Title: Exploiting Differential Signal Filtering (DSF) and Image Structure Filtering (ISF) methods for untargeted Mass Spectrometry Imaging of bacterial metabolites.

Authors: Pierre Burguet[1], Raphaël La Rocca[1], Christopher Kune[1], Déborah Tellatin[2], Nudzejma Stulanovic[2], Augustin Rigolet[3], Johann Far[1], Marc Ongena[3], Sébastien Rigali[2], Loïc Quinton[1,*]

[1] Mass Spectrometry Laboratory, MolSys Research Unit, University of Liège, 4000 Liège, Belgium.

[2] InBioS – Center for Protein Engineering, University of Liège, 4000 Liège, Belgium.

[3] Gembloux Agro-Bio Tech, University of Liege, 5030 Gembloux, Belgium.

Abstract:

Matrix Assisted Laser Desorption/Ionization (MALDI) mass spectrometry imaging (MSI) is a label-free technique, producing images where pixels contain mass spectra. The technique allows the visualization of the spatial distribution of (bio)molecules from metabolites to proteins, on surfaces such as tissues sections or bacteria culture media. One particularly exciting example of MALDI-MSI use rest on its potential to localize ionized compounds produced during microbial interactions and chemical communication, offering a molecular snapshot of metabolomes at a given time. The huge size and the complexity of generated MSI data make the processing of the data challenging, which requires the use of computational methods. Despite recent advances, currently available commercial software relies mainly on statistical tools to identify patterns, similarities, and differences within datasets. However, grouping *m/z* values unique to a given dataset according to microbiological contexts, such as co-culture experiments, still requires tedious manual analysis. Here we propose a non-targeted method exploiting the differential signals between negative controls and tested experimental conditions, i.e. a differential signal filtering (DSF), and a scoring of the ion images using an image structure filtering (ISF) coupled with a fold change score between the controls and the conditions of interest. These methods were firstly applied to co-culture experiments involving *Escherichia coli* and *Streptomyces coelicolor*, revealing specific MS signals during the bacterial interaction. Two case studies were also investigated (i) the cellobiose-mediated induction for the pathogenicity of *Streptomyces scabiei*, the causative agent of common scab on root and tuber crops, and (ii) the iron-repressed production of siderophores of *S. scabiei*. This report proposes guidelines for

MALDI-MSI data treatment applied in the case of microbiology contexts, with enhanced ion peak annotation in specific culture conditions. The strengths and weaknesses of the methods are discussed.

Introduction:

Matrix Assisted Laser Desorption/Ionization mass spectrometry imaging (MALDI-MSI) is a label-free technique used to visualize the spatial distribution of molecules[1–3], such as metabolites[4], peptides[5], lipids[6], or proteins[7] on various surfaces. Information obtained from comprehensive studies based on mass spectrometry (MS) and the determination of the spatial distributions of molecules in samples offer significant knowledge about the molecular characterization of biological samples, ranging from tissues sections (organ biopsies, plant slices…) to biopolymer thin films[8]. Over the last decades, MSI methods have been increasingly explored in microbiology to produce images of metabolite synthesis and distribution by various interacting microorganisms, usually grown on thin-layer of agar based medium[2,9–11]. The molecular 'snapshot' of the metabolome of strains grown under controlled experimental conditions allows inferring molecular crosstalk and induced metabolites expression within and between microorganisms. Competition for resources and inhibition or elimination of antagonist species would greatly benefit of *in situ* and real-time monitoring, making MALDI-MSI well suited for biomarkers, chemical communication elucidation, or new molecules discovery[12].

Despite the recent advances of MSI applied on the field of microbiology, concerning the reproducibility of sample preparation, homogeneous matrix deposition or data acquisition[12], the MSI data generated have a huge memory footprint and are challenging to process and analyze. Actually, each image file contains thousands of spectra (and consequently thousands of pixels), each of which displaying thousands of mass *m/z* signals (peaks), especially when MALDI-MSI and High-Resolution Mass Spectrometry (HRMS) are engaged. These *m/z* signals can have diverse origins. In the low mass range, they are for instance mainly due to matrix ions, adducts, and aggregates[13]. The mass spectra also contain signals from salts, agar and the unconsumed nutrients (amino acids, sugars, ...) contained in the culture media that further increase the chemical background and may cause ion suppression effects, impairing the detection of ions specifically implied in the investigated metabolic activities of the microorganisms. Most of these molecules can be ionized and desorbed in positive ionization mode as protonated species, or cationized with sodium, ammonium, and potassium adducts, and in various oxidation states due to the ionization processes. In negative ionization mode, the metabolites and their oxidized forms are mainly detected as deprotonated species. Overall, the number of signals contained in each image can easily exceed several thousands of *m/z* peaks, making untargeted and discovery studies by manual data analysis unrealistic even if it could lead to the highlight of a few (sometimes arbitrarily selected) ions of interest[8]. Therefore, the data generated by MSI often requires advanced software or computational approaches to be analyzed. Computational tools can handle the complexities of MSI datasets, such as spatial segmentation, component analysis, hypothesis tests or statistical tools based on co-localized features. They have been developed for identifying patterns, similarities, and differences within datasets. Spatial segmentation appears to be an appropriate approach for assessing similarities between spectra and grouping them into clusters. This allows differentiation of specific areas within the sample, based on their molecular compositions[14]. The so-called co-localized value approach reveals *m/z* peaks that are primarily localized within a specific region of the image, proving highly beneficial for evaluating the spatial relationship among various molecular species in a sample[15]. Principal Component Analysis (PCA) remains one of the most popular methods to determine the variance within a data set. These approaches are typically used to reduce the complexity of the dataset by identifying patterns and highlighting variations between different samples[16]. Finally, univariate statistical hypothesis and discriminative tests such as t-test are suitable to determine whether an *m/z* value can be used to confidently discriminate between two or more groups[17].

However, none of the currently available software, including both academic and commercial codes used for analyzing MSI data, provide satisfactory data processing and still requires tedious efforts to refine the processed data by hand, especially for features which are unique to a peculiar dataset. This is though pertinent across various fields, including microbiology, and of great importance when studying secondary metabolism of bacteria. By modifying the culture conditions, it is possible to trigger global changes in the regulation of the metabolic pathways but also to induce the expression of cryptic genes responsible for the synthesis of unknown metabolites[18–21]. The development of our own in-house computational and statistical tools was proposed not only for improving data treatment in the context of microbiology-based studies, but also for speeding up the MSI data analysis while enhancing the quality of the results. Similarly, the development of in-house software by other laboratories has already been demonstrated significant to improve analytical capabilities, showcasing the broad applicability and potential of custom solutions in scientific research[22–24].

Here we propose a complete workflow based on 2 algorithms used separately or in combination to filter and extract a maximum of ions of interest with a great confidence. The first one is exploiting the differential signals detected between two or more datasets, the differential signal filtering (DSF). The second is a scoring of the ion image structure (i.e. evaluation of the degree of coherence concerning the repartition of the ion image in the whole image) using an image structure filtering (ISF) coupled with a fold change score (i.e. how much the score changed between controls and the condition of interest).

We challenged our method using three case studies to demonstrate its robustness and the increased speed of the data analysis. In the first case study, we propose guidelines on how MSI data treatment should be performed in the frame of MALDI-MSI applied to metabolomic response of bacterial co-culture in interaction between *Escherichia coli* and *Streptomyces coelicolor* to improve the detection and the identification of ions. The strengths and weaknesses of DSF and ISF methods – applied individually or in combination – is also critically discussed, especially in regards of a tedious manual analysis. Finally, two other case studies were applied using our proposed methods: (i) the cellobiose-mediated induction of the pathogenicity of *Streptomyces scabiei*, the causative agent of common scab on root and tuber crops, and (ii) the iron-repressed production of siderophores of *Streptomyces scabiei*.

Methods

Chemicals. α-cyano-4-hydroxycinnamic acid (HCCA), 9-aminoacridine (9-AA), and formic acid (FA) were purchased from Sigma-Aldrich (Belgium). Acetonitrile were LC-MS grade solvents obtained from Biosolve (Belgium).

Bacterial Strains, Medium, and Culture Conditions. The bacterial strains used in this study are *Streptomyces coelicolor* M145, *Streptomyces scabiei* 87-22, and *Escherichia coli* DH5α. Cultivation media were prepared as follows (for 1 liter): International Streptomyces Project number 2 (ISP2) medium (Yeast extract (VWR™) 4 g, Bacto™ Malt Extract (Gibco™) 10 g, D(+)-Glucose monohydrate (Carl Roth®) 4 g, and Bacto Agar (BD Difco™) 20 g) ; Peptone Yeast Malt (PYM, Yeast extract (VWR™) 3 g, Bacto™ Malt Extract (Gibco™) 3 g, D(+)-Glucose monohydrate (Carl Roth®) 10 g, Bacteriological Peptone (Condalab) 5 g, and Bacto Agar (BD Difco™) 20 g); Thaxtomin Defined Medium (TDM, $(NH_4)_2SO_4$ 2 g, $MgSO_4 \cdot 7H_2O$ 0.6 g, K_2HPO_4 2.6 g, sucrose 1,71 g, 1 ml of minor elements solution ($FeSO_4 \cdot 7H_2O$ 1 g, $ZnSO_4 \cdot 7H_2O$ 1 g, $MnCl_2 \cdot 4H_2O$ 1 g, $CaCl_2$ 1 g, in 1 liter, and Bacto Agar (BD Difco™) 20 g). Co-culture experiments were carried out by inoculating 30 µl of spores (10^7 CFU/µl) of *S. coelicolor* M145 in line on a thin layer on ISP2 medium[16]. After 24h of incubation at 28°C, 5 µl of an overnight culture of *E. coli* DH5α (OD_{600nm} set to 0.1) was spotted in contact with *S. coelicolor* and incubated at 28°C for 2 days. Induction of the virulome of *S. scabiei* 87-22 was monitored on TDM medium supplemented with sucrose (5 mM), and with either glucose (2.5 mM) or cellobiose (2.5 mM) as additional carbon sources. Finally, the effect of iron ($FeCl_3$ 10 µM) on the production of siderophores produced by *S. scabiei* 87-22 was monitored on Peptone Yeast Malt (PYM) agar medium.

Sample preparation for Mass Spectrometry imaging. Microbial cultures grown on the solid medium were excised directly from the agar plate and transferred to a microscope slide (VWR, USA), previously covered with a double-sided conductive copper tape (3M, Reference no. 1182/25). The slides were air-dried in an oven at 40°C for 2 h. For MALDI MSI experiments using positive ion mode, a HCCA matrix solution was prepared at 5 mg/mL in 80/20 ACN/water doped with 0.1% FA. 10 layers of matrix were sprayed onto the slides with the SunCollect instrument (SunChrom, Friedrichsdorf, Germany). The first layer was sprayed at a flow rate of 10 µL/min. The flow rate was increased by steps of 10 µL/min after each layer until it reached 60 µL/min. The 4-last layers of sprayed matrix was done using our maximum flow rate of 60 µL/min. The procedure of matrix deposition was repeated once again, which means that a total of 20 layers of matrix were sprayed per slides.

FT-ICR Mass Spectrometry. MSI data were acquired on dual source ESI/MALDI Solarix xR 9.4T (Bruker Daltonics, Bremen, Germany) using a mass range of 200-1600 m/z in amplitude mode and a file size of 2M (FWMH ± 300,000 @ 400*m/z*). This solariX xR 9.4T is an FT-ICR fitted with a high field supraconductive magnet, dynamically harmonized ICR cell ("paracell"), and 1ω detection capable of extreme mass resolving power (FWMH of 1 million and better at *m/z* 400) and mass accuracy better than 0.5 ppm. For each mass spectrum (pixel) of the image, the beam focus was set to "small" to reach a lateral resolution of the pixel's image around 100µm. 20 laser shots was performed at a repetition rate of 200 Hz. The laser power was tuned to produce a stable ion current between pixel as described elsewhere[25]. In this work, the laser power was set to 14%. Before imaging the sample, the mass spectrometer was externally mass calibrated from 200 *m/z* to 1600 *m/z* before each analysis using red phosphorus suspension in pure acetone and spotted directly onto the slides covered with the conductive tape. The laser power was adapted to produce a similar amount of ions (total ion current) than those obtained from the sample. FlexImaging v5 (Bruker Daltonics, Bremen, Germany) software was used to design the imaged areas for MALDI-MSI acquisition, using a pixel step size for the surface raster set to 200 µm. For supplementary sets of parameters of the FT-ICR used for performing MSI, see supplementary data (see Figure S1).

Data pre-processing and filtering. The data were first exported into the imzML files[26] of each image with Fleximaging 5.0 (Bruker). Each continuum mass spectrum was then individually (i.e., each pixel) converted to centroid mass spectrum. The mass spectra of one image of reference, were recalibrated using the m/z of HCCA clusters and then the average spectrum of the reference image is generated and the m/z of the peaks in this spectrum are used to recalibrate all the other images using algorithm published by La Rocca et al[27]. The MS images have been separately analyzed by the different filtration algorithms differential signal filtering (DSF) and image structure filtering (ISF) methods (discussed below). All images reported in this work were generated using SciLs™ 2016b (Bruker Daltonics, Bremen, Germany). All figures reported in this work will adhere to the SMART reporting standards to enhance the transparency and reproducibility of the experiments[28].

The differential signal filtering (DSF) method. DSF algorithm has been integrated in Mass Spectrometry Kendrick Filter (MSKF), an in-house Python3 software suite allowing the filtration of mass spectrometry data based on the Kendrick mass defect[29]. MSKF software reprocesses the imzML data into ICK files. Briefly the .ICK files are Python object files that contained a DataFrame of all signals detected in the MSI data conserving the pixel localization (i.e., X- and Y-position), the mass-to-charge m/z, and the signal intensity. All the MSI data in the imzML format were recalibrated using the software from La Rocca and coworkers[27] before the conversion into to .ICK files, as described elsewhere[30–32]. A noise reduction, based on a minimum intensity threshold, was applied to remove the background noise in each MS spectrum (each pixel). This intensity threshold was set high enough to remove the residual MS noise, but without suppressing the potential weakest signals intensity of the species of interest. After manually reviewing randomly selected pixels on a few images, a minimum intensity of 10^{+5} counts per scan (cps) was retained (maximum signal intensity in a pixel was around 5.10^{+7} cps). The MSKF differential signals filtering tool was used to extract differential signals between a set of .ICK files (e.g. condition A vs condition B, ...) and generated a new filtered .ICK file. This algorithm is based on a signal (*m/z* peaks) subtraction between data sets (e.g., condition A – condition B – condition C), retaining only the *m/z* peaks detected in the first data (unique signal in condition A in regards of condition B and condition C) disregarding their intensity values. The software requests an additional user-defined *m/z* tolerance window to consider the possible mass shift events and decide if the peaks corresponded to the same ions or not. Additionally, the filtering could also impose an additional parameter defined by a minimum number of pixels that the signal has to occur on the whole image (e.g. the m/z peak need to be detected in more than 100 pixels), minimizing the risk of considering chemical noise due to e.g. single-pixel hot spot.

The image structure filtering (ISF) method. ISF algorithm is an in-house software combining a python 3 script and R script. From the imzML recalibrated MSI dataset, the mean spectra of the MSI experiments for each case studied were computed. Peak picking was performed within contiguous mass window of at least 0.01 *m/z*. In the case of several *m/z* peaks are found in this window, only the most intense one is conserved. A list of *m/z* containing the 10^{+4} most intense *m/z* peaks detected in the mean spectrum is generated. An ion image is generated for each of these peaks and for each MS image to be compared. This is performed by iterating through each pixel and reporting the maximum intensity value detected within the +/- 0.01 *m/z* window around the *m/z* from the peak list. Once all the MSI pixels have been parsed, a list of intensity values that can be re-arranged into an ion image from the MSI pixel coordinates (X- and Y-positions) is obtained. The images generated are scored between 0 and 1 to assess the topological coherence of the signal, also described as structure, using an algorithm that evaluates the level of structure of the ion image. A score of 0 corresponds to a complete absence of structure and a totally random distribution of pixel intensities across the analyzed space, while a score closed to 1 corresponds to a structured image, where large regions of pixels have very close intensity values (see examples in Figure 4). The current approach to compute the level of structure is highly based on the measure of spatial chaos algorithm, developed by A. Palmer and coworkers[33], which was initially inspired by the concept of persistent homology[34]. The score can be exploited to evaluate the variation of spatial distribution between 2 ion images by computing the fold change of structure scores (score of the tested condition divided by the score of the control condition).

To provide context to the processing time of the resulting data, the algorithm was run on a single processor workstation containing an Intel(R) Core(TM) i9-9920X CPU @ 3.50GHz. For one imzML file, it takes less than 40 minutes to complete the processing. The algorithm runs independently without requiring supervision, allowing time to focus on other tasks.

Code Availability

Software to perform the image structure filtering are implemented in an R package. The software and installation instructions are available on GitHub at https://github.com/LaRoccaRaphael/ISF.

Results and Discussion

The currently available software, including both academic and commercial codes, mainly focuses on processing the MS image data through targeted mode, untargeted mode based on statistical analyses of the data (multivariate analysis, data clustering, ROI, etc.), or employs brute force peak annotation (requiring an existing and compatible database). None of these approaches provide satisfactory data processing when applied to the field of chemical communication and chemical interaction between bacteria, fungi, and even plants, particularly in the context of studies to discover new compounds. Indeed, these solutions are not specifically designed for identifying *m/z* signals that are uniquely present in one dataset compared to control samples. This specific capability is essential for accurately distinguishing the unique chemical signals emerging from interactions within these biological systems. Scientists have to then use the basic features of the MSI software such as data visualization, data segmentation, and peak list generation, that need to be manually reviewed and processed. These are tedious and time-consuming tasks where human errors can likely occur due to, among other things, the large (thousands of *m/z* peaks) peak list to handle, significant chemical background noise, disparity of the cations adducts, or ion suppression effects. Here, the idea is based on the fact that unique chemical signals are emitted exclusively during interaction between two bacterial (or other) species and not in monocultures. So, to isolate these interaction-specific signals, the list of *m/z* features identified under monoculture conditions (serving as a negative control) is compared from the list obtained under co-culture conditions. This makes it possible to distinguish only the differential signals specific to the interaction. Moreover, one of the strengths of MSI lies in its ability to discern between compounds and the random chemical background, leveraging the principle that compounds produced during interactions are not scattered randomly but are organized into structured ionic images. This structured nature of the compounds ionic images serves as a foundational concept for isolating signals specific to interactions, thereby enhancing the accuracy of the identified chemical signatures unique to molecular crosstalk.

Methodology. Co-culture experiments between *Escherichia coli* and *Streptomyces coelicolor*, two well-known bacterial species, were carried out in 5 replicates. Both species have been extensively studied and therefore benefit from well-established metabolomic databases[35–37]. Furthermore, preliminary assays have revealed that specific compounds were generated as a consequence of the interaction between these two-model species.

The differential signal filtering (DSF) and image structure filtering (ISF) methods were challenged for identifying specific signals in comparison to the manual reviewing of the dataset of MS images. A detailed and methodical process is described in Figure 1. The initial step involves an exhaustive manual reviewing of all ions present in the images using SciLs™. This manual analysis generates one reference lists that contains all the *m/z* features corresponding to all ions detected in one specific condition, the interactions between the two bacterial strains. The *m/z* features, grouped in the list A, are retained only if the *m/z* peaks were detected in at least 4 out of the 5 replicates. These manual analyses were conducted for benchmarking purposes, providing a reference to compare and validate the performance of the two methods. The comprehensive nature of manual analysis ensures that we have a robust dataset against which to measure the effectiveness of our automated approaches. Thus, the automated workflows described in our study are designed to streamline the process, offering a more scalable solution by reducing the time and effort required for analyzing MS images. We plan next to apply these methods in the two next study cases, allowing us to further validate their effectiveness and demonstrate their practicality in various experimental contexts.

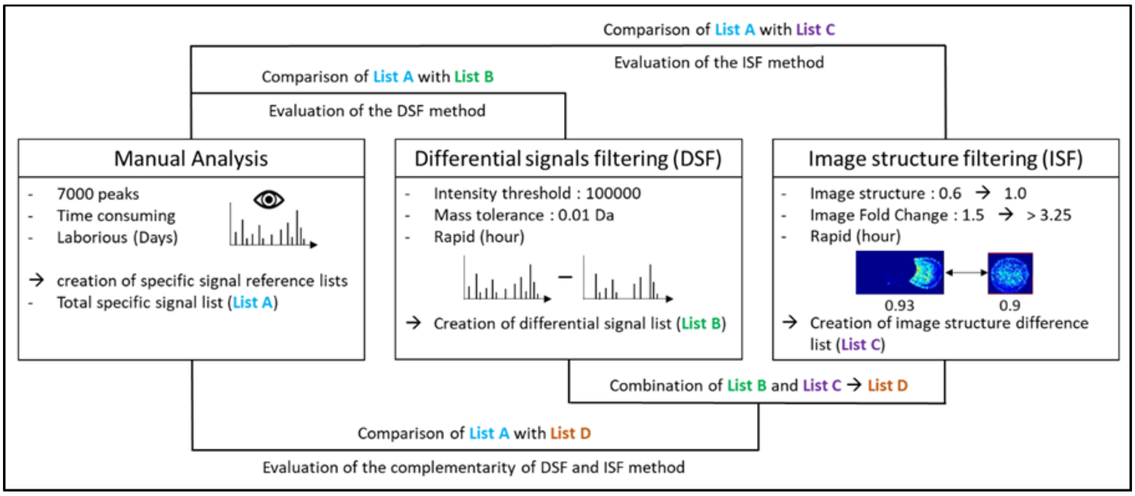

Figure 1: Methodology for filtering and comparing specific signals depending on the filtration method (DSF and ISF). Manual analysis generates the List A (blue), which serves as reference lists for methods evaluation. Differential signals filtering (DSF) and image structure filtering (ISF) generates respectively List B and C, allowing the comparison to be made. Finally, list D is created by combining List B and C to evaluate the complementarity of the methods to identify specific signals.

Manual analysis of the interaction between *Streptomyces coelicolor* and *Escherichia coli*:

In a first phase of the study, we identified in the MS images all (or at least considered as such) the signal (i.e. *m/z* peaks) specifically induced during the bacterial interaction. To do so, the MS images obtained from the three controls (non-inoculated ISP2 medium, and ISP2 inoculated either with *S. coelicolor* or with *E. coli* alone) and from the tested condition (co-culture of *S. coelicolor* and *E. coli* on ISP2 medium) were first processed with SciLs™ software. The average spectrum corresponding to the co-culture condition was then investigated to look for signals specifically induced by the interaction between bacteria with high confidence level. For this purpose, the following criteria were required to postulate about the specificity of a signal implied during the interaction:

- The ion must be exclusively present in the mass spectra of the tested condition (interaction).
- The resulting ion image had to be recorded in a minimum of 4 out of 5 biological replicates, the topological coherence of the ion image structures was visually assessed by the operator.

MSI is generally speaking not a quantitative method, especially considering the label-free techniques, despite substantial efforts have been done on tissue sections during the last years[38–43]. In the absence of quantitative method that could operate directly on agar media, we decided to retain the signals detected only during the interaction between the bacteria. Consequently, we omitted the signals variation linked to the overexpression or repression of the compounds, initiated by the co-culture conditions, but in return, we preserved our data from biases and ambiguities.

Figure 2A shows the optical images obtained for five biological replicates of samples from control (ISP2 medium), from monocultures (*E. coli* and *S. coelicolor* grown alone), and from the interaction between the two bacteria. The colonies are delimited using dashed lines for convenience. Figure 2B and 2C show the ionic images of two ions, m/z 588.263 and m/z 609.329, which are present in the interaction samples but not in the control conditions. These ions are part of list A (the peak list of all detected ions in the interaction samples but not in the controls). The signal detected at m/z 588.263, secreted by *S. coelicolor*, and m/z 609.329, produced by *E. coli*, were exclusively detected when the two bacteria are co-cultured.

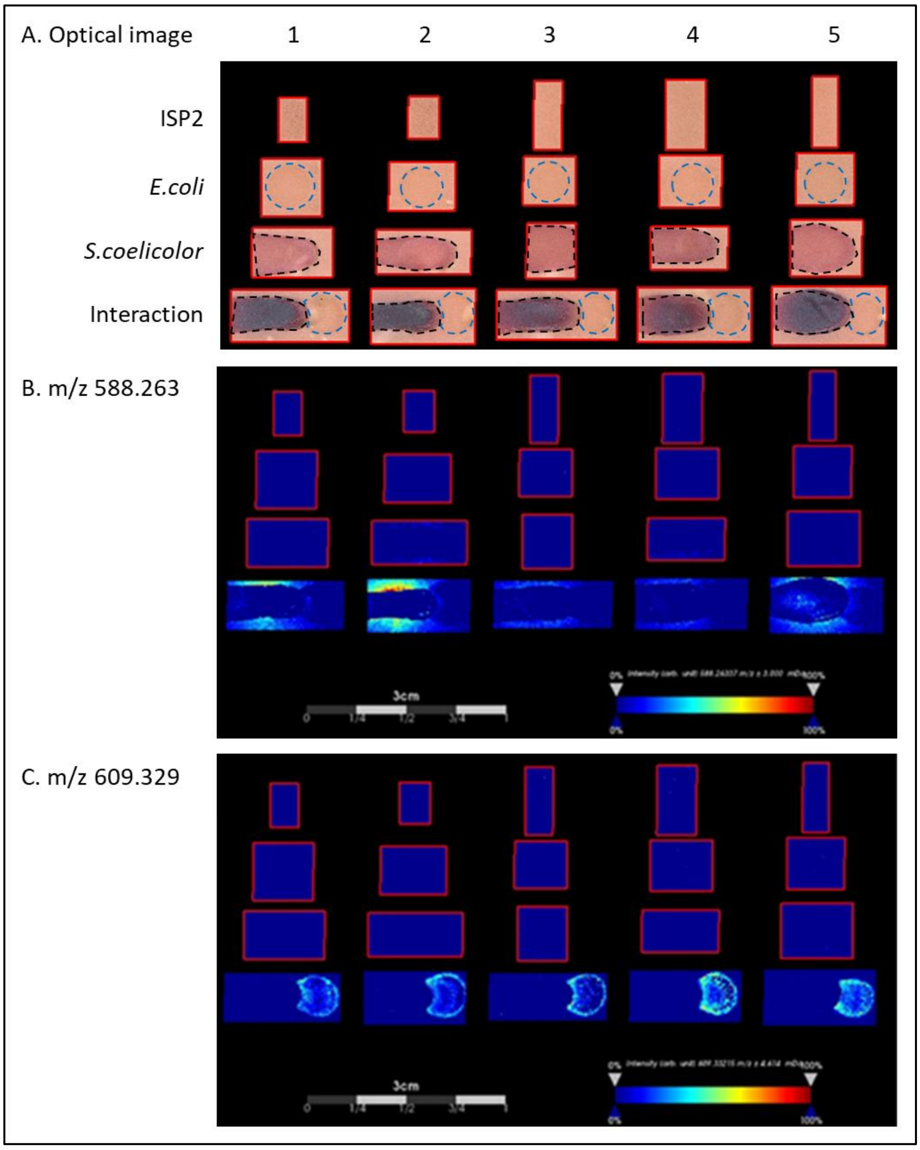

Figure 2: study of the interaction between *S. coelicolor* and *E. coli* by MALDI mass spectrometry imaging. A. Optical images obtained for the five biological replicates. Colonies are delimited using dashed lines for convenience B. m/z 588.263 and C. m/z 609.329, representing two of the ions preserved after manual filtration. Scale bar (grey and white) represent the real size of the images. The color bar is coding the relative ion intensity (from blue as 0% to red as 100%) of the extracted ions for each pixel of the ionic images. Bacteria colony limits are represented by dotted blue lines for *E.coli* and black lines for *S.coelicolor*

About 7,000 peaks were present in the average spectrum of the MS images generated in SciLs™. 205 m/z signals were specific to the tested condition when compared against controls (medium and mono-culture), and these were compiled into List A which represents about only 3% of all signals manually filtered. It is important to emphasize that, within this list of signals, a large proportion present highly structured ionic images, offering notable consistency in relation to the samples analyzed. This consistency reflects the specificity of chemical interactions within the cultures studied, but also underlines the reliability of the selection of signals for List A. However, it is also crucial to recognize that some signals within this list are less structured, generally characterized by lower intensity. All the images are present in the supplementary files (Supplementary file_2). Thus, the manually processed data leading to List A constitutes the reference set allowing the benchmarking of our proposed filtering bioinformatics tools in terms of performance and accuracy.

Evaluation of the differential signals filtering (DSF) method. The DSF method, which drastically speedup the data filtering, was initially used to extract the ions of interest, appearing in the context of the bacterial interaction between *E. coli* and *S. coelicolor*. The MSI data corresponding to this interaction were filtered with successive subtractions of the MSI data of the blank (non-inoculated ISP2 culture medium) and negative controls (monocultures). For example, the signals of the MALDI matrix, present in all data sets, were automatically removed during the filtering steps. Figure 3 illustrates the signal subtraction processes on a restricted mass range (*m/z* from 390 to 400) for clarity. After the subtractions of the m/z peaks from the three controls (non-inoculated medium and the two monocultures, Figure 3B, 3D, and 3F), only signals corresponding to the *m/z* peaks specific to the interaction between both microorganisms remain (Figure 3G). Each of these peaks resulting from the co-culture can then be extracted and their corresponding images reconstructed in SCiLs™. The metabolite detected at *m/z* 394.286 is illustrated as example (Figure 3H), demonstrating the capacity of the filtering method to highlight interaction-specific signals.

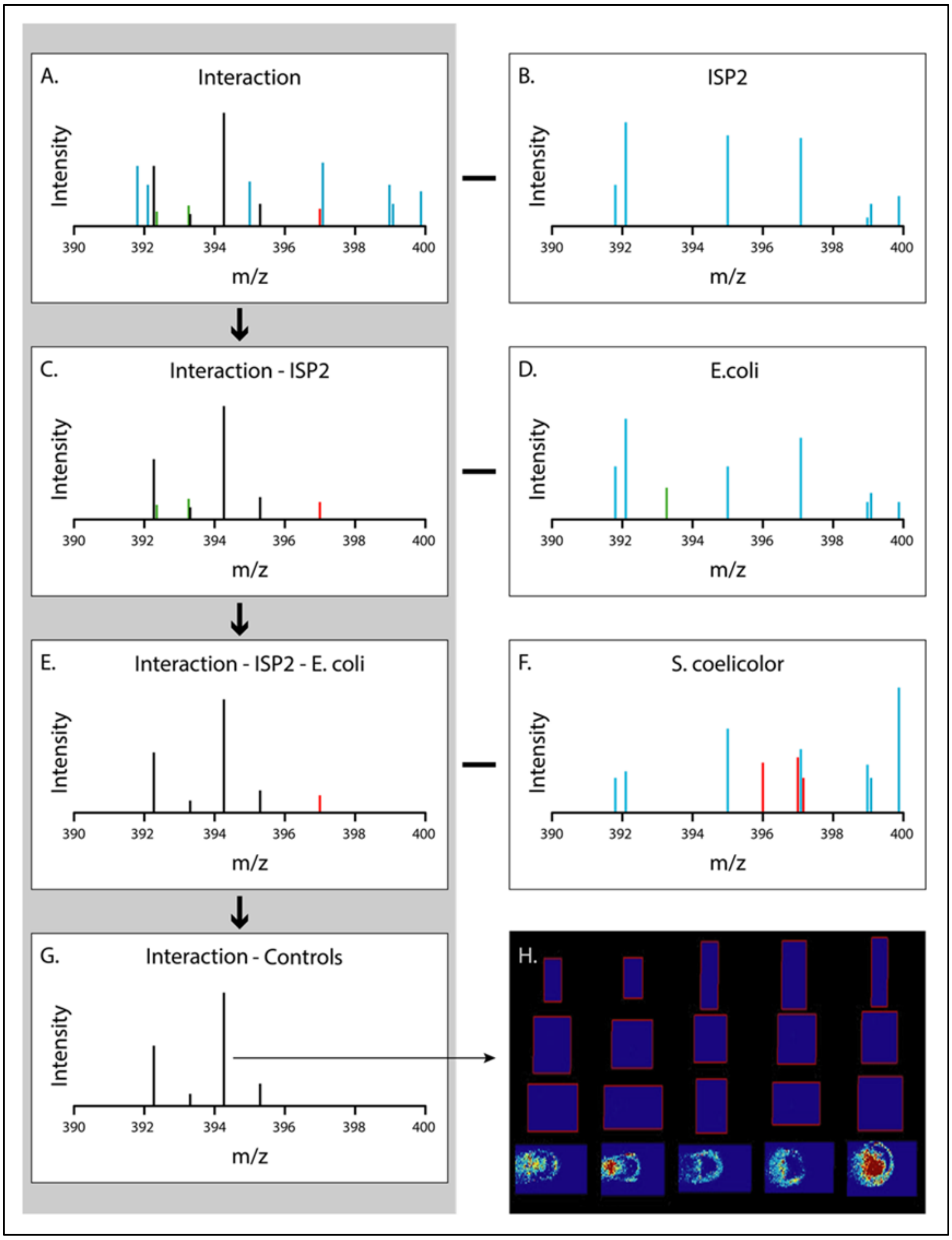

Figure 3: Overview of the differential signal filtering method processing. A. Zoom of the MSI average mass spectrum between m/z 390 to 400 of the interaction between *S. coelicolor* and *E. coli.* B., D., F. panels present the mass spectra of the control (no bacterial interactions) conditions i.e., the non-inoculated ISP2 culture medium, the monoculture of *E. coli* on ISP2, and the monoculture of *S. coelicolor* on ISP2, respectively. C., E., G. panels present the processed mass spectra after successive subtraction of controls presented in panels B, D, and F, respectively. Signals (m/z peaks) are colored based as their function of origins: blue, culture medium ISP2; Green, *E. coli* monoculture; Red, *S. coelicolor* monoculture; Black, signal specific to bacterial co-culture (interaction). H. MSI image of a specific signal (394.286 m/z) obtained after filtration where relative intensity of the signal is color coded from blue (0%) to red (100%).

Each of the five biological replicates generates so its own mass lists, containing exclusively the compounds produced during the bacterial interaction for a given replica.

To maintain consistency with the manual analysis approach, the same criteria of selection was kept and only the *m/z* signals retrieved in a minimum of four replicas were retained for further data analysis. This method resulted in the selection of 107 signals, constituting the List B (Figure 1) (Supplementary file_1). From these 107 *m/z* peaks, 87 are also among the 205 signals of List A (*total specific signal list)*.

Remember that the DSF method applied an intensity signal threshold and a mass shift tolerance during the filtration of the MSI data (see Material and Methods), the less intense ions could be lost due to the applied threshold, and closely related *m/z* values from 2 different MSI dataset to be subtracted could be considered as a unique and same feature by the algorithm included in the home-made MSKF software, depending on the user settings. The introduction of false negatives and false positives can be avoided using a better evaluation of the signal threshold in regards of the background noise in the MSI data and the reduction of the applied mass tolerance. The range of the mass tolerance window has to be considered in regards of the MS instrumentation, MS methods, and MS performance during the acquisition of the MSI data. The mass tolerance can be greatly reduced but it requires a high quality of data such as delivered by MALDI-Orbitrap and MALDI FT-ICR instruments fitted with high field supraconductive magnet, dynamically harmonized ICR cell ("paracell"), and 1ω or 2ω detection, which limited the mass shift and provide sub ppm or better mass accuracy (sub ppm mass accuracy ensures 6 significant digits of the floating-point *m/z* value). However, it is noteworthy that data realignment or recalibration serves as an effective strategy to minimize the mass shift effect and improve the mass accuracy, respectively[44]. The evaluation of the spectrum noise background is evaluated by manual evaluation of the average noise around dominant m/z peaks observed in some ion image pixels randomly selected from the MS image.

Evaluation of the image structure filtering method (ISF). The second filtration method adopted is based on an approach particularly well adapted to MSI dataset. This method, based on the topological coherence of the image (also defined as structure of the image), assigns a score to the structure of each feature (i.e. an extracted m/z value from the ionic image). See in the material and methods for details. Briefly the score tends to 0 in the absence of structure and reaches a maximum of 1 for a very structured ion image. For illustrating this, Figure 4 shows the structure scores of three different *m/z* values in a biological replica. These scores are then used to calculate a new parameter, called the "fold change structure", representing the ratio of the highest score of the interaction to the highest score obtained in the controls. This calculation is performed on the 10^{+4} most intense features, generating a table including all masses (*m/z* peaks) according to two criteria: the image structure of the interaction and the fold change (FC).

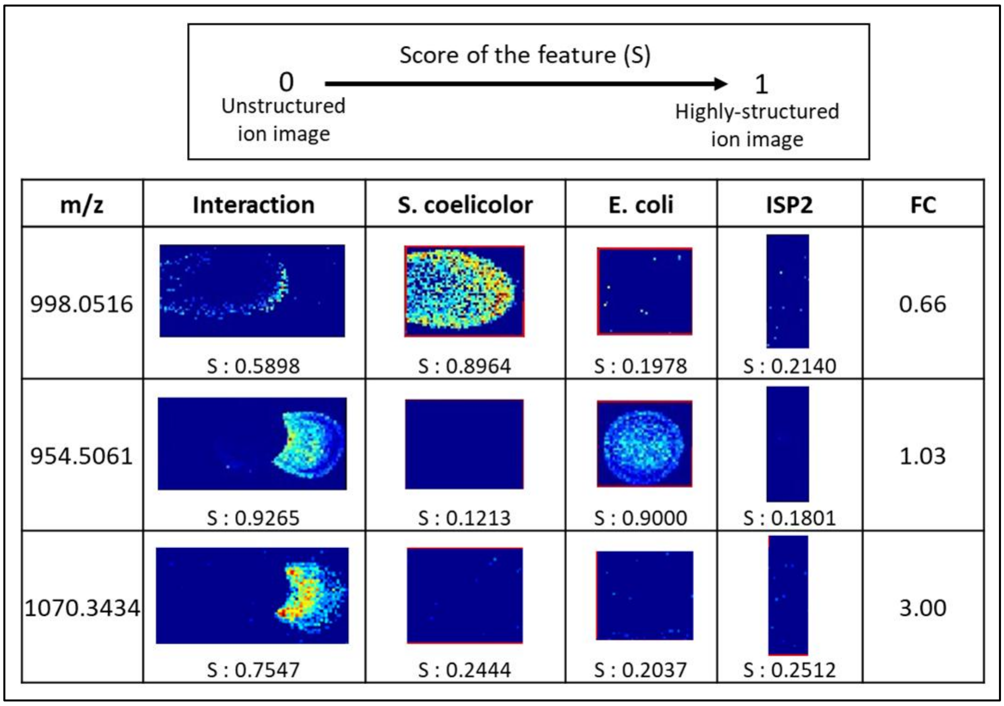

Figure 4: Overview of the image structure filtering method. The legend ranging from 0 to 1 show that scores close to 0 correspond to unstructured ion image and 1, highly structured ion image. To illustrate this, three m/z were picked (m/z 998.052, 954.506 and 1070.343) and the images and scores (below the image) of every condition (Interaction, *S. coelicolor*, *E. coli* and ISP2) for one biological replicate are shown. Finally, Fold Change score (FC) is computed by comparing scores of the interaction over the scores of the highest control condition.

Figure 4 additionally shows the ionic images of selected *m/z* values to illustrate the varying fold changes that can be obtained by comparing their structural scores. Indeed, for the first *m/z* 998.052, a fold change (FC) of 0.66 is calculated. As this score is below 1, the image structure is better in at least one of the negative controls, here *Streptomyces coelicolor*, compared to that of the interaction (co-culture). This suggests that the metabolite is potentially less abundant in the co-culture condition compared to the monoculture (bacterium grown separately). Indeed, the intensity usually reflects the abundance of the metabolites and can contribute to higher image structure, as it often affects the topological coherence of the signal. For m/z 954.506, a FC 1.03 is calculated, which indicates no significant change in the abundance of this metabolite between the sample resulting from the interaction experiment and the bacterial monoculture. Indeed, the associated images do not show a notable difference, confirming that the interaction does not affect significantly the relative intensity of the detected ion corresponding to this metabolite. For the m/z 1070.343 an FC value of 3.00 is reached, suggesting an induced production of this compound during the co-culture, which is indeed validated by the visual evaluation of the images.

The underlying hypothesis of this filtering method is that if a signal is specific to a condition under test, it will have a high structure score, and the ratio of fold change (FC) will indicate to which condition the improvement of the structure score is associated. In evaluating these 10,000 features, we can retain only signals with a structure score for the condition tested of our interest in regard of to their fold change.

Thus, Figure 5 shows the results obtained using the ISF filtering method, conserving the images with structure score higher than 0.6, 0.7, 0.8, and 0.9 only. The blue line indicates the number of peaks classified into the reference Lists A, representing thus the target to achieve. Considering a structure score greater or equal than 0.9 and a fold change greater or equal than 1.5, only 29 significant signals remain, all of them initially included in the List A. This drastic structure fold change scores exemplifies how the methodology can automatically identify peaks exclusively specific to the interaction condition. By tuning the structure and fold change parameters (for instance signals with a structure score greater than 0.7 and a fold change greater than 1.5), the number of signals obviously increases. For these new arbitrarily selected parameters, 160 significant signals are identified (List C, see Supplementary file_1) among which 115 were already present in the List A (*list of specific signals determined with a manual analysis*). These signals account for 56% of all candidate ions characterized by a production response unique to the bacterial interaction. This filtration method is therefore of great interest not only because it automatically extracts most of the significant signals from the huge amount of data, but also because it allows the parameters to be modulated by the operator, offering a flexible and efficient approach to the analysis process.

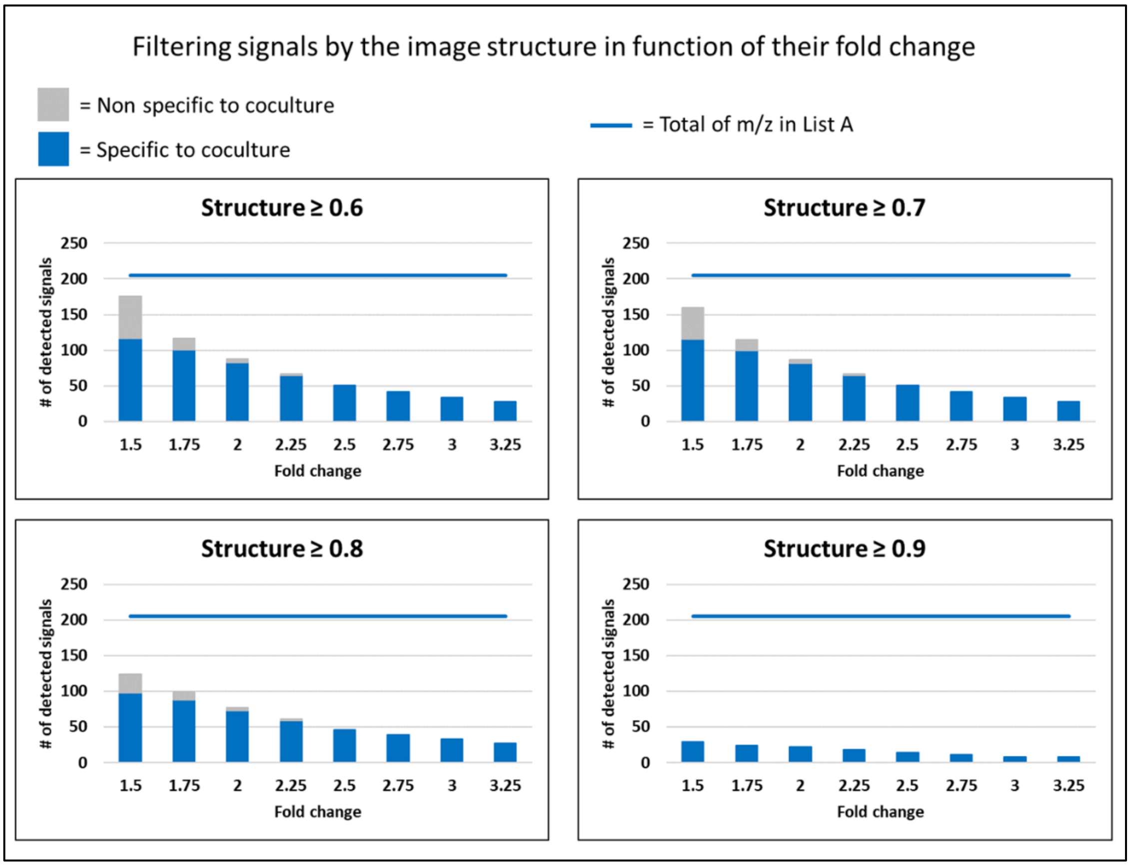

Figure 5: Histograms Comparing the total specific m/z depending on the image structure in function of the fold change. The blue line represents the number of signals in the List A. The histograms show the number of signals (# of m/z) retained as a function of the parameters applied (fold change of structure). The blue part of each histogram indicates the signals thus the m/z present in the List A, while the gray parts show the signals considered as false positives.

Obviously, when the Fold Change exceeds 1, the number of ions specific to the bacterial interaction increases, suggesting that a higher FC effectively filters out false positives. In the examples shown in Figure 5, a FC score of below than 2.25 makes false positives signals appear (signals in grey in Figure 5). Furthermore, when the base score of the structure is high (0.8 and 0.9), the total number of ions specific to the co-culture is of course lower but strongly relevant. Here is now the tradeoff for the users: Choosing higher structure scores ensures reliable results but could lead to the exclusion of relevant ions that would have been selected at lower thresholds. Conversely, lower structure scores can identify a larger number of ions, but this also includes a higher likelihood of detecting false positives, thus necessitating careful evaluation of the MSI data.

The ISF method exhibits thus certain limitations, particularly regarding false negatives and false positives, due mainly to too selective signal criteria. On one hand, false negatives can be highlighted for two primary reasons. First, the calculation of the Structural Fold Change, a critical parameter for distinguishing significant *m/z*, relies on comparing the highest structural score of the interactions among the 5 replicas against the highest score of all the control experiments. If a structural score in a control is abnormally high, it can lead to false negatives. Secondly, the ISF method was designed to retain the 10,000 most intense signals. This constraint could induce biases by ignoring minor *m/z* peaks in the MS data. On the other hand, several factors also explain the occurrence of false positives. In contrast to false negatives, if the structure scores in the controls are too low, the ratio between the structure scores can lead to false positives. Moreover, the current ISF method does not consider our last criteria which is an ion having a structure should appear on 4 of our 5 replicates. For example, the compounds detected at *m/z* 600.3204 was computed with descent ISF and FC scores but was detected in only 3 of our 5 biological replicates (see Figure S2) This compound was assigned as false positive, maybe wrongly.

Synergy between the Differential Signal Filtering and the Image Structure Filtering methods.

Ultimately, our aim was to analyze whether these two filtering methods were complementary and led to a more efficient filtering of the detected features in the MSI dataset. The 107 m/z signals obtained by the differential subtraction filtration (DSF) were merged with the 160 masses obtained by the image structure filtration (ISF) by applying a threshold structure parameter greater than 0.7 and a Fold-change greater than 1.5. This combination resulted in a total of 204 m/z signals, referred to as List D. Among these 204 m/z signals in List D, 139 were also found in List A (specific signal list, manual filtering) (Supplementary file_1).

Moreover, more of these specific signals occurred if isotopologues were also considered. The disparity in signal intensity between the ion at the monoisotopic mass and the respective isotopologues typically results in a lower percentage of covering of the list A by the different filtering method. Indeed, isotopologues have a greater chance to exhibit lower structural score in ISF method or are most likely to be excluded by the threshold applied in DSF method than the corresponding monoisotopic ion. It is important to note that the DSF-ISF combined approach filtered and retrieved within a three to four hours, 79.0 % of the signals specific to the case study, whereas the manual data filtering required days of continuous workdays. Among these identified ions through the filtering process, this mainly corresponds to signals presenting a certain intensity with ionic images possessing notable characteristics, such as a relative abundance around or inside the colonies, and diffusion from the colonies. The ions missing from this filtration compared to the reference list A, primarily, correspond to signals of low intensities, which typically have a weaker structure and are generally less visually interpretable highlighting the efficiency and precision of the DSF-ISF approach in prioritizing signals with significant biological relevance.

The Figure 6 proposes a Venn diagram illustrating the synergic effect of both filtering method. Combined DSF-ISF filtering shares 63 common features with List A, while additionally providing their own masses. The 24 and 52 signals unique to the DSF and ISF lists, respectively, are not false positives, as they are present in List A. These signals could only be detected by one of the filtering methods, highlighting the need for orthogonal methods to extract the maximum information from the complex data obtained from mass spectrometry images. While the filtering methods have proven efficient in reducing the data from several thousand features to only a couple of hundred m/z, they are not perfect. Some unique features might be lost due to parameters such as the intensity threshold applied in the DSF method or the fold change threshold chosen in the ISF method. It doesn't feel like there is a specific trend in the unique features found by the two methods. This underscores the importance of using complementary approaches to ensure comprehensive data analysis.

However, the 20 and 45 m/z signals that are fully unique to the DSF and ISF lists, respectively, are indeed considered false positives. This can be explained by several aspects of the filtering methods. For the DSF method, the false positives can arise from the signal threshold applied during filtration. While, on the other hand, in the ISF method, false positives may result from a too low structure score in the control samples or failing to meet one of the criteria, which requires that an ion with a structure should appear in 4 out of our 5 replicates (see Figure S2).

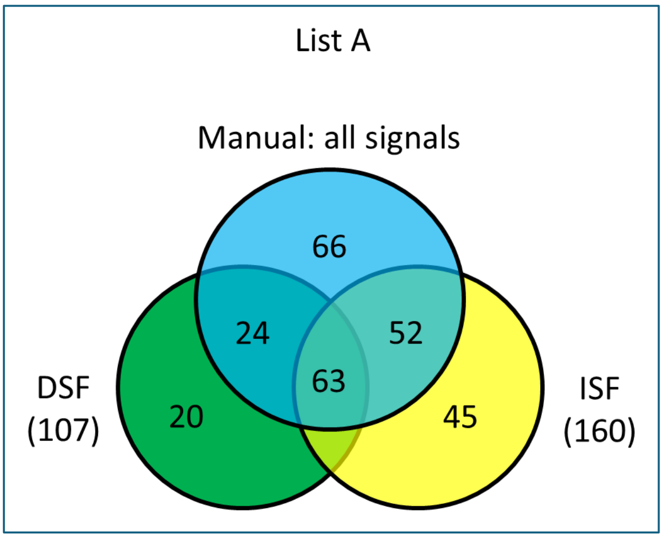

Figure 6: Venn diagrams compile the two new methods (DSF an ISF) of filtration in comparison of manual data analysis (all signal from the averaged mass spectrum in the MSI data).

Challenging the methodology with well-described case studies. A major challenge in modern microbiology is to identify the environmental elicitors and nutrients that modulate the expression of genes associated with the secondary/specialized metabolism of bacteria and fungi. Finding these elicitors and the characterization of the metabolomic response could also lead to the identification of novel molecules associated with the numerous cryptic biosynthetic gene clusters (BGCs). In that context, we challenged the DFS and ISF methods on two well-described culture conditions known for either activating or repressing the production of specialized metabolites. The production of these molecules is a costly process that requires tight expression control of biosynthetic gene cluster (BGCs) which transcription is activated or repressed upon sensing specific environmental signals. The two chosen on/off models are: (i) the cellobiose-mediated induction of the pathogenic lifestyle of *Streptomyces scabiei*, the causative agent of common scab on root and tuber crops, and (ii) the iron-repressed production of siderophores of *S. scabiei*.

(i) First case study: Cellobiose-mediated induction of *S. scabiei*'s pathogenicity. Thaxtomin phytotoxins are the main virulence determinants of *Streptomyces scabiei*. Their production is mandatory for pathogenicity and is induced by the transport of the cellulose byproducts: cellobiose and cellotriose[45,46]. Deflandre and coworkers[47] recently studied the composition of the virulome of *S. scabiei* in response to cellobiose. This study revealed by RNA sequencing that, in addition to the BGC responsible for producing thaxtomins, the expression of nine cryptic BGCs was activated by cellobiose.

We applied both DSF and ISF approaches to evaluate how they facilitate the identification of thaxtomins and to possibly discover candidate ions of new metabolites of cryptic BGCs which production would also be under control of cellobiose.

S. scabiei was grown on agar medium to permit the bacteria colonizing the entire surface of the plate. The production of BGCs was induced by diffusion of cellobiose from a soaked square-shaped piece made of paper deposited directed on the surface of the agar medium. Finally, water- and glucose-soaked papers were used as negative controls (Figure 8 A). The combined DSF-ISF method allowed to retain 393 ions from a total of more than 4,000 ions. Most of them was identified to be specific to the cellobiose condition (Figure 7 and Supplementary file_3-4). Among these signals, various ionization adducts of thaxtomins A and B were easily identified (Figure 7 B-C.) which confirmed that our filtering method allowed to retain ions of molecules regarded as positive control for cellobiose-induced metabolites.

This synergistic data filtering also revealed specific signals that could not be attributed to any known metabolite produced by *S. scabiei* upon cellobiose-mediated induction, such as the metabolite detected at *m/z* 934.220 ($[M+Na]^+$) and *m/z* 950.194 ($[M+K]^+$) (Figure 7D). These metabolites could actually be a combination of cellobiose and HCCA matrix, with some sodium/potassium adducts indicating a complex interplay of molecular interactions and ionization processes to the conditions under study.

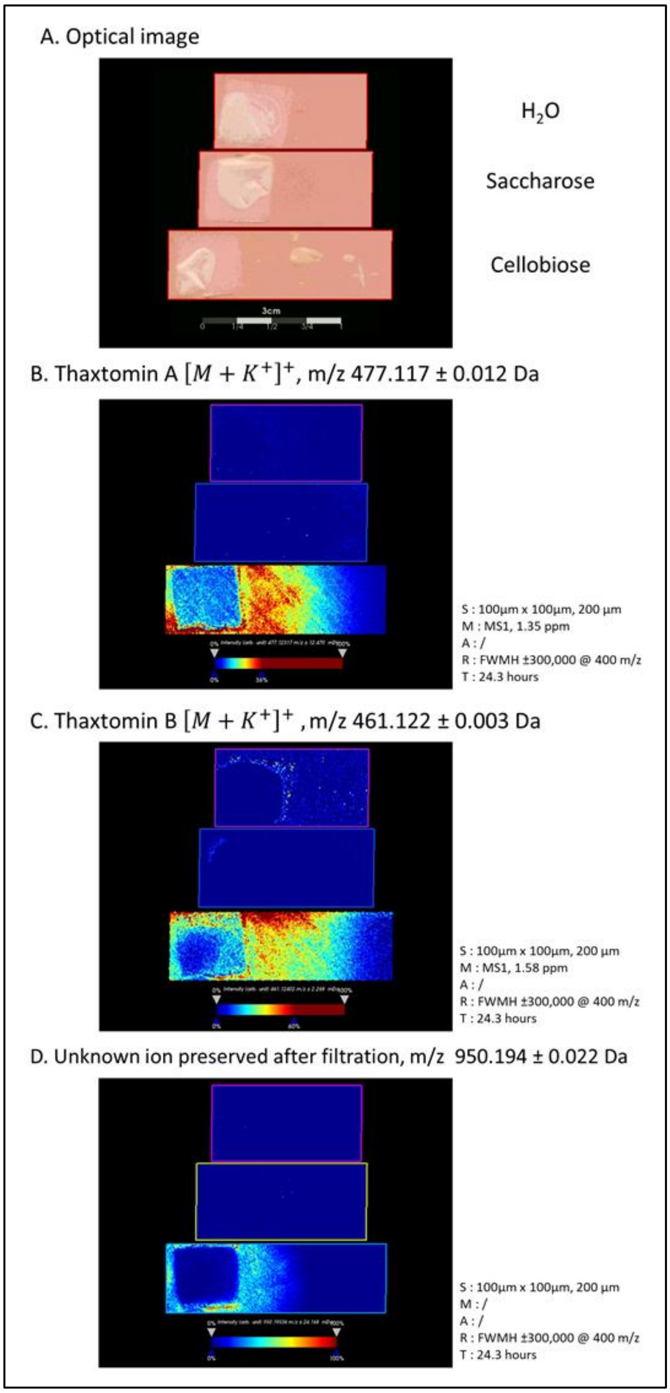

Figure 7: A. Optical image of *Streptomyces scabiei* grown on minimal agar medium, covering the entire, with a piece of paper soaked with water, saccharose or cellobiose. B. 477.117 m/z corresponding to [Thaxtomin A + K]+ C. 461.122 m/z corresponding to [Thaxtomin B + K]+ and D. 950.19 m/z. corresponding to an unknown ion appearing preserved after filtration.

(ii) Second case study: The iron-mediated repression production of siderophores of *S. scabiei*. Our second case study consisted of examining how iron regulates metabolism with *S. scabiei*, specifically looking at siderophore production in relation to iron presence in the medium. Although many siderophores of *S. scabiei* have already been identified (desferrioxamines, pyochelin, scabichelin/turgichelin), others, not yet described as siderophores, have been predicted through genome mining[47]. Indeed, siderophore production is classically controlled by the availability of iron. Indeed, iron depleted environment triggers their biosynthesis while iron overload represses their production[48]. In this context, we evaluated the combined DSF-ISF approach by MALDI mass spectrometry imaging (Figure 8) for screening the presence of features associated with siderophores productions by *S. scabiei*, cultivated under conditions either triggering or repressing the production of iron chelating molecules. The strain was cultivated on peptone – yeast – malt extracts (PYM) medium without or with iron (FeCl3 10 μM). Applying the DSF-ISF filtering method led to a total of 749 selected signals (out of more than 6,000 features) that disappear in the presence of iron. A majority of these signals display the expected contrasted production pattern, i.e. no signal in the presence of iron (Supplementary file_5-6). Among them, ions corresponding to *S. scabiei* siderophores desferrioxamine B (Fig. 8B), desferrioxamine E (Fig. 8C), and scabichelin (Fig. 8D) were easily detected.

Moreover, we observe ions such as m/z 498.266 (Fig. 8E) and m/z 514.240 corresponding, respectively, to a not yet identified compound with a sodium and potassium adduct, which both display iron-repressed production patterns similar to the siderophore produce by S. scabiei but cannot be related to any known siderophore of this strain when looking in database such as NPatlas[49] or StreptomeDB[35]. The distinctive ion signatures and their iron-repressed production patterns, similar to known siderophores yet unidentifiable in current databases, suggest these compounds could represent novel siderophores synthesized by S. scabiei. This possibility underlines the importance of further investigation into these compounds.

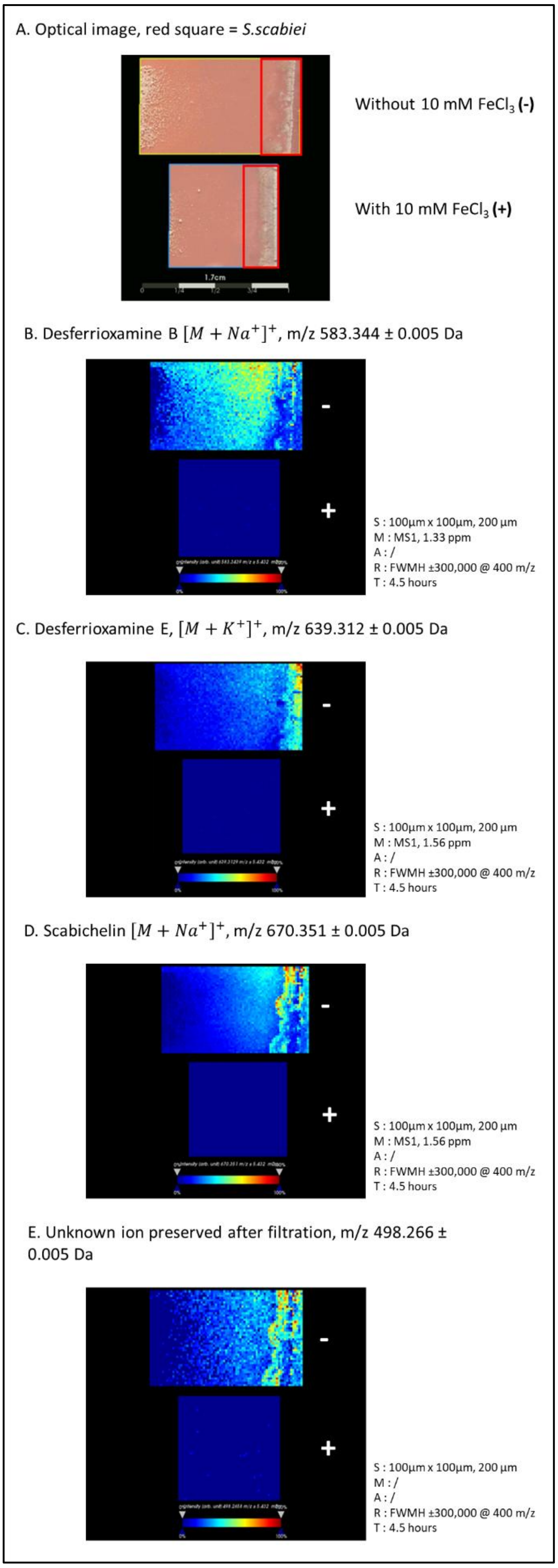

Figure 8: A. Optical image of *Streptomyces scabiei* cultivated on peptone – yeast – malt extracts (PYM) supplemented without (top) and with 10µM of FeCl3 (down) B. m/z 583.344 corresponding to [Desferrioxamine B + Na]+ ion C. m/z 639.312 corresponding to [Desferrioxamine E + K]+ ion D. m/z

670.351 corresponding to [Scabichelin + Na]+ ion and E. m/z 498.266 corresponding potentially to an unknown siderophore with a sodium adduct.

Conclusions and perspectives:

In this study, two computational methods for filtering mass spectrometry imaging (MSI) data were introduced and evaluated, with a particular focus on untargeted microbiological applications. The differential signal filtering (DSF) and image structure filtering (ISF) methods were meticulously assessed through co-culture experiments and the appropriate negatives controls, revealing their significant potential in streamlining the data analysis process of MSI, especially in the context of molecule discovery and microbiological applications. Our method was evaluated in regards of a set of data manually filtered, which could itself includes biases due to human errors. Additionally, we used "only" one manually filtered dataset as a reference dataset to benchmark our method. Confronting the method to a larger set of trusted reference data would be desirable. If we could use more dataset with high confidence level of the annotated compounds, we are envisioning a better refinement of the user-defined parameters used by our algorithms and limit as much as possible the presence of false positives and false negatives.

The efficiency of these filtration methods has been demonstrated, particularly when used concomitantly, which allowed notably rapid screening and delineation of specific signals in atypical biological samples for mass spectrometry imaging. The time-saving aspect and easy operation to process these MSI data in regards of the manual processing or other multivariate statistical or data clustering tools (see Figure S3) included in other software suite is therefore most welcome. In the absence of quantitative *in situ* method by MALDI MSI applied for such agar-based culture medium, no induction or repression factor is yet accessible. This work offers new insights and perspectives in the study of microbiological samples, such as the secondary metabolisms of bacteria and data filtration of complex data can be particularly helpful for the adduct deconvolution of signals to identify metabolites, with an example of metabolite identification such as lyso-phosphatidylethanolamine 16:0 (see Figure S4-6).

We aim to further enhance the capabilities of these computational methods. The development pipeline includes the implementation of isotope deconvolution, a feature that will further assist the users during the metabolite identification process. This addition is anticipated to offer an even deeper understanding of the complex data generated by MSI, potentially broadening the scope of potential applications.

Acknowledgments

The authors are grateful for the financial support from the University of Liège by the mean of the Concerted Action Research MIND (2020-2023) and for the FEDER BIOMED HUB (number 2.2.1/996) that funded the MALDI FT-ICR SolariX XR and the TimsTOFpro.

Supporting Information

Parameters for the solariX XR 9.4T to perform MSI (Figure S1), Example of some false positive obtained by ISF (Figure S2), Data treatment using ROC between class 1 (controls) and class 2 (interaction between *E. coli* and *S. coelicolor*) (Figure S3), Filtration enables better adduct deconvolution of signals to identify metabolites (Figure S4), Images of lyso-phosphatidylethanolamine 16:0 with different adduct and sharing similar image (Figure S5), Experimental m/z and possible mono-isotopic mass considering different adducts scenario where M1 = 475.2651 Da and M2 = 453.2848 (Table 1) and Identification of lyso-phosphatidylethanolamine 16:0 by LC-MS/MS (Figure S6).

<u>**References**</u>

(1) Ho, Y. N.; Shu, L. J.; Yang, Y. L. Imaging Mass Spectrometry for Metabolites: Technical Progress, Multimodal Imaging, and Biological Interactions. *Wiley Interdiscip Rev Syst Biol Med* **2017**, *9* (5). https://doi.org/10.1002/wsbm.1387.

(2) Feucherolles, M.; Frache, G. MALDI Mass Spectrometry Imaging: A Potential Game-Changer in a Modern Microbiology. *Cells* **2022**, *11* (23). https://doi.org/10.3390/cells11233900.

(3) Zaima, N.; Hayasaka, T.; Goto-Inoue, N.; Setou, M. Matrix-Assisted Laser Desorption/Ionization Imaging Mass Spectrometry. *Int J Mol Sci* **2010**, *11* (12), 5040–5055. https://doi.org/10.3390/ijms11125040.

(4) Sturtevant, D.; Lee, Y. J.; Chapman, K. D. Matrix Assisted Laser Desorption/Ionization-Mass Spectrometry Imaging (MALDI-MSI) for Direct Visualization of Plant Metabolites in Situ. *Curr Opin Biotechnol* **2016**, *37*, 53–60. https://doi.org/10.1016/j.copbio.2015.10.004.

(5) Mitra, A.; Kant Choubey, S.; Das, R.; Rout, P.; Sridhar, T. S.; Mishra, D.; Kumar Mandal, A. Differential Expression of Proteins in Human Prostate Cancer Tissues Probed by MALDI Imaging Mass Spectrometry. **2020**.

(6) Carter, C. L.; McLeod, C. W.; Bunch, J. Imaging of Phospholipids in Formalin Fixed Rat Brain Sections by Matrix Assisted Laser Desorption/Ionization Mass Spectrometry. *J Am Soc Mass Spectrom* **2011**, *22* (11), 1991–1998. https://doi.org/10.1007/s13361-011-0227-4.

(7) Rauser, S.; Marquardt, C.; Balluff, B.; Deininger, S. O.; Albers, C.; Belau, E.; Hartmer, R.; Suckau, D.; Specht, K.; Ebert, M. P.; Schmitt, M.; Aubele, M.; Hǒfler, H.; Walch, A. Classification of HER2 Receptor Status in Breast Cancer Tissues by MALDI Imaging Mass Spectrometry. *J Proteome Res* **2010**, *9* (4), 1854–1863. https://doi.org/10.1021/pr901008d.

(8) Alexandrov, T. MALDI Imaging Mass Spectrometry: Statistical Data Analysis and Current Computational Challenges. *BMC Bioinformatics* **2012**, *13 Suppl 1* (Suppl 16). https://doi.org/10.1186/1471-2105-13-s16-s11.

(9) Watrous, J. D.; Dorrestein, P. C. Imaging Mass Spectrometry in Microbiology. *Nat Rev Microbiol* **2011**, *9* (9), 683–694. https://doi.org/10.1038/nrmicro2634.

(10) Gonzalez, D. J.; Haste, N. M.; Hollands, A.; Fleming, T. C.; Hamby, M.; Pogliano, K.; Nizet, V.; Dorrestein, P. C. Microbial Competition between Bacillus Subtilis and Staphylococcus Aureus Monitored by Imaging Mass Spectrometry. *Microbiology (N Y)* **2011**, *157* (9), 2485–2492. https://doi.org/10.1099/mic.0.048736-0.

(11) Dunham, S. J. B.; Ellis, J. F.; Li, B.; Sweedler, J. V. Mass Spectrometry Imaging of Complex Microbial Communities. *Acc Chem Res* **2017**, *50* (1), 96–104. https://doi.org/10.1021/acs.accounts.6b00503.

(12) Yang, J. Y.; Phelan, V. V.; Simkovsky, R.; Watrous, J. D.; Trial, R. M.; Fleming, T. C.; Wenter, R.; Moore, B. S.; Golden, S. S.; Pogliano, K.; Dorrestein, P. C. Primer on Agar-

Based Microbial Imaging Mass Spectrometry. *J Bacteriol* **2012**, *194* (22), 6023–6028. https://doi.org/10.1128/JB.00823-12.

(13) Baquer, G.; Sementé, L.; García-Altares, M.; Lee, Y. J.; Chaurand, P.; Correig, X.; Ràfols, P. RMSIcleanup: An Open-Source Tool for Matrix-Related Peak Annotation in Mass Spectrometry Imaging and Its Application to Silver-Assisted Laser Desorption/Ionization. *J Cheminform* **2020**, *12* (1), 1–13. https://doi.org/10.1186/s13321-020-00449-0.

(14) Lagarrigue, M.; Lavigne, R.; Guével, B.; Palmer, A.; Rondel, K.; Guillot, L.; Kobarg, J. H.; Trede, D.; Pineau, C. Spatial Segmentation and Metabolite Annotation Involved in Sperm Maturation in the Rat Epididymis by MALDI Imaging Mass Spectrometry. *Journal of Mass Spectrometry* **2020**, *55* (12). https://doi.org/10.1002/jms.4633.

(15) Mittal, P.; Price, Z. K.; Lokman, N. A.; Ricciardelli, C.; Oehler, M. K.; Klingler-Hoffmann, M.; Hoffmann, P. Matrix Assisted Laser Desorption/Ionization Mass Spectrometry Imaging (MALDI MSI) for Monitoring of Drug Response in Primary Cancer Spheroids. *Proteomics* **2019**, *19* (21–22), 1–4. https://doi.org/10.1002/pmic.201900146.

(16) Wang, S.; Bai, H.; Cai, Z.; Gao, D.; Jiang, Y.; Liu, J.; Liu, H. MALDI Imaging for the Localization of Saponins in Root Tissues and Rapid Differentiation of Three Panax Herbs. *Electrophoresis* **2016**, *37* (13), 1956–1966. https://doi.org/10.1002/elps.201600027.

(17) Keller, C.; Maeda, J.; Jayaraman, D.; Chakraborty, S.; Sussman, M. R.; Harris, J. M.; Ané, J. M.; Li, L. Comparison of Vacuum Maldi and Ap-Maldi Platforms for the Mass Spectrometry Imaging of Metabolites Involved in Salt Stress in Medicago Truncatula. *Front Plant Sci* **2018**, *9* (August), 1–15. https://doi.org/10.3389/fpls.2018.01238.

(18) Barrios-González, J. Secondary Metabolites Production. *Current Developments in Biotechnology and Bioengineering* **2018**, 257–283. https://doi.org/10.1016/b978-0-444-63990-5.00013-x.

(19) Fouillaud, M.; Dufossé, L. Microbial Secondary Metabolism and Biotechnology. *Microorganisms* **2022**, *10* (1), 1–9. https://doi.org/10.3390/microorganisms10010123.

(20) Craney, A.; Ahmed, S.; Nodwell, J. Towards a New Science of Secondary Metabolism. *Journal of Antibiotics* **2013**, *66* (7), 387–400. https://doi.org/10.1038/ja.2013.25.

(21) Rigali, S.; Anderssen, S.; Naômé, A.; van Wezel, G. P. Cracking the Regulatory Code of Biosynthetic Gene Clusters as a Strategy for Natural Product Discovery. *Biochemical Pharmacology*. 2018. https://doi.org/10.1016/j.bcp.2018.01.007.

(22) Bemis, K. D.; Harry, A.; Eberlin, L. S.; Ferreira, C.; Van De Ven, S. M.; Mallick, P.; Stolowitz, M.; Vitek, O. Cardinal: An R Package for Statistical Analysis of Mass Spectrometry-Based Imaging Experiments. *Bioinformatics* **2015**, *31* (14), 2418–2420. https://doi.org/10.1093/bioinformatics/btv146.

(23) Race, A. M.; Palmer, A. D.; Dexter, A.; Steven, R. T.; Styles, I. B.; Bunch, J. SpectralAnalysis: Software for the Masses. *Anal Chem* **2016**, *88* (19), 9451–9458. https://doi.org/10.1021/acs.analchem.6b01643.

(24) Janda, M.; Seah, B. K. B.; Jakob, D.; Beckmann, J.; Geier, B.; Liebeke, M. Determination of Abundant Metabolite Matrix Adducts Illuminates the Dark Metabolome of MALDI-Mass Spectrometry Imaging Datasets. *Anal Chem* **2021**, *93* (24), 8399–8407. https://doi.org/10.1021/acs.analchem.0c04720.

(25) Tiquet, M.; La Rocca, R.; Kirnbauer, S.; Zoratto, S.; Van Kruining, D.; Quinton, L.; Eppe, G.; Martinez-Martinez, P.; Marchetti-Deschmann, M.; De Pauw, E.; Far, J. FT-ICR Mass Spectrometry Imaging at Extreme Mass Resolving Power Using a Dynamically Harmonized ICR Cell with 1ω or 2ω Detection. *Anal Chem* **2022**, *94* (26), 9316–9326. https://doi.org/10.1021/acs.analchem.2c00754.

(26) Schramm, T.; Hester, A.; Klinkert, I.; Both, J. P.; Heeren, R. M. A.; Brunelle, A.; Laprévote, O.; Desbenoit, N.; Robbe, M. F.; Stoeckli, M.; Spengler, B.; Römpp, A. ImzML - A Common Data Format for the Flexible Exchange and Processing of Mass Spectrometry Imaging Data. *J Proteomics* **2012**, *75* (16), 5106–5110. https://doi.org/10.1016/j.jprot.2012.07.026.

(27) Rocca, R. La; Kune, C.; Tiquet, M.; Stuart, L.; Eppe, G.; Alexandrov, T.; de Pauw, E.; Quinton, L. Adaptive Pixel Mass Recalibration for Mass Spectrometry Imaging Based on Locally Endogenous Biological Signals. *Anal Chem* **2021**, *93* (8), 4066–4074. https://doi.org/10.1021/acs.analchem.0c05071.

(28) Xi, Y.; Sohn, A. L.; Joignant, A. N.; Cologna, S. M.; Prentice, B. M.; Muddiman, D. C. SMART: A Data Reporting Standard for Mass Spectrometry Imaging. *Journal of Mass Spectrometry* **2023**, *58* (2), 1–5. https://doi.org/10.1002/jms.4904.

(29) Kune, C.; McCann, A.; Raphaël, L. R.; Arias, A. A.; Tiquet, M.; Van Kruining, D.; Martinez, P. M.; Ongena, M.; Eppe, G.; Quinton, L.; Far, J.; De Pauw, E. Rapid Visualization of Chemically Related Compounds Using Kendrick Mass Defect As a Filter in Mass Spectrometry Imaging. *Anal Chem* **2019**, *91* (20), 13112–13118. https://doi.org/10.1021/acs.analchem.9b03333.

(30) Hustin, J.; Kune, C.; Far, J.; Eppe, G.; Debois, D.; Quinton, L.; De Pauw, E. Differential Kendrick's Plots as an Innovative Tool for Lipidomics in Complex Samples: Comparison of Liquid Chromatography and Infusion-Based Methods to Sample Differential Study. *J Am Soc Mass Spectrom* **2022**, *33* (12), 2273–2282. https://doi.org/10.1021/jasms.2c00232.

(31) Müller, W. H.; Verdin, A.; Kune, C.; Far, J.; De Pauw, E.; Malherbe, C.; Eppe, G. Dual-Polarity SALDI FT-ICR MS Imaging and Kendrick Mass Defect Data Filtering for Lipid Analysis. *Anal Bioanal Chem* **2021**, *413* (10), 2821–2830. https://doi.org/10.1007/s00216-020-03020-w.

(32) McCann, A.; Kune, C.; La Rocca, R.; Oetjen, J.; Arias, A. A.; Ongena, M.; Far, J.; Eppe, G.; Quinton, L.; De Pauw, E. Rapid Visualization of Lipopeptides and Potential Bioactive Groups of Compounds by Combining Ion Mobility and MALDI Imaging Mass Spectrometry. *Drug Discov Today Technol* **2021**, *39*, 81–88. https://doi.org/10.1016/j.ddtec.2021.08.003.

(33) Palmer, A.; Phapale, P.; Chernyavsky, I.; Lavigne, R.; Fay, D.; Tarasov, A.; Kovalev, V.; Fuchser, J.; Nikolenko, S.; Pineau, C.; Becker, M.; Alexandrov, T. FDR-Controlled Metabolite Annotation for High-Resolution Imaging Mass Spectrometry. *Nat Methods* **2016**, *14* (1), 57–60. https://doi.org/10.1038/nmeth.4072.

(34) Edelsbrunner, H.; Harer, J. Persistent Homology—a Survey. **2008**, No. January, 257–282. https://doi.org/10.1090/conm/453/08802.

(35) Moumbock, A. F. A.; Gao, M.; Qaseem, A.; Li, J.; Kirchner, P. A.; Ndingkokhar, B.; Bekono, B. D.; Simoben, C. V.; Babiaka, S. B.; Malange, Y. I.; Sauter, F.; Zierep, P.; Ntie-Kang, F.; Günther, S. StreptomeDB 3.0: An Updated Compendium of Streptomycetes Natural Products. *Nucleic Acids Res* **2021**, *49* (D1), D600–D604. https://doi.org/10.1093/nar/gkaa868.

(36) Guo, A. C.; Jewison, T.; Wilson, M.; Liu, Y.; Knox, C.; Djoumbou, Y.; Lo, P.; Mandal, R.; Krishnamurthy, R.; Wishart, D. S. ECMDB: The E. Coli Metabolome Database. *Nucleic Acids Res* **2013**, *41* (D1), 625–630. https://doi.org/10.1093/nar/gks992.

(37) Sajed, T.; Marcu, A.; Ramirez, M.; Pon, A.; Guo, A. C.; Knox, C.; Wilson, M.; Grant, J. R.; Djoumbou, Y.; Wishart, D. S. ECMDB 2.0: A Richer Resource for Understanding the Biochemistry of E. Coli. *Nucleic Acids Res* **2016**, *44* (D1), D495–D501. https://doi.org/10.1093/nar/gkv1060.

(38) Vandenbosch, M.; Mutuku, S. M.; Mantas, M. J. Q.; Patterson, N. H.; Hallmark, T.; Claesen, M.; Heeren, R. M. A.; Hatcher, N. G.; Verbeeck, N.; Ekroos, K.; Ellis, S. R. Toward Omics-Scale Quantitative Mass Spectrometry Imaging of Lipids in Brain Tissue Using a Multiclass Internal Standard Mixture. *Anal Chem* **2023**, *95* (51), 18719–18730. https://doi.org/10.1021/acs.analchem.3c02724.

(39) Han, Y.; Qu, X.; Geng, H.; Wang, L.; Zhu, Z.; Zhang, Y.; Cui, X.; Lu, H.; Wang, X.; Chen, P.; Wang, Q.; Sun, C. Isotope-Coded On-Tissue Derivatization for Quantitative Mass Spectrometry Imaging of Short-Chain Fatty Acids in Biological Tissues. *Anal Chem* **2023**, *95* (48), 17622–17628. https://doi.org/10.1021/acs.analchem.3c03308.

(40) Holbrook, J. H.; Kemper, G. E.; Hummon, A. B. Quantitative Mass Spectrometry Imaging: Therapeutics & Biomolecules. *Chemical Communications* **2024**, 2137–2151. https://doi.org/10.1039/d3cc05988j.

(41) Nagano, E.; Odake, K.; Shimma, S. An Alternative Method for Quantitative Mass Spectrometry Imaging (q-MSI) of Dopamine Utilizing Fragments of Animal Tissue. *Mass Spectrometry* **2023**, *12* (1). https://doi.org/10.5702/massspectrometry.A0128.

(42) Unsihuay, D.; Mesa Sanchez, D.; Laskin, J. Quantitative Mass Spectrometry Imaging of Biological Systems. *Annu Rev Phys Chem* **2020**, *72*, 307–329. https://doi.org/10.1146/annurev-physchem-061020-053416.

(43) Lamont, L.; Hadavi, D.; Viehmann, B.; Flinders, B.; Heeren, R. M. A.; Vreeken, R. J.; Porta Siegel, T. Quantitative Mass Spectrometry Imaging of Drugs and Metabolites: A Multiplatform Comparison. *Anal Bioanal Chem* **2021**, *413* (10), 2779–2791. https://doi.org/10.1007/s00216-021-03210-0.

(44) Eriksson, J. O.; Sánchez Brotons, A.; Rezeli, M.; Suits, F.; Markó-Varga, G.; Horvatovich, P. MSIWarp: A General Approach to Mass Alignment in Mass Spectrometry Imaging. *Anal Chem* **2020**, *92* (24), 16138–16148. https://doi.org/10.1021/acs.analchem.0c03833.

(45) Jourdan, S.; Francis, I. M.; Kim, M. J.; Salazar, J. J. C.; Planckaert, S.; Frère, J. M.; Matagne, A.; Kerff, F.; Devreese, B.; Loria, R.; Rigali, S. The CebE/MsiK Transporter Is a Doorway to the Cello-Oligosaccharide-Mediated Induction of Streptomyces Scabies Pathogenicity. *Sci Rep* **2016**, *6* (May), 1–12. https://doi.org/10.1038/srep27144.

(46) Francis, I. M.; Bergin, D.; Deflandre, B.; Gupta, S.; Salazar, J. J. C.; Villagrana, R.; Stulanovic, N.; Ribeiro Monteiro, S.; Kerff, F.; Loria, R.; Rigali, S. Role of Alternative Elicitor Transporters in the Onset of Plant Host Colonization by Streptomyces Scabiei 87-22. *Biology (Basel)* **2023**, *12* (2). https://doi.org/10.3390/biology12020234.

(47) Deflandre, B.; Stulanovic, N.; Planckaert, S.; Anderssen, S.; Bonometti, B.; Karim, L.; Coppieters, W.; Devreese, B.; Rigali, S. The Virulome of Streptomyces Scabiei in Response to Cello-Oligosaccharide Elicitors. *Microb Genom* **2022**, *8* (1). https://doi.org/10.1099/mgen.0.000760.

(48) Pecoraro, L.; Wang, X.; Shah, D.; Song, X.; Kumar, V.; Shakoor, A.; Tripathi, K.; Ramteke, P. W.; Rani, R. Biosynthesis Pathways, Transport Mechanisms and Biotechnological Applications of Fungal Siderophores. *Journal of Fungi* **2022**, *8* (1). https://doi.org/10.3390/jof8010021.

(49) Van Santen, J. A.; Jacob, G.; Singh, A. L.; Aniebok, V.; Balunas, M. J.; Bunsko, D.; Neto, F. C.; Castaño-Espriu, L.; Chang, C.; Clark, T. N.; Cleary Little, J. L.; Delgadillo, D. A.; Dorrestein, P. C.; Duncan, K. R.; Egan, J. M.; Galey, M. M.; Haeckl, F. P. J.; Hua, A.; Hughes, A. H.; Iskakova, D.; Khadilkar, A.; Lee, J. H.; Lee, S.; Legrow, N.; Liu, D. Y.; Macho, J. M.; McCaughey, C. S.; Medema, M. H.; Neupane, R. P.; O'Donnell, T. J.; Paula, J. S.; Sanchez, L. M.; Shaikh, A. F.; Soldatou, S.; Terlouw, B. R.; Tran, T. A.; Valentine, M.; Van Der Hooft, J. J. J.; Vo, D. A.; Wang, M.; Wilson, D.; Zink, K. E.; Linington, R. G. The Natural Products Atlas: An Open Access Knowledge Base for Microbial Natural Products Discovery. *ACS Cent Sci* **2019**, *5* (11), 1824–1833. https://doi.org/10.1021/acscentsci.9b00806.

For Table of Contents Use Only

Exploiting Differential Signal Filtering (DSF) and Image Structure Filtering (ISF) methods for untargeted Mass Spectrometry Imaging of bacterial metabolites.

Pierre Burguet[1], Raphaël La Rocca[1], Christopher Kune[1], Déborah Tellatin[2], Nudzejma Stulanovic[2], Augustin Rigolet[3], Johann Far[1], Marc Ongena[3], Sébastien Rigali[2], Loïc Quinton[1,*]

[1] Mass Spectrometry Laboratory, MolSys Research Unit, University of Liège, 4000 Liège, Belgium.

[2] InBioS – Center for Protein Engineering, University of Liège, 4000 Liège, Belgium.

[3] Gembloux Agro-Bio Tech, University of Liege, 5030 Gembloux, Belgium.

The graphic illustrates the process of filtering mass spectrometry imaging (MSI) data using differential signal filtering (DSF) and image structure filtering (ISF) methods to find unique features in a dataset. The left panel shows the full spectrum of m/z (mass-to-charge ratio) values with numerous peaks. By applying these filtering methods, we obtain a filtered spectrum, indicating the retention of specific m/z values. Finally, the red arrow highlight the identification of unique features specific to bacterial interaction.