and further increase sorting accuracy by not carrying any unlabeled cells into the collection stream. Another way to enhance Raman scattering, and thus sensitivity, is to bring molecules of interest into very close contact with rough metal surfaces in a process called surface-enhanced Raman spectroscopy (SERS). An increase in the detection sensitivity by specialized Raman microspectroscopy setups, such as stimulated Raman spectroscopy (SRS)[90] or coherent anti-Stokes Raman spectroscopy (CARS)[91], would likely result in dramatically reduced acquisition times, strongly increased throughput, and would render the RACS system suitable for selection of cells based on a wider range of other chemical characteristics that are detectable only as weaker bands in their Raman spectra. Finally, sorting could be accelerated through the use of multiple-tweezers systems, at the cost, however, of significantly increased complexity of the RACS setup.

Code availability. All of the custom codes used in this study can be accessed upon request from the corresponding author. MATLAB GUI (graphical user interface) software for the operation of RACS platform is provided in **Supplementary Files 1** and **2**. R code for the calculation in **Supplementary Figure 18** is provided in **Supplementary File 3**.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The data that support the finding of this study are available from the corresponding author upon request. 16S rRNA gene sequence data has been deposited in the NCBI Sequence Read Archive (SRA) under SRP144990. Metagenomic data has been deposited in the NCBI under SRP144778. MAGs have been deposited as Whole Genome Shotgun projects at DDBJ/ENA/GenBank under the accessions RYVY00000000-RYWW00000000. All accession numbers with information on the associated samples are provided in **Supplementary Table 6**.

References

1. Rinke, C. *et al.* Insights into the phylogeny and coding potential of microbial dark matter. *Nature* **499**, 431-437 (2013).
2. Marcy, Y. *et al.* Dissecting biological "dark matter" with single-cell genetic analysis of rare and uncultivated TM7 microbes from the human mouth. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 11889-11894 (2007).
3. Blainey, P. C. The future is now: single-cell genomics of bacteria and archaea. *FEMS Microbiol. Rev.* **37**, 407-427 (2013).
4. Xu, L., Brito, I. L., Alm, E. J. & Blainey, P. C. Virtual microfluidics for digital quantification and single-cell sequencing. *Nat. Methods* **13**, 759-762 (2016).
5. Ackermann, M. A functional perspective on phenotypic heterogeneity in microorganisms. *Nat. Rev. Microbiol.* **13**, 497-508 (2015).

6. Ackermann, M. Microbial individuality in the natural environment. *ISME J.* **7**, 465-467 (2013).

7. Elowitz, M. B., Levine, A. J., Siggia, E. D. & Swain, P. S. Stochastic gene expression in a single cell. *Science* **297**, 1183-1186 (2002).

8. Stocker, R. Marine microbes see a sea of gradients. *Science* **338**, 628-633 (2012).

9. Yoon, H. S. *et al.* Single-cell genomics reveals organismal interactions in uncultivated marine protists. *Science* **332**, 714-717 (2011).

10. Labonte, J. M. *et al.* Single-cell genomics-based analysis of virus-host interactions in marine surface bacterioplankton. *ISME J.* **9**, 2386-2399 (2015).

11. Wagner, M., Nielsen, P. H., Loy, A., Nielsen, J. L. & Daims, H. Linking microbial community structure with function: fluorescence *in situ* hybridization-microautoradiography and isotope arrays. *Curr. Opin. Biotechnol.* **17**, 83-91 (2006).

12. Neufeld, J. D., Wagner, M. & Murrell, J. C. Who eats what, where and when? Isotope-labelling experiments are coming of age. *ISME J.* **1**, 103-110 (2007).

13. Lee, N. *et al.* Combination of fluorescent *in situ* hybridization and microautoradiography-a new tool for structure-function analyses in microbial ecology. *Appl. Environ. Microbiol.* **65**, 1289-1297 (1999).

14. McGlynn, S. E., Chadwick, G. L., Kempes, C. P. & Orphan, V. J. Single cell activity reveals direct electron transfer in methanotrophic consortia. *Nature* **526**, 531-535 (2015).

15. Krupke, A. *et al.* The effect of nutrients on carbon and nitrogen fixation by the UCYN-A-haptophyte symbiosis. *ISME J.* **9**, 1635-1647 (2015).

16. Berry, D. *et al.* Host-compound foraging by intestinal microbiota revealed by single-cell stable isotope probing. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 4720-4725 (2013).

17. Wang, Y., Huang, W. E., Cui, L. & Wagner, M. Single cell stable isotope probing in microbiology using Raman microspectroscopy. *Curr. Opin. Biotechnol.* **41**, 34-42 (2016).

18. Jing, X. *et al.* Raman-activated cell sorting and metagenomic sequencing revealing carbon-fixing bacteria in the ocean. *Environ. Microbiol.* https://doi.org/10.1111/1462-2920.14268.

19. Li, M. *et al.* Rapid resonance Raman microspectroscopy to probe carbon dioxide by single cells in microbial communities. *ISME J.* **6**, 875-885 (2012).

20. Huang, W. E. *et al.* Raman-FISH: combining stable-isotope Raman spectroscopy and fluorescence *in situ* hybridization for the single cell analysis of identity and function. *Environ. Microbiol.* **9**, 1878-1889 (2007).

21. Wagner, M. Single-cell ecophysiology of microbes as revealed by Raman

microspectroscopy or secondary ion mass spectrometry imaging. *Annu. Rev. Microbiol.* **63**, 411-429 (2009).

22. Rinke, C. *et al.* Obtaining genomes from uncultivated environmental microorganisms using FACS-based single-cell genomics. *Nat. Protoc.* **9**, 1038-1048 (2014).
23. Dyksma, S. *et al.* Ubiquitous Gammaproteobacteria dominate dark carbon fixation in coastal sediments. *ISME J.* **10**, 1939-1953 (2016).
24. Dupont, C. L. *et al.* Genomic insights to SAR86, an abundant and uncultivated marine bacterial lineage. *ISME J.* **6**, 1186-1199 (2012).
25. Woyke, T. *et al.* Assembling the marine metagenome, one cell at a time. *PLoS One* **4**, e5299 (2009).
26. Hatzenpichler, R. *et al.* Visualizing in situ translational activity for identifying and sorting slow-growing archaeal-bacterial consortia. *Proc. Natl. Acad. Sci. U.S.A.* **113**, E4069-E4078 (2016).
27. Berry, D. *et al.* Tracking heavy water (D_2O) incorporation for identifying and sorting active microbial cells. *Proc. Natl. Acad. Sci. U.S.A.* **112**, E194-E203 (2015).
28. Huang, W. E., Ward, A. D. & Whiteley, A. S. Raman tweezers sorting of single microbial cells. *Environ. Microbiol. Rep.* **1**, 44-49 (2009).
29. Zhang, P. *et al.* Raman-activated cell sorting based on dielectrophoretic single-cell trap and release. *Anal. Chem.* **87**, 2282-2289 (2015).
30. Mcllvenna, D. *et al.* Continuous cell sorting in a flow based on single cell resonance Raman spectra. *Lab Chip* **16**, 1420-1429 (2016).
31. Johansson, M. E. V. *et al.* The inner of the two Muc2 mucin-dependent mucus layers in colon is devoid of bacteria. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 15064-15069 (2008).
32. Schroeder, B. O. *et al.* Bifidobacteria or fiber protects against diet-induced microbiota-mediated colonic mucus deterioration. *Cell Host Microbe* **23**, 27–40.e7 (2018).
33. Desai, M. S. *et al.* A dietary fiber-deprived gut microbiota degrades the colonic mucus barrier and enhances pathogen susceptibility. *Cell* **167**, 1339–1353.e21 (2016).
34. Png, C. W. *et al.* Mucolytic bacteria with increased prevalence in IBD mucosa augment in vitro utilization of mucin by other bacteria. *Am. J. Gastroenterol.* **105**, 2420–2428 (2010).
35. Sonnenburg, E. D. & Sonnenburg, J. L. Starving our microbial self: the deleterious consequences of a diet deficient in microbiota-accessible carbohydrates. *Cell Metab.* **20**, 779–786 (2014).
36. Tailford, L. E., Crost, E. H., Kavanaugh, D. & Juge, N. Mucin glycan foraging in the human gut microbiome. *Front. Genet.* **6**, 81 (2015).

37. Bae, S. C., Lee, H., Lin, Z. & Granick, S. Chemical imaging in a surface force apparatus: confocal Raman spectroscopy of confined poly(dimethylsiloxane). *Langmuir* **21**, 5685-5688 (2005).
38. Dochow, S. *et al.* Quartz microfluidic chip for tumour cell identification by Raman spectroscopy in combination with optical traps. *Anal. Bioanal. Chem.* **405**, 2743-2746 (2013).
39. Dholakia, K. & Reece, P. Optical micromanipulation takes hold. *Nano Today* **1**, 18-27 (2006).
40. Lagkouvardos, I. *et al.* The Mouse Intestinal Bacterial Collection (miBC) provides host-specific insight into cultured diversity and functional potential of the gut microbiota. *Nat. Microbiol.* **1**, 16131 (2016).
41. Stepanauskas, R. & Sieracki, M. E. Matching phylogeny and metabolism in the uncultured marine bacteria, one cell at a time. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 9052–9057 (2007).

References in Methods

42. Qin, D., Xia, Y. & Whitesides, G. M. Soft lithography for micro- and nanoscale patterning. *Nat. Protoc.* **5**, 491-502 (2010).
43. Schuster, K. C., Reese, I., Urlaub, E., Gapes, J. R. & Lendl, B. Multidimensional information on the chemical composition of single bacteria cells by confocal Raman microspectroscopy. *Anal. Chem.* **72**, 5529-5534 (2000).
44. Carlo, D. D., Irimia, D., Tompkins, R. G. & Toner, M. Continuous inertial focusing, ordering, and separation of particles in microchannels. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 18892-18897 (2007).
45. Ha, B. H., Lee, K. S., Jung, J. H. & Sung, H. J. Three-dimensional hydrodynamic flow and particle focusing using four vortices Dean flow. *Microfluid. Nanofluid.* **17**, 647-655 (2014).
46. Chu, H., Doh, I. & Cho, Y. A three-dimensional (3D) particle focusing channel using the positive dielectrophoresis (pDEP) guided by a dielectric structure between two planar electrodes. *Lab Chip* **9**, 686-691 (2009).
47. Shi, J. *et al.* Three-dimensional continuous particle focusing in a microfluidic channel via standing surface acoustic waves (SSAW). *Lab Chip* **11**, 2319-2324 (2011).

48. Rhee, M. *et al.* Synthesis of size-tunable polymeric nanoparticles enabled by 3D hydrodynamic flow focusing in single-layer microchannels. *Adv. Mater.* **23**, H79-H83 (2011).

49. Ozkumur, E. *et al.* Inertial focusing for tumor antigen-dependent and -independent sorting of rare circulating tumor cells. *Sci. Transl. Med.* **5**, 179ra47 (2013).

50. Svoboda, K. & Block, S. M. Biological applications of optical forces. *Annu. Rev. Biophys. Biomol. Struct.* **23**, 247-285 (1994).

51. Dienerowitz, M. *et al.* Optically trapped bacteria pairs reveal discrete motile response to control aggregation upon cell-cell approach. *Curr. Microbiol.* **69**, 669-674 (2014).

52. Chang, C. B., Huang, W.-X., Lee, K. H. & Sung, H. J. Optical levitation of a non-spherical particle in a loosely focused Gaussian beam. *Opt. Express* **20**, 24068-24084 (2012).

53. Lee, K. S. *et al.* Radiation forces on a microsphere in an arbitrary refractive index profile. *J. Opt. Soc. Am. B* **29**, 407-414 (2012).

54. Koppel, D. E., Axelrod, D., Schlessinger, J., Elson, E. L. & Web, W. W. Dynamics of fluorescence marker concentration as a probe of mobility. *Biophys. J.* **16**, 1315-1329 (1976).

55. Kiefer, J., Ebel, N., Schlucker, E. & Leipertz, A. Characterization of *Escherichia coli* suspensions using UV/Vis/NIR absorption spectroscopy. *Anal. Methods.* **2**, 123-128 (2010).

56. Lucker, S., Nowka, B., Rattei, T., Spieck, E. & Daims, H. The genome of *Nitrospina gracilis* illuminates the metabolism and evolution of the major marine nitrite oxidizer. *Front. Microbiol.* **4**, 27 (2013).

57. Patzold, R. *et al.* In situ mapping of nitrifiers and anammox bacteria in microbial aggregates by means of confocal resonance Raman microscopy. *J. Microbiol. Methods* **72**, 241-248 (2008).

58. Milucka, J. *et al.* Zero-valent sulphur is a key intermediate in marine methane oxidation. *Nature* **491**, 541-546 (2012).

59. Song, Y. e*t al.* Single-cell genomics based on Raman sorting reveals novel carotenoid-containing bacteria in the Red Sea. *Microb. Biotechnol.* **10**, 125-137 (2017).

60. Griffiths, R. I., Whiteley, A. S., O'Donnell, A. G. & Bailey, M. J. Rapid method for coextraction of DNA and RNA from natural environments for analysis of ribosomal DNA- and rRNA-based microbial community composition. *Appl. Environ. Microbiol.* **66**, 5488-5491 (2000).

61. Herbold, C. W. *et al.* A flexible and economical barcoding approach for highly multiplexed amplicon sequencing of diverse target genes. *Front. Microbiol.* **6**, 731 (2015).

62. Hamady, M., Walker, J., Harris, J., Gold, N. & Knight, R. Error-correcting barcoded primers allow hundreds of samples to be pyrosequenced in multiplex. *Nat. Methods* **5**, 235–237 (2008).

63. Schubert, M., Lindgreen, S. & Orlando, L. AdapterRemoval v2: rapid adapter trimming, identification, and read merging. *BMC Res. Notes* **9**, 88 (2016).

64. Bankevich, A. *et al.* SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **19**, 455–477 (2012).

65. Kang, D. D. *et al.* MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. *PeerJ* **3**, e1165 (2015).

66. Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* **25**, 1043–1055 (2015).

67. Olm, M. R., Brown, C. T., Brooks, B. & Banfield, J. F. dRep: a tool for fast and accurate genome de-replication that enables tracking of microbial genotypes and improved genome recovery from metagenomes. *ISME J.* **11**, 2864-2868 (2017).

68. Aziz, R. K. *et al.* The RAST Server: rapid annotations using subsystems technology. *BMC Genomics* **9**, 75 (2008).

69. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2 -- approximately maximum-likelihood trees for large alignments. *PLoS ONE* **5**, e9490 (2010).

70. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312-1313 (2014).

71. Richter, M. & Rosselló-Móra, R. Shifting the genomic gold standard for the prokaryotic species definition. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 19126-19131 (2009).

72. Ravcheev, D. A. & Thiele, I. Comparative genomic analysis of the human gut microbiome reveals a broad distribution of metabolic pathways for the degradation of host-synthetized mucin glycans and utilization of mucin-derived monosaccharides. *Front. Genet.* **8**, 111 (2017).

73. Abu-Ali, G. S. *et al.* Metatranscriptome of human faecal microbial communities in a cohort of adult men. *Nat. Microbiol.* **3**, 356-366 (2018).

74. Arike, L., Holmén-Larsson, J. & Hansson, G. C. Intestinal Muc2 mucin O-glycosylation is affected by microbiota and regulated by differential expression of glycosyltranferases. *Glycobiology* **27**, 318-328 (2017).

75. Evans, C. C. *et al.* Exercise prevents weight gain and alters the gut microbiota in a mouse model of high fat diet-induced obesity. *PLoS ONE* **9**, e92193 (2014).

76. Rooks, M. G. *et al.* Gut microbiome composition and function in experimental colitis during active disease and treatment-induced remission. *ISME J.* **8**, 1403-1417 (2014).

77. Ormerod, K. L. *et al.* Genomic characterization of the uncultured Bacteroidales family S24-7 inhabiting the guts of homeothermic animals. *Microbiome* **4**, 36 (2016).

78. Eichorst, S. A. *et al.* Advancements in the application of NanoSIMS and Raman microspectroscopy to investigate the activity of microbial cells in soils. *FEMS Microbiol. Ecol.* **91**, fiv106 (2015).

79. Ashkin, A. Forces of a single-beam gradient laser trap on a dielectric sphere in the ray optics regime. *Biophys. J.* **61**, 569-582 (1992).

80. Novotny, L., Bian, R.X. & Xie, X.S. Theory of nanometric optical tweezers. *Phys. Rev. Lett.* **79**, 645-648 (1997).

81. Samek, O. *et al.* Quantitative Raman spectroscopy analysis of polyhydroxyalkanoates produced by *Cupriavidus necator* H16. *Sensors* **16**, 1808 (2016).

82. Gruber-Vodicka, H. R. *et al*. *Paracatenula*, an ancient symbiosis between thiotrophic *Alphaproteobacteria* and catenulid flatworms. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 12078-12083 (2011).

83. Majed, N., Matthaus, C., Diem, M. & Gu, A. Z. Evaluation of intracellular polyphosphate dynamics in enhanced biological phosphorus removal process using Raman microscopy. *Environ. Sci. Technol.* **43**, 5436-5442 (2009).

84. Kumar, B. N. V., Kampe, B., Rosch, P. & Popp, J. Characterization of carotenoids in soil bacteria and investigation of their photodegradation by UVA radiation via resonance Raman spectroscopy. *Analyst* **140**, 4584-4593 (2015).

85. Lutz, M. Resonance Raman spectra of chlorophyll in solution. *J. Raman Spectrosc.* **2**, 497-516 (1974).

86. Salama, S. & Spiro, T. G. Visible and near-ultraviolet resonance Raman spectra of photolabile vitamin B_{12} derivatives with a rapid-flow technique. *J. Raman Spectrosc.* **6**, 57-60 (1977).

87. Spiro, T. G. & Strekas, T. C. Resonance Raman spectra of heme proteins. Effects of oxidation and spin state. *J. Am. Chem. Soc.* **96**, 338-345 (1974).

88. Palings, I. *et al.* Assignment of fingerprint vibrations in the resonance Raman spectra of rhodopsin, isorhodopsin, and bathorhodopsin: implications for chromophore structure and environment. *Biochemistry* **26**, 2544-2556 (1987).

89. Bjerg, J. T. *et al.* Long-distance electron transport in individual, living cable bacteria. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 5786-5791 (2018).
90. Cheng, J. X. & Xie, X. S. Vibrational spectroscopic imaging of living systems: an emerging platform for biology and medicine. *Science* **350**, aaa8870 (2015).
91. Evans, C. L. & Xie, X. S. Coherent anti-Stokes Raman scattering microscopy: chemical imaging for biology and medicine. *Annu. Rev. Anal. Chem.* **1**, 883-909 (2008).

Acknowledgments

We acknowledge support by a US Department of Energy Joint Genome Institute Emerging Technologies Opportunity grant DE-AC02-05CH11231 (to R.S. and M.W.). M.W. and M.P. were also supported by the European Research Council via the Advanced Grant project NITRICARE 294343. R.S. acknowledges support from a Gordon and Betty Moore Marine Microbial Initiative Investigator Award (#3783). F.M. was supported by the Engineering and Physical Sciences Research Council (EP/R035350/1, EP/S001921/1). F.C.P. and D.B. were supported by the Austrian Science Fund (FWF; P26127-B20) and the European Research Council (Starting Grant: FunKeyGut 741623). F.C.P. was also supported by the European Union's Horizon 2020 Framework Programme for Research and Innovation (grant No. 658718). We thank Ruben Weiß and Natalia Ivleva for collaboration on SERS, Craig Herbold for bioinformatics assistance and input, and Tanja Woyke and Rex Malmstrom for many helpful discussions.

Author contributions

K.S.L. and R.S. created the software for the RACS. K.S.L., D.B. and R.S. performed the numerical calculations. K.S.L., M.P., V.I.F. and F.M. developed the RACS system, and designed and performed the pure culture RACS experiments. K.S.L., F.C.P. and D.B. designed and performed the mouse colon sample experiments. K.S.L., A.M. and H.D.

designed and performed the marine enrichment sample experiments. M.W. and R.S. supervised the project. All authors wrote the manuscript.

Competing interests

The authors declare no competing financial interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/xxxxxxx

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and request for materials should be addressed to R.S.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Figure legends

Figure 1 | Design and working principle of the RACS system. a, Cells from a sample fluid are flown into a microfluidic device and focused in the vertical and horizontal directions by two sheath flows. Cells not captured by the optical tweezers (solid green circle) flow directly into the waste outlet, whereas cells captured by the optical tweezers are moved to the sample-free stream (dashed green circle) and have their Raman spectrum measured. Cells of interest (for example, active cells labeled by deuterium) are released into the collection outlet, while other cells are carried back to the sample stream and released into the waste outlet (see **Supplementary Fig. 11** for microfluidic tubing connections for sample input and cell collection). **b,** A rod-shaped microbial cell confined in the optical tweezers aligns along the laser propagation direction (vertical). **c,** Normalized intensity distribution of the Raman signal (color scale) coming from the cell and from the fluid (see **Methods** for calculations). Under our experimental conditions and for a 2.0 µm × 0.5 µm rod-shaped bacterium, $I_{\mathrm{cell}}/I_{\mathrm{fluid}} = 6.22$. **d,** The optofluidic setup: two laser sources for the Raman measurement (532 nm; 15mW; light green path) and the optical tweezers (1064 nm; 400 mW; purple path) are focused at the same position in the microfluidic device by a 60× objective. The scattered light is directed by a removable beam splitter to either CCD camera #1 for initial laser alignment in the microfluidic device or to a spectrometer for measurement of the Raman spectrum. A notch filter suppresses the excitation wavelengths before the spectrometer, then a grating divides the scattered light into discrete wavelengths, which are measured by the spectrometer. CCD camera #2 is used to monitor the RACS process while the system is in Raman measurement mode (i.e., CCD camera #1 is deactivated), with a ring illuminator (470 nm) as a light source, a 10× objective, an optical array yielding 3.5× magnification, and a shortpass filter (cutoff wavelength: 500 nm) to prevent laser-induced damage of the camera.

Figure 2 | RACS operation (see also **Supplementary Video 4**). **a,** The analysis region in the microfluidic sorter. The low-angle (30°) ring illuminator (**Fig. 1d**) results in dark-field images in which cells appear as bright spots on a dark background. Because of reflections between the objectives above (60× water-immersion) and below (10× air) the microfluidic device, circular patterns surround the optical tweezers (OT). The optical tweezers are moved between their default, capture location (CL; solid green circle) at $y = 50$ µm from the channel sidewall, and the evaluation location (EL; dashed green circle) at $y = 320$ µm. **b-e,** The four main steps involved in RACS for a deuterium-labeled cell (images show the purple dashed region of panel **a**). The further magnified insets (green squares) show the surroundings of the optical tweezers as they move, with the background intensity subtracted for clearer visualization of captured cells (white arrows). When the software recognizes that a cell is captured in the optical tweezers based on changes in the Raman spectrum at the capture location (**b**), the optical tweezers move (**c**) to the evaluation location in the cell-free stream (EL in panel **a**), where a Raman spectrum with an extended acquisition time is acquired to determine if the cell is deuterium-labeled. If so, the cell is released (**d**) and the flow carries it into the collection outlet. The optical tweezers then return to the capture position and the program starts over (**e**). **f-i**, As for **b-e**, but for a deuterium-unlabeled cell. Steps **f** and **g** are identical to steps **b** and **c**, respectively. The white background around the captured cell in the inset of panel **g** is due to reflections between the objectives above and below. If the cell is identified as unlabeled, the optical tweezers move it back to the capture location (**h**), where it is released (**i**) and carried by the flow into the waste outlet, and the program then starts over. The cells analyzed here were *S. typhimurium*. Scale bars, 50 µm.

Figure 3 | RACS of reference strains and performance criteria. a, Representative Raman spectra of deuterium-labeled and unlabeled *E. coli* cells in the flowing fluid, as well as of the background fluid. This experiment was repeated independently at least 20 times with similar results. Color bars and insets show the regions of the spectrum used to define the 'cell index' P_C (1620–1670 cm^{-1}) to distinguish cells from background, and the 'labeling index' P_L (1850–1900 cm^{-1} and 2040–2300 cm^{-1}) to distinguish labeled from unlabeled cells for automated RACS. **b,** Normalized Raman intensity (color scale) of single deuterium-labeled *E. coli* cells in the flowing fluid (averaged over 10 cells) for different acquisition times ranging from 0.1 to 5.0 s (in 0.1-s increments), and determination of the acquisition times necessary for robust calculation of P_C and P_L. The red and blue shaded areas represent the s.d. for P_C and P_L, respectively. P_i refers to either P_C or P_L, and superscripts +/- refer to adjacent time steps. Note that a converged signal for P_C (red) and P_L (blue) is obtained for acquisition times of 2 s (dashed red line) and 5 s (dashed blue line), respectively. **c**, P_C and **d,** P_L values of four strains (*E. coli*, green; *B. subtilis*, blue; *S. typhimurium*, red; *M. adhaerens*, violet) calculated from Raman spectra in order to determine thresholds to use in the automated RACS. Measurements of deuterium-labeled (filled symbols) and unlabeled cells (empty symbols) were conducted separately in the flowing fluid. The number of cells analyzed was n = 31/19 (labeled/unlabeled), n = 24/24, n = 31/28, n = 22/16 for *E. coli*, *B. subtilis*, *S. typhimurium*, and *M. adhaerens*, respectively. **e,** Raman spectra for each *E. coli* cell (n = 185; 100 selected / 85 rejected) analyzed with automated RACS of a 1:1 mixture of deuterium-labeled and unlabeled cells. The spectrum of each cell was normalized (color bar) by its average intensity in the 1850–1900 cm^{-1} region (used to compute P_L). 'Captured and selected' and 'captured and rejected' refer to the cells identified as deuterium-labeled and unlabeled, respectively. Cell loss refers to the cell loss during the movement of the optical tweezers between the capture location and the evaluation location (see **Methods**; **Supplementary Fig. 4**).

Figure 4 | RACS analysis and targeted mini-metagenomics of mucin-degrading bacteria from a mouse colon microbiota. **a,** P_L values of cells from the mouse colon samples, suspended in 0.2 M glycerol. Control (n = 106, blue) and test sample (n = 132, red; 'B10-MucB' in **Supplementary Table 2**) denote cells from colon microbiota incubated for 6 h in non-D_2O-containing PBS supplemented with 2.5 mg/mL of glucose and 50% D_2O-containing PBS supplemented with 2.5 mg/mL of mucin, respectively. In this test sample, 16% (21 out of 132) cells showed $P_L > 6.19$ (chosen to exclude values in the control; dashed black line) and were sorted. **b,c,** Genome-encoded *O*-glycan-degrading capability of mucin-stimulated RACS cells. **b,** Scheme showing a representative mucin *O*-glycan chain and the sites of action of common *O*-glycan-degrading enzymes. **c,** Presence (filled squares) or absence (empty squares) matrix of genes encoding enzymes involved in mucin degradation in the annotated MAGs. Filled symbols indicate that at least one copy of a gene encoding the enzyme of interest is present in the annotated MAG (filled squares) or in the genome of the closest relative organism (filled circles).