Title

Evaluation of Data Analysis Platforms and Compatibility with MALDI-TOF Imaging Mass Spectrometry Data Sets

Journal

Journal of The American Society for Mass Spectrometry, 31(6)

ISSN

1044-0305

Authors

Luu, Gordon T
Condren, Alanna R
Kahl, Lisa Juliane
et al.

Publication Date

2020-06-03

DOI

10.1021/jasms.0c00039

Peer reviewed

Published in final edited form as:

J Am Soc Mass Spectrom. 2020 June 03; 31(6): 1313–1320. doi:10.1021/jasms.0c00039.

Evaluation of data analysis platforms and compatibility with MALDI-TOF imaging mass spectrometry data sets

Gordon T. Luu[1,¤], **Alanna R. Condren**[1,¤], **Lisa Juliane Kahl**[2], **Lars E.P. Dietrich**[2], **Laura M. Sanchez**[1,*]

[1]Department of Pharmaceutical Sciences, University of Illinois at Chicago, Chicago, IL 60612

[2]Department of Biological Sciences, Columbia University, New York, NY 10027

Abstract

Imaging mass spectrometry (IMS) has proven to be a useful tool when investigating the spatial distributions of metabolites and proteins in a biological system. One of the biggest advantages of IMS is the ability to maintain the 3D chemical composition of a sample and analyze in a label free manner. However, acquiring the spatial information leads to an increase in data size. Due to the increased availability of commercial mass spectrometers capable of IMS, there has been an exciting development of different statistical tools that can help decipher the spatial relevance of an analyte in a biological sample. To address this need, software packages like SCiLS and the open source R package Cardinal have been designed to perform unbiased spectral grouping based on the similarity of spectra in an IMS data set. In this note we evaluate SCiLS and Cardinal compatibility with MALDI-TOF IMS data sets of the Gram-negative pathogen *Pseudomonas aeruginosa* PA14. Both software were able to perform unsupervised segmentation with similar performance. There were a few notable differences which are discussed related to the identification of statistically significant features which required optimization of preprocessing steps, region of interest, and manual analysis.

Graphical Abstract

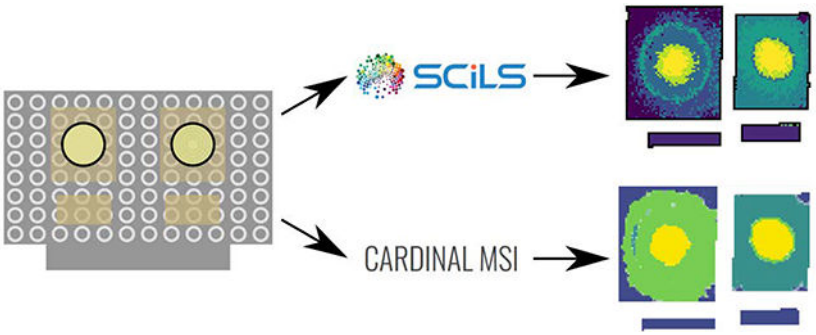
SCiLS and Cardinal MSI workflow comparison for MALDI-TOF IMS data sets.

*Corresponding Author, sanchelm@uic.edu.
¤Denotes equal contribution

Associated Content

The Supporting Information is available free of charge at https://pubs.acs.org/

Complete list of SCiLS and Cardinal outputs for each data set can be found in Supplemental File 1. Organized by the date data was collected, includes *m*/*z*'s of statistically significant ($p<0.05$) signals and their respective mass error, and classification of signal localization and regulation.



Keywords

imaging mass spectrometry; data processing; data analysis; SCiLS; Cardinal; MALDI-TOF

Introduction

The increase of information gained via the ability to visualize the spatial distribution of metabolomic and proteomic systems using imaging mass spectrometry (IMS) has led to a surge of publications incorporating this valuable analytical tool in targeted and untargeted biological studies.[1] The field of natural products has eagerly incorporated IMS, however, these biologically active compounds can elude researchers since they are typically produced in low titers, thus requiring high sensitivity for detection. There are several forms of ionization techniques that support the acquisition of spatial information from a biological sample with one of the most widely utilized techniques being matrix assisted laser desorption/ionization time-of-flight imaging mass spectrometry (MALDI-TOF-IMS).[2] Regardless of ionization method for IMS analysis, biological and technical replicates can experience experimental variation due to differences in co-crystallization with matrix and varying laboratory conditions such as humidity, temperature, and airflow during sample preparation and/or data acquisition.[3, 4] Therefore, there is a need for powerful, easy-to-use software that is able to analyze IMS data to (1) identify statistically significant low intensity features and (2) determine significant differences between treatments or conditions.

Currently, there are several commercial and open-source software packages available to analyze IMS data written in a variety of languages in online or offline formats, including but not limited to MSiReader, msIQuant, SpectralAnalysis, flexImaging, SCiLS, and Cardinal. [2, 5–11] Additionally, there are many in-house algorithms used that have not been formally adapted into a program. For example, consensus spectra calculated using unbiased peak selection and QT clustering were used to identify several significant differentially expressed gangliosides (GM1, GM2, and GM3) in 7-week-old wild-type $Npc1^{+/+}$ and mutant $Npc^{-/-}$ mice.[12]

Two programs in particular have been specifically designed to perform spatial segmentation on IMS data sets. The first is SCiLS, a commercial software available from Bruker Daltonics Inc. that supports vendor-neutral data analysis of IMS data sets. SCiLS is capable of multiple sample comparative analyses including designation of p-values to significant

features, generation of principal component analyses or box plots identifying variations in features from imaging datasets, and visualization of samples in 2D or 3D (supervised analysis). Focusing on the spatial component of IMS data sets, SCiLS can mine the data set via spatial segmentation (unsupervised analysis) which visually maps out the location of similar or different spectra across a biological specimen.[13, 14] SCiLS has proven to be a reliable resource to rapidly process and analyze IMS data sets, however, adoption of SCiLS into data analysis workflows can be a significant financial investment and the algorithms performed during data analysis can be opaque due to the proprietary nature of the software.

Cardinal is an open source R package that has been developed as a preprocessing and statistical analysis pipeline with a focus on image segmentation (unsupervised analysis) and image classification (supervised analysis).[5] Since its release, Cardinal has made substantial improvements, including the implementation of new segmentation algorithms, and has been increasingly incorporated into analysis of IMS data (Table S1). While users are expected to be able to code in the R programming language, access to documentation and support from the developer/community reduces the barrier to implementation compared to low-level programming languages such as C/C++.

Since both software packages have proven to successfully perform robust spatial segmentation on biologically diverse datasets, we were interested in comparing the two packages using the same IMS data sets. SCiLS was exclusively designed for TOF and FT-ICR IMS data sets whereas Cardinal's open format structure has allowed for it to be used to analyze IMS data sets from a variety of mass spectrometers. Table 1 contains a non-exhaustive list of each program's advantages. Here we investigate the similarities and differences of segmentation performed by SCiLS and Cardinal and their ability to identify statistically significant features in MALDI-TOF IMS data sets of the Gram-negative pathogen *Pseudomonas aeruginosa* PA14. We also evaluate the ability of these workflows to segment spectra within light or dark-grown PA14 biofilms due to variation in light exposure with the goal of visualizing changes in metabolite production between light and dark growth conditions.[15]

Experimental Section

PA14 imaging mass spectrometry experiments

Bacterial culture conditions, growth, and preparation for IMS experiments was performed as previously reported with the exception of one dataset that was run with a raster size of 200 μm rather than 500 μm.[16]

Imaging mass spectrometry on PA14 Δphz colonies grown in the light or dark

PA14 Δ*phz* was grown in LB media for 13 h at 37 °C, shaking at 250 rpm.[17] Cultures were then diluted into fresh Lysogeny broth (LB) medium (1:100) and subcultures were grown for 2.5 h at 37 °C until mid-exponential growth was reached (OD of ~0.4–0.6 at 500 nm). Colony agar was composed of 1% tryptone, 1.25% agar mixture (Teknova) and 90 mL of autoclaved and cooled agar were poured per 120 mm×120 mm×17 mm plate (Greiner Bio-One). Two microliters of subculture were spotted. Plates were stored at 25 °C and at

high humidity (+90%) in a Percival CU-22LC9 incubation chamber with pre-installed lighting system for four days. Over the 4 d, the colonies were exposed to white light (light tubes: Philips F17T8/TL841/ALTO, calibrated to 95 μmol photons m^{-2} s^{-1}) or kept in the same chamber in constant darkness. After 4 d, images of colony biofilms were taken using an iPod attached to the ocular of a Zeiss AxioZoom.V16 Fluorescence zoom Stereomicroscope. Colony biofilms were shipped overnight in constant darkness for IMS analysis and prepared for IMS analysis following the same procedure used for PA14 WT experiments described.[16] Data was collected on a MALDI-TOF mass spectrometer (Bruker Autoflex LRF Speed) in positive reflectron mode and the following settings were used for IMS data acquisition: detector gain: 15.4x; laser power: 73%; range of detection: 100 Da to 3,500 Da; ion suppression: 50 Da; raster size: 300 μm; shots per raster: 200; laser size: large (4).

Data Processing and Unsupervised Segmentation in SciLS

To manually analyze imaging data sets, FlexImaging v.4.1 (Bruker) software was used with no post processing steps. Reduced spectra (.dat file) were imported into SCiLS software (Bruker, version 2015b) with no baseline removal for statistical analysis. Segmentation was performed using the following settings; normalization: RMS, Denoising: weak, and minimal interval width: ∓0.2 Da. SCiLS segmentation map was separated into seven groups of similar spectra as seen in Figure 1. The same settings were used for segmentation of the PA14 light vs dark colonies with the exception of using TIC for normalization.

Data Export to .imzML Format

All acquired IMS data was exported from FlexImaging v4.1 (Bruker) using its built in export function. No background subtraction, binning, or processing was performed before export (“Export spectra from disk”).

Data Preprocessing and Multivariate Unsupervised Segmentation in Cardinal

Each data set was imported into Cardinal 2.4.0.9000. Normalization, signal smoothing, and baseline correction were applied to the individual data sets followed by peak picking, alignment, and filtering. All import and processing settings can be found in Table S2. Multivariate unsupervised segmentation was performed using the spatial shrunken centroid (SSC) algorithm.[18] The spatial neighborhood radius (r) was set to 1, 2, or 3. The number starting of segments (k) was set between 2–20 in increments of 2. The shrinkage parameter (s) was set between 0–15 in increments of 3. SSC was run with every permutation of these three parameters. To find the optimal set of parameters for r, k, and s in each dataset (Table S3), a heuristic algorithm based on the criteria described by Bemis *et al.* was written (Supplemental Information).[18] All other *spatialShrunkenCentroids* parameters were left at their default values.

Identifying Statistically Significant Features in SciLS

To identify significant features in each of the three PA14 IMS datasets the “Find discriminative plots” feature was used in SCiLS as detailed in Condren *et al.*[16]

Identifying Statistically Significant Features in Cardinal

All biological replicate data sets, excluding the partial sample replicate, were imported into Cardinal 2.4.0. The metadata in each data set was updated to include region of interest (ROI) labels and treatment condition information using an in-house R script utilizing a Spot List exported from FlexImaging v4.1 (Bruker). All four data sets were combined into a single large data set using Cardinal's *combine* function. The combined data set was processed using the same workflow as described above with all settings found in Table S2. The *meansTest* function used means based testing to identify features that were differentially expressed between conditions by comparing regions of interest (ROIs) belonging to different treatment conditions to all other ROIs in the combined data set. The *spatialDGMM* function followed by the *segmentationTest* function were used to apply univariate unsupervised segmentation to each individual feature to determine whether they were differentially expressed between conditions by comparing ROIs belonging to different treatment conditions to all other ROIs in the combined data set. Ion images of select features were generated in Cardinal using a false color scheme with "linear" normalization, "Gaussian" image smoothing, and contrast enhancement using histograms applied.

Code Availability

All code for the analysis performed and figures generated using Cardinal as well as details of our branch of Cardinal can be found at https://github.com/gtluu/ims_processing_evaluation. Several in-house functions used during data processing/analysis can be found and installed from https://github.com/gtluu/cardinalscripts. Cardinal 2.4.0.9000, our branch of Cardinal based on Cardinal 2.4.0 containing minor changes, was only used for multivariate segmentation and can be found at https://github.com/gtluu/Cardinal/tree/v2.4.0.9000.

Data Accessibility

Massive accession for raw MALDI-TOF MS data: MSV000084061

Results and Discussion

Spatial Segmentation of *P. aeruginosa* PA14 IMS Experiments

Determining where to begin when analyzing IMS data sets can be difficult due to their size and untargeted nature. Users can manually examine the computationally averaged spectrum but this takes a significant amount of time and the analysis is subjective. By utilizing software programs such as SCiLS and Cardinal, unsupervised spatial segmentation maps can be generated to separate the thousands of spectra composing an imaging data set into spectral groups based on feature similarity. This allows the user to rapidly visualize differences in their sample and determine if a group of spectra are localized in a biologically relevant location. To compare segmentation maps between the two software packages, we used the same IMS data sets of *Pseudomonas aeruginosa* PA14 treated with the biofilm inhibiting agent taurolithocholic acid (TLCA; Figure 1).[16] In SCiLS, each spectral group is assigned a false color and the user is able to further divide spectral groups to distinguish distinct features (Figure 1B). While maintaining the spatial integrity of the IMS data set,

SCiLS was able to separate spectra into seven spectral groups: six bacterial associated groups (3,693 spectra) and one agar control associated group (1,121 spectra). The segmentation also showed distinct localization of spectra correlated to biologically relevant spaces such as the interior of the bacterial colony (yellow), the proliferating outer edge (teal), and features secreted into the agar (dark blue).

While SCiLS performs a single segmentation and allows the user to interactively determine how many segments are present in each data set via a “segmentation tree”, Cardinal’s implementation of the SSC algorithm takes an alternative approach. The user initially specifies one or multiple sets of parameters and segmentation models for each set of user-defined parameters are generated. Since a range of parameters must be chosen beforehand, it is necessary to be able to determine the optimal set of parameters for generating a segmentation model. Thus, a new algorithm was written to determine the optimal values for r, k, and s in each data set from a given set of parameters (Supplemental Information). Figure 1C shows segmentation performed on the same data set as Figure 1B using the optimal parameters identified by this algorithm. SSC was able to segment spectra into ten spectral groups: eight bacteria associated groups (3,510 spectra) and two agar control associated groups (1,304 spectra). Similarly to SCiLS, the interior of the bacterial colony (yellow), proliferating outer edge (yellow-green/green), and features secreted into the agar (teal/blue/dark blue) can be distinguished.

The same workflow in SCiLS and Cardinal was applied to four other biological replicates. Figure S1 & Figure S2 show two more replicates were imaged using the same raster size of 500 μm across a full sample. Interestingly, the replicate in Figure S1C had many more segments using SSC than our first replicate (Figure 1C). The agar segments in our third replicate appeared to be compromised and resulted in less informative segmentation, most likely due to the inclusion of a fifth matrix ROI and/or possible artifacts as a result of sample preparation that confounded the segmentation algorithms in SCiLS and Cardinal (Figure S2B & S2C). Figure S3 shows a fourth replicate imaged at a higher spatial resolution (raster size: 200 μm) that showed a similar segmentation pattern to our first replicate, though with a higher number of spectral groups in both programs. Figure S4 shows the last replicate in which only half of the PA14 control colony was imaged. Again, a similar segmentation pattern was seen from this dataset.

Our analyses show that both SCiLS and Cardinal’s SSC algorithms are able to segment different data sets with varying spatial resolutions and experimental design in a similar capacity. Sample preparation, and to a lesser degree experimental design, are the most crucial factors affecting downstream analysis. It should be noted that while a higher spatial resolution did not make any major differences in segmentation using this data set, it is more computationally expensive to analyze. Therefore, available computational resources should also be taken into consideration when designing an IMS experiment.

Identification of Significant Features due to TLCA treatment of *P. aeruginosa* PA14

Unsupervised segmentation of spectra in an IMS data set is valuable when a researcher is interested in the spatial component of particular features, however, segmentation cannot differentiate changes in intensity of a specific feature between two conditions. For example,

identifying significant differences in metabolite production when *P. aeruginosa* was treated with biofilm inhibitor TLCA could provide insight into how the pathogen responds to exogenous stimuli. Thus, we sought to examine the ability of SCiLS and Cardinal to identify significant features that were either up- or down-regulated from TLCA treatment. A more in-depth description of algorithms used in both programs can be found in the Supplemental Information.

In our PA14 IMS experiments, when SCiLS or Cardinal identified a feature that had a higher intensity in the PA14 TLCA treated colony, we refer to that feature as up-regulated. When a feature was determined to have a higher intensity in the vehicle-treated PA14 control, we refer to that feature as down-regulated since the TLCA treated colony had a lower intensity of the feature than the PA14 control (Supplemental File 1). To directly compare the two software packages, in Figure 2 we highlighted three features which correspond to specialized metabolites produced by *P. aeruginosa*: the phenazines pyocyanin (*m/z* 211) and phenazine-1,6-dicarboxylic acid (*m/z* 269), and an uncharacterized metabolite (*m/z* 609). These metabolites represent examples of up or down-regulated features and displayed various spatial distributions in the bacterial colonies.

SCiLS identified *m/z* 211 as down-regulated from TLCA treatment whereas *m/z* 269 and *m/z* 609 were up-regulated (Figure 2B; $p<0.05$). Thus, SCiLS provides statistical support to the argument that TLCA treatment is leading to a downregulation of the *P. aeruginosa* toxin, pyocyanin, which is produced during biofilm formation, colonization, and infection.[19–21] It is important to note that among the list of significantly altered features found between the ROIs, SCiLS determined *m/z* 211 and *m/z* 609 to be significantly altered features, but not *m/z* 269. When re-running the analysis and instructing SCiLS to only compare the pixels specific to the bacterial colonies, SCiLS identified *m/z* 269 as one of several significantly altered features (Supplemental File 1). The *m/z* 269 feature was concentrated within the bacterial colony so by analyzing the entire ROI, the overall intensity of the feature was not accurate, leading to features such as *m/z* 269 not being identified as significant. Therefore, when mining ROIs for discriminate *m/z* values, it is important to take into consideration the spatial distribution of the analyte of interest. SCiLS offers users the ability to define an unlimited number of ROIs post acquisition, allowing for the direct comparison of regions within the imaging data set.

Using Cardinal's *meansTest* function to compare the entire ROIs between conditions, only *m/z* 609 was identified as up-regulated in PA14 treated with TLCA (Supplemental File 1; $p < 0.05$). Interestingly, while SCiLS only identified *m/z 269* when comparing colonies, when re-analyzing the data using *spatialDGMM* followed by *segmentationTest*, *m/z* 269 was identified by Cardinal as being upregulated in PA14 treated with TLCA even when analyzing an entire ROI. Additionally, *segmentationTest* was able to identify many more statistically significant features than *meansTest* in our combined data set (Supplemental File 1). Univariate segmentation seemed to be a much more sensitive method for identifying significant features, though unique features could be found in analyses from either algorithm. Therefore, we recommend users run both algorithms to obtain a more comprehensive list of features. Both algorithms in Cardinal were unable to detect *m/z* 211. Since *meansTest and segmentationTest* usage requires merging multiple replicate data sets,

the *tolerance* parameter in the *peakAlignment* function must be optimized so that variation/drift between runs and sample preparation can be properly accounted for. If the tolerance is too low, peaks corresponding to the same feature will not be grouped, and if the tolerance is too high, peaks corresponding to different features will be grouped together. Users should optimize this parameter to the best of their ability. By default, *peakAlignment* attempts to estimate the tolerance, and in this case, its tolerance estimation (4 Da) was too large, resulting in a noisy ion image at *m/z* 211 (Figure 2C). Cardinal is also capable of analyzing user defined ROIs post acquisition to compare specific sections of the data set, but users must estimate the coordinates of the desired new ROI or define the ROI interactively in FlexImaging, export an updated Spot List, and update the metadata using an R script. Therefore, a certain level of coding knowledge is required to perform more in-depth analyses. The *updateMetadata* function in our in-house R package cardinalscripts attempts to help with defining ROIs and assigning conditions to data sets.

These three features highlight the capabilities and limitations of these two programs to identify differences in regulation of specific *m/z*'s from TLCA treatment and the importance of ROI selection when analyzing IMS data sets. This highlights that some manual analysis is often inevitable and necessary when analyzing IMS data sets even with these software packages.

Spatial Segmentation of Expanding *P. aeruginosa* PA14 Colonies

P. aeruginosa PA14 alters its biofilm morphology when grown in the dark or light.[15] We hypothesized that light exposure may induce changes in specialized metabolism. We performed IMS experiments to visualize potential differences in expanding colonies grown in the light or dark (Figure S5). To test the spatial segmentation limits of these software packages, we used an IMS data set of light- or dark-grown *P. aeruginosa* colonies to determine if SCiLS and Cardinal could segment features within the expanding colonies when only a section of the sample was imaged at a higher spatial resolution unlike our other WT PA14 data sets.

As observed in our previous experiments, SCiLS was able to differentiate spectra that were specific to bacterial metabolites (five groups; 760 spectra) from agar controls (one group; 415 spectra). Interestingly, within the six groups representing spectra correlated to bacterial metabolites, SCiLS successfully grouped spectra that formed patterns along the expanding colony that was exposed to light, but less so in a colony that was grown in the dark (Figure 3B). Through spatial segmentation, SCiLS was capable of visualizing differences in spectral groups between the two conditions and promotes that light exposure may alter specialized metabolite production within PA14 biofilms.

Segmentation in Cardinal was able to detect the colony and secreted metabolites in both colonies, with four groups corresponding to the bacterial colony and secreted metabolites (1,084 spectra) and one group corresponding to the agar (384 spectra; Figure 3C). Here we can see that secreted metabolites comprising the light blue segment seem to be slightly up-regulated when PA14 was grown in the dark, which again indicates a change in specialized metabolite production. Unlike in SCiLS, there appeared to be no significant difference between both conditions. Overall, we found that SCiLS and Cardinal were both able to

segment this data set and identify spatial differences of spectra in each region despite the relatively small size of the region that was imaged. In the end, the most optimal segmentation will depend on the biological question being posed.

Performance of SCiLS and Cardinal

In addition to instrumentation and experimental setup, available computational hardware, cost of said hardware, and cost of software are also important considerations in IMS experiments due to the multi-dimensional nature of the data acquired. We benchmarked how time and resource intensive both programs were during our analyses. All computation was done on a desktop with an AMD Ryzen 1600 3.2GHz processor, 32 GB 3000MHz memory, and an Inland Professional solid-state drive.

In SCiLS, each task took no longer than several minutes to run, with larger data sets such as our 200 μm raster size data set taking longer. We found that the processor and hard drive speeds mainly affected computation time. Our workflow in Cardinal took much longer than SCiLS (several hours to several days) and showed that having an adequate amount of memory was crucial. Most analyses were performed without parallelization due to lack of available memory. Since Cardinal is compatible with Unix systems, users can take advantage of available in-house, institutional, or commercial servers (i.e. Amazon Web Services) to allow for faster data processing/analysis, as servers often provide faster processors and more available memory. In both programs, hard drive speed was also a major factor in analysis time. While traditional mechanical hard drives are better for data storage due to more efficient $/GB, use of solid-state drives was preferred when processing and analyzing data due to its ability to read and write data more quickly. Cardinal uses the matter R package which allows data to be read from the hard drive as needed to reduce the amount of memory used, but consequently, hard drive speed plays a larger role in computational time.[22]

Conclusion

With the steadily growing incorporation of IMS experiments into scientific exploration, comes the need to develop new techniques for manual and statistical analysis of big data. Thus, software like Bruker's SCiLS and the R package Cardinal have been designed to perform spatial segmentation and statistical analysis on IMS datasets. To evaluate the advantages of these two software packages, we tested the segmentation capabilities of SCiLS and Cardinal using IMS data sets with varying spatial resolutions from *P. aeruginosa* PA14 and found that both can successfully segment spectra that correspond to distinct zones within the bacterial colony and surrounding agar, even when only a partial sample was present. Comparing feature intensities between two conditions (no treatment vs TLCA treatment of PA14), SCiLS and Cardinal were both able to identify a list of statistically significant up- or down-regulated features. Some features were complementary as highlighted in Figure 2 and Supplemental File 1. Lastly, we tested the spatial segmentation limits of SCiLS and Cardinal to determine if these programs could spatially differentiate between metabolite distributions within PA14 colony biofilms grown in the light or dark and found that both programs spatially segment differences in specialized metabolite production within a PA14 biofilms even when only analyzing a small portion of the sample. It is

important that normalization plays a large role in the output of these analyses and we found that root mean squared (RMS) works best for large ROI's (Figure 1 & 2) and total ion count (TIC) is best for smaller ROI's (Figure 3). As summarized in Table 1, we observed several advantages to both programs and would consider a listed advantage for one program to be a current limitation to the other.

We also presented a novel algorithm implemented through an in-house R function to help guide Cardinal users in selecting "optimal" parameters for segmentation maps and demonstrated its utility (Figure 1C, 3C, S1C, S2C, S3C, S4C). While our algorithm was able to assist in the selection of parameters in our data sets, because our parameter optimization algorithm relies on the nature of the "lines" derived from plotting a predicted number of segments for a pair of r and k versus s (Figure S6), it is important to pick appropriate ranges of parameters when initializing r, k, and s for SSC. During analysis using SSC, it is possible that the most optimal parameters for r, k, and s could be absent in the initial vectors, preventing their selection. Even when they are present, the parameters selected may be mathematically optimal but uninformative depending on the question being posed. For example, in our second PA14 replicate, the optimal SSC parameters were even able to show a difference between the agar (negative control) and TLCA treated agar (positive control) (Figure S1C). Detailed segmentation may be useful to identify minute differences between conditions, but it is unnecessary if one is only looking for a general difference between regions of interest. Therefore, manual verification of segmentation maps generated from each set of parameters should still be performed since our algorithm only seeks to provide a recommended set of parameters and not an absolute 'best' set. While Cardinal is unable to interactively change the number of segments in a segmentation image, in-house R scripts can be used to generate a report with images for each segmentation model from SSC.

Overall, we found both software packages to be extremely thorough and efficient in their analyses. SCiLS predates Cardinal, having been available since 2014, and has had continual commercially available updates every six months to incorporate more features and thereby improving analysis and user experience. Meanwhile, Cardinal was released in 2015, and had a major update in 2019, and has even been partially adapted into a Galaxy workflow.[23] To utilize Cardinal, a certain level of coding knowledge in the R language as well as familiarity of open source mass spectrometry data formats is necessary as users may encounter situations that require troubleshooting. Although it may have a steep learning curve, Cardinal has the advantage of being a cost friendly solution for data structures and workflows that can be used to incorporate new algorithms. Cardinal shows much promise in its ability to compete with other IMS packages, both current and future.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We would like to thank Mingxun Wang, Dan Guo, and the Vitek Lab for helpful discussions. Funding was provided by the University of Illinois at Chicago Startup Funds (L.M.S.); American Society for Pharmacognosy research startup grant (L.M.S.); NIH/NIAID grant R01AI103369 and an NSF CAREER award (L.E.P.D) NIH/NCCIH grant

T32-AT007533 (G.T.L.). ARC was supported in part by the National Science Foundation Illinois Louis Stokes Alliance for Minority Participation Bridge to the Doctorate Fellowship (grant number 1500368) and a UIC Abraham Lincoln retention fellowship.

References

1. Doerr A: Mass spectrometry imaging takes off. Nat. Methods. 15, 32–32 (2018)
2. Spraker JE, Luu GT, Sanchez LM: Imaging mass spectrometry for natural products discovery: a review of ionization methods. Nat. Prod. Rep. (2019). 10.1039/c9np00038k
3. Wiangnon K, Cramer R: Sample preparation: a crucial factor for the analytical performance of rationally designed MALDI matrices. Anal. Chem. 87, 1485–1488 (2015) [PubMed: 25597381]
4. Umemura H, Kodera Y, Nomura F: Effects of humidity on the dried-droplet sample preparation for MALDI-TOF MS peptide profiling. Clin. Chim. Acta. 411, 2109–2111 (2010) [PubMed: 20709043]
5. Bemis KD, Harry A, Eberlin LS, Ferreira C, van de Ven SM, Mallick P, Stolowitz M, Vitek O: Cardinal: an R package for statistical analysis of mass spectrometry-based imaging experiments. Bioinformatics. 31, 2418–2420 (2015) [PubMed: 25777525]
6. Race AM, Palmer AD, Dexter A, Steven RT, Styles IB, Bunch J: SpectralAnalysis: Software for the Masses. Anal. Chem. 88, 9451–9458 (2016) [PubMed: 27558772]
7. Källback P, Nilsson A, Shariatgorji M, Andrén PE: msIQuant--Quantitation Software for Mass Spectrometry Imaging Enabling Fast Access, Visualization, and Analysis of Large Data Sets. Anal. Chem. 88, 4346–4353 (2016) [PubMed: 27014927]
8. Bokhart MT, Nazari M, Garrard KP, Muddiman DC: MSiReader v1.0: Evolving Open-Source Mass Spectrometry Imaging Software for Targeted and Untargeted Analyses. J. Am. Soc. Mass Spectrom. 29, 8–16 (2018) [PubMed: 28932998]
9. Robichaud G, Garrard KP, Barry JA, Muddiman DC: MSiReader: an open-source interface to view and analyze high resolving power MS imaging files on Matlab platform. J. Am. Soc. Mass Spectrom. 24, 718–721 (2013) [PubMed: 23536269]
10. Buchberger AR, DeLaney K, Johnson J, Li L: Mass Spectrometry Imaging: A Review of Emerging Advancements and Future Insights. Anal. Chem. 90, 240–265 (2018) [PubMed: 29155564]
11. Eliceiri KW, Berthold MR, Goldberg IG, Ibáñez L, Manjunath BS, Martone ME, Murphy RF, Peng H, Plant AL, Roysam B, Stuurman N, Swedlow JR, Tomancak P, Carpenter AE: Biological imaging software tools. Nat. Methods. 9, 697–710 (2012) [PubMed: 22743775]
12. Tobias F, Olson MT, Cologna SM: Mass spectrometry imaging of lipids: untargeted consensus spectra reveal spatial distributions in Niemann-Pick disease type C1. J. Lipid Res. 59, 2446–2455 (2018) [PubMed: 30266834]
13. Kriegsmann J, Kriegsmann M, Kriegsmann K, Longuespée R, Deininger S-O, Casadonte R: MALDI Imaging for Proteomic Painting of Heterogeneous Tissue Structures. Proteomics Clin. Appl. 13, e1800045 (2019) [PubMed: 30471204]
14. Zink KE, Tarnowski DA, Mandel MJ, Sanchez LM: Optimization of a minimal sample preparation protocol for imaging mass spectrometry of unsectioned juvenile invertebrates. J. Mass Spectrom. (2019). 10.1002/jms.4458
15. Kahl LJ, Price-Whelan A, Dietrich LEP: Light-mediated decreases in cyclic di-GMP levels are potentiated by pyocyanin and inhibit structure formation in Pseudomonas aeruginosa biofilms, bioRxiv (2019)
16. Condren AR, Kahl LJ, Boelter G, Kritikos G, Banzhaf M, Dietrich LE, Sanchez LM: Biofilm inhibitor taurolithocholic acid alters colony morphology, specialized metabolism, and virulence of Pseudomonas aeruginosa. ACS Infect Dis. (2019). 10.1021/acsinfecdis.9b00424
17. Dietrich LEP, Price-Whelan A, Petersen A, Whiteley M, Newman DK: The phenazine pyocyanin is a terminal signalling factor in the quorum sensing network of Pseudomonas aeruginosa. Mol. Microbiol. 61, 1308–1321 (2006) [PubMed: 16879411]
18. Bemis KD, Harry A, Eberlin LS, Ferreira CR, van de Ven SM, Mallick P, Stolowitz M, Vitek O: Probabilistic Segmentation of Mass Spectrometry (MS) Images Helps Select Important Ions and Characterize Confidence in the Resulting Segments. Mol. Cell. Proteomics. 15, 1761–1772 (2016) [PubMed: 26796117]

19. Das T, Kutty SK, Kumar N, Manefield M: Pyocyanin facilitates extracellular DNA binding to Pseudomonas aeruginosa influencing cell surface properties and aggregation. PLoS One. 8, e58299 (2013) [PubMed: 23505483]

20. Ran H, Hassett DJ, Lau GW: Human targets of Pseudomonas aeruginosa pyocyanin. Proc. Natl. Acad. Sci. U. S. A. 100, 14315–14320 (2003) [PubMed: 14605211]

21. Allen L, Dockrell DH, Pattery T, Lee DG, Cornelis P, Hellewell PG, Whyte MKB: Pyocyanin production by Pseudomonas aeruginosa induces neutrophil apoptosis and impairs neutrophil-mediated host defenses in vivo. J. Immunol. 174, 3643–3649 (2005) [PubMed: 15749902]

22. Bemis KA, Vitek O: matter: an R package for rapid prototyping with larger-than-memory datasets on disk. Bioinformatics. 33, 3142–3144 (2017) [PubMed: 28633357]

23. Föll MC, Moritz L, Wollmann T, Stillger MN, Vockert N, Werner M, Bronsert P, Rohr K, Grüning BA, Schilling O: Accessible and reproducible mass spectrometry imaging data analysis in Galaxy. Gigascience. 8, (2019). 10.1093/gigascience/giz143

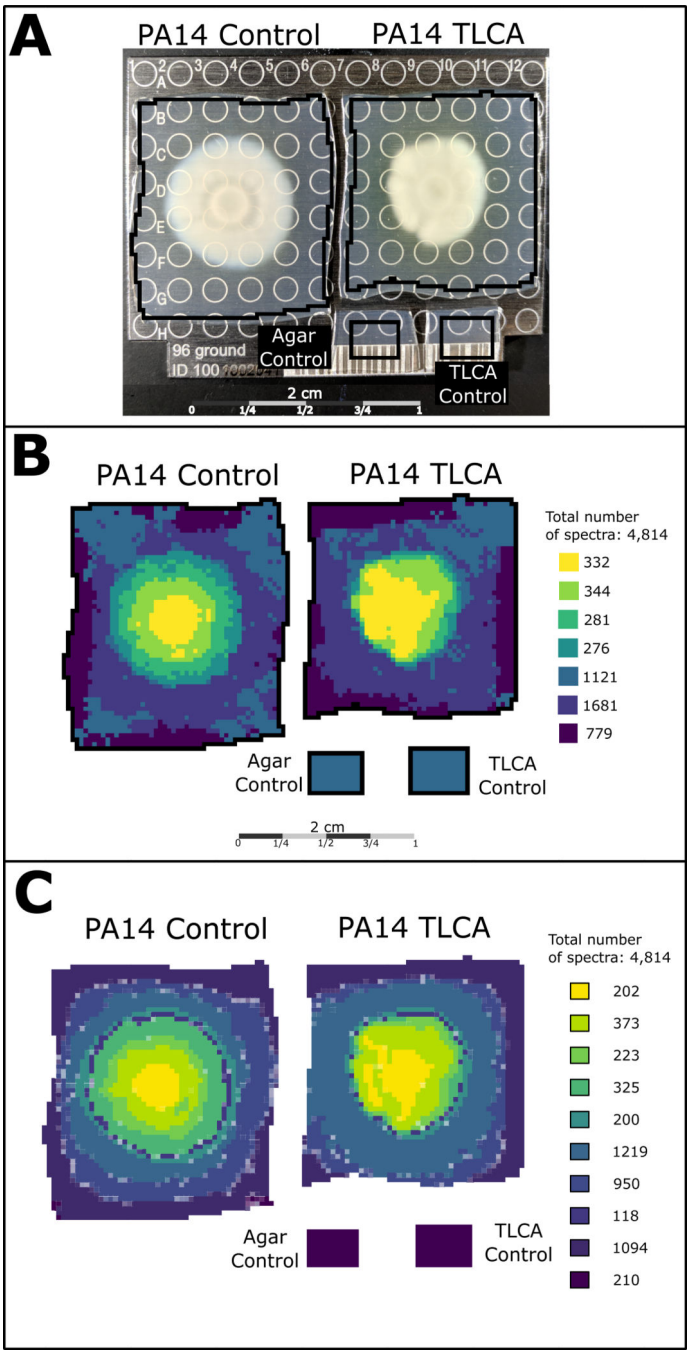


Figure 1:

SCiLS vs Cardinal SpatialShrunkenCentroids Segmentation of WT PA14 Colonies

(A) Optical image of WT PA14 IMS plate. (B) Segmentation in SCiLS. (C) Segmentation in Cardinal using SSC with optimal parameters. Similar segmentation patterns can be seen in the bacterial colonies when using SCiLS and Cardinal. However, features secreted into the agar further away from the colonies are segmented into more circular segments as opposed to the “patchy” segments seen in SCiLS.

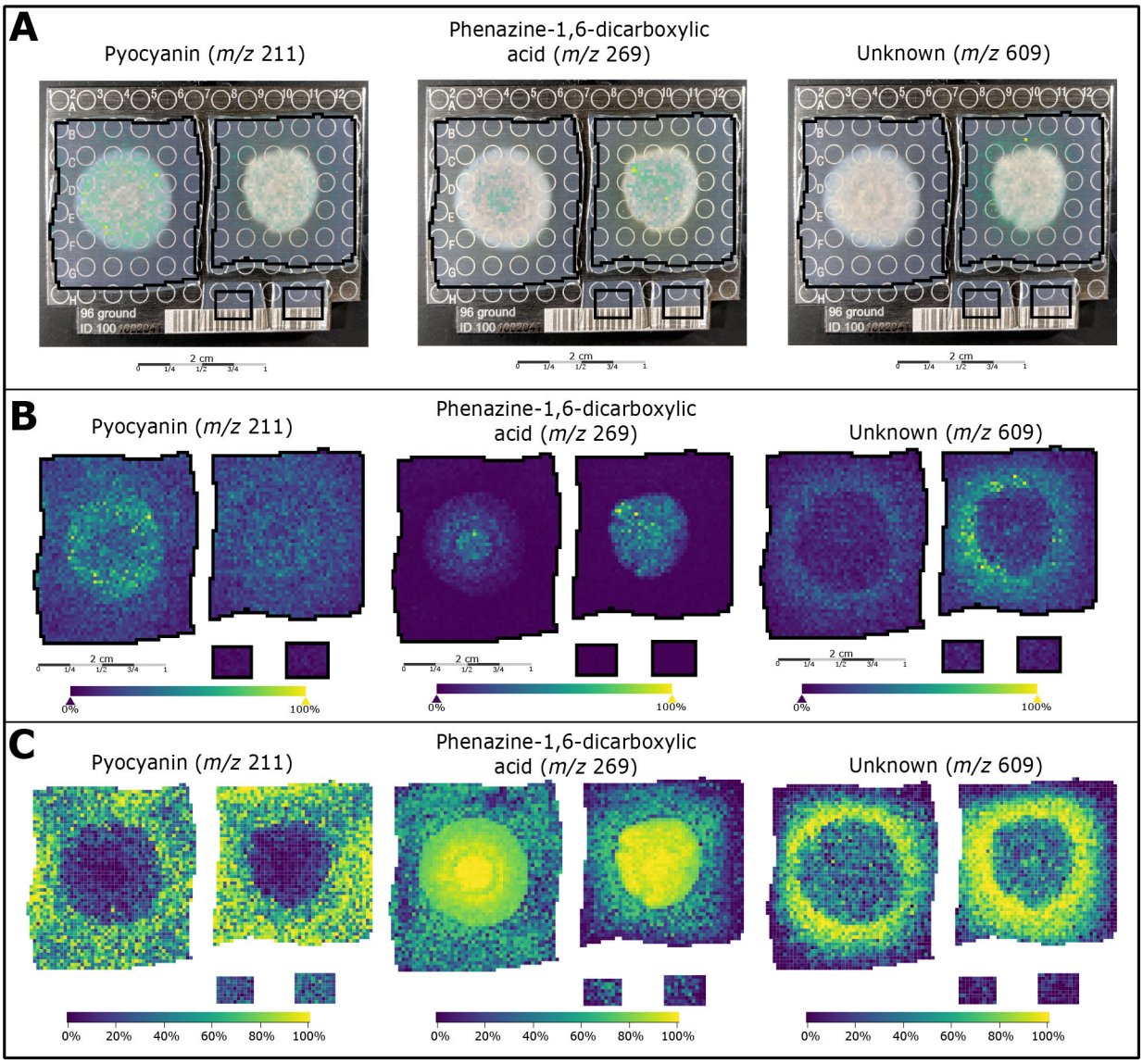


Figure 2:

Significant features identified through SCiLS and Cardinal SpatialShrunkenCentroids (A) Optical images of statically significant features ($p<0.05$). Ion images generated (B) in SCiLS and (C) in Cardinal. Both programs identified pyocyanin and the unknown metabolites as statistically significant between the two growth conditions (untreated vs TLCA-treated), while phenazine-1,6-dicarboxylic acid was only identified as statistically significant when analyzing the colony pixels. It should be noted that in Cardinal, pyocyanin was identified as *m/z* 210.9 and phenazine-1,6-dicarboxylic acid was identified as *m/z* 270.1, with both values being within a ∓0.2 and ∓0.4 Da tolerance, respectively.

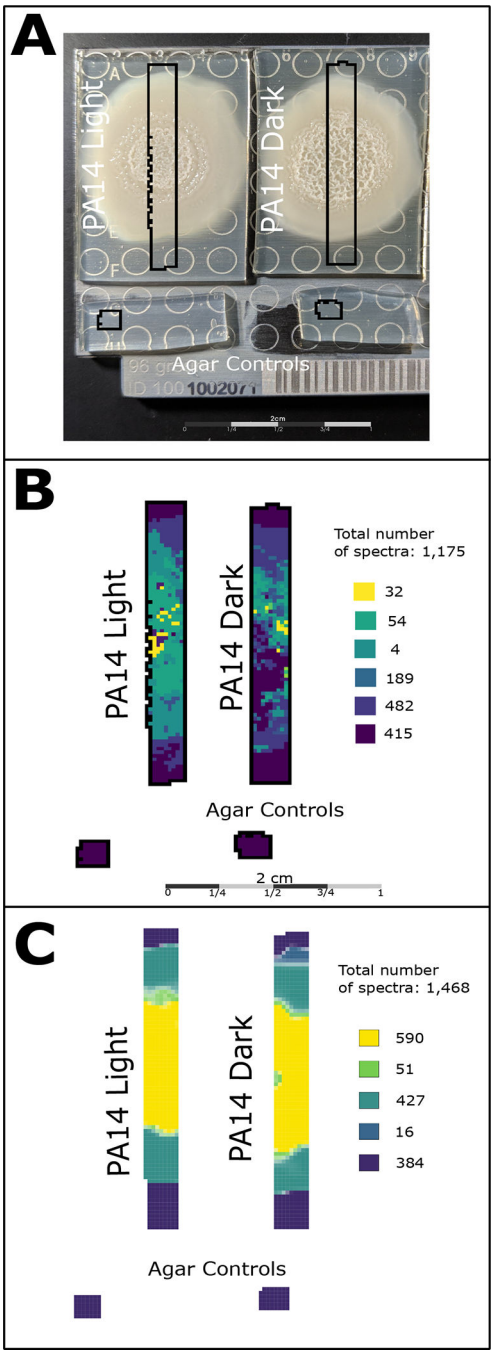


Figure 3:

SCiLS vs Cardinal SpatialShrunkenCentroids Segmentation of PA14 Colonies grown in the light and dark.

(A) Optical image of IMS plate with PA14 colonies that were exposed to light or grown in the dark. Spatial segmentation maps of the PA14 colonies generated in (B) SCiLS and (C) Cardinal using SSC with optimized parameters.

Table 1:

Advantages of SCiLS and Cardinal. The advantage listed by one program should be read as a current limitation of the other program.

SCiLS Advantages	**Cardinal Advantages**
No coding experience needed	Free and open-source software (FOSS) analysis package
Commercial customer support	In-house functions/scripts can be integrated into Cardinal workflows
Analysis can be done on a single biological replicate	Source code can be modified for bug fixes or added functionality
Analysis performed on compressed binary data resulting in reduced file sizes	Well documented functions and workflows
Easily define custom regions of interest without use of other software	Ability to automate analysis queue and generate reports using user written R scripts
Much faster processing time	Data structures using *matter* allow for analysis on lower power hardware
	Compatible with Windows, Mac OS, and Linux