Phenotypic heterogeneity within microbial populations arises even when the cells are exposed to putatively constant and homogeneous conditions. The outcome of this phenomenon can affect the whole function of the population, resulting in, for example, new “adapted” metabolic strategies and impacting its fitness at given environmental conditions. Accounting for phenotypic heterogeneity becomes thus necessary, due to its relevance in medical and applied microbiology as well as in environmental processes. Still, a comprehensive evaluation of this phenomenon requires a common and unique method of quantitation, which allows for the comparison between different studies carried out with different approaches. Consequently, in this study, two widely applicable indices for quantitation of heterogeneity were developed. The heterogeneity coefficient (HC) is valid when the population follows unimodal activity, while the differentiation tendency index (DTI) accounts for heterogeneity implying outbreak of subpopulations and multimodal activity. We demonstrated the applicability of HC and DTI for heterogeneity quantitation on stable isotope probing with nanoscale secondary ion mass spectrometry (SIP–nanoSIMS), flow cytometry, and optical microscopy datasets. The HC was found to provide a more accurate and precise measure of heterogeneity, being at the same time consistent with the coefficient of variation (CV) applied so far. The DTI is able to describe the differentiation in single-cell activity within monoclonal populations resolving subpopulations with low cell abundance, individual cells with similar phenotypic features (e.g., isotopic content close to natural abundance, as detected with nanoSIMS). The developed quantitation approach allows for a better understanding on the impact and the implications of phenotypic heterogeneity in environmental, medical and applied microbiology, microbial ecology, cell biology, and biotechnology.

Keywords: phenotypic heterogeneity, single-cell resolution, SIP–nanoSIMS, anabolic activity, flow cytometry, multimodality, heterogeneity quantitation, Zipf’s law
INTRODUCTION

Ecosystems and microbial communities have traditionally been studied as an assembly of different monoclonal populations. Each population has been considered to be composed of genetically identical cells performing the same metabolic function. However, techniques with single-cell resolution were developed over the past decades, allowing us to narrow down the vision at the level of individual cells (Brehm-Stecher and Johnson, 2004; Pumphrey et al., 2009; Vasdeki and Stephanopoulos, 2015; Gao et al., 2016). This facilitates the understanding of monoclonal population’s physiology, strengthening the concept that genetically identical cells can show cell-to-cell variability even when sharing the same environmental and nutritional conditions and phenomenon, generically known as phenotypic heterogeneity (Avery, 2006; Ackermann, 2015; Davis Kimberly and Isberg Ralph, 2016). By definition, the phenotype is every observable feature of an organism originating from the complex interaction between genotype, cellular biochemical mechanisms, and environment. Although a univocal definition of phenotypic heterogeneity is currently missing in literature, this usually embraces all intrinsic and extrinsic cellular noise (Simpson et al., 2009) as well as environment-induced differences arising between single cells belonging to a monoclonal isogenic population under homogeneous environmental conditions. Phenotypic heterogeneity arises from stochastic effects in random molecular and biochemical processes (Elowitz et al., 2002; Kærn et al., 2005; Kussell and Leibler, 2005; Kiviet et al., 2014), inequalities in gene expression, and transcriptional regulatory networks (Newman et al., 2006; Maheshri and O’shea, 2007; Fraser and Kærn, 2009) as well as constitutive single-cell features like cell cycle or cell aging (Sumner and Avery, 2002; Levy et al., 2012). All these effects contribute to the so-called cellular noise (Avery, 2006; Samoilov et al., 2006; Tsimring, 2014), which is extensively used in literature referring to cell-to-cell variability due to molecular mechanisms, regulatory pathways, and genetic cues (Balázsi et al., 2011; Levchenko and Nemenman, 2014). The unequal segregation of DNA hyper-structures transmitted to the daughter cells during division and the variation in populations’ growth rate lead to the diversification in substrate assimilation by single cells in batch and chemostat cultures (Kopf Sebastian et al., 2015; Gangwe Nana et al., 2018). Metabolic/functionual diversifications help microbial populations to cope better with stresses and fluctuations in their surrounding environment (Avery, 2006; Ackermann, 2015; Bódi et al., 2017). Indeed, differences at the metabolic level occur when cells respond to niche perturbations with different metabolic strategies gaining new ecological functions from which the whole population benefits (West and Cooper, 2016; Bódi et al., 2017). The environment affects considerably the metabolic fluxes inside bacterial cells via activation of different regulatory pathways of the central carbon metabolism, thus leading to different phenotypes within an isogenic population (Kotte et al., 2014). Yeast cells reveal multiple metabolic states under certain environmental conditions, regulating dynamically the glycolysis pathway and thus preventing metabolic imbalances (Van Heerden et al., 2014). The occurrence of cell-to-cell differences in metabolic traits results from an interplay of ecological, intracellular and extracellular factors, usually referred to as metabolic heterogeneity (Takhaveev and Heinemann, 2018). Thus, the term phenotypic heterogeneity encompasses all aspects of cell-to-cell variability including epigenetic, regulatory, metabolic, physical, or physiological aspects observed at the single-cell level.

A challenge in the resolution of single-cell phenotypic features as well as in the tracking of intracellular metabolic fluxes remains the detection limits of the applied experimental techniques and their throughput (Takhaveev and Heinemann, 2018). Nanoscale secondary ion mass spectrometry (nanoSIMS) provides high lateral resolution for single microbial cell imaging together with a mass resolving power (MRP) sufficient for tracking the assimilation of isotope-labeled substrates (Lechene et al., 2006). A combination of stable isotope probing (SIP) with nanoSIMS (SIP–nanoSIMS) has been applied to shed light on single-cell metabolic activity within microbial populations in environmental setups as well as under laboratory conditions (Musat et al., 2008; Pumphrey et al., 2009; Pett-Ridge and Weber, 2012; Sheik et al., 2015; Zimmermann et al., 2015; Jiang et al., 2016; Schreiber et al., 2016; Nikolic et al., 2017; Nuñez et al., 2017). Studies applying SIP–nanoSIMS have been focused on assimilation activity in the frame of specific metabolic functions, considering exclusively the heterogeneity in terms of isotopic enrichment at the single-cell level. Metabolic heterogeneity in monoclonal populations has been shown to increase under limitation of nutrients and/or electron donors (Zimmermann et al., 2015, 2018; Schreiber et al., 2016); nutritional and temperature upshifts as well as carbon-source competition enhance metabolic specialization among monoclonal cells (Sheik et al., 2015; Nikolic et al., 2017; Simşek and Kim, 2018). The above-mentioned studies have quantified the uptake of isotope-labeled substrates normalizing it with output from bulk measurements and applying empirical assimilation expressions. It is difficult to compare the results on metabolic activity from different studies, even when derived from SIP–nanoSIMS experiments, if a unified approach for quantitation of single-cell assimilation is not followed. A SIP–nanoSIMS-based approach has recently introduced the “relative assimilation” (K d) as a measure to quantify the single-cell assimilation activity (Stryhanyuk et al., 2018). However, being exceptionally powerful in quantitation of isotopic composition at the single-cell level, nanoSIMS-based approaches require a long time to acquire chemical maps for a limited number of single

**Abbreviations:** SIP–nanoSIMS, stable isotope probing combined with nanoscale secondary ion mass spectroscopy; K A, fraction of an element incorporated by a single cell during the incubation with isotope-labeled growth substrates; CV, coefficient of variation (equal to the ratio of standard deviation to the mean); CQD, coefficient of quartile deviation; COD, coefficient of dispersion; Qm, quantile with percentile m; SK, skewness, indication of asymmetry in the distribution of a population; DW, distribution width on the histogram representing the distribution of a population; HC, heterogeneity coefficient developed in the present study; CSR, counting statistics error; HC corr, heterogeneity coefficient corrected for counting statistics error coming from nanoSIMS measurement; CSV A, CSV B, counting statistics variations of isolate ratios (R) and relative assimilation (K A); CSV HC, counting statistics heterogeneity of anabolic activity; DTI, differentiation tendency index defined as a slope (s) in a rank–activity distribution; CDTI, cumulative differentiation tendency index; SDT, strong differentiation tendency; WDT, weak differentiation tendency.
cells (∼1–10^2), implying thus a considerable limitation in the method throughput.

Flow cytometry and fluorescence time-lapse microscopy have also been applied to study phenotypic heterogeneity (Davey and Kell, 1996; Müller and Babel, 2003; Balaban et al., 2004; Nikolic et al., 2013, 2017; Lieder et al., 2014; Pratt et al., 2019). The output of these methods’ application comprises multidimensional datasets providing information on, for example, gene expression, physiology, metabolism, and morphology of individual cells (Davey and Kell, 1996; Müller and Babel, 2003; Balaban et al., 2004; Nikolic et al., 2013, 2017; Lieder et al., 2014; Pratt et al., 2019). Notably, the data are acquired within a short time while analyzing thousands of cells (∼10^3–10^5). With single-cell resolution and high sensitivity, flow cytometry is however not an imaging technique, and it cannot be applied to resolve intracellular compartments and single cells in spatial arrangements.

Physical properties of single cells (e.g., their morphology, intracellular and extracellular features, and viability) and the 3-D structure of their arrangement can be characterized with optical, atomic force, or electron/ion probe microscopy techniques (Hawkes and Spence, 2007; Hlawacek and Gözlhäuser, 2016). The microscopy imaging delivers information on single-cell phenotypes (geometry, size, volume, etc.), showing that the size of individual cells differs from the population average, which in turn is modulated by the environment and the growth conditions (Grover and Woldringh, 2001; Taheri-Araghi et al., 2015; Nikolic et al., 2017; Westfall and Levin, 2018).

To our knowledge, a general approach to quantify phenotypic heterogeneity has not been developed to date. Nowadays, the interest on phenotypic heterogeneity is increasing due to its implications in medical microbiology, microbial ecology, and biotechnology. Phenotypic heterogeneity has been shown to have an impact on medical care, concerning antibiotic resistance, treatment persistence, and biofilm formation (Turner et al., 2000; Sumner and Avery, 2002; Balaban et al., 2004; Grote et al., 2015; Dhar et al., 2016; Sadiq et al., 2017; Van Den Bergh et al., 2017) but also in biomedical research for drug discovery, cancer therapy, and diagnostics due to the variable effectiveness of care treatments on different phenotypes (Almendro et al., 2013; Gough et al., 2017). Understanding cell-to-cell heterogeneity in bioprocessing is currently a big challenge due to its implications in processes’ performance and therefore product yield in large-scale production (Delvigne et al., 2014, 2017). Hence, the quantitation of heterogeneity becomes relevant for an efficient process optimization in every field of biotechnology.

Phenotypic heterogeneity has been so far expressed with the coefficient of variation (CV) for its comparison under different experimental conditions (Grover and Woldringh, 2001; Kopf Sebastian et al., 2015; Schreiber et al., 2016; Nikolic et al., 2017). The CV is calculated as the ratio of standard deviation over the arithmetic mean value, and its application as a measure of heterogeneity implies therefore a normal distribution of single cells in their activity or function. However, this is not always the case for a heterogeneous population, in which a large dispersion of data from the centroid is observed and the distribution shape is often asymmetrical (skewed). Measures of dispersion, such as coefficient of quartile deviation (CQD) (Bonett, 2006) or coefficient of dispersion (COD) (Bonett and Seier, 2006), are likewise used to calculate the scattering of the data around the centroids, but they are affected differently by fluctuations (outliers) of observations in a dataset. Additionally, there are currently no indices accounting for one of the distinctive features of phenotypic heterogeneity, that is, the uprising of clonal subpopulations (multimodality phenomenon) with different metabolic activities (Arnoldini et al., 2014; Li et al., 2018; Simšek and Kim, 2018). An attempt to quantify the growth-related heterogeneity of microbial populations has been recently done in a flow cytometry experiment (Heins et al., 2019), where authors considered the slope of cumulative distribution function in combination with skewness, peak width, and CV. The suggested combination of distribution parameters may represent an average or describe a most abundant subpopulation but does not provide a robust measure of heterogeneity for a multimodal distribution.

By extending the concept of CV and considering the distribution of single-cell anabolic activity measured in $K_A$ (Stryhanyuk et al., 2018), we developed a novel approach for heterogeneity expression which returns the “heterogeneity coefficient” (HC) involving the correction for the counting statistics error (CSE) coming from nanoSIMS analysis. Since neither the CV nor the HC accounts for the outcome of subpopulations, the challenge of quantitating heterogeneity in multimodal single-cell activity had to be tackled. With this purpose, a modification of power Zipf’s (1935) law describing the rank–frequency distribution of words in literary texts (Voloshynovska, 2011) was applied. The slope in rank–activity distribution of single cells, interpreted as differentiation tendency index (DTI), was suggested for heterogeneity quantitation. The DTI is independent of the population size, normalization procedures, and measure units, making its use general and universal.

In this study, the applicability of both developed indices, HC and DTI, was first tested on SIP–nanoSIMS datasets acquired for two different bacterial strains. The HC facilitates routine quantitation of heterogeneity for monoclonal populations following unimodality and can be easily calculated for any single-cell-resolved dataset with Supplementary Table S1 provided in the Supplementary Material. The DTI allows for recognition of subpopulations and expresses the heterogeneity showing multimodality. To show the wide applicability of the HC index and DTI, heterogeneity was also quantitated for datasets of flow cytometry and optical microscopy experiments. To our knowledge, this is the first time that such an approach was applied broadly for quantitation and comparison of phenotypic heterogeneity. Indeed, the two indices can be calculated for any numerically expressed feature at the single-cell level, since they are not bound to either any unit of measure or any particular technique.

**MATERIALS AND METHODS**

**Cultivation and Stable Isotope Labeling**

_Pseudomonas putida_ strain KT2440 (DSM 6125) and _Pseudomonas stutzeri_ (environmentally isolated) were grown
under aerobic conditions with mineral salts medium (MSM) containing (g/L) NH_4Cl (0.3), KH_2PO_4 (0.2), NaCl (1), MgSO_4·7H_2O (0.5), and CaCl_2·2H_2O (0.1) supplemented with NaHCO_3 (30 ml/L) as a buffer system. The medium was supplemented with 1 ml of a vitamin mix solution in phosphate buffer [10 mM NaH_2PO_4·H_2O, pH 7.1, containing, in milligrams per liter, 4-aminobenzoic acid, 40; D(+)-biotin, 10; nicotinic acid, 100; Ca-D(-)-pantothenate, 50; pyridoxine dihydrochloride, 150; folic acid, 40; and lipoic acid, 15] and with 1 ml of chelated trace element solution (containing, in milligrams per liter, Na_2EDTA, 5,200; H_3BO_3, 10; MnCl_2·4H_2O, 5; FeSO_4·7H_2O, 2,100; CoCl_2·6H_2O, 190; NiCl_2·6H_2O, 24; CuCl_2·2H_2O, 10; and ZnSO_4·7H_2O, 144; pH adjusted to 6.0 with 1 M NaOH). The vitamin mix and trace element solutions were prepared as previously described (Widdel, 2010). Both bacterial cultures were prepared by inoculating 1 ml of bacterial suspension in 29 ml of mineral medium in a sealed 140-ml serum bottle. The cultures were grown at 30°C with orbital shaking (100 rpm) with 5 mM acetate as growth substrate. Acetate is an intermediate of many cellular biosynthetic pathways, therefore a good candidate to track single-cell anabolic activity with the SIP–nanoSIMS approach. Stable isotope labeling was performed with [13C_2]-acetate (Sigma Aldrich) to reach 20% of 13C relative to the total carbon in the growth substrate. A stock solution of the isotope-labeled acetate was prepared in MSM and filter-sterilized before use. Cultures were supplied with [13C_2]-acetate at the onset of their exponential growth phase. Samples were collected at three different time points within the exponential growth phase and prepared for nanoSIMS analysis.

**Sample Preparation for NanoSIMS Analysis**

Each sample was fixed overnight with 3% glutaraldehyde in sodium cacodylate buffer (0.2 M, pH 7.4, EM Science) at 4°C. Filters (0.22 μm pore size, 25 mm, GTTP type, Merck) were coated with a 30 nm gold–palladium layer and mounted inside a stainless steel syringe 25 mm filter holder (Sartorius). Bacterial cells were filtered; washed twice with cacodylate buffer; incubated with 1% H_2O_2 in cacodylate buffer, for 30 min; dehydrated via increasing concentrations of ethanol in water (30 → 50 → 70 → 80 → 90 → 95 → 100); and dried upon 20 cycles in a critical point drying device (Leica EM CPD 300a, Germany).

**NanoSIMS Measurement**

NanoSIMS analysis parameters were optimized in order to achieve the appropriate precision at the single-cell level [Supplementary Information (SI)]. From each filter of *P. putida* and *P. stutzeri*, a piece of 5 mm in diameter was cut and mounted on the 24-holes-holder for nanoSIMS measurement. Single-cell analysis was performed using a nanoSIMS-50L instrument (Cameca) acquiring the following seven molecular ion species (masses) 18O^−, 12C_2^−, 13C_2^12C^−, 12C_14N^−, 13C_14N^−, 35S^−, and 31P_16O_2^−. The MRP was set above 7,000. Samples areas of 100 × 100 μm² were pre-implemented with 100 pA of a 16 keV cesium (Cs⁺) primary ion beam for 5 min before measurements. Smaller areas within 20 × 20 μm² field of view (FoV) were analyzed with 4 pA primary Cs⁺ ion beam, a nominal size of 100 μm for the entrance slit, 40 μm exit slits, and an energy slit cutting 20% of secondary ion intensity at the high-energy distribution tail. Samples were scanned with a 512 × 512-pixel raster size and a dwell time of 2 ms/pixel. In total, 30 planes were acquired, ensuring complete sample consumption in each FoV analyzed; 11 of those were accumulated and corrected for lateral drifting using the Look@NanoSIMS (LANS) software (Polerecky et al., 2012). Regions of interest (RoI) were drawn manually around each single cell using the 12C\(^{14}N^−\) map as a biomass distribution template supported by the cell image acquired in secondary electron signal. Isotope ratio data were exported and further processed with OriginPro 2019 software for statistical analysis and graphing. The relative assimilation (K_a) values were calculated with the template provided in Stryhan yuk et al. (2018) and used for further numerical analysis and graphical representations.

**Cell Cultivation and Preparation for Flow Cytometric Analysis**

After pre-cultivation on peptone medium (Medium 1, DSMZ), *P. putida* KT2440 was inoculated in Erlenmeyer flasks with 50 ml of M9 leucine medium and 1 g/L of acetate as carbon and energy source and grown at 30°C and 125 rpm. The strain grew with a μ_max of 0.62 h⁻¹. After 24 h of incubation, acetate was added to the cultures again reaching a concentration of 1 g/L. Cell suspensions (1–3 ml) were sampled hourly and centrifuged for 10 min (3,200 × g at 4°C). The resulting pellets were resuspended in 4 ml of formaldehyde solution [1 ml of 8% (wt/vol) formaldehyde/phosphate-buffered saline (PBS) in 3 ml of PBS] and incubated at 4°C for 30 min. After a final centrifugation, the pellets were resuspended in 4 ml of 70% (vol/vol) ethanol/bi-distilled H_2O and stored at −20°C.

For flow cytometric analysis, the cells were stained according to Koch et al. (2013). In short, 1 ml of the fixed sample was centrifuged, and the pellet was resuspended in 2 ml of PBS (6 mM Na_2HPO_4·2H_2O, 1.8 mM NaH_2PO_4, and 145 mM NaCl with bi-distilled H_2O, pH 7). The optical density (OD) of the well-mixed samples was measured (d = 0.5 cm; λ = 700 nm) and adjusted to 0.04 with PBS. After centrifugation of 1 ml of this solution for 10 min (3,200 × g at 4°C), the supernatant was discarded. The cell pellet was resuspended in 0.5 ml of permeabilization buffer (0.1 M citric acid, 4.1 mM Tween 20, and bi-distilled H_2O) and incubated at room temperature for 20 min. After a further centrifugation step, the supernatant was discarded, and the cells were resuspended in 1 ml of DNA dye solution for overnight staining at room temperature until cytometric measurement [0.68 μM 4’,6-diamidino-2-phenylindole (DAPI), Sigma-Aldrich, in 417 mM Na_2HPO_4·NaH_2PO_4 buffer, 289 mM Na_3HPO_4, and 128 mM NaH_2PO_4 with bi-distilled H_2O, pH 7].

**Flow Cytometric Analysis**

The DAPI-stained *P. putida* KT2440 single cells were analyzed with a MoFlo Legacy cell sorter (Beckman-Coulter, Brea, California, USA) equipped with two lasers. The forward scatter (FSC) and side scatter (SSC) signals were measured using a blue laser (488 nm, 400 mW, Genesis MX488-500 STMODS, Coherent, Santa Clara, California, USA). The FSC signal (band-pass filter 488 ± 5 nm, neutral density filter 1.9) is an optical
characteristic related to cell size, whereas the SSC signal (trigger signal, band-pass filter 488 ± 5 nm, neutral density filter 1.9) is related to cell density. The DAPI fluorescence (band-pass filter 450 ± 32.5 nm) was excited by a UV laser (355 nm, 150 mW, Xcyte CY-355-150, Lumentum, Milpitas, California, USA). The resulting scatter and fluorescence signals were detected by photomultiplier tubes (Hamamatsu Photonics, models R928 and R3896; Hamamatsu City, Japan). The fluidic system was run at 56 psi (3.86 bar) with sample overpressure at a maximum of 0.3 psi (0.02 bar) and a 70 µm nozzle. The sheath fluid consisted of 10× sheath buffer (19 mM KH2PO4, 38 mM KCl, 166 mM Na2HPO4, and 1.39 M NaCl with bi-distilled H2O) that was diluted to a 0.2× working solution with 0.1 µm filtrated bi-distilled H2O. The instrument was calibrated in the linear and logarithmic ranges prior to all measurements. Blue fluorescent 1 µm beads [FluoSpheres F8815 (350/440), lot no.: 69A1-1] and 2 µm yellow-green fluorescent beads [FluoSpheres F8827 (505/515), lot no.: 1717426, both from Molecular Probes, Eugene, Oregon, USA] were used for linear calibration. Blue fluorescent 0.5 and 1 µm beads [both Fluoresbrite BB carboxylate microspheres (360/407), lot nos.: 552744 and 499344, Polyscience, Niles, Illinois, USA] and red fluorescent 1 µm beads [FluoSpheres F8816 (625/645), lot no.: 24005W] were used for calibration in the logarithmic scale and added to every sample to secure instrument stability. For every sample, ~10^5 cells were analyzed. Cell gate definition is shown in Supplementary Figure S11. The analog current signal delivered by the photomultiplier tubes was amplified and converted into a voltage signal within a range of 0–10 V in a preamplifier. This voltage signal was then amplified onto a 4-decade logarithmic range. The analog-to-digital converter assigns each event in this signal to one of 1,024 intensity channels according to the event peak height.

RESULTS AND DISCUSSION

In the present study, the heterogeneity quantitation approach was developed with considerations of single-cell anabolic activity derived from nanoSIMS data. The applicability of the suggested heterogeneity indices was proved for the analysis of (i) cellular DNA content measured with flow cytometry of DAPI-stained single cells and (ii) length of single cells acquired with fluorescence microscopy.

Quantification of Single-Cell Anabolic Activity and Evaluation of Activity Distribution

The two Pseudomonas strains were grown in batch cultures with defined starting concentrations of 13C-labeled and unlabeled acetate. Afterwards, SIP–nanoSIMS experiments were performed to follow the biosynthetic (anabolic) activity and to compare the metabolic differences within the two monoclonal populations. Single-cell activity was expressed as relative assimilation (K_A, Equation 1) (Stryhanyuk et al., 2018):

\[
K_A = \frac{R_f - R_i}{(R_i + 1) \times (D_{pi} \times (R_f + 1) - R_f)}
\]

where R_f and R_i are the final and initial cellular isotope ratios, respectively, and D_{pi} is the fraction of heavy isotope in the growth substrate during incubation.

The distribution of single cells in K_A was considered for the analysis of anabolic heterogeneity within each population. The applicability of existing dispersion expressions for quantitation of heterogeneity is limited by the shape of the distribution. When in a certain population single cells fit a normal distribution in their anabolic activity, the dispersion in K_A (i ∈ [1, n]) values around the mean ( \overline{K_A} ),

\[
\text{mean (}K_A\text{)} \equiv \overline{K_A} = \frac{\sum_{i=1}^{n} K_{Ai}}{n},
\]

is expressed for a population of n cells as standard deviation (σ):

\[
\sigma = \sqrt{\frac{\sum_{i=1}^{n} (K_{Ai} - \overline{K_A})^2}{n}}.
\]

The 2σ range represents the distribution width (DW) comprising 68.2% of the population distributed equally (±34.1%) around \overline{K_A}. The probability density function of a normal distribution reveals the symmetrical bell-shaped profile with the probability maximum (mode) and the median centered at the mean value (\overline{K_A}).

The variability inside the population can be thus quantified by the CV calculated as

\[
\text{CV} = \frac{\sigma}{\overline{K_A}}.
\]

In previous studies, the CV has been applied for the evaluation of population heterogeneity and biological noise (Grover and Woldringh, 2001; Bar-Even et al., 2006; Newman et al., 2006; Simpson et al., 2009; Schreiber et al., 2016; Nikolic et al., 2017). However, the biosynthetic activity of individual cells may deviate from a normal distribution even in monoclonal populations, which calls for more comprehensive approaches to evaluate heterogeneity at the single-cell level.

Quantitation of the Distribution Width for Heterogeneity Measurement

Even in actively growing populations, the single-cell anabolic activity distribution cannot always be described with a symmetric probability density function. When a distribution of single cells in anabolic activity is asymmetric, the \overline{K_A} is displaced from the distribution maximum (mode, Figure 1), rendering the CV (Equation 4) unsuitable. The median value

\[
\tilde{K}_A = \text{median (}K_A\text{)}, \quad i \in [1, n],
\]

is more appropriate since it represents exactly the centroid of an asymmetric distribution rather than the average of its terms (like \overline{K_A} does). Thus, to express a robust measure of heterogeneity, considering the distribution asymmetry, the centroid of the activity distribution derived as \tilde{K}_A has to substitute the \overline{K_A} in the denominator of Equation (4). Additionally, to keep the quantitation of heterogeneity consistent with the CV, half of
The distribution width (i.e., DW/2) has to be considered in the nominator, exactly as half of the $2\sigma$ range is considered in the nominator of the CV expression (Equation 4). With the fulfillment of these two objectives, that is, robustness and consistency, the HC has been developed. By means of this index, the anabolic heterogeneity of a population can be expressed with the DW in $K_A$ normalized to the $\tilde{K}_A$ as follows:

$$HC = \frac{1}{2} \times \frac{DW}{\tilde{K}_A}. \quad (6)$$

The DW can be measured in different ways, for example, as standard deviation ($\sigma$); median absolute deviation (MAD); and min–max, interquantile, or interquartile range. With the MAD, the deviation of a single-cell $K_{Ai}$ value from the median centroid ($\tilde{K}_A$) of activity distribution is expressed as

$$MAD = \text{median} \left( |K_{Ai} - \tilde{K}_A| \right), \quad i = 1, \ldots, n. \quad (7)$$

In the case of normal distribution, the $2 \times$ MAD range ($MAD \approx 0.67449 \times \sigma$) comprises 68% of the population equally distributed ($\pm 34.1\%$) around $\tilde{K}_A$. With the DW measured as $2 \times$ MAD, the HC expression (Equation 6) returns the COD (Bonett and Seier, 2006):

$$COD = \frac{MAD}{\tilde{K}_A}. \quad (8)$$

Thus, as a measure of heterogeneity, COD substantially reduces sensitivity to the distribution asymmetry.

Expressing the DW in terms of quantiles $Q_{(p)}$ provides the possibility of tuning the sensitivity of a heterogeneity index to the distribution asymmetry. The $Q_{(p)}$ quantile splits the distribution into two parts comprising $p\%$ and $100 - p\%$ of a population. A $Q_{(p)}$ quantile equals the $K_A$ value, which is higher than those values revealed by $p\%$ of cells in the population. The percentile $p\%$ varies within a range of 0–100%, where 100% corresponds to the whole population. Thus, quantile $Q_{(50)}$ of $K_A$ distribution equals the median value $\tilde{K}_A$. With a set of three quantiles ($Q_1$, $Q_2$, and $Q_3$; Figure 1),

$$Q_1 = Q_{(p)}; \quad Q_2 = Q_{(50)}; \quad Q_3 = Q_{(100 - p)},$$

the DW can be expressed as an interquartile range: $\Delta_{i-j} = Q_j - Q_i$.

When the percentile $P$ equals 25%, the quantiles $Q_1$, $Q_2$, and $Q_3$ become $Q_{(25)}$, $Q_{(50)}$, and $Q_{(75)}$, respectively, and are called quartiles dividing a cell population into four quarters with an equal cell number. The following measures of DW are expressed in terms of quartiles:

- interquartile range, $\Delta_{1-3} = Q_{(75)} - Q_{(25)}$;
- semi-interquartile deviation, $Q_A = 0.5 \times (Q_{(75)} - Q_{(25)})$; and
- CQD = $(Q_{(75)} - Q_{(25)})/(Q_{(75)} + Q_{(25)})$.

The expression of DW in quartiles provides a reduced sensitivity to the distribution asymmetry, since it considers just 50% of the population within the interquartile range. To tune the asymmetry contribution into a heterogeneity index, one can expand or reduce the interquartile range by varying the percentile $P$-value. Hence, the maximum sensitivity is achieved with DW expressed with $Q_{(10)}$ and $Q_{(100)}$ as the [min–max] range. The calculation of heterogeneity indices with different definitions of DW is implemented in Supplementary Table S1. However, to render the HC (Equation 6) consistent with the CV (Equation 4), the interquantile range $\Delta_{1-3} = Q_{(84)} - Q_{(16)}$ was considered as DW in Equation (6) for further steps of the heterogeneity quantitation. In the case of normal distribution, the HC value
equals approximately the CV (Equation 4). Indeed, with a limited number of single cells, quantiles are defined by real single-cell values, whereas the calculated mean value does not necessarily match any experimentally observed one (because it represents the average of the observed values). For a discrete distribution, the quantiles $Q_{(16)}$, $Q_{(50)}$, and $Q_{(84)}$ return single-cell activity values that are not matching exactly the mean and the borders of the $2\sigma$ range (i.e., the $Q_{(16)}$ and $Q_{(84)}$ values are approximately matching the $\pm 34.1\%$ points around the mean); therefore, CV and HC values cannot be exactly the same.

**HC for Skewed Distributions**

The heterogeneity in the anabolic activity of monoclonal population causes the asymmetry of the $K_A$ distribution profile (Figure 1). For this reason, an asymmetry measure was considered in the expression of anabolic heterogeneity. The asymmetry is revealed as a distribution skew toward higher values (positive skew) or lower values (negative skew). The skew causes the displacement of the centroid ($Q_2$) from the maximum (mode), gaining the difference $\Delta = \Delta_{2-3} - \Delta_{1-2}$ between the interquantile distances (Figure 1). The degree of distribution asymmetry is measured with skewness ($Sk$) expressed in interquantile distances as

$$Sk = \frac{\Delta_{2-3} - \Delta_{1-2}}{\Delta_{1-3}}. \quad (10)$$
According to the interquantile range considered in the nominator of Equation (9), the difference \( \Delta \) is equal to
\[
\Delta = \Delta_{2-3} - \Delta_{1-2} = (Q_{84} - Q_{50}) - (Q_{50} - Q_{16})
\] (11)
Taking Equation (10) into account, \( \Delta \) can be expressed in terms of \( Sk \) and \( \Delta_{1-3} \) as follows:
\[
\Delta = |\Delta_{2-3} - \Delta_{1-2}| = |Sk| \times \Delta_{1-3}
\] (12)
To operate with positive \( \Delta \) values and thus make Equation (12) applicable for positively and negatively skewed distributions, the absolute value of skewness \( |Sk| \) was considered.
With the difference \( \Delta \) (Equation 12), the skewness was taken into account for the quantitation of heterogeneity. Expression of DW as the sum of the interquantile range \( \Delta_{1-3} \) and the \( \Delta \) value allows for the enhancement of the HC sensitivity to the data points scattered over the distribution tails (Figure 1).
\[
DW = \Delta_{1-3} + \Delta
\]
\[
\Delta_{1-3} = Q_{84} - Q_{16}
\] (13)
The \( \Delta \) expression in Equation (12) allows for rewriting of the DW (Equation 13) in terms of \( Sk \) and \( \Delta_{1-3} \) as follows:
\[
DW = \Delta_{1-3} + |Sk| \times \Delta_{1-3} = (1 + |Sk|) \times \Delta_{1-3}
\] (14)
With the DW expressed in terms of skewness and interquantile distance (Equation 14), the HC expression (Equation 6) can be now rewritten as
\[
HC = \frac{1}{2} \times \frac{DW}{\bar{K}_A} = \frac{(1 + |Sk|) \times \Delta_{1-3}}{2 \times Med}
\] (15)

or
\[
HC = \frac{(1 + |Sk|) \times (Q_3 - Q_1)}{2 \times Q_2}
\] (15’)
The introduction of the \( Sk \) weighting factor \( \varepsilon \) into Equations (15 and 15’) as
\[
HC = \frac{(1 + \varepsilon \times |Sk|) \times \Delta_{1-3}}{2 \times Q_2} = \frac{(1 + \varepsilon \times |Sk|) \times (Q_3 - Q_1)}{2 \times Q_2}
\] (16)
provides the possibility of adjusting the HC sensitivity to the skew by varying the DW within the \( \Delta_{1-3} \pm \Delta \) range with \( \varepsilon \in [-1; 1] \). With \( \varepsilon = 1 \), the DW = \( \Delta_{1-3} + \Delta \) (Figure 1) and Equation (16) turns to Equation (15). By setting the weighting factor as \( \varepsilon = -1 \) (Figure 1), we can narrow down the DW to \( \Delta_{1-3} - \Delta \), and the sensitivity is reduced. With \( \varepsilon = 0 \), the term \( \varepsilon \times |Sk| \) is nullified, and Equation (16) turns to Equation (9); even in this case though, it has to be considered that the skewness affects the HC value since \( \Delta_{1-3} \) and \( \bar{K}_A \) values are intrinsically influenced by the distribution skew.

If the \( K_A \) distribution is normal \( (\varepsilon \times |Sk| = 0, Q_2 = \bar{K}_A, \Delta_{1-3} = Q_3 - Q_2 = 2\sigma) \), the HC expressed with Equation (16) becomes approximately equal to CV (Equation 4).

**Correction of the HC for the CSE**

In the calculation of anabolic activity \( (K_A) \) from SIP–nanoSIMS data, the changes in cellular isotopic composition upon incubation with stable isotope labels are considered at the single-cell level (Sstryhanyuk et al., 2018). During nanoSIMS analysis, the quantification of cellular isotopic composition is based on the isotope ion counts accumulated in imaging mode for each pixel within a single cell. Therefore, it is important to evaluate the effect of CSE of the acquired data on the derived \( K_A \) values considered for quantitation of heterogeneity.

The CSE is derived as the square-root of secondary ion counts \( (N) \) (Sprawls, 1995).
\[
CSE = \sqrt{N}
\]
Taking as example Carbon (C) isotopes, the calculation of their ratio \( (R) \) considers the counts of \( ^{13}\text{C}^- \) and \( ^{12}\text{C}^- \) isotope ions as
\[
R = \frac{N_{^{13}\text{C}}}{N_{^{12}\text{C}}}
\]
and implies the propagation of CSE (Fitzsimons et al., 2000) into the Counting Statistics Variations of isotope ratios \( (CSV_R) \).
\[
CSV_R = \sqrt{\left( \frac{\partial R}{\partial N_{^{13}\text{C}}} \times CSE_{^{13}\text{C}} \right)^2 + \left( \frac{\partial R}{\partial N_{^{12}\text{C}}} \times CSE_{^{12}\text{C}} \right)^2} = \sqrt{\left( \frac{CSE_{^{13}\text{C}}}{N_{^{12}\text{C}}} \right)^2 + \left( \frac{N_{^{13}\text{C}} \times CSE_{^{12}\text{C}}}{N_{^{12}\text{C}}^2} \right)^2}
\] (17)
Thus, the presence of the isotope ratios \( (R_i \) and \( R_f \) in the expression of \( K_A \) (Equation 1) implicates the CSE propagation (Fitzsimons et al., 2000) into the variation of calculated \( K_A \) values \( (CSV_{K_A}) \).
\[
CSV_{K_A} = \sqrt{\left( \frac{\partial K_A}{\partial R_f} \times CSV_{R_f} \right)^2 + \left( \frac{\partial K_A}{\partial R_i} \times CSV_{R_i} \right)^2} = \sqrt{\left( \frac{K_A}{R_f} \times CSV_{R_f} \right)^2 + \left( \frac{K_A}{R_i} \times CSV_{R_i} \right)^2}
\] (18)
Consequently, the propagation of CSE introduces the term of Counting Statistics Heterogeneity \( (CSH_{K_A}) \) as an error into the anabolic activity derived from the SIP–nanoSIMS experiment that can be expressed as:
\[
CSH_{K_A} = \frac{CSV_{K_A}}{K_A}
\] (19)
The correction for the CSE propagation is especially necessary when the HC value (Equation 16), derived for the cell anabolism \( (K_A) \), approaches the \( CSH_{K_A} \) term (Equation 19).
FIGURE 2 | Counting statistics variation $CSV_{KA}$ (Equation 18, dashed line) and counting statistics heterogeneity $CSH_{KA}$ (Equation 19, solid line) calculated for the $^{13}$C isotope fraction $D_{GS}$ of the growth substrate corresponding to 10 at% (A) and 20 at% (B). Two values of the total carbon ion counts are represented: $N = 15 \times 10^3$ (thin lines) and $N = 5 \times 10^4$ (thick lines). The initial isotope ratio $R_i = 0.011$, corresponding to 1% of $^{13}$C cellular abundance, and the $CSV_{R_i} = 0$ are considered.

The $CSV_{KA}$ and the corresponding $CSH_{KA}$ depend on $KA$ (Figure 2) and are defined with

i) accumulated ion counts ($N$) for both isotopes,

ii) isotope-labeled substrate content ($D_{GS}$) in the growth substrate.

Considering different $CSV_{KA}$ values for $Q_i$ quantiles of a certain activity distribution, the CSE propagation into the HC can be expressed more precisely with $HC$ in the following way:

$$\Delta HC = \frac{\sum_{i=1}^{3} \left[ \frac{\partial HC}{\partial Q_i} \times CSV \{Q_i\} \right]^2}{2}$$  \hspace{1cm} (19')

where $CSV \{Q_i\}$ are the $CSV_{KA}$ values (Equation 18) corresponding to $Q_i$ quantiles ($i = 1, 2, 3$) in an experimentally derived distribution of cellular anabolic activity. With the denotation of constant term

$$\frac{(1 + \varepsilon \times |Sk|)}{2} \equiv C$$

in the $HC$ expression (Equation 16), the partial derivatives are calculated as follows:

$$\frac{\partial HC}{\partial Q_1} = \frac{C}{Q_2};$$

$$\frac{\partial HC}{\partial Q_2} = \frac{C \times (Q_3 - Q_1)}{(Q_2)^2};$$

$$\frac{\partial HC}{\partial Q_3} = \frac{C \times C}{Q_2}.$$

The correction of $HC$ for the CSE propagation was implemented via the subtraction of the $CSV_{KA}$ from the dispersion of experimental data in the following steps:

1) calculation of the $CSV_{KA}$ for $KA$ values corresponding to $Q_1$ and $Q_3$ quantiles (i.e., $CSV \{Q_1\}$ and $CSV \{Q_3\}$) with Equation (18) accounting for the isotope ion counts ($N^{12}$C and $N^{13}$C) and the fraction of labeling isotope ($D_{GS}$) in the growth substrate;

2) calculation of the corrected interquantile range $\Delta_{1-3corr}$ by subtracting the $CSV$ defined by $CSV \{Q_1\}$ and $CSV \{Q_3\}$ (derived in the previous step)

$$\Delta_{1-3corr} = \sqrt{(Q_3 - Q_1)^2 - [(CSV \{Q_1\})^2 + (CSV \{Q_3\})^2]}$$  \hspace{1cm} (20)

from the interquantile range $\Delta_{1-3} = Q_3 - Q_1$ in the following way:

$$\Delta_{1-3corr} = \sqrt{(Q_3 - Q_1)^2 - [(CSV \{Q_1\})^2 + (CSV \{Q_3\})^2]}$$  \hspace{1cm} (20)

3) calculation of the $HC_{corr}$ with Equation (16) substituting the $\Delta_{1-3}$ with the $\Delta_{1-3corr}$ in the numerator and considering the experimental median $Q_2 = KA$ in the denominator:

$$HC_{corr} = \frac{(1 + \varepsilon \times |Sk|) \times \Delta_{1-3corr}}{2 \times KA}.$$
The HC_{corr} error due to the CSE propagation into the HC_{corr} calculation can be expressed in the following way:

\[ \Delta H_{corr} = \sqrt{\sum_{i=1}^{3} \left( \frac{\partial H_{corr}}{\partial Q_i} \times CSV \{Q_i\} \right)^2}. \]  

(21')

\[ \frac{\partial H_{corr}}{\partial Q_1} = \frac{C}{Q_2} \times \sqrt{(Q_3 - Q_1)^2 - [(CSV\{Q_1\})^2 + (CSV\{Q_3\})^2]} \]

\[ \frac{\partial H_{corr}}{\partial Q_2} = -\frac{C}{Q_2} \times \sqrt{(Q_3 - Q_1)^2 - [(CSV\{Q_1\})^2 + (CSV\{Q_3\})^2]} \]

\[ \frac{\partial H_{corr}}{\partial Q_3} = \frac{C}{Q_2} \times \sqrt{(Q_3 - Q_1)^2 - [(CSV\{Q_1\})^2 + (CSV\{Q_3\})^2]} \]

Figure 2 shows the comparison of CSH_{K_A} (solid lines) and CSV_{K_A} (dashed lines) for different experimental conditions (acquired ion counts N and concentrations of isotope-labeled growth substrate D_{gs}). The CSH_{K_A} and the CSV_{K_A} terms are enhanced with lower ion counts (N = 1.5 \times 10^4, thin lines) and lower D_{gs} (10 at%, Figure 2A) as compared with the case when higher N and D_{gs} values are considered, that is, N = 5 \times 10^4 (thick lines) and D_{gs} = 20 at% (Figure 2B). For K_A > 0.5, the steepness of the CSV_{K_A} and CSH_{K_A} profiles increases with the reduction of N and D_{gs}. High CSH_{K_A} values at K_A < 0.5 are due to the division of CSV_{K_A} by a small K_A value (Equation 19). With the K_A approaching zero, the CSV_{K_A} becomes comparable with the DW and the CSH_{K_A} reaches its asymptote. These trends cause a huge error in HC calculation rendering the correction for CSE impossible and the HC value unreliable, when small changes in cellular heavy-isotope enrichment (small R_f - R_e difference in Equation 1; K_A \rightarrow 0) are considered. Consequently, the anabolic heterogeneity cannot be expressed in terms of HC when cellular isotope content is close to its natural abundance (e.g., short-time incubation or time prior incubation with isotope-labeled substrates). Therefore, in this study HC values were not provided for samples before incubation with 13C-labeled acetate (hereafter T0).

The increase in CSV_{K_A} and CSH_{K_A} is steeper when the cellular isotope-label content (D_f) approaches the D_{gs} value,

\[ D_f = \frac{R_f}{(R_f + 1)}. \]  

(22)

In biological systems, the cellular isotope enrichment D_f may approach but not exceed the D_{gs} because the cells retain the original unlabeled material along cell division. The only exception is possible when metabolic reactions reveal an isotopic fractionation factor smaller than the unit (\( R_f = \frac{R_f}{1} = \alpha < 1 \)) (Stryhanyuk et al., 2018). The values of K_A, CSV_{K_A}, and CSH_{K_A} show a strong increase at higher D_f especially when approaching their asymptote at D_f = D_{gs}. It is therefore not recommended to quantify the anabolic activity and the anabolic heterogeneity when the cellular isotope-label content D_f exceeds 0.6 \times D_{gs} (Stryhanyuk et al., 2018). An increase of D_f requires high amount of labeled substrate to be assimilated when D_f approaches the asymptote at D_{gs}. The D_f proximity to the corresponding K_A asymptote causes a strong increase in \[ \frac{\partial K_A}{\partial Q}. \]

Equation (18) providing a higher error in K_A calculation as well as an enhancement of the CSE propagation into the CSV_{K_A}. In turn, the increase in CSV_{K_A} intensifies the CSH_{K_A} term (Equation 19) of heterogeneity inaccuracy originated from the CSE.

HC Applicability to NanoSIMS Data

To test the applicability of the derived HC (Equation 16) on quantitation of anabolic heterogeneity, two Pseudomonas strains were incubated with 13C-labeled acetate, sampled at different time points during their exponential growth and analyzed by nanoSIMS. The acquired isotope distribution maps were used to calculate the single-cell isotope content D_f.

The dependence of K_A on D_f (Equations 1 and 22) is not linear (dotted line in Figure 3A, right Y-axis); for example, to increase the 13C content (D_f) from 7 to 10 at%, the cells have to assimilate twice (\( \times 2.08 \)) more carbon (higher steepness causing larger K_A interval), as compared with the same (3 at%) increase in D_f from 2 to 5 at% (lower steepness causing smaller K_A interval). To illustrate this, the distribution of P. stutzeri cells in their 13C-fraction and K_A (Figures 3A,B) were shown. Due to this non-linearity, the dispersion of cells (DW) in K_A scale (Figure 3B) does not reproduce the one in D_f scale (Figure 3A) although the same P. stutzeri cells are represented in both D_f and K_A plots (Figures 3A,B). In comparison with the D_f distribution, the K_A distribution appears more compressed (Figures 3A,B) in the range of small K_A values (Figure 3A, 15 min); instead it is more stretched and better reveals the intra-population metabolic diversifications between cells with higher K_A values (Figure 3A, 60 min).

To compare the heterogeneity in anabolic activity of the two strains, the K_A distribution of P. putida cells (Figure 3C) were plotted with the same integration interval (0.02) as the one used for P. stutzeri histogram plots (Figure 3B). The histograms display different distribution in cellular anabolic activity between the two bacterial populations; however, this representation does not provide any quantitative measure of the heterogeneity in single-cell activity. For this reason, the developed HC-computation approach was applied for quantitation of heterogeneity and used afterwards for comparison between incubation time points of both strains.

As one would expect from an actively growing culture, the single-cell K_A values increase over the time of incubation with the isotope-labeled substrate. Meanwhile, the diversity in cellular anabolic activity causes the broadening of K_A distribution; that is, the DW varies over time. An increase in DW does not necessary indicate any rise of heterogeneity; namely, the HC value (Equation 15) is preserved when DW increases or decreases proportionally to the change in K_A centroid. The scattering of single-cell anabolic activity (K_A) revealed as DW increase and skew of K_A distribution, is indeed induced by the diversification.
in single-cell anabolism; that is, each cell incorporates a different amount of substrate at that specific time point.

The DW values were derived from the experimental data according to Equation (14). To account for the CSE, the corrected interquantile range \( \Delta_{1-3,corr} \) (Equation 20) was considered in Equation (14) (instead of \( \Delta_{1-3} \)). The CSE does depend neither on the skew nor on the width of the distribution, but is defined by the CSV\( K_A \) (Equation 18) depending on \( K_A \) term, collected isotope-ion counts \( (N_{12C} + N_{13C} \text{, Equation 17}) \) and content of heavy isotope \( (D_{gs}) \) in the growth substrate (Figure 2, dashed lines). Therefore, to account for CSE, the CSV\( K_A \) values (CSV \( \{Q_1\} \) and CSV \( \{Q_3\} \)) were subtracted from \( \Delta_{1-3} \) interquantile range (Equation 20) without taking the skewness into account. The CSV\( K_A \) dependence on \( K_A \) corresponding to the nanoSIMS applied experimental conditions for both \( Pseudomonas \) strains \( (N = N_{12C} + N_{13C} = 15 \times 10^3 \text{ counts}, D_{gs} = 20 \text{ at\%}) \) is shown in Figure 2B with the thin dashed line; thin solid line (Figure 2B) represents the corresponding \( CSH_{K_A} \) dependence. For the CSE correction, the CSV\( K_A \) values (Equation 18) were derived from experimental \( K_A \) values corresponding to \( Q_1 \) and \( Q_3 \) quantiles (i.e., CSV \( \{Q_1\} \) and CSV \( \{Q_3\} \)) in the activity distributions of both strains at every specific time point. The CSE-related error \( \Delta \text{HC} \) (Table 1) of the uncorrected HC (Equation 16) can be expressed with Equation (19'). Considering the \( \Delta_{1-3,corr} \) values, the HC\( \text{corr} \) and HC\( \text{corr} \) (Equations 21 and 21') were calculated with \( \varepsilon = 0 \) to demonstrate numerically the effect of CSE correction in the resulted heterogeneity index HC\( \text{corr} \) (Table 1).

Despite small CSV\( K_A \) values (0.01–0.04 for \( \tilde{K}_A \in [0.15, 0.80] \)) throughout constant applied nanoSIMS experimental conditions, the CSE correction results in HC\( \text{corr} \) deviating from the uncorrected HC in different extent for the analyzed strains.
TABLE 1 | Effect of counting statistics error (CSE) correction on the heterogeneity coefficient (HC).

| Incubation time (min) | \( \hat{K}_A \) | \( HC \pm \Delta HC \) | \( \Delta 1-3 \) | \( CSV_{K_A} \) | \( HC_{corr} \pm \Delta HC_{corr} \) | CV* |
|-----------------------|----------------|----------------|----------------|----------------|----------------|-------|
|                       |                | \( HC \) | \( \Delta HC \) | \( CSV(Q_1) \) | \( CSV(Q_3) \) | \( HC_{corr} \) | \( \Delta HC_{corr} \) |       |
| **Pseudomonas putida** |                |             |               |                |                |             |               |       |
| 60                    | 0.2451         | 0.9822      | 0.0878        | 0.4815         | 0.0042         | 0.9799       | 0.0879       | 0.9728 |
| 90                    | 0.4627         | 0.7434      | 0.0663        | 0.6882         | 0.0046         | 0.7417       | 0.0663       | 0.7953 |
| 120                   | 0.7291         | 0.6119      | 0.0524        | 0.8923         | 0.0042         | 0.6104       | 0.0624       | 0.6234 |
| **Pseudomonas stutzeri** |              |             |               |                |                |             |               |       |
| 15                    | 0.1847         | 0.1377      | 0.0462        | 0.0509         | 0.0105         | 0.1301       | 0.0487       | 0.3375 |
| 30                    | 0.3837         | 0.1530      | 0.0367        | 0.1174         | 0.0158         | 0.1488       | 0.0376       | 0.3437 |
| 60                    | 0.7403         | 0.0886      | 0.0364        | 0.1312         | 0.0313         | 0.0809       | 0.0398       | 0.2119 |

*Coefficient of variation (CV) values were added for comparison with \( HC_{corr} \) value proposed in the present study.

The effect of CSE correction is negligible (0.23–0.25%) for \( P. \ putida \) whereas it becomes considerable (2.75–8.70%) for \( P. \ stutzeri \) strain. Such a difference in the correction effect can be explained by the \( \Delta 1-3 \) in the nominator of \( HC_{corr} \) (Equation 21). The difference between HC and \( HC_{corr} \) is large when \( CSV_{K_A} \) (i.e., \( CSV(Q_1) \) and \( CSV(Q_3) \) values) are subtracted with Equation (20) from a relatively small \( \Delta 1-3 \) interquartile range, as observed for \( P. \ stutzeri \) (Figure 3B; Table 1). Instead \( P. \ putida \) revealed strong dispersion (large \( \Delta 1-3 \)) of single-cell anabolic activity (Figure 3C; Table 1) values exceeding considerably the \( CSV_{K_A} \), thus making the CSE-correction effect negligible.

The reduction of CSE is possible via optimization of the SIP-nanoSIMS experimental conditions (\( D_{SIP} \), \( D_f \), \( Ni_{2C} \) and \( Ni_{1C} \)) in order to achieve minimal \( CSV_{K_A} \) and \( CSV_{K_A} \). Increase of raster density and current of primary ions (PI) improve the counting statistics. However, the increase of PI current causes also the reduction in focus quality and may render the derived isotope-ratio values unreliable when the counting of major isotopes reaches saturation. Thus, a compromise in SIP-nanoSIMS conditions has to be found to deliver heterogeneity values with acceptable errors; that is, \( HC_{corr} \pm \Delta HC_{corr} \) (SI, Supplementary Figures S1, S2).

The HC calculation developed in the present study for nanoSIMS-derived data, including CSE correction, were incorporated in the up-to-date version of the supplementary Excel template (Supplementary Table S1). The developed HC expression can be applied as a measure of heterogeneity not only of cellular anabolic activity, but also of any parameter measured at the single-cell level within a population: for example, length, volume, fluorescence yield, and gene expression (Nikolic et al., 2013; Gangwe Nana et al., 2018; Simšek and Kim, 2018; Heyse et al., 2019). The supplementary Excel template was therefore extended for HC calculation on the data acquired with other single-cell-resolved techniques.

**Dispersion of Metabolic Activity in Monoclonal Subpopulations**

Different studies dealing with phenotypic heterogeneity have shown that a population splits in subpopulations with different functions and/or activities (Simpson et al., 2009; Arnoldini et al., 2014; Kotte et al., 2014; Lieder et al., 2014; Li et al., 2018; Simšek and Kim, 2018). The same outcome was found in our results, in particular for \( P. \ putida \) strain. The histograms of \( K_A \) distribution of both strains investigated here (Figures 3B,C) revealed several peaks (subpopulations) with different centroids of anabolic activity. For example, for \( P. \ putida \) cells after 60 min of incubation (Figure 3C), several subpopulations could be clearly resolved in the histogram (with \( \hat{K}_A \) at around 0.03, 0.27, 0.48, 0.76, and 0.95). Even without considering the outer two subpopulations, the cell number in the first three subpopulations is comparable, rendering the overall DW of the entire distribution not representative for the \( K_A \) dispersion of the single subpopulations. Also, the consideration of a single centroid value (\( \hat{K}_A \)) renders the HC calculation inappropriate when the cell activity is distributed over several subpopulations. Thus, a reliable approach for quantitation of heterogeneity upon multimodal anabolic activity is necessary, ensuring the elucidation of (i) the activity dispersion in monoclonal subpopulations and (ii) its propagation into the entire microbial population.

Empirical approximation can be applied to study the relation between cellular properties, surrounding environments and experimentally observed development of heterogeneity in single-cell functions. Empirical analysis has been employed in allometry, for example, to describe the relation between animal body mass and its metabolic rate with the power function according to Kleiber’s law (Kleiber, 1947). Another example of empirical approximation is the relation between cell dry mass and its volume described with a power function (Loferer-Krössbacher et al., 1998).
Furthermore, the distribution of single-characteristic values ranked in their descending order has been also described with the power function. George K. Zipf had shown (Equation 23) (Zipf, 1935) that the frequency \( f \) of a word in a text of natural language is inversely proportional to the rank \( r \) of a word in the word frequency table sorted in frequency descending order (Powers, 1998),

\[
f(r; s, N) = \frac{1}{\sum_{n=1}^{N} \left( \frac{1}{r} \right)^s}
\]  

(23)

where \( N \) is the number of words in a text.

The power function \( f(r; s, N) \) Equation (23) represents the Zipf’s law and thanks to the variable power index \( s \) is applicable in different fields whenever the power law is obeyed. George U. Yule had applied the analysis of rank–frequency distribution to study new genera development from monotypic genus during evolution (Yule, 1925). Power law approximation has been employed also in ecological applications to quantify the relative abundance of species in different ecosystems (Mouillot and Lepretre, 2000). The statistical properties of selfish spreading DNA repeats were also studied with a power-law approach, explaining how the high abundance of long DNA elements does not depend on the coded functions but rather on the ability of selfishly spreading in the host genome (Sheinman et al., 2016). The rank–frequency distribution of guanine/cytosine nucleotides in mitochondrial DNA has been approximated with Zipf’s law to distinguish families and genera for taxonomic classification (Rovenchak, 2018).

In the present study, we applied the power law approach (i) to measure the ability of single cells to differentiate from each other in their anabolic activity and (ii) to account for the subpopulations inside a monoclonal population. The derivation of the \( s \) index in single-cell rank–activity distribution allowed for quantitation of anabolic heterogeneity as explained below.

**Differentiation Tendency in Cellular Anabolic Activity**

The development of metabolic heterogeneity was resolved with the rank–activity distribution of single cells plotted in double-logarithmic scale. The analyzed cells were ranked according to their anabolic activity (\( K_A \)) sorted in descending order; the cells with a low \( r \) show higher anabolism as compared to those at high \( r \) ranges.

The rank–activity plot of *P. putida* cell at 60-min time point (Figure 4, open circles) shows the distribution of single cells along multiple steps. Each step shows a specific slope describing the tendency of the cells to differentiate in their anabolic activity. This differentiation tendency of single cells causes the population heterogeneity and can be measured for each subpopulation by the steepness of the corresponding slope in the rank–activity distribution. In the present study, common trend in the activity differentiation, that is, fitting a single slope, was considered as criterion for assignment of single cells to a certain subpopulation. The \( K_A \) distribution slope revealed in the double-logarithmic coordinates plot (log \( K_A \) : log \( r \) ) can be described with \( s \) exponent of a power function as:

\[
K_A \propto C \times r^{-s}; \quad C \equiv \text{constant}
\]

\[
\log(K_A) = \log(C) - s \times \log(r)
\]  

(24)

predicting the anabolic activity \( (K_A) \) of single cells with their rank \( r \) in the cells series sorted in \( K_A \) descending order.

The \( K_A \) stays constant \( (K_A = C) \) with the power index \( s = 0 \) (Equation 24) describing the case in which cells grow without differentiation in their anabolism and their heterogeneity is approaching zero. Higher power index values \( s \) indicate a steeper slope of rank–activity distribution, implying an increase in the differentiation of anabolic activity and a heterogeneity gain. Thus, the power index \( s \) was considered as a measure of anabolic heterogeneity and is referred hereafter as DTI \( (s) \).

**Quantitation of Differentiation Tendency in Anabolic Activity**

Rank distributions of real experimental data, following a power law, show usually a single slope in the range of intermediate \( r \) values, but its profile in the range of low and high \( r \) values shows a considerable deviation from a single slope. Numerous modifications of the Zipfian function (Yule, 1944; Mandelbrot, 1953; Simon, 1955; Lavalette, 1996) have been suggested for the improvement of experimental data fit at either low or high \( r \) values. Poor fit accuracy and lack of meaningful parameters’ interpretation limit the application of power law functions for approximation of experimental data. In the present study, the expression of the Zipfian function Equation (25), previously suggested by Voloshynovska for the analysis of word frequencies in texts (Voloshynovska, 2011), was adopted...
for the approximation of the rank–activity distribution of our experimental data (Figure 4):\[ K_A(r; q, s, N) = C \times \left( \frac{N \times [r + q]}{(N - [r + q] + 1)} \right)^{-s} ; \quad C \equiv \text{constant.} \quad (25) \]

The \( r + q \) term, suggested by Mandelbrot, provides the possibility to improve the Zipfian fit in the low-rank range (small \( r \) and high \( K_A \) values). Lavalette’s term \((N^{\frac{q}{r+1}})\) facilitates a more accurate fit of rank–activity distribution in the high \( r \) range (high \( r \) approaching \( N \) and low \( K_A \) values) and takes explicitly the size \((N)\) of a population (or a subpopulation) into account.

In the case of cell rank–activity distribution, the \( s \) in the power exponent (Equations 24 and 25) stays invariant to linear normalization and to population size (Voloshynovska, 2011), allowing therefore for the following:

i) comparison of heterogeneity in subpopulations with different \( K_A \) centroids;

ii) quantitation of differentiation tendency in anabolic activity of subpopulations with low cell abundance;

iii) resolution of subpopulations possessing close median values; and

iv) elucidation of differentiation in anabolic activity even after a short incubation with stable-isotope-labeled growth substrates.

Importantly, the CSE propagation does not influence the slope of rank–activity distribution as it does on the DW considered in the HC expression (Equation 6). Therefore, DTI is less sensitive to the CSV\(_{K_A}\) variation (Equation 18) as long as \( D_f \) is kept well below \( D_{gi} \), and a sufficient number of isotope ion counts is accumulated during the data acquisition with nanoSIMS; otherwise, high CSV\(_{K_A}\) value smears the rank–activity distribution and may enhance the uncertainty in the slope determination, lowering the accuracy of the experimental data fit (Figure 4) considered below.

For a unimodal anabolic activity, the rank–activity distribution was expected to show a single slope, namely, without a subpopulation outcome. To underpin this assumption, simulated normal distributions (Supplementary Figures S3A, S4) were approximated with the Zipfian function (Equation 25). Normal distributions were simulated, keeping the mean value fixed and changing the \( \sigma \) values; then DTI and CV values were calculated for each of the distribution, revealing the same trend in \( \sigma \) (Supplementary Figure S3B). This behavior of DTI provided proof of its suitability as a heterogeneity index under unimodal activity. Moreover, the Zipfian approximation delivers the slope values \( s = 0.1624 \pm 0.0006 \) for 1,000 cells, \( s = 0.1631 \pm 0.0011 \) for 200 cells, and \( s = 0.1619 \pm 0.0082 \) for 50 cells, proving the invariance of \( s \) (DTI) to the population size, as shown in Supplementary Figure S5.

For the instance in which the population is split in multiple subpopulations (multimodal distribution, Figure 4), the rank–activity distribution was implemented here with the \( K_A(r; q, s, N) \) expression (Equation 25) represented as a combination of subpopulation activities \( K_{Ai} \) \( (i \in \{1, m\}) \):

\[
K_A(r; q, s, d, n) = \sum_{d_i \leq r < d_{i+1}} \left( \frac{n_i}{n_i - [r - d_i + q_i] + 1} \right)^{-s_i} \times \left( \frac{n_i}{n_i - [r - d_i + q_i] + 1} \right)^{-s_i}.
\]

\( m \): number of subpopulations;

\( d_i \): rank of cells with the highest activity in a subpopulation \( (i \in \{1, m\}) \);

\( C_i \): scaling constant of the activity in a subpopulation \( (i \in \{1, m\}) \);

\( q_i \): Mandelbrot’s parameter for a subpopulation \( (i \in \{1, m\}) \); \( s_i \): slope of rank–activity distribution in a subpopulation \( (i \in \{1, m\}) \);

\( r \): rank \( (r \in \{1, N\}) \);

\( N \): total number of cells in the whole population \( (N = \sum_{i=1}^{m} n_i) \).

The approximation of an experimental rank–activity distribution with \( K_A(r; q, s, d, n) \) in this multicomponent function (Equation 26) provides a set of slope values \( s_i \) each of them characterizing the DTI in the anabolic activity of a subpopulation. In order to obtain a unique value, able to describe how the differentiation in anabolic activity of single subpopulations is propagated into the entire monoclonal population, the CDTI \( (S) \) was introduced. The CDTI was calculated as the distance in the \( m \)-dimensional Euclidean space considering \( s_i \) values of \( m \) subpopulations as differentiation propagation vectors using the following expression:

\[
S = \sqrt{\sum_{i=1}^{m} (\gamma_i \times s_i)^2} \quad (i \in \{1, m\})
\]

The coefficients \( \gamma_i \) were introduced to weight the contribution of each subpopulation into the cumulative differentiation. The weighting term \( \gamma_i \) is expressed as a logarithm of a cell number \( (n_i) \) in a subpopulation relative to the logarithm of the total number of cells in the whole population \( (N) \).

\[
\gamma_i = \frac{\log(n_i)}{\log(N)} \quad (i \in \{1, m\})
\]

Taking the weighting term into account, the CDTI was expressed in the following way:

\[
S = \sqrt{\sum_{i=1}^{m} \left( \frac{\log(n_i)}{\log(N)} \times s_i \right)^2} \quad (i \in \{1, m\}).
\]

The propagation of the fit errors \( \Delta s_i \) into the calculated CDTI (Equation 27) results in \( \Delta S \) expressed as

\[
\Delta S = \sqrt{\sum_{i=1}^{m} \left( \frac{\partial S}{\partial s_i} \times \Delta s_i \right)^2} = \sqrt{\sum_{i=1}^{m} \frac{[\gamma_i^2 \times s_i \times \Delta s_i]^2}{\sum_{i=1}^{m} (\gamma_i \times s_i)^2}}.
\]
From NanoSIMS Data

Power Law Fit of Rank–Activity Distribution Derived From NanoSIMS Data

The CDTI calculations (S and ΔS) developed in the present study were implemented in the supplementary Excel template (Supplementary Table S2).

The rank–activity distributions of P. putida cells after different incubation times were approximated with the multicomponent Zipfian function (Equation 26; Supplementary Figures S6A, S7). First, DTI \( s_i \pm \Delta s_i \) values were derived as fitting parameters for all the revealed subpopulations; then the CDTI \( S \pm \Delta S \) values (Equations 27 and 28; Supplementary Table S2) were calculated and compared with the corresponding HCcorr values (Figure 5).

Even when calculated with \( \varepsilon \in [-1; 1] \) (Equation 21), the HCcorr (Figure 5A) did reveal neither a clear correlation nor a common trend with the corresponding CDTI (Figure 5B) for P. putida. To clarify this discrepancy, the rank–activity plot was overlaid with the corresponding histogram (Figure 4). The representation of \( K_A \) distribution with a rank–activity plot allows for distinction between cells with close \( K_A \) values, providing a single-cell resolution, which is not achievable with a histogram plot. The profiles of rank–activity distribution (Figure 4; Supplementary Figures S7, S9) show the cells dispersed into two main groups. In general, the cells in the lower-rank range (high \( K_A \) values) are fitting moderate slopes, whereas those in the high-rank range show considerably steeper slopes. Thus, the population is clearly split into two cell–activity trends: (i) weak differentiation tendency (WDT) and (ii) strong differentiation tendency (SDT). The SDT cells (steeper slopes) can also be observed in the lower-rank range (high \( K_A \) values) as revealed by P. putida populations at time points of 90 and 120 min (Supplementary Figure S7). The SDT cells showing high anabolic activity (high \( K_A \) values) are metabolically different from those showing low anabolic activity (low \( K_A \) values). Therefore, high-SDT cells (high \( K_A \) values) were differentiated from low-SDT cells (low \( K_A \) values), and the CDTI values were calculated separately for the corresponding SDT subpopulations of P. putida at 90- and 120-min time points (Figure 5C). At a time point of 60 min, P. putida populations did not show high-SDT cells; consequently, the CDTI values were calculated for WDT and low-SDT cells. Considering the CDTI calculated separately (Figure 5C), the SDT subpopulations reproduced the descending trend of the HCcorr derived with \( \varepsilon \leq 0 \) (Figure 5A), providing the major contribution to the heterogeneity of the entire population (CDTI, Figure 5B).

The activity of P. stutzeri single cells (Figure 3B), contributing to the main peak in the histogram, was considered to be unimodal. Those cells were ascribed to the WDT subpopulation, and their rank–activity distribution was initially approximated with the single-component Zipfian function (Equation 25), delivering DTI as \( s \pm \Delta s \) values. However, this approximation provided considerably high relative errors \( \Delta s/s \times 100 \% \); Supplementary Figure S8). Consequently, the rank–activity distributions of entire P. stutzeri populations were approximated (Supplementary Figures S6B, S9) with the multicomponent Zipfian function (Equation 26), and the CDTI values were derived as \( S \pm \Delta S \) (Supplementary Table S2) according to Equations 27 and 28 (Figure 6B). The low-SDT and WDT subpopulations were recognized in the rank–activity distribution, and the corresponding CDTI values were calculated separately (Figure 6C). The increase in CDTI from 15- to 30-min time points is driven by low-SDT subpopulations (Figure 6C, triangles), which correspond to the single cells showing low \( K_A \) values in the histogram (Figure 3A). Those cells are included into the HCcorr calculation when \( \varepsilon \geq 0 \) (Figure 6A), resulting in the HC trend following the one revealed by CDTI (Figure 6B) for the entire P. stutzeri population. Hence, the SDT subpopulations were shown to contribute mostly to the overall heterogeneity in anabolic activity of P. putida and P. stutzeri. The CDTI trends over time of both Pseudomonas strains are compared in Supplementary Figure S10.

As mentioned in Correction of the HC for the CSE, the calculation of HC on SIP–nanoSIMS data is unreliable for T0
The corresponding heterogeneity indices derived for *Pseudomonas stutzeri* after different incubation times. (A) Comparison of the HC$_{corr}$ (Equation 21) calculated with different $\varepsilon$ values; error bars represent the $\Delta$HC$_{corr}$ (Equation 21'). (B) Cumulated differentiation tendency index (CDTI) trend of entire populations (represented as $S$ $\pm$ $\Delta$S; Equations 27 and 28); error bars show the $\pm$ $\Delta$S intervals. (C) CDTI calculated (Equations 27 and 28) for cell subpopulations revealing weak differentiation tendency (WDT) and strong differentiation tendency at low $K_A$ (low SDT).

**Method Applicability**

To demonstrate the broad applicability of the suggested heterogeneity indices, HC and DTI/CDTI were derived for two different datasets: (i) the distribution of DAPI-stained single cells of *P. putida* acquired with flow cytometry and (ii) the distribution of *Escherichia coli* single-cell length upon different growth conditions reported by Nikolic et al. (2017).

**Heterogeneity Quantitation on Flow Cytometry Data**

The *P. putida* cells were analyzed with flow cytometry at eight time points upon cultivation in batch for 26 h. The cellular DNA content was measured as DAPI fluorescence intensity and visualized in dot plots (Figure 7A; Supplementary Figure S11) together with the cell-size-related FSC intensity. The cells were distributed within a subpopulations' pattern that changed dynamically during cultivation. The boundaries between the five subpopulations G1, G2, G3, G4, and Gx were defined with this pattern according to the local minima in the DAPI fluorescence intensity histogram (Figure 7B) at inoculation (0 h). The distribution of cells along this histogram represents the different chromosome numbers per cell that can occur during different states of cell growth. During the cell cycle, cells increase in size (frequently, but not always) while duplicating their DNA, and consequently, their FSC intensity increases (Müller, 2007).

The observed changes in cellular DNA contents result in heterogeneous distributions that can be described by the HC index and the CDTI. For this purpose, the flow cytometric raw data in FCS 3.0 format (Seamer et al., 1997) were treated as described in Supplementary Figure S11. In short, the values in logarithmic scale were translated into equally distributed 1,024 channels. In a next step, only the range of channels that represented cells was plotted. Finally, the cells were ranked according to their DAPI fluorescence intensity, and the multicomponent Zipfian function (Equation 26) was fitted to the resulting rank–distribution curves (Figure 7). The changes in the Zipfian slope (Figure 7C) matched the established subpopulation segregation, deduced from the dot plots (Figure 7A) and shown by the corresponding histograms (Figure 7B). The heterogeneity in cellular DNA content was expressed with the HC index (despite the multimodality) and the CDTI calculated (Supplementary Table S2) with the results of multicomponent Zipfian approximation. The development over the 26-h growth curve of both indices is represented in Figure 8.

For the initial and final growth time points, the HC was highly dependent on the $\varepsilon$ value, which can be used to adjust the index sensitivity to the distribution skew. The HC decreased from the time of inoculation, when the population is constituted by cells with varying chromosome numbers, to the 1-h time point with most cells containing only one chromosome. During cell growth, the HC showed the heterogeneity boost due to an increasing number of cells harboring two, four, and more chromosomes.

Interestingly, the HC values do not change much with the variation of $\varepsilon$ values during the exponential phase. The $\varepsilon$ value weights the asymmetry (skew) contribution in the HC. The reduced effect of $\varepsilon$ can be considered...
FIGURE 7 | Heterogeneity dynamics of *Pseudomonas putida* population during growth over 26 h. (A) DAPI fluorescence intensity (related to DNA content) vs. forward scatter (FSC) intensity (related to cell size) dot plots of the 0-, 2-, 6-, and 24-h samples after data transformation (details about the data transformation and dot plots of the remaining samples in Supplementary Figures S11, S12). (B) Histograms of DAPI fluorescence intensity distribution used to define the boundaries between the subpopulations G1–Gx, which correspond to the chromosome number in the cells. (C) Cells ranked according to their fluorescence intensity (open circles) together with the multicomponent Zipflan fit (solid red line).
as an evidence for a negligible skewness. In the case of multimodality, a small skewness value may be achieved with the centroid (median, Q3), located equidistant between two quantiles (Q1 and Q5), even though a single Q2 value is not representative in the case of a multimodal distribution. Finally, at the 24-h time point, with the population in the stationary phase, the HC values differentiate according to the chosen ε values again. In the present study, the error in graded fluorescence intensity was estimated to be 2 of 1,024 channels (<0.5%) and therefore neglected upon the HC calculations.

The HC describes the general dynamic of heterogeneity during the population growth, but its explanatory power is restricted in case of subpopulation segregation displayed by the P. putida culture. In this case, the CDTI is more suitable for the heterogeneity quantitation since it accounts for multimodality. The CDTI was initially calculated for the entire population with all slope values derived from the Zipfian fit (Figure 8B, rectangles). In order to avoid over-weighting the influence of low-probability events in Gx that likely comprise cell aggregates (0 h, 6.60%; 1 h, 0.45%; 2 h, 0.18%; 4 h, 4.44%; 6 h, 6.80%; 8 h, 9.49%; 24 h, 6.00%; and 26 h, 3.28%), the CDTI(G1-G4) (Figure 8B, circles) was additionally calculated for the cells in the clearly segregated subpopulations G1–G4. The CDTI value increased from 0 to 1 h when cells started to duplicate their chromosomes after a short lag phase. At 2 h, most cells have doubled their DNA content and subsequently start the duplication resulting in (i) the majority of cells having a single chromosome and (ii) heterogeneity decrease. In the exponential growth phase, the population showed increasing CDTI at 4 and 6 h with higher chromosome numbers and uncoupled DNA synthesis. The confinement of cell distribution in the G1–G4 subpopulations resulted in the reduction of CDTI at 8 h when the culture gets into the stationary phase. In the advanced stationary phase at 24 h, the cell distribution pattern is very similar to the one upon inoculation (0-h time point). This pattern similarity is supported by the close CDTI values derived for 0- and 24-h time points. The restarting cell activity after the acetate feeding at 24 h was captured by the slight increase of CDTI at 26 h. Thus, the CDTI (Figure 8B, circles) describes the dynamics of the population heterogeneity, whereas the HC (Figure 8A) represents its general trend missing the resolution of subpopulations.

**Heterogeneity Quantitation on Cell Length Distribution**

The heterogeneity quantitation was also applied to cell length distributions from a published dataset (Nikolic et al., 2017). In their work, E. coli cells, grown in the presence of two different carbon sources, did not specialize in a bimodal fashion but rather followed unimodality, consuming both substrates simultaneously at different rates. The highest degree of heterogeneity of the monoclonal population (measured with CV) was found during co-feeding with glucose (Glc) and arabinose (Ara) under carbon limitation as compared with the growth with one single carbon source under nitrogen limitation (Nikolic et al., 2017). This behavior was a consequence of different gene expression and transcription levels as well as different rates of single-cell growth in chemostat. Nutrient-dependent changes in the cultivation environment influence the growth rate, which in turn influences the cell size (Chien et al., 2012; Westfall and Levin, 2018). Bacterial cells have to find a compromise between the maintenance of a certain DNA amount and their cytoplasmic size in order to keep the DNA-to-cell mass ratio constant upon different growth rates. However, when external conditions start to be unfavorable, while cells continue to duplicate their DNA via multiple duplication forks, the mechanism of division via the FtsZ ring and other intracellular molecules is inhibited, thus resulting in a larger size of cells (Chien et al., 2012). The data on cell length provided by Nikolic et al. (2017), confirmed these experimental observations. This dataset was analyzed (Figure 9) to elucidate the heterogeneity in cell size upon different conditions reported in their study (Nikolic et al., 2017). The derived CV, HC index, and DTI are shown in Figure 10. The highest median value of cell length (Figure 10A,
FIGURE 9 | The distribution of single cells in their length represented for the different growth conditions reported in Nikolic et al. (2017) with histograms and rank–length plots.
circles) was found with the “3 μM Ara + 3 μM Glc” substrate composition, which represents the strongest carbon limitation. The HC value calculated with $-1 \leq \varepsilon \leq 0$ (Figure 10B) followed the CV trend (Figure 10A, rectangles), as expected for unimodal distribution. The approximation of rank–length distributions with the single-component Zipfian function delivered the slope values ($s$, DTI; Figure 10C), bringing two interesting outcomes: (i) the heterogeneity of population in cell size upon “3 μM Ara + 3 μM Glc” is not much higher than that upon “10 μM Ara + 10 μM Glc” as revealed with HC ($\varepsilon = 0$; Figure 10B) although this difference was overestimated with the CV (Figure 10A) and (ii) the length of single cells showed a unimodal distribution confirmed by small $\Delta s$ (and high goodness of fit), which is in agreement with the unimodality reported in Nikolic et al. (2017).

Guidelines on the Heterogeneity Quantitation

The summary of the approaches (HC and DTI/CDTI) developed in this study for heterogeneity quantitation is provided in Table 2. An accurate quantitation of cellular anabolic activity requires a reliable data acquisition with single-cell resolution. The aspects described for the SIP–nanoSIMS experiment (step 1 in Table 2) are, in general, relevant for experiments implying data acquisition in counting mode.

Each step in Table 2 contains a link to the corresponding section, equation, or figure in the main text or in SI. The calculation of HC including the CSE correction is implemented in the Excel template (Supplementary Table S1). The results of nanoSIMS data processing with the LANS software together with other input parameters (like $D_{po}$, $D_{fl}$, and $N$) have to be pasted into the appropriate green-marked fields, and the Excel template calculates the outputs including $K_A$, HC, and $HC_{corr}$ values. If the analyzed population of single cells follows unimodal anabolic activity, heterogeneity can be measured with HC or $HC_{corr}$.

Unimodal rank–activity distribution can be also approximated with the single-component Zipfian function expressing the heterogeneity with the Zipfian slope as $s \pm \Delta s$. The approximation of multimodal rank–activity distribution with single-component Zipfian delivers poor fit accuracy, resulting in large $\Delta s$ values. In such a case, the rank–activity distribution has to be approximated with the multicomponent Zipfian function, and heterogeneity is measured with CDTI ($S \pm \Delta S$). The CDTI calculation was also implemented in the corresponding Excel template (Supplementary Table S2). In this study, the OriginPro 2019 software was used for Zipfian approximation of rank–activity distributions.

When heterogeneity has to be quantitated on a set of epifluorescence microscopy data, the error in the HC and DTI/CDTI indices may arise from the following factors: (i) non-linearity of fluorescence signal detection; (ii) contribution of fluorescence background; (iii) presence of contaminants and extracellular substances. These factors may modulate the fluorescence intensity affecting the resulting HC and DTI values. Both indices are influenced by background subtraction while the DTI remains invariant to linear normalization and cell number.

CONCLUSIONS

In the present study, indices to measure metabolic heterogeneity were subjected to a comprehensive consideration. The expression of HC was developed for the quantitation of heterogeneity within a single-cell dataset. The adjustment of HC sensitivity to the distribution asymmetry was discussed and implemented in HC expression. When calculated with the same “sensitivity settings,” HC allows for heterogeneity comparison among different datasets. Importantly, the implemented HC expression returns the CV value if applied on datasets that reveal normal distribution. For the HC calculated from the nanoSIMS data, the correction for CSE was implemented. The influence of experimental conditions on CSE propagation
into the calculated HC was discussed, and appropriate recommendations for a SIP–nanoSIMS experiment setup were made.

Anabolic heterogeneity of monoclonal populations can be also measured with the slope \( s \) of a rank–activity distribution, characterizing the tendency of cells to differentiate in their activity within several subpopulations. The DTI \( (s) \) is derived for each subpopulation as an exponent of the Zipfian power law approximation function. The CDTI \( (S) \) is derived from DTI values of a subpopulation set and measures the anabolic heterogeneity of the entire population. The power exponent \( s \) (DTI) is invariant to linear normalization and to the number of cells in a subpopulation/pool. Both HC and DTI can be used as heterogeneity indices when a population shows unimodal anabolic activity. If the anabolic activity becomes multimodal, the DTI of each subpopulation \( (s_i) \) has to be first derived with multicomponent Zipfian approximation, and thereafter, the calculation of CDTI \( (S) \) delivers a measure of anabolic heterogeneity of the entire monoclonal population. Besides SIP–nanoSIMS, the applicability of both developed indices was shown on flow cytometry and epifluorescence microscopy datasets. The HC and DTI/CDTI were proven to be robust indices for quantitative and comparative analyses of heterogeneity in single-cell-resolved studies. Thus, these indices are not restricted to particular techniques, have a broad range of applications, and measure phenotypic heterogeneity in all its aspects: metabolic, physiological, and morphological.

### DATA AVAILABILITY STATEMENT

The flow cytometry dataset generated and analyzed in this study is available in the FlowRepository (ID: FR-FCM-Z2AS).

### AUTHOR CONTRIBUTIONS

FC, IV, and HS conceived the developed methods and developed the quantitation of differentiation tendency (DTI/CDTI). FC, NM, FM, SM, HR, and HS designed the experiments. FC performed the growth experiments and isotope labeling and processed the nanoSIMS data. FC and HS performed the nanoSIMS analysis, developed the mathematical expression for the heterogeneity coefficient (HC), and wrote the manuscript with contribution from all authors. SM and JL performed the flow cytometry experiment. SM, FC, JL, and HS processed the flow cytometry data. MS, LW, SM, JL, and MT provided advices and critical inputs. All authors contributed to manuscript revision and approved the submitted version.

### FUNDING

This work was funded by the Helmholtz Centre for Environmental Research (UFZ) and contributes to the integrated project Controlling Chemicals’ Fate (CCF) of the Research Topic Chemical in the Environment (CITE).
within the Research Program Terrestrial Environment of the Helmholtz Association.

ACKNOWLEDGMENTS

The authors acknowledge ProVIS Centre for Chemical Microscopy by the Helmholtz Centre for Environmental Research (UFZ) and the funds for its establishment provided within Europäischer Fonds für regionale Entwicklung (EFRE) and the Freistaat Sachsen Program. FM acknowledges the support from the Helmholtz Association of German Research Centers.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2019.02814/full#supplementary-material

REFERENCES

Ackermann, M. (2015). A functional perspective on phenotypic heterogeneity in microorganisms. Nat. Rev. Microbiol. 13:497. doi: 10.1038/nrmicro3491

Almendro, V., Marusyk, A., and Polyak, K. (2013). Cellular heterogeneity and molecular evolution in cancer. Annu. Rev. Pathol. Mech. Dis. 8, 277–302. doi: 10.1146/annurev-pathol-020712-163923

Arnoldini, M., Vizcarra, I. A., Peña-Miller, R., Stocker, N., Diard, M., Vogel, V., et al. (2014). Bistable expression of virulence genes in salmonella leads to the formation of an antibiotic-tolerant subpopulation. PLoS Biol. 12:e1001928. doi: 10.1371/journal.pbio.1001928

Avery, S. V. (2006). Microbial cell individuality and the underlying sources of heterogeneity. Nat. Rev. Microbiol. 4:577. doi: 10.1038/nrmicro1460

Balaban, N. Q., Merrin, J., Chait, R., Kowalik, L., and Leibler, S. (2004). Cell size control in bacteria.

Elowitz, M. B., Levine, A. J., Siggia, E. D., and Swain, P. S. (2002). Stochastic gene expression in a single cell. Science 297:1183. doi: 10.1126/science.1070919

Fitzsimons, I. C. W., Harte, B., and Clark, R. M. (2000). SIMS stable isotope measurement: counting statistics and analytical precision. Mineral. Mag. 64, 59–83. doi: 10.1108/002666100549139

Fraser, D., and Kærn, M. (2009). A chance at survival: gene expression noise and phenotypic diversification strategies. Mol. Microbiol. 71, 1333–1340. doi: 10.1111/j.1365-2958.2009.06605.x

Gangwe Nana, G. Y., Ripoll, C., Cabin-Flaman, A., Gibouin, D., Delaune, A., Janniere, L., et al. (2018). Division-based, growth rate diversity in bacteria. Front. Microbiol. 9:849. doi: 10.3389/fmicb.2018.00849

Gao, D., Huang, X., and Tao, Y. (2016). A critical review of NanoSIMS in analysis of microbial metabolic activities at single-cell level. Crit. Rev. Biotechnol. 36, 884–890. doi: 10.3109/07388551.2015.1057550

Gough, A., Stern, A. M., Maier, J., Lezon, T., Shun, T.-Y., Chennubhotla, C., et al. (2017). Biologically relevant heterogeneity: metrics and practical insights. SLAS Disocn. 22, 213–237. doi: 10.1177/2472555216682725

Grote, J., Kryszczak, D., and Streit, W. R. (2015). Phenotypic heterogeneity, a phenomenon that may explain why quorum sensing does not always result in truly homogeneous cell behavior. Appl. Environ. Microbiol. 81, 5280–5289. doi: 10.1128/AEM.00900-15

Grover, N. B., and Woldringh, C. L. (2001). Dimensional regulation of cell-cycle events in Escherichia coli during steady-state growth. Microbiology147, 171–181. doi: 10.1099/00221287-147-1-171

Hawkes, P. W., and Spence, C. H. J. (2007). Science of Microscopy, Vol. I. New York: Springer-Verlag. doi: 10.1007/978-0-387-49762-4

Heins, A. L., Johanson, T., Han, S., Lundin, L., Carlquist, M., Gernaey, K. V., et al. (2019). Quantitative flow cytometry to understand population heterogeneity in response to changes in substrate availability in Escherichia coli and Saccharomyces cerevisiae. Front. Microbiol. 10:12752. doi: 10.3389/fmicb.2019.00187

Heyse, J., Buyschaert, B., Props, R., Rubbens, P., Skirtach, A. G., Waegeman, W., et al. (2019). Coculturing bacteria leads to reduced phenotypic heterogeneities. Appl. Environ. Microbiol. 85, e02814-02818. doi: 10.1128/AEM.02814-18

Hlawacek, G., and Golzhäuser, A. (2016). Helium Ion Microscopy. Springer International Publishing. doi: 10.1007/978-3-319-41990-9

Jiang, H., Kilburn, M. R., Decelle, J., and Musat, N. (2016). NanoSIMS chemical imaging combined with corelative microscopy for biological sample analysis. Curr. Opin. Biotechnol. 41, 130–135. doi: 10.1016/j.copbio.2016.06.006

Kærn, M., Elston, T. C., Blake, W. J., and Collins, J. J. (2005). Stochasticity in gene expression: from theories to phenotypes. Nat. Rev. Genet. 6, 451–464. doi: 10.1038/nrg1615

Kiviet, D. J., Nghe, P., Walker, N., Boulouin, S., Sunderlïkova, V., and Tans, S. J. (2014). Stochasticity of metabolism and growth at the single-cell level. Nature 514:376. doi: 10.1038/nature13582

Kleiber, M. (1947). Body size and metabolic rate. Physiol. Rev. 27, 511–541. doi: 10.1152/physrev.1947.27.4.511

Koch, C., Günther, S., Desta, A. F., Hübschmann, T., and Müller, S. (2013). Cytometric fingerprinting for analyzing microbial intracommunity structure variation and identifying community substructure. Nat. Protoc. 8:190. doi: 10.1038/nprot.2012.149

Kopf Sebastian, H., McGlynn Shawn, E., Green-Saxena, A., Guan, Y., Newman Dianne, K., and Orphan Victoria, J. (2015). Heavy water and 15N labelling with NanoSIMS analysis reveals growth rate-dependent metabolic heterogeneity in chemostats. Environ. Microbiol. 17, 2542–2556. doi: 10.1111/1462-2920.12752
Van Den Bergh, B., Fauvart, M., and Michiels, J. (2017). Formation, physiology, ecology, evolution and clinical importance of bacterial persisters. *FEMS Microbiol. Rev.* 41, 219–251. doi: 10.1093/femsre/fux001

Van Heerden, J. H., Wortel, M. T., Bruggeman, F. J., Heijnen, J. J., Bollen, Y. J. M., Planqué, R., et al. (2014). Lost in transition: start-up of glycolysis yields subpopulations of non-growing cells. *Science* 343, 1245114. doi: 10.1126/science.1245114

Vasdekis, A. E., and Stephanopoulos, G. (2015). Review of methods to probe single cell metabolism and bioenergetics. *Metab. Eng.* 27, 115–135. doi: 10.1016/j.ymben.2014.09.007

Voloshynovska, I. A. (2011). Characteristic features of rank-probability word distribution in scientific and belletristic literature. *J. Quant. Linguist.* 18, 274–289. doi: 10.1080/09296174.2011.583405

West, S. A., and Cooper, G. A. (2016). Division of labour in microorganisms: an evolutionary perspective. *Nat. Rev. Microbiol.* 14:716. doi: 10.1038/nrmicro.2016.111

Westfall, C. S., and Levin, P. A. (2018). Comprehensive analysis of central carbon metabolism illuminates connections between nutrient availability, growth rate, and cell morphology in *Escherichia coli*. *PLoS Genet.* 14:e1007205. doi: 10.1371/journal.pgen.1007205

Widdel, F. (2010). “Cultivation of anaerobic microorganisms with hydrocarbons as growth substrates,” in *Handbook of Hydrocarbon and Lipid Microbiology*, ed K. N. Timmis (Heidelberg: Springer), 3787–3798.

Yule, G. U. (1925). *II-A mathematical theory of evolution, based on the conclusions of Dr. J. C. Willis, F. R. S.* Philosophical Transactions of the Royal Society of London. Series B, Containing Papers of a Biological Character 213, 21–87.

Yule, G. U. (1944). *The Statistical Study of Literary Vocabulary*. Cambridge, MA: The University of California Press.

Zimmermann, M., Escrig, S., Hübschmann, T., Kirsch, M. K., Brand, A., Inglis, R. F., et al. (2015). Phenotypic heterogeneity in metabolic traits among single cells of a rare bacterial species in its natural environment quantified with a combination of flow cell sorting and NanoSIMS. *Front. Microbiol.* 6:243. doi: 10.3389/fmicb.2015.00243

Zimmermann, M., Escrig, S., Lavik, G., Kuypers, M. M. M., Meibom, A., Ackermann, M., et al. (2018). Substrate and electron donor limitation induce phenotypic heterogeneity in different metabolic activities in a green sulphur bacterium. *Environ. Microbiol. Rep.* 10, 179–183. doi: 10.1111/1758-2229.12616

Zipf, G. K. (1935). *The Psycho-Biology of Language; an Introduction to Dynamic Philology*. Boston: Houghton Mifflin company.

**Conflict of Interest:** IV was employed by company le-tex publishing services GmbH.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Calabrese, Voloshynovska, Musat, Thullner, Schlömann, Richnow, Lambrecht, Müller, Wick and Stryhanyuk. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.