Analytical and Bioanalytical Chemistry
https://doi.org/10.1007/s00216-021-03505-2

RESEARCH PAPER

Exploiting targeted and untargeted approaches for the analysis of bacterial metabolites under altered growth conditions

Delphine Zanella[1] · Tiffany Liden[2] · Jamie York[2] · Flavio A. Franchina[1] · Jean-François Focant[1] · Kevin A. Schug[2,3]

Received: 10 May 2021 / Revised: 22 June 2021 / Accepted: 25 June 2021

Abstract

In the host, pathogenic microorganisms have developed stress responses to cope with constantly changing environments. Stress responses are directly related to changes in several metabolomic pathways, which could hamper microorganisms' unequivocal identification. We evaluated the effect of various in vitro stress conditions (acidic, basic, oxidative, ethanolic, and saline conditions) on the metabolism of *Staphylococcus aureus*, *Bacillus cereus*, and *Pseudomonas aeruginosa*, which are common lung pathogens. The metabolite profiles of the bacteria were analyzed using liquid chromatography coupled to triple quadrupole and quadrupole time-of-flight mass spectrometry. The advantages of targeted and untargeted analysis combined with univariate and multivariate statistical analysis (principal component analysis, hierarchical cluster analysis, partial least square discriminant analysis, random forest) were combined to unequivocally identify bacterial species. In normal in vitro conditions, the targeted methodology, based on the analysis of primary metabolites, enabled the rapid and efficient discrimination of the three bacteria. In changing in vitro conditions and specifically in presence of the various stressors, the untargeted methodology proved to be more valuable for the global and accurate differentiation of the three bacteria, also considering the type of stress environment within each species. In addition, species-specific metabolites (i.e., fatty acids, polysaccharides, peptides, and nucleotide bases derivatives) were putatively identified. Good intra-day repeatability and inter-day repeatability (< 10% RSD and < 15% RSD, respectively) were obtained for the targeted and the untargeted methods. This untargeted approach highlights its importance in unusual (and less known) bacterial growth environments, being a powerful tool for infectious disease diagnosis, where the accurate classification of microorganisms is sought.

Keywords Liquid chromatography · Metabolomics · Multivariate analysis · Quadrupole time-of-flight mass spectrometer · Triple quadrupole mass spectrometer

Introduction

The identification of microorganisms is of importance in numerous clinical, biological, and food applications, such as chronic infectious disease diagnosis, antimicrobial resistance, and quality control in food industry [1]. The accurate diagnosis of infectious diseases, from human or foodborne pathogens, would enable to decrease their incidence and improve human health. To survive in constantly changing environments, pathogenic microorganisms have developed adaptative defense mechanisms, known as stress responses, that permit their survival and play an important role in their evolution. Responses to environmental stress, such as pH changes, salinity, nutrient availability, or oxidation, are directly related to changes in several cellular metabolic pathways, which could hamper the unequivocal identification of microorganisms [2, 3]. Therefore, the development of methods enabling the discrimination between various bacterial species, regardless of the growth environment, is necessary and, in addition, the understanding of stress responses would provide insights into bacterial pathogenesis.

✉ Kevin A. Schug
kschug@uta.edu

[1] Molecular System, Organic & Biological Analytical Chemistry Group, University of Liege, 11 Allee du Six Aout, 4000 Liege, Belgium

[2] Department of Chemistry and Biochemistry, The University of Texas at Arlington, 700 Planetarium Place, Box 19065, Arlington, TX 76019, USA

[3] Affiliate of Collaborative Laboratories for Environmental Analysis and Remediation, The University of Texas at Arlington, Arlington, TX 76019, USA

Published online: 12 July 2021

Metabolomics has been considered as a promising approach for the characterization of metabolites in biological samples [4]. Metabolomics approach enables the comprehensive profiling and identification of cellular metabolites that are able to reflect ongoing biological processes, and/or altered states compared to normal conditions. Mass spectrometry (MS)–based metabolomics hyphenated with gas chromatography (GC-MS) or liquid chromatography (LC-MS) has been successfully used for the fingerprinting of various bacteria [5–7]. However, only few studies focused on the characterization of bacterial species under stress environments [8, 9].

LC enables the detection of a large set of metabolites, characterized by various physicochemical properties, without the need for extensive sample preparation. LC coupled to triple quadrupole MS in multiple reaction monitoring (MRM) mode enables sensitive and specific quantitative analysis of targeted metabolites with high reproducibility [10]. Recently, the potential of such LC-MS/MS approach has been demonstrated for the simultaneous targeted analysis of 113 primary metabolites [7]. Since primary metabolites, such as amino acids, organic acids, nucleotides, and nucleosides, are essential for energy production, normal growth, and development of living organisms, their analysis could provide valuable insights into the metabolism of numerous bacterial species. However, the use of a targeted methodology limits metabolite coverage. On the other hand, LC hyphenated with high-resolution (HR) MS enables the untargeted analysis of metabolites present in biological systems; therefore, it can increase metabolite detection with no need of a priori knowledge [10, 11]. The accurate mass measurement, together with MS/MS spectral acquisition, also enables the investigation of unknown metabolites. Nevertheless, untargeted metabolomics often requires multivariate statistical analysis to reduce data complexity, and identification of metabolites remains a challenge [12].

In this work, the advantages of targeted and untargeted analysis, respectively exploiting high-performance liquid chromatography (HPLC) triple quadrupole (QqQ) MS and ultra-high-performance liquid chromatography (UHPLC) quadrupole time-of-flight (QTof) MS platforms, were evaluated and combined to investigate the metabolism of three bacterial species (i.e., *Staphylococcus aureus*, *Bacillus cereus*, and *Pseudomonas aeruginosa*). *S. aureus* and *P. aeruginosa* are common pathogens that cause chronic lung infections. *P. aeruginosa* is the most prevalent pathogen associated with cystic fibrosis [3]. *B. cereus* bacteria are mainly responsible for food poisoning, but they can also cause various systemic infections, such as lung infections, pneumonia, and septicemia [13].

The three bacterial species were grown under various stress conditions (i.e., acidic and basic pH, oxidative, ethanolic, and saline conditions) to determine their effect on extracellular metabolites present in the culture medium. Univariate and multivariate statistical analyses were used to reduce data complexity and identify species-specific metabolites. The performances of the two MS detectors for the bacteria discrimination, and the complementarity of the approaches were highlighted.

Experimental section

Chemical reagents and standards

A standard solution (1 ppm) used to evaluate the HPLC-QqQ MS method performance was composed of the following standards: 2-aminoadipic acid, 2-ketoisovaleric acid, 4-hydroxyphenyllactic acid, adenine, adenosine, adenosine monophosphate, ascorbic acid, biotin, citrulline, cystathionine, cytidine, cytidine monophosphate, ethylenediamine, fumaric acid, glucosamine, glycyl-glutamine, guanine, guanosine, guanosine monophosphate, histamine, inosine, isocitric acid, malic acid, *n*-acetylaspartic acid, nicotinic acid, *o*-phosphoethanolamine, pipecolic acid, putrescine, pyridoxal, succinic acid, thymine, uracil, uric acid, uridine, xanthine, and xanthosine. A standard solution (1 ppm) used to evaluate the UHPLC-QToF MS method performance was composed of the following standards: 2-ketoisobvaleric acid, 4-hydroxyphenyllactic acid, ascorbic acid, fumaric acid, guanine, isocitric acid, malic acid, succinic acid, uric acid, and xanthine. The internal standard used was 2-isopropylmalic acid. The standards were purchased from Alfa Aesar (Ward Hill, MA), Merck (Darmstadt, Germany), Sigma-Aldrich (St. Louis, MO), Supelco (Bellefonte, PA), Tci America (Portland, OR), and Thermo Fisher Scientific (Fair Lawn, NJ) (see Table S1 in the Supplementary Information (ESM)). As a mobile phase additive, formic acid was used (Sigma-Aldrich). LC-MS grade water and acetonitrile were purchased from Honeywell Burdick & Jackson (Muskegon, MI).

Bacteria culture condition and stress assay

Three bacterial species were used in this study: *Pseudomonas aeruginosa* (ATCC BAA-26), *Bacillus cereus*, and *Staphylococcus aureus*. *B. cereus* and *S. aureus* were isolates obtained from groundwater and identified using matrix-assisted laser desorption ionization time-of-flight mass spectrometry and RNA sequencing [14]. All three species were pre-cultured aerobically overnight at 37 °C and 200 rpm in 125-mL conical flasks with 50 mL of nutrient broth (NB, Merck, Darmstadt, Germany) and then inoculated (1:1000) under identical conditions into 50 mL of fresh NB. At the log phase of the growth, the bacteria were transferred to 125-mL conical flasks containing either 50 mL of acidified NB (HCl, pH = 2), 50 mL of basified NB (NaOH, pH = 9), or NB supplemented with diluted H_2O_2 (1 mM), 4% *v/v* of

ethanol, or 4% *v/v* of salt. As a positive control, bacteria were grown in NB without stressors. Finally, fresh NB and NB supplemented with the different stressors were used as negative controls. Flasks were incubated for 12 h at 37 °C and 200 rpm corresponding to the stationary phase of growth. Optical density measurements were performed at 600 nm using a BioSpec-mini™ spectrophotometer (Shimadzu Scientific Instruments, Inc., Columbia, MD).

Sample preparation

At the stationary phase of growth, flasks were submerged on ice to quench cellular metabolism and centrifuged at 15,000 rpm for 10 min. Then, 100 μL of supernatant was transferred under sterile conditions in a 1.5-mL Eppendorf tube containing 200 μL of acetonitrile. The sample was spiked with the internal standard to a final concentration of 5 ppm. Following the centrifugation at 15,000 rpm for 10 min, 100 μL of supernatant was transferred to an LC vial and diluted (1:10) with ultrapure water prior to analysis. Six biological replicates were prepared and analyzed for each stress condition. For each biological replicate, two technical replicates were prepared for the analysis by HPLC-QqQ MS and by UHPLC-QToF MS.

Targeted methodology (HPLC-QqQ MS)

A Shimadzu LCMS-8050 triple quadrupole instrument (Shimadzu Scientific Instruments) equipped with LC-20AD solvent delivery pumps, DGU-20A5 degassing unit, SIL-20AC XR autosampler, CTO-20AC column oven, and CBM-20A system controller was used for the targeted analysis. Instrument control and data acquisition were performed using LabSolutions software v.5.97 (Shimadzu Corp., Tokyo, Japan). Injections of 1 μL were used. Separation was achieved in reversed phase mode using a Discovery HS F5-3 column (150 mm × 2.1 mm × 3 μm) (Shimadzu), made of a pentafluorophenylpropyl (PFPP) stationary phase. This column provides unique retention characterized by various interactions such as electrostatic dipole-dipole interactions, pi-pi interactions, and hydrophobic forces. The potential of such approach has been reported for simultaneous analysis of a broad range of metabolites [7].

Mobile phases consisted of water with 0.1% formic acid (solvent A) and acetonitrile with 0.1% formic acid (solvent B). Gradient elution was performed at a flow rate of 0.35 mL min^{-1}, as follows: 0–1.4 min: 0% B; 1.4–3.5 min: 25% B; 3.5–7.5 min: 35% B; 7.5–10.30 min: 95% B; 10.30–13.70 min: 95% B; 13.8–17 min: 0% B. The temperature of the autosampler tray and the temperature of the oven were set at 5 °C and 40 °C, respectively. Electrospray ionization (ESI) was performed in the negative and positive ionization mode. The MS data was collected under the following ESI conditions: nitrogen nebulizing gas and drying gas flows were 3 L min^{-1} and 10 L min^{-1}, respectively; the desolvation line temperature was 250 °C and the heat block temperature was 400 °C; the positive and negative interface voltages were 4.5 kV and −3.5 kV, respectively. Each analyte of interest was monitored by optimized multiple reaction monitoring (MRM). Representative precursor ions, product ions, collision energies, and retention times for each compound are listed in Table S1 of the ESM.

The acquired data were processed using LabSolutions software and the obtained peak list containing the retention times and relative peak areas of the targeted metabolites in each sample was used for further statistical analysis.

Untargeted methodology (UHPLC-QToF MS)

A Shimadzu Nexera X2 ultra-high-performance liquid chromatograph coupled to a LCMS-9030 quadrupole time-of-flight (QToF) mass spectrometer (Shimadzu Scientific Instruments, Inc., Columbia, MD) equipped with LC-30AD solvent delivery pumps, DGU-20A5R degassing unit, SIL-30AC autosampler, CTO-20AC column oven, and CBM-20A system controller was used for the untargeted analysis. Instrument control and data acquisition were performed using LabSolutions software v.5.97 (Shimadzu Corp., Tokyo, Japan). Injections of 5 μL were used. Separation was achieved using a Discovery HS F5-3 column (150 mm × 2.1 mm × 3 μm) (Shimadzu) using identical mobile phases and gradient elution as for the analysis using the LCMS-8050 (see the "Targeted methodology (HPLC-QqQ MS)" section). The temperature of the autosampler tray and the temperature of the oven were set at 5 °C and 40 °C, respectively. Electrospray ionization (ESI) was performed in the negative ionization mode using data-independent acquisition (DIA) mode, with a mass-to-charge (*m/z*) range set to 50–1000 Da, and a collision energy of 25 eV (with 17-eV CE spread). The MS data was collected under the following ESI conditions: Nitrogen nebulizing gas and drying gas flows were 2 L min^{-1} and 10 L min^{-1}, respectively; the desolvation line temperature was 250 °C and the heat block temperature was 400 °C; the interface voltage was −3.5 kV.

The acquired data were exported from LabSolutions as mzML files and imported into MS Dial (v. 4.00, Yokohama City, Japan). Data pre-processing (noise setting, baseline correction, peak picking and alignment) was performed using MS Dial. The following parameter settings were used: minimum peak height, 100 counts; mass width, 0.1 Da; mass tolerance (MS1), 0.02 Da. Results were further exported as a .txt file from MS Dial for further statistical analysis.

Data processing and statistical analysis

Prior to statistical analysis, the data were normalized using the internal standard signal and auto-scaled (mean-centered and divided by the standard deviation of each variable). No further data transformation was performed.

Two-tailed equal variance t-test, one-way analysis of variance (ANOVA), and fold-change were used to determine significant features characterized by a p value < 0.05 and a fold-change > 1.5. The p values obtained from the one-way ANOVA were corrected for multiple testing using false discovery rate adjustment. For the untargeted analysis, multivariate statistical approaches (i.e., partial least square discriminant analysis (PLS-DA) and random forest (RF)) were used to highlight the most influential metabolites, in addition to t-test and fold-change. Thresholds of 1.5 for variable importance in projection (VIP) scores and 0.001 for mean decrease accuracy (MDA) were defined. Metabolites characterized by a p value < 0.05, a fold-change > 1.5, VIP scores > 1.5, and MDA > 0.01 were selected as the most influential metabolites.

RStudio (v.3.3.2) and MetaboAnalyst 5.0 online (Quebec, CA) were used to perform statistical analysis. The R packages tidyverse, pheatmap, and caret were used to generate principal component analysis (PCA), hierarchical cluster analysis (HCA), and random forest (RF). MetaboAnalyst was used to perform partial least square discriminant analysis (PLS-DA).

MS Finder (v.3.20, Yokohama City, Japan) was used to predict the formula of the most influential metabolites as well as to predict chemical ontologies. Molecular formulas were predicted from the precursor ion using the accurate mass, isotope ratio, and product ion information. Atoms included for the molecular formula search were C, H, O, N, P, and S. In addition, a 20% isotopic tolerance was defined. The experimental MS/MS spectra were then compared to theoretical fragments calculated on known metabolites retrieved from structure databases. Searched databases included the Chemical Entities of Biological Interest (ChEBI), the Human Metabolome Database (HMDB), and the Metabolic In Silico Network Expansion (MINE) databases.

Results and discussion

Targeted analysis of primary metabolites to discriminate bacterial species

The profile of the three bacteria, resulting from the analysis of the targeted 96 primary metabolites using HPLC-QqQ MS, is displayed in the overlaid multiple reaction monitoring (MRM) chromatograms in Fig. 1. The analysis resulted in the detection of 78 metabolites, 86 metabolites, and 71 metabolites in *P. aeruginosa*, *S. aureus*, and *B. cereus*, respectively. Considering the six biological replicates analyzed for each species, a satisfactory average repeatability (21 %RSD for *S. aureus*, 13 %RSD for *B. cereus*, and 18 %RSD for *P. aeruginosa*) was obtained for the targeted metabolites detected.

Principal component analysis (PCA) and hierarchical cluster analysis (HCA) were used to evaluate the potential of the primary metabolites to distinguish between the bacterial species. To ensure that the observed metabolic changes were specific to the bacterial species, bacteria-free nutrient broth was considered as negative control (see the "Bacteria culture condition and stress assay" section). The PCA of Fig. 2a displays the discrimination between the negative control and the three bacterial species, as well as the clear clustering between the species, using the two first principal components (PCs) expressing 61.2% of the total variance within the dataset. The first PC, accounting for 39.5% of the variance, enabled the discrimination of the bacteria from the nutrient broth (negative control), while the second PC, accounting for 21.7% of the variance, enabled the discrimination of *S. aureus* from *B. cereus* and *P. aeruginosa*.

Discrimination between the three species was expected since they present distinct cell wall structures and shapes. Indeed, *P. aeruginosa* is a Gram-negative rod-shaped species, while *S. aureus* and *B. cereus* are Gram-positive round- and rod-shaped species, respectively. In addition, *B. cereus* has the potential to produce endospores [15]. Nevertheless, the clustering of the bacterial species observed in the PCA of Fig. 2a demonstrated the potential of the selected primary metabolites (listed in ESM Table S1) to reflect the unique metabolism of bacterial species.

The heatmap of Fig. 2b shows the qualitative distribution of the metabolites among the samples. The consumption of 11 nutrients from the growth media by the bacteria can be observed (Fig. 2b, top left). These nutrients were mainly amino acids (alanine, arginine, asparagine, aspartic acid, glutamic acid, glycine, serine, and threonine), together with adenosine, cystathionine, and cysteine, and their consumption suggest the importance of these metabolites for normal in vitro bacterial growth. Previous studies highlighted the importance of these amino acids for optimal yeast, bacterial, and mammalian cell growth [16]. In addition, the analysis of *S. aureus*, *B. cereus*, and *P. aeruginosa* revealed metabolic differences and higher amounts of some metabolites for each species were identified. Regarding *P. aeruginosa*, adenosine monophosphate, pantothenic acid, and ornithine were specifically more abundant than in the other species. For *S. aureus*, 4-hydroxyproline, guanine, histamine, and hypoxanthine were specifically abundant. *B. cereus* was characterized by a low amount of all targeted metabolites compared to the two other bacteria. Indeed, among the

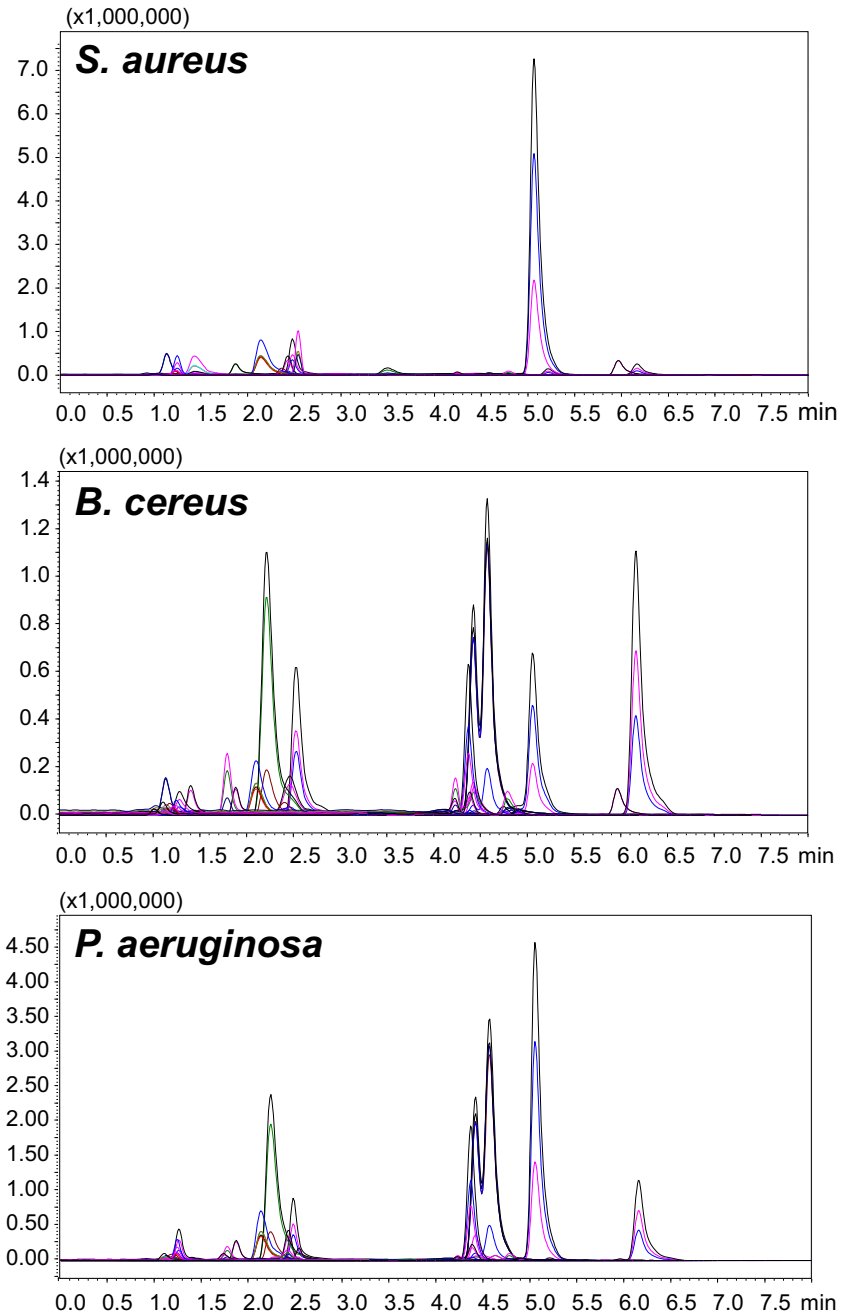

Fig. 1 Multiple reaction monitoring (MRM) chromatograms of *S. aureus*, *B. cereus*, and *P. aeruginosa* grown under normal in vitro conditions. The black trace is the TIC; the colored traces represent the different MRM transitions for each target analyte. For the details of the MRMs, refer to Table S1 in the ESM

96 targeted metabolites, uric acid was the only metabolite detected in higher amounts in *B. cereus*.

Ultimately, among the 96 targeted primary metabolites, 45 metabolites were statistically significant, i.e., characterized by a p value < 0.05 using one-way ANOVA, to distinguish the bacterial species. These metabolites were not initially present in the bacteria-free nutrient broth. The discrimination between the bacterial species was further improved when considering only the significant metabolites, as can be visualized on the PCA of Fig. S1 (see ESM), where the two first PCs account for 79.8% of the variance.

Effect of stress conditions on primary metabolite profiles

Adaptative response to stress environments, including changes in pH, salinity, nutrient scarcity, and oxidation, allows pathogenic bacteria to resist host defense and plays an important role in their evolution [2]. It is important to evaluate the potential of primary metabolites to discriminate between bacterial strains even under stress conditions. Therefore, the targeted LC-MS/MS method was evaluated to discriminate *S. aureus*, *B. cereus*, and *P. aeruginosa* under five different

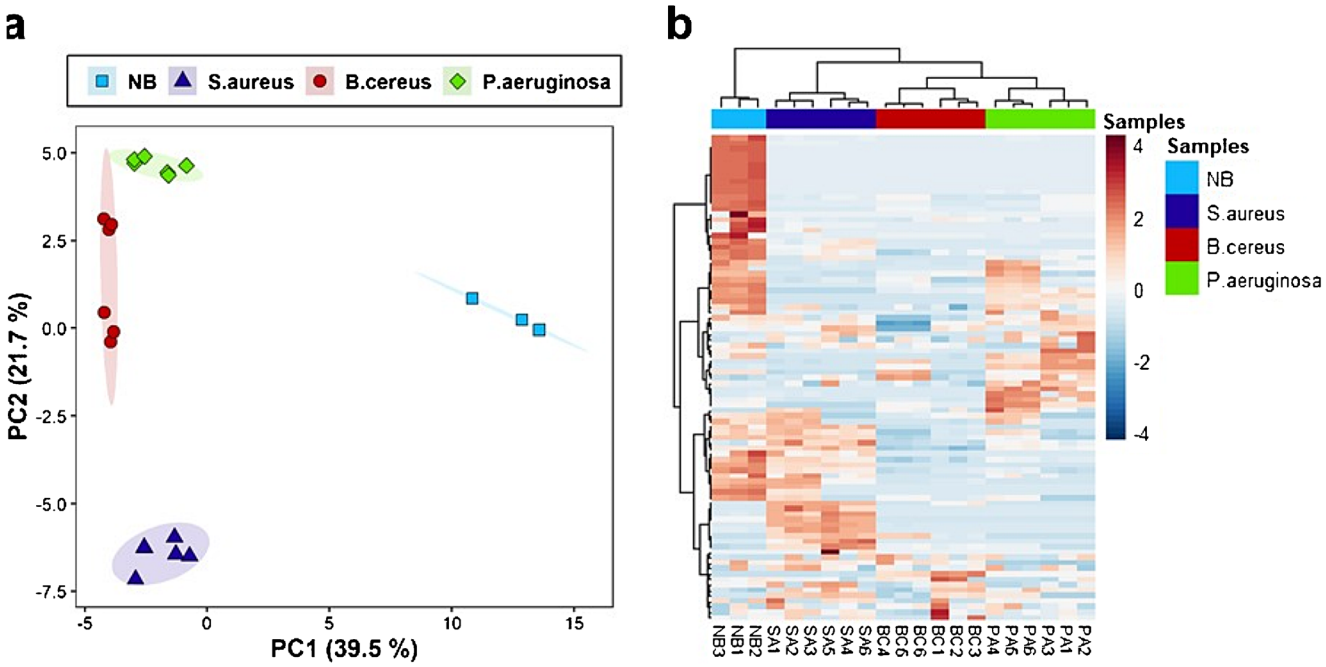

Fig. 2 **a** Principal component analysis of the 3 bacterial species grown under normal in vitro conditions and the nutrient broth (negative control) evaluated by HPLC-QqQ MS using the 96 targeted metabolites, and **b** heatmap generated using hierarchical cluster analysis of the samples and the targeted metabolites

stress conditions: acidic, basic, oxidative, saline, and ethanolic environments.

Bacteria growth rates were altered according to the type of stress. Values of optical density measurements at 600 nm (OD_{600}) are reported in Table 1. Acidic and oxidative stress decreased the growth rate of *S. aureus* and *B. cereus*, while the growth rate of *P. aeruginosa* was decreased by acidic and ethanolic stress. Indeed, the measurements of OD_{600} resulted in values lower than 0.5, corresponding to the early logarithmic phase of growth, while the measurement of OD_{600} of the non-stressed bacterial culture resulted in an average value of 1.63, corresponding to the stationary phase of growth.

Table 1 Values of optical density measurements at 600 nm of *P. aeruginosa*, *S. aureus*, and *B. cereus* grown in normal (positive control) and in five different stress conditions

	S. aureus	*B. cereus*	*P. aeruginosa*
Stress types	Optical density (600 nm)		
None (positive control)	1.73	1.77	1.39
Oxidative	0.10	0.11	1.09
Acidic	0.39	0.27	0.14
Basic	1.79[a]	1.64[a]	1.06
Ethanolic	1.27	1.30	0.24
Saline	1.81[a]	2.05[a]	1.89

[a] OD values not statistically different from the positive control

Metabolic changes associated with the stress responses hampered the discrimination of the three bacterial species using the targeted primary metabolites, as can be seen on the PCA of Fig. 3. Only the acidic and oxidative stress for *S. aureus* and *B. cereus*, and the acidic and ethanolic stresses for *P. aeruginosa* could be clearly distinguished from the

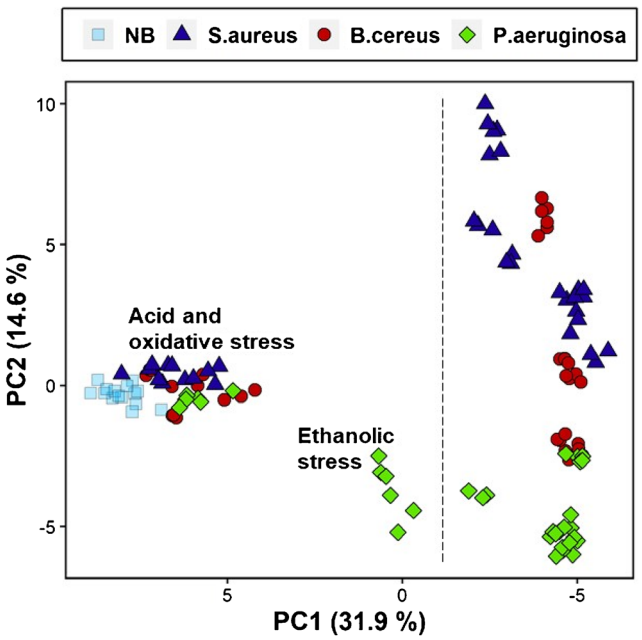

Fig. 3 Principal component analysis of the three bacterial samples grown in normal in vitro conditions and in five different stress conditions using the 96 targeted primary metabolites (HPLC-QqQ MS) (NB, nutrient broth used as negative control)

other conditions (Fig. 3, dotted line). This clustering most likely corresponds to the altered growth rate of the bacteria strains, as observed by the OD_{600} measurements. It has been shown that in acidic environments, a decrease in intracellular pH is observed which alters the structure of the bacterial membrane, decreases the activity of several pH-dependent enzymes, and therefore alters bacterial growth and the metabolites [15]. In addition, when plotting in the PCA the negative control samples (bacteria-free nutrient broth with and without the stressors), they cluster closely with the acidic and oxidative conditions, as can be seen in Fig. 3. The nutrient broth samples (negative controls) and the bacteria grown under acidic and oxidative conditions were both characterized by a higher amount of alanine, arginine, asparagine, aspartic acid, glutamic acid, glycine, serine, and threonine. These amino acids, as previously mentioned, are present specifically in the nutrient broth and further consumed by the three bacterial species for their normal growth. Therefore, these metabolites are not bacterial stress response markers of acidic and/or oxidative conditions but rather, their presence represents altered bacterial growth rate.

The targeted analysis of the selected primary metabolites was very valuable for identification of and discrimination between the growth metabolism of the different bacterial species, and as a reflection of variations/alterations in the normal bacterial growth in vitro. However, the distinction between bacterial species using the primary metabolites was more complicated in the simulated situations in which a stress response was induced.

Untargeted analysis of bacterial metabolites under stress conditions

Untargeted MS-based metabolomics enable a more comprehensive analysis of the metabolome and therefore is a more suitable tool/technique to identify bacterial species regardless of the growth environments. The analysis of *S. aureus*, *B. cereus*, and *P. aeruginosa* stress responses resulted in the detection of a total of 24,000 features using UHPLC-QToF MS (see data processing information in the "Untargeted methodology (UHPLC-QToF MS)" section). To elucidate bacterial stress responses, while identifying potential markers of the different bacterial species, data reduction was performed. Features identified in the blanks (mobile phase and extraction solvents) together with features characterized by RSD > 30% were removed from the dataset, reducing the number of features to 1832.

PCA and HCA were performed to evaluate the discrimination between the three bacterial species grown in normal in vitro conditions. Similar clustering as for the 96 primary target metabolites (see the "Targeted analysis of primary metabolites to discriminate bacterial species" section) was observed when using these 1832 features, and biological replicates clustered within the 95% confidence interval (see ESM Fig. S2).

To further extract meaningful information, univariate and multivariate statistical approaches were applied, as displayed in Fig. 4. Bacteria were investigated in a "one versus all" classification method (i.e., *S. aureus* vs others, *P. aeruginosa* vs others, and *B. cereus* vs others using volcano plot (i.e., t-test and fold-change), partial least square discriminant analysis (PLS-DA), and random forest (RF)). The features shared by the three statistical approaches were further investigated manually to ensure a Gaussian peak profile. Combining univariate (volcano plot) and multivariate (PLS-DA and RF) statistical approaches aimed at eliminating potential bias in the feature selection process [17]. These data reduction and feature selection approaches led to a final dataset of 129 features, characterized by unique retention times and accurate masses.

The combination of these 129 features enabled the complete separation of the three bacterial species, as can be visualized on the PCA score plot of Fig. 5a. Clear clustering of the three species, regardless of the stress environment, can be observed using only the two first PCs accounting for 88.8% of the total variance within the dataset, indicating clear differences between the bacterial species. To further illustrate metabolic differences between the three bacteria, the 108 samples were subjected to HCA. As can be observed in the resulting heatmap of Fig. 5a (on the right), the metabolites were clustered in three main groups. The first cluster (i) includes 66 metabolites among which 46 are more abundant in *P. aeruginosa*. The second cluster (ii) contains 22 metabolites that are more abundant in *B. cereus*. The third cluster (iii) counts 41 metabolites among which 32 are more abundant in *S. aureus*. Although the three microorganisms are clearly separated on the PCA, *P. aeruginosa* in ethanolic environment shares a more similar profile with *S. aureus*, dictating the misclassification of 2 replicates in the HCA (Fig. 5a).

In addition, as can be seen in Fig. 5b, for each bacterial species, most replicates were well clustered into their respective growth environments. This highlights the potential of the 129 features to identify the bacteria species together with their growth conditions. Groups further away from the positive control (the bacteria grown in normal in vitro environment) revealed significant alterations in the expression of the metabolites compared to the groups closer to the control group.

As observed with the analysis of the primary metabolites (Fig. 3), the acidic and oxidative environments created distinctive bacterial profiles. Better visible in Fig. 5b, the growth of *P. aeruginosa* in acidic and ethanolic environments yielded the most distinct profiles, separated on PC1 from the other stress conditions. Unlike *S. aureus* and *B. cereus*, *P. aeruginosa* exhibits an oxidative stress response and pigment production Regulator (ospR) gene that senses oxidative stress and regulates multiple pathways to enable its survival in

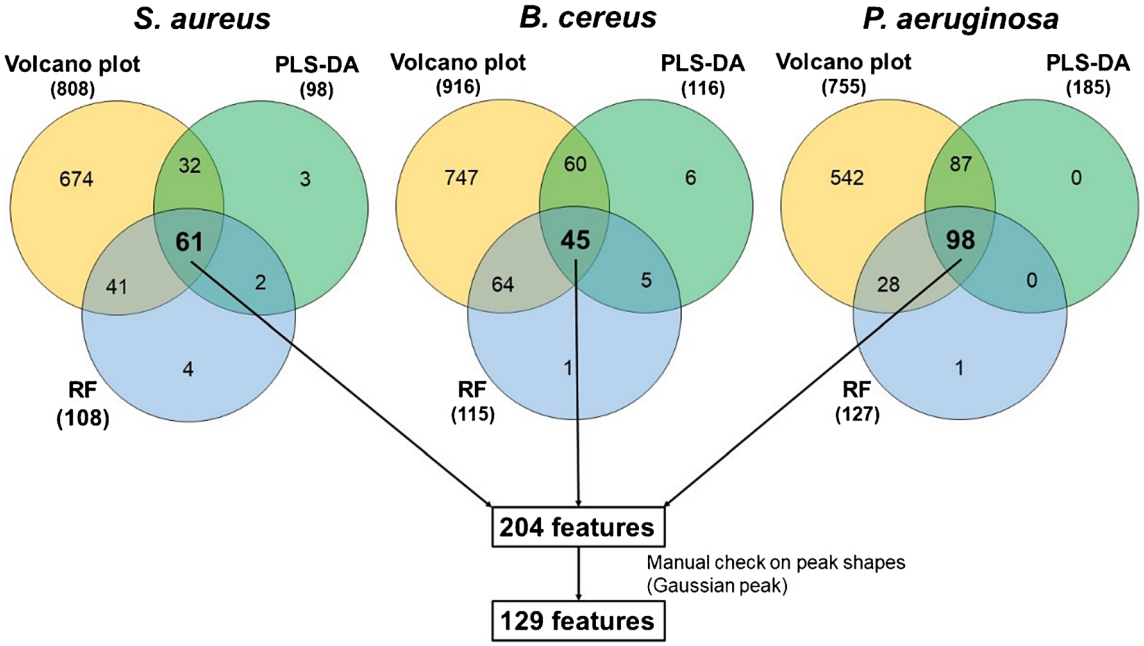

Fig. 4 Feature selection workflow, using volcano plot, PLS-DA, and RF, to identify influential features able to discriminate the bacterial species regardless of the growth environment

oxidative environments [18]. However, it has been previously shown that *P. aeruginosa*'s resistance to ethanol is lower than *S. aureus* and *B. cereus*. Indeed, reduced growth rate and motility have been observed when *P. aeruginosa* was grown in ethanolic environments, even at low concentration [19]. This explains the clustering of *P. aeruginosa* under oxidative stress with the control, while the ethanolic stress clusters closer to the acidic stress.

For the three bacterial species, the saline growth condition significantly impacted the metabolite profile and the saline stress was distinguished from the other conditions by the second PC. This could be explained by the fact that most of the bacterial species present a strong response to hyperosmotic salt conditions and can rapidly adapt to such environments. Recently, it has been shown on a broad range of species that salt stress alters hundreds of metabolites associated with the central carbon metabolism and heme biosynthesis [20].

Finally, the basic environment (pH=9) yielded a similar profile as the positive controls (i.e., the bacteria grown without stressor), indicating rapid adaptations to higher alkalinity. Recent studies reported normal growth rate of *S. aureus* and *P. aeruginosa* until pH 9, while the optimal growth of *B. cereus* was observed at pH 9 [21].

Tentative identification of bacterial species-specific metabolites

From the untargeted analysis of bacterial metabolites, species-specific features were detected. Such features present in one bacteria type, regardless of the environment, and absent in others are of interest as potential bacterial markers, which could be used to diagnose their presence/occurrence in infections. For *S. aureus*, 5 specific features were detected, while for *B. cereus* and *P. aeruginosa*, 8 species-specific features were detected.

Although the use of data-independent acquisition (DIA) enabled the generation of MS/MS spectra for almost all metabolic features, only 5% of MS/MS spectra are available in databases, which makes the identification of metabolites in untargeted analysis challenging [11]. Since no reference mass spectra were available for the species-specific features highlighted, molecular formulas were predicted from the precursor ion using the accurate mass, isotope ratio, and product ion information. Molecular formulae presenting the highest score and the lowest mass error (< 2 mDa) were selected and are reported in Table 2. In addition, theoretical fragmentation was generated for all predicted molecular formulas, supporting the structure elucidation process, and therefore enabling the structure-based classification of the species-specific features. Practically, each experimental MS/MS spectrum was compared to theoretical fragments calculated on known compounds retrieved from structure databases. This approach is often used for the annotation of metabolites with unknown MS/MS spectra, i.e., that are not available in mass-spectral databases and enables the interpretation and identification of biochemically relevant pathways [12, 22].

The production of fatty acids (FAs) by *B. cereus* could be explained by their major role in growth, adaptation, and survival. Unlike other Gram-positive bacteria, *Bacillus* species are known to produce mainly branched-chain FAs, increasing fluidity, and to display unusual FAs (cyclic or hydroxy), characterized by antimicrobial properties [23]. In addition, it has been shown that, in response to environmental changes, the

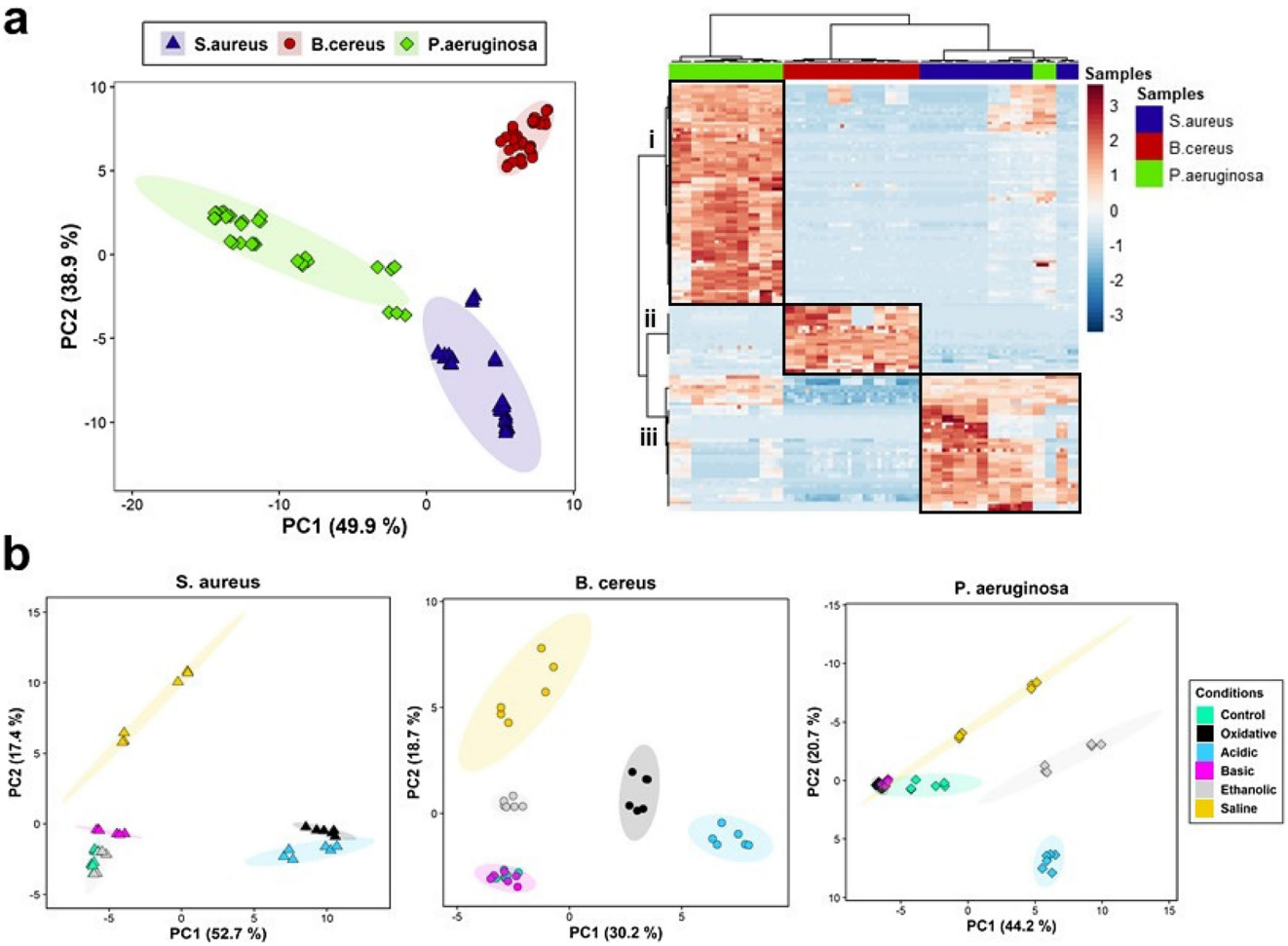

Fig. 5 **a** Principal component analysis of the 3 bacterial species grown under normal in vitro conditions and in five different stress conditions evaluated by UHPLC-QToF MS using 129 features (left) and heatmap generated using hierarchical cluster analysis of the samples and the 129 features, from which 3 clusters were highlighted. **b** Principal component analysis of *S. aureus* (left), *B. cereus* (middle), and *P. aeruginosa* (right) grown under different environments: control (spring green), oxidative (black), acidic (light blue), basic (pink), ethanolic (gray), and saline (yellow)

FA composition of *Bacillus* species is regulated to ensure survival by maintaining cell membrane homeostasis. Regarding *P. aeruginosa*, this species has the ability to form biofilms, which requires polysaccharides, proteins, and nucleic acids [24, 25]. Therefore, the identification of polysaccharides and peptides that are specific to *P. aeruginosa* might be explained by its capacity to form biofilms. Finally, it has been highlighted that under stress conditions, the purine metabolism of *S. aureus* is altered. Specifically, under osmotic stress, an increase of purine and glutamine has been highlighted, which might reflect survival mechanisms [26].

Method performances

The intra-day repeatability of the targeted HPLC-QqQ MS method, in both ESI+ and ESI− mode, was evaluated using the area response of the 1 ppm standard solution containing 36 standards, while the intra-day repeatability of the untargeted UHPLC-QToF MS method in ESI− mode was evaluated using the area response of the 1 ppm standard solution containing 10 standards. The intra-day repeatability was evaluated on the analysis of four replicates of the standard solution in water and in the nutrient broth. The standard solution spiked in water was used to assess initially the repeatability of the targeted and untargeted methodologies prior to their use on the bacterial cultures. In addition, the same standard solution was spiked in the growth media (nutrient broth) and it was used to monitor the repeatability during the analysis of the three bacterial cultures.

Good intra-day repeatability was obtained for both the targeted and untargeted methods. The targeted method gave average RSD of 7.2% and 7.9% from the analysis of the standard solution, in water and in nutrient broth, respectively. Concerning the untargeted method, average RSD of 5.6% and 5.0% were obtained from the analysis of the standard solution, in water and in nutrient broth, respectively. The

Table 2 Tentative identification of bacterial specific features detected by UHPLC-QToF MS

Bacterial species	RT (min)	Accurate mass	Assigned adduct	Molecular formula	Error (mDa)	Chemical ontology/structural classification
B. cereus	3.52	227.1146	$[M-H]^-$	$C_9H_{16}N_4O_3$	0.4	Arginine and derivatives
	5.36	313.0738	$[M-H]^-$	$C_{14}H_{18}O_6S$	1.3	Benzopyrans
	5.45	215.0923	$[M-H]^-$	$C_{10}H_{16}O_5$	0.2	Fatty acid esters
	5.65	173.0811	$[M-H]^-$	$C_8H_{14}O_4$	0.8	Fatty acid esters
	6.18	252.1459	$[M-H]^-$	$C_{11}H_{19}N_5O_2$	0.7	Amino acids and derivatives
	7.39	201.1113	$[M-H]^-$	$C_{10}H_{18}O_4$	1.9	Fatty acid esters
	8.02	279.9172	$[M-H]^-$	$C_5H_3N_3O_5S_3$	−1.0	Unknown
	10.63	387.2864	$[M-H]^-$	$C_{20}H_{40}N_2O_5$	0.0	Amino acids and derivatives
P. aeruginosa	5.34	437.1664	$[M-H]^-$	$C_{23}H_{26}N_4O_3S$	−1.1	Amino acids and derivatives
	5.93	279.0980	$[M-H]^-$	$C_{13}H_{16}N_2O_5$	0.6	Dipeptides
	6.12	548.2118	$[M-H]^-$	$C_{27}H_{35}NO_{11}$	1.9	Polysaccharides
	6.15	412.2199	$[M-H]^-$	$C_{18}H_{31}N_5O_6$	0.3	Peptides
	6.21	270.1455	$[M-H]^-$	$C_{12}H_{21}N_3O_4$	0.4	Alpha amino acid esters
	8.10	490.1933	$[M-H]^-$	$C_{26}H_{29}N_5O_3S$	−1.5	Phenylpyridines
	8.70	537.3045	$[M-H]^-$	$C_{25}H_{42}N_6O_7$	−0.3	Dipeptides
	9.31	470.2594	$[M-H]^-$	$C_{25}H_{37}N_5O_2S$	0.1	Phenylpyrazoles
S. aureus	2.12	159.0762	$[M-H]^-$	$C_6H_{12}N_2O_3$	1.3	Glutamine and derivatives
	3.08	133.0148	$[M-H]^-$	$C_5H_2N_4O$	0.8	Purine and derivatives
	3.12	150.0409	$[M-H]^-$	$C_5H_5N_5O$	1.2	Purine and derivatives
	3.16	286.0168	$[M-H]^-$	$C_{14}H_9NO_4S$	1.2	Benzothiazoles
	5.25	547.1398	$[M-H]^-$	$C_{33}H_{24}O_8$	0.0	Arylbenzofuran derivatives

obtained average areas and %RSD for all the single standards used in the mix are listed in the ESM Table S2 (for the HPLC-QqQ MS method) and Table S3 (for the UHPLC-QToF MS method). The higher response of the spiked standards in the nutrient broth compared to water is due to their presence in the original composition of the growth media. The investigation of a possible contribution of the matrix was not within the scope of the present study and can be assessed in a future validation study. In any case, the presence of a matrix effect would not affect the observations and conclusions of the present study.

The inter-day repeatability was assessed using the area response of the internal standard (2-isopropylmalic acid). It was evaluated in four consecutive days using the 126 analyses consisting of the three bacteria types grown under six environments and analyzed in six replicates and the six nutrient broth samples analyzed in triplicates. Similar inter-day repeatability for the internal standard was obtained in the targeted and untargeted methods with an average RSD of 12% for the targeted method, and of 13% for the untargeted method.

Conclusion

The combination of targeted and untargeted metabolomics proved to be highly valuable for the differentiation of bacterial species under normal and under stress conditions. The targeted analysis of primary metabolites, using HPLC-QqQ MS, was efficient to rapidly identify and discriminate between the *P. aeruginosa*, *S. aureus*, and *B. cereus* microorganisms under normal in vitro growth conditions. However, in altered environments, the distinction between the bacterial species, based on the initial target primary metabolites using HPLC-QqQ MS, was no longer achievable.

In various stress environments (acidic, basic, oxidative, saline, ethanolic), the untargeted UHPLC-QToF MS method combined with univariate (t-test, fold-change) and multivariate statistical analyses (PCA, PLS-DA, HCA, RF) proved to be highly valuable for the global and accurate differentiation of the bacterial species together with the various stress responses within each species. With this approach, 21 species-specific metabolites were highlighted (Table 2), regardless of the growth environment. The use of high-resolution MS, with data-independent acquisition, enabled the prediction of molecular formulas and the structure-based classification of the metabolites, belonging to fatty acids, polysaccharides, peptides, and nucleotide base derivatives. Good intra-day repeatability and inter-day repeatability (<10% RSD and <15% RSD, respectively) were obtained for both the targeted and the untargeted methods.

Further studies would require the validation of the identification of the species-specific metabolites with standards, as

well as with a correlation of the metabolomic pathways leading to their production. The chemical validation of these metabolites would also enable the development of a targeted method, using MRM transitions, allowing a faster and easier profiling of microorganisms under the various growth environments. Such an approach could be then applied in numerous matrices, from medical to environmental, and food applications, where the accurate classification and differentiation of microorganisms is sought.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00216-021-03505-2.

Acknowledgements The authors would like to thank Shimadzu Scientific Instruments, Inc., for their instrumental support.

Funding The PhD fellowship of Delphine Zanella was funded by the Fund for Industry and Agricultural Research (FRIA). Partial monetary support for research was also provided by Restek Corporation. We also thank the University Foundation Leon Fredericq for the financial support.

Data availability All the data are described within the manuscript. The datasets generated and analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Conflict of interest The authors declare no competing interests.

References

1. AlMasoud N, Muhamadali H, Chisanga M, Alrabiah H, Lima CA, Goodacre R. Discrimination of bacteria using whole organism fingerprinting: the utility of modern physicochemical techniques for bacterial typing. Analyst. 2021;146(3):770–88.
2. Fang FC, Frawley ER, Tapscott T, Vazquez-Torres A. Bacterial stress responses during host infection. Cell Host Microbe. 2016;20(10):133–43.
3. Cullen L, McClean S. Bacterial adaptation during chronic respiratory infections. Pathogens. 2015;4(1):66–89.
4. Klassen A, Faccio AT, Canuto GAB, da Cruz PLR, Ribeiro HC, Tavares MFM, et al. Metabolomics: definitions and significance in systems biology. In: Sussulini A, editor. Metabolomics: from fundamentals to clinical applications. Cham: Springer International Publishing; 2017. p. 3–17.
5. Franchina FA, Purcaro G, Burklund A, Beccaria M, Hill JE. Evaluation of different adsorbent materials for the untargeted and targeted bacterial VOC analysis using GC×GC-MS. Anal Chim Acta. 2019;1066:146–53.
6. Davis TJ, Karanjia AV, Bhebhe CN, West SB, Richardson M, Bean HD. Pseudomonas aeruginosa volatilome characteristics and adaptations in chronic cystic fibrosis lung infections. mSphere. 2020;5(5):1–19.
7. Takenaka M, Yoshida T, Hori Y, Bamba T, Mochizuki M, Vavricka CJ, et al. An ion-pair free LC-MS/MS method for quantitative metabolite profiling of microbial bioproduction systems. Talanta. 2021;222(September 2020):121625.
8. Santos IC, Chaumette A, Smuts J, Hildenbrand ZL, Schug KA. Analysis of bacteria stress responses to contaminants derived from shale energy extraction. Environ Sci Process Impacts. 2019;21(2):269–78.
9. Favre L, Ortalo-Magné A, Greff S, Pérez T, Thomas OP, Martin JC, et al. Discrimination of four marine biofilm-forming bacteria by LC-MS metabolomics and influence of culture parameters. J Proteome Res. 2017;16(5):1962–75.
10. Madji Hounoum B, Blasco H, Emond P, Mavel S. Liquid chromatography-high-resolution mass spectrometry-based cell metabolomics: experimental design, recommendations, and applications. TrAC Trends Anal Chem. 2016;75:118–28.
11. Guo J, Huan T. Comparison of full-scan, data-dependent, and data-independent acquisition modes in liquid chromatography-mass spectrometry based untargeted metabolomics. Anal Chem. 2020;92(12):8072–80.
12. Tsugawa H, Kind T, Nakabayashi R, Yukihira D, Tanaka W, Cajka T, et al. Hydrogen rearrangement rules: computational MS/MS fragmentation and structure elucidation using MS-FINDER software. Anal Chem. 2016;88(16):7946–58.
13. Didelot X, Barker M, Falush D, Priest FG. Evolution of pathogenicity in the Bacillus cereus group. Syst Appl Microbiol. 2009;32(2):81–90.
14. Martin MS, Santos IC, Carlton DD, Stigler-Granados P, Hildenbrand ZL, Schug KA. Characterization of bacterial diversity in contaminated groundwater using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. Sci Total Environ. 2018;622–623:1562–71.
15. Willey J, Ka S, Wood D. Prescott's microbiology. 11th ed: McGraw-Hill Education; 2020.
16. Hanko VP, Rohrer JS. Determination of amino acids in cell culture and fermentation broth media using anion-exchange chromatography with integrated pulsed amperometric detection. Anal Biochem. 2004;324(1):29–38.
17. Saccenti E, Hoefsloot HCJ, Smilde AK, Westerhuis JA, Hendriks MMWB. Reflections on univariate and multivariate analysis of metabolomics data. Metabolomics. 2014;10(3):361–74.
18. Lan L, Murray TS, Kazmierczak BI, He C. Pseudomonas aeruginosa OspR is an oxidative stress sensing regulator that affects pigment production, antibiotic resistance and dissemination during infection. Mol Microbiol. 2010;75(1):76–91.
19. Lewis KA, Baker AE, Chen AI, Harty CE, Kuchma SL, Hogan DA. Ethanol decreases Pseudomonas aeruginosa flagellar motility through the regulation of flagellar stators. J Bacteriol. 2019;201(18):1–17.
20. Sévin DC, Stählin JN, Pollak GR, Kuehne A, Sauer U. Global metabolic responses to salt stress in fifteen species. PLoS One. 2016;11(2):1–21.
21. Okanlawon BM, Ogunbanwo ST, Okunlola AO. Growth of Bacillus cereus isolated from some traditional condiments under different regimens. Afr J Biotechnol. 2010;9(14):2129–35.
22. Blaženović I, Kind T, Ji J, Fiehn O. Software tools and approaches for compound identification of LC-MS/MS data in metabolomics. Metabolites. 2018;8(2):31. https://doi.org/10.3390/metabo8020031.
23. Diomandé SE, Nguyen-The C, Guinebretière MH, Broussolle V, Brillard J. Role of fatty acids in Bacillus environmental adaptation. Front Microbiol. 2015;6(AUG):1–20.

24. Ryder C, Byrd M, Wozniak DJ. Role of polysaccharides in Pseudomonas aeruginosa biofilm development. Curr Opin Microbiol. 2007;10(6):644–8.
25. Mielko KA, Jabłoński SJ, Milczewska J, Sands D, Łukaszewicz M, Młynarz P. Metabolomic studies of Pseudomonas aeruginosa. World J Microbiol Biotechnol. 2019;35(11):1–11.
26. Alreshidi MM. Selected metabolites profiling of Staphylococcus aureus following exposure to low temperature and elevated sodium chloride. Front Microbiol. 2020;11(May):1–13.