the plant journal

SEB Society for Experimental Biology

The Plant Journal (2018) **93,** 193–206 doi: 10.1111/tpj.13763

TECHNICAL ADVANCE

Secondary ion mass spectrometry imaging and multivariate data analysis reveal co-aggregation patterns of *Populus trichocarpa* leaf surface compounds on a micrometer scale

Purva Kulkarni[1,2], Mina Dost[3], Özgül Demir Bulut[4], Alexander Welle[4], Sebastian Böcker[1], Wilhelm Boland[3] and Aleš Svatoš[2,*]

[1]*Lehrstuhl für Bioinformatik, Friedrich Schiller University, Ernst-Abbe-Platz 2, 07743, Jena,Germany,*

[2]*Research Group Mass Spectrometry, Max Planck Institute for Chemical Ecology, Hans-Knöll-Strasse 8, 07745, Jena, Germany,*

[3]*Department of Bioorganic Chemistry, Max Planck Institute for Chemical Ecology, Hans-Knöll-Strasse 8, 07745, Jena, Germany, and*

[4]*Institute of Functional Interfaces and Karlsruhe Nano Micro Facility, Karlsruhe Institute of Technology (KIT), 76344, Eggenstein-Leopoldshafen, Germany*

Received 23 August 2016; revised 13 October 2017; accepted 23 October 2017; published online 2 November 2017.
*For correspondence (e-mail svatos@ice.mpg.de).

SUMMARY

Spatially resolved analysis of a multitude of compound classes has become feasible with the rapid advancement in mass spectrometry imaging strategies. In this study, we present a protocol that combines high lateral resolution time-of-flight secondary ion mass spectrometry (TOF-SIMS) imaging with a multivariate data analysis (MVA) approach to probe the complex leaf surface chemistry of *Populus trichocarpa*. Here, epicuticular waxes (EWs) found on the adaxial leaf surface of *P. trichocarpa* were blotted on silicon wafers and imaged using TOF-SIMS at 10 μm and 1 μm lateral resolution. Intense $M^{+\bullet}$ and $M^{-\bullet}$ molecular ions were clearly visible, which made it possible to resolve the individual compound classes present in EWs. Series of long-chain aliphatic saturated alcohols (C_{21}–C_{30}), hydrocarbons (C_{25}–C_{33}) and wax esters (WEs; C_{44}–C_{48}) were clearly observed. These data correlated with the ^{7}Li-chelation matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) analysis, which yielded mostly molecular adduct ions of the analyzed compounds. Subsequently, MVA was used to interrogate the TOF-SIMS dataset for identifying hidden patterns on the leaf's surface based on its chemical profile. After the application of principal component analysis (PCA), a small number of principal components (PCs) were found to be sufficient to explain maximum variance in the data. To further confirm the contributions from pure components, a five-factor multivariate curve resolution (MCR) model was applied. Two distinct patterns of small islets, here termed 'crystals', were apparent from the resulting score plots. Based on PCA and MCR results, the crystals were found to be formed by C_{23} or C_{29} alcohols. Other less obvious patterns observed in the PCs revealed that the adaxial leaf surface is coated with a relatively homogenous layer of alcohols, hydrocarbons and WEs. The ultra-high-resolution TOF-SIMS imaging combined with the MVA approach helped to highlight the diverse patterns underlying the leaf's surface. Currently, the methods available to analyze the surface chemistry of waxes in conjunction with the spatial information related to the distribution of compounds are limited. This study uses tools that may provide important biological insights into the composition of the wax layer, how this layer is repaired after mechanical damage or insect feeding, and which transport mechanisms are involved in deploying wax constituents to specific regions on the leaf surface.

Keywords: SIMS imaging, multivariate analysis, data analysis, leaf surface, *Populus trichocarpa*, secondary ion mass spectrometry, co-localization, leaf surface compounds.

INTRODUCTION

The physicochemical properties of plant surfaces have been the subject of many studies (Koch *et al.*, 2008; Samuels *et al.*, 2008; Koch and Barthlott, 2009). Leaf surfaces are composed of cuticular waxes, in which spatial differentiation exists (Haas and Rentschler, 1984), and contain an inner layer that is composed of intracuticular waxes, and an outer layer that is composed of epicuticular waxes (EWs; Figure 1a). These possess considerable ultrastructural and chemical diversity (Koch *et al.*, 2009).

Mass spectrometry imaging (MSI) has emerged as one of the leading spatial analysis technologies whose use can unveil the complexity of the underlying surface chemistry and organization of such biological surfaces (Boughton and Hamilton, 2017). MSI strategies have the potential to analyze a variety of analytes across a wide range of masses with diverse molecular specificity and high sensitivity, as compared with other existing imaging modalities. Although matrix-assisted laser desorption/ionization mass spectrometry imaging (MALDI-MSI), a widely applied MSI method, provides chemical information with high spatial resolution and has been successfully employed in plant science (Vrkoslav *et al.*, 2010), the lateral resolution does not exceed tens of micrometers (Svatoš, 2010). Time-of-flight secondary ion mass spectrometry (TOF-SIMS) overcomes this limitation and has acquired increasing importance over the past few years. It has emerged as a powerful technique due to its sub-micrometer lateral resolution, high mass resolution, wide mass range and high sensitivity (Biesinger *et al.*, 2002). Traditionally used in material sciences to analyze semi-conductors and polymers, TOF-SIMS has the potential to become more common in the study of biological samples, especially the detection of small biomolecules (Winograd, 2005; Fletcher *et al.*, 2011). The advantages of TOF-SIMS (Touboul and Brunelle, 2016) make it possible to study the localization of compounds present on plant surfaces at a lateral resolution of 1 μm or better.

The strength of surface analysis using TOF-SIMS imaging can be fully understood only when the huge wealth of information generated by a single experiment is interpreted correctly. A TOF-SIMS experiment usually generates a large data cube with two spatial dimensions and one *m/z* dimension. Depending on the biological question, if one is interested in observing the spatial distribution of only a few specific ions of interest individually, then the data cube can generate two-dimensional (2-D) molecular ion intensity maps. A 2-D ion intensity map for a specific mass of interest is generated using the intensities of that particular mass peak across all pixels. It is also possible to visualize the spectral features corresponding to a known region of interest using the data cube and then further correlate these with the histological features observed using classical microscopy (Chaurand *et al.*, 2004; Römpp *et al.*, 2010). However, if the goal is to understand the overall chemical composition of the sample or to analyze correlations between regions and study multiple analytes, then a different approach is necessary.

As a first step, a mean spectrum of the data is usually generated to identify major peaks. However, the mean spectrum at times under-represents mass peaks that are present in only a small portion of the spectra and, as a result, peaks that are biologically significant may be overlooked (Alexandrov *et al.*, 2010a). It becomes important to perform exploratory analyses of the spatial distribution and co-localization patterns of several compounds by generating hundreds of ion intensity maps. Such an analysis is extremely tedious to carry out as well as difficult to draw inferences from when performed manually. The total number of TOF-SIMS spectra acquired – where each spectrum corresponds to a pixel – makes interpretation difficult,

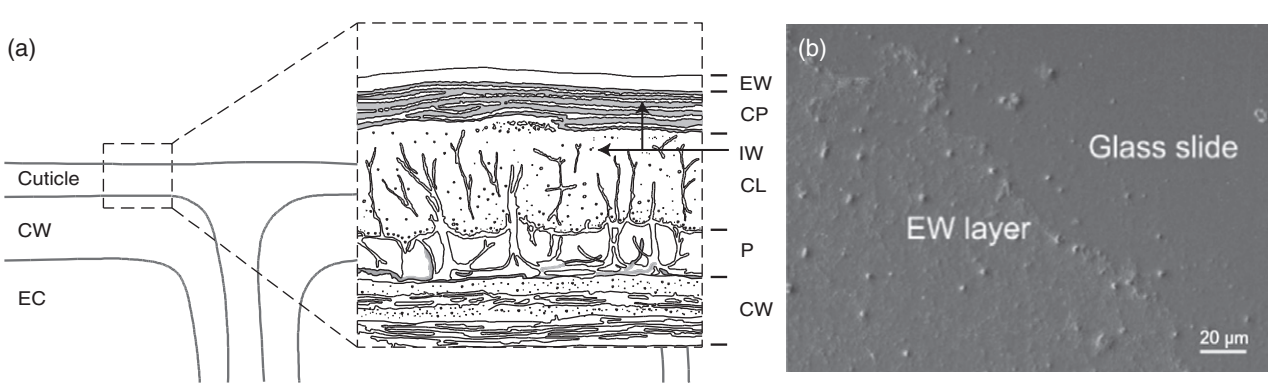

Figure 1. (a) Cross-section of cuticle layers and upper epidermis (EWs, epicuticular waxes; IWs, intracuticular waxes; CP, cuticle proper; CL, cuticle layer; P, pectinaceous layer and middle lamella; CW, cell wall; EC, epidermal cell), modified from Jeffree (1996) and Jetter *et al.* (2000). (b) Scanning electron micrograph of EWs from the surface of adaxial leaves of *Populus trichocarpa* isolated using the cryo-adhesive tape embedding method.

leading to the possibility that much significant information related to hundreds of mass peaks remains hidden (Hook *et al.*, 2015). It has also been observed that within a typical TOF-SIMS spectrum, multiple peaks are generated from the same surface molecules, and their relative yields are often related.

Owing to the complex nature of the data, multivariate data analysis (MVA) approaches. such as principal component analysis (PCA), multivariate curve resolution (MCR), exploratory factor analysis, neural networks and mixture models are useful for identifying chemically significant areas on 2-D ion intensity maps (Tyler, 2003; Graham *et al.*, 2006; Tyler *et al.*, 2007; Park *et al.*, 2009; Graham and Castner, 2012). MVA and cluster analysis techniques have been used extensively to distinguish spatial structures and establish correlation patterns for data obtained from the SIMS imaging of several biological samples (Boxer *et al.*, 2009). These include proteins (Aoyagi *et al.*, 2004), lipids (Biesinger *et al.*, 2006; Brulet *et al.*, 2010), biomaterials (Tyler *et al.*, 2007) and single cells (Colliver *et al.*, 1997).

In this study, we apply TOF-SIMS imaging followed by MVA to investigate the chemistry of EWs present on the leaf surface of *Populus trichocarpa*. Black cottonwood, *P. trichocarpa* (Torr. & A. Gray), is an economically and ecologically relevant tree and the first woody plant whose genome has been sequenced (Wullschleger *et al.*, 2002). Nowadays, *P. trichocarpa* is considered a model for long-lived trees (Bradshaw *et al.*, 2000; Taylor, 2002; Brunner *et al.*, 2004). However, knowledge is still lacking regarding how much of its leaf surface is represented by EWs, as well as the chemical characterization of EWs and how they interact with herbivorous insects. The crystallization of EWs on leaf surfaces is also not fully understood. From the literature, we know that after the crystallization of EWs, properties of the leaf's surface appear to be dramatically modified in comparison to properties of an amorphous EW layer (Jeffree, 1996; Post-Beittenmiller, 1996). Crystallization may be initiated as a mono-layer self-assembly, later rising above the original layer and supported by an underflow of EWs at the center. The crystallization of EWs has also been studied for pure compounds, as well as for isolated and partially purified EW mixtures (Jetter and Schäffer, 2001). Crystals, which tend to form diverse geometric shapes, have been shown to manage the water flow and minimize the droplet adhesion on the surface (lotus effect) or to act as plant defenses against herbivory (Alfaro-Tapia *et al.*, 2007).

To study the crystallization of EWs, the acquired TOF-SIMS imaging data were preprocessed and then analyzed using MVA and clustering approaches. This data analysis pipeline helped to distinguish spatial structures as well as to establish the distribution and covariance patterns of ions of interest on the leaf surface.

Principal components analysis was applied to the preprocessed data to identify discriminating factors and detect underlying structures based on similarities or differences among the mass spectra. Because PCA results can be difficult to interpret, an MCR model was applied to the preprocessed data to directly correlate the discriminating factors with the SIMS spectra and identify important interactions among multiple compounds.

We also applied unsupervised classification approaches, such as hierarchical clustering analysis (HCA), in order to group ions based on their spatial distribution patterns and partitioning approaches – for example, divisive cluster analysis (DCA) and spatial segmentation; we hoped to find sets of spectrally related pixels. Using this approach, we established the spatial distribution of selected alcohols (Alc), alkanes (Alk) and wax esters (WEs; Dost, 2015), and identified the elements responsible for distinct crystal formation patterns.

It should be noted that the approach discussed in this work is applied to a subset of the data that contains only a few selected ions of interest (often referred to as 'targeted analysis'; Dunn *et al.*, 2011). When there is no prior knowledge about the surface chemistry of the analyzed sample, then the full spectrum data have to be used. In either case, our approach should be equally adaptable for exploratory/untargeted analyses. The MVA methods applied here consider all the peak intensities that help reveal salient chemical patterns and different correlations on the surface – patterns that at times can remain hidden – when a targeted analysis is performed. However, selecting robust preprocessing approaches that address acquisition artifacts in the full mass range and provide a good signal-to-noise separation ratio is crucial.

As the TOF-SIMS method is now widely available and several institutes are running SIMS service centers, implementing this state-of-the art method in plant sciences is increasingly feasible. The approach may help shed light on several crucial biological mechanisms in plants, such as mapping the flow of cuticular lipids arising from the anticlinal walls, and many more. The aim of this work is to present the applicability of high-resolution TOF-SIMS imaging in plants and to show how various bioinformatics approaches can be combined to precisely analyze the resulting high-dimensional data.

RESULTS

Prior to TOF-SIMS imaging, the composition of blotted EWs was studied using MALDI-TOF MS. Positive ion mode MALDI-TOF MS spectra $[M + Li]^+$ from the adaxial surface of *P. trichocarpa* leaves showed three characteristic series (Figure 2). Ions within each series were separated by 14 Da (CH_2), indicating the presence of consecutive homologs within the group. Their masses were compared with standards and previously published data (Cvačka and Svatoš,

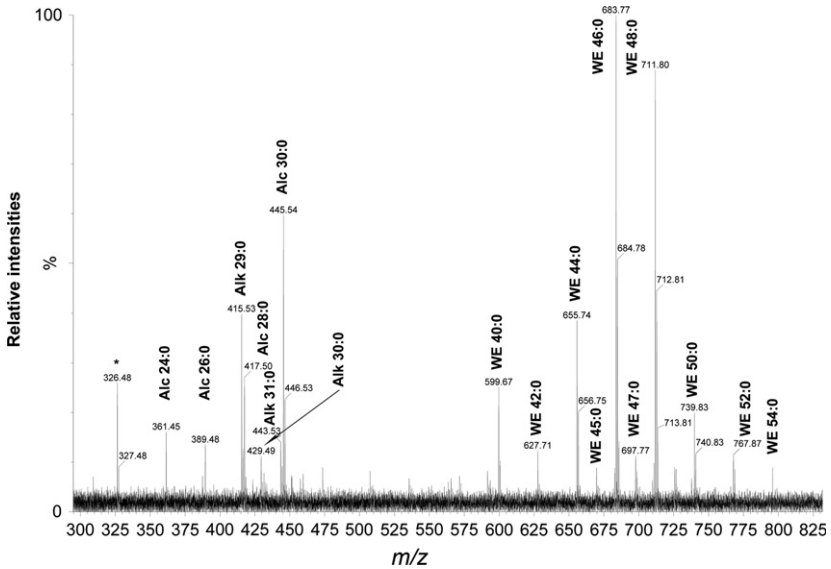

Figure 2. Matrix-assisted laser desorption/ionization time-of-flight mass spectroscopy (MALDI-TOF MS) spectrum of epicuticular waxes (EWs) isolated from the surface of adaxial leaves of *Populus trichocarpa* using the cryo-adhesive isolation method. Alk, alkanes; Alc, alcohols; WEs, wax esters.

2003; Vrkoslav *et al.*, 2010). The first series observed was assigned to C_{29}–C_{31} hydrocarbons; the next series, with an increment of 28 or 14 Da, was assigned to C_{24}–C_{33} alcohols; and the third series was assigned to long-chain C_{40}–C_{54} saturated WEs. The observed intensities in the MALDI-TOF spectrum might have been influenced by the fact that WEs will bind Li^+ ions strongly to hydrocarbons (Cvačka and Svatoš, 2003), and therefore the intensities of WEs might be biased. Within this series, additional signals with a decrement of 2 Da were observed and assigned to unsaturated esters (Table S1).

Later, an extensive TOF-SIMS imaging analysis using two lateral resolutions of 10 μm and 1 μm in both positive and negative ion modes was performed to validate MALDI-TOF MS data. Molecular ions $M^{+\bullet}$ and $M^{-\bullet}$ for the 19 selected compounds of interest, including 11 alcohols, four hydrocarbons and four WEs, are shown in Table 1. Here, some signal intensities were biased, and the low abundance of WEs observed could be explained by their low stability under SIMS experimental conditions. On the other hand, Alk and Alc series were well represented but had higher intensities of the corresponding ions in MALDI-TOF MS spectra (Figure 2; Table S1).

Table 1 Selected ions of interest for multivariate and classification analysis from TOF-SIMS imaging data

Chemical class	Chemical formula	Monoisotopic mass (Da)
Alcohol (Alc)	$C_{21}H_{44}O$	312.3392
Alc	$C_{22}H_{46}O$	326.3548
Alc	$C_{23}H_{48}O$	340.3705
Alkane (Alk)	$C_{25}H_{52}$	352.4069
Alc	$C_{24}H_{50}O$	354.3861
Alc	$C_{25}H_{52}O$	368.4018
Alc	$C_{27}H_{56}O$	396.4331
Alk	$C_{29}H_{60}$	408.4695
Alc	$C_{28}H_{58}O$	410.4487
Alk	$C_{30}H_{62}$	422.4852
Alc	$C_{29}H_{60}O$	424.4644
Alk	$C_{31}H_{64}$	436.5008
Alc	$C_{30}H_{62}O$	438.4800
Alc	$C_{31}H_{64}O$	452.4957
Alc	$C_{33}H_{68}O$	480.5270
Wax ester (WE)	$C_{44}H_{88}O_2$	648.6784
WE	$C_{46}H_{92}O_2$	676.7097
WE	$C_{47}H_{94}O_2$	690.7253
WE	$C_{48}H_{96}O_2$	704.7410

MVA captures co-aggregation patterns on the surface of *Populus trichocarpa* leaves

PCA. After mean centering and Poisson scaling the raw TOF-SIMS imaging data, we performed PCA on the preprocessed data. For this study, we selected six principal components (PCs) based on the total variance captured (Figure S1). PC 1 captures 42% of the total variation, while PCs 2–6 capture a total of 36%, making a total of 78% variance captured by the selected PCs. Details related to the percentage of contribution by selected PCs are provided in Table 2. The variance captured by each factor decreases quickly for factors that contain chemical features, and then reaches a gently declining slope for factors that describe noise variations. As PCs higher than 6 do not provide any additional information and do not seem to contain any clear systematic structure, it is appropriate to consider them as reflecting noise. The typical computation time is less than 5 sec due to the small size of the data matrix. The score and loading plots for all the selected PCs are shown in Figure 3.

The score plot for PC 1 (Figure 3a) captures the overall variation in intensity arising from the topography of the

Table 2 Percent variance captured by PCA performed on the pre-processed data obtained from TOF-SIMS of *Populus trichocarpa* leaf surface

PC	Eigenvalue	Variance captured by PC (%)	Cumulative variance (%)
1	1.83e + 01	42.02	42.02
2	7.93e + 00	18.23	60.25
3	2.36e + 00	5.43	65.68
4	2.11e + 00	4.85	70.53
5	1.80e + 00	4.14	74.67
6	1.28e + 00	2.94	77.61

PC, principal component.

leaf surface. Such topography includes the lateral leaf vein, background and small islets (here, 'crystals'), and represents the largest source of variance in the data. The loading plot, which indicates the contribution of each ion to PC 1 (Figure 3b), shows high positive loadings for pentacosanol C_{25}-Alc, *m/z* 368, and C_{23}-Alc, *m/z* 340, with negative loadings for C_{21}-Alc, *m/z* 312, and C_{22}-Alc, *m/z* 326.

The score plot for PC 2 (Figure 3c) shows an enhanced chemical contrast between the leaf surface background and certain localized pixels ('crystals'). The high-contrast yellow regions in the score plot can be mainly attributed to the increased contribution from C_{29}-Alc; for *m/z* 424, on the other hand, the pixels in black can be correlated with the high negative loadings exhibited by C_{23}-Alc, *m/z* 340, as shown in the loading plot (Figure 3d).

The score plot for PC 3 (Figure 3e) mainly highlights the other crystal-forming regions (shown in yellow) on the leaf surface. These can be attributed to increased C_{23}-Alc, *m/z* 340, in loadings for PC 3 (Figure 3f). The sharp negative loadings for C_{25}-Alc, *m/z* 368, can be attributed to the black pixels in the upper right corner of each leaf's surface area in its corresponding score plot. The individual crystal patterns from PC 2 and PC 3 show different spatial organization accompanied by distinct chemistry. On calculating the distance in pixels for the crystal patterns observed for contribution of C_{29}-Alc, *m/z* 424, in PC 2, and C_{23}-Alc, *m/z* 340, in PC3, it was found that the distance among the crystals in the two patterns differs distinctly (about 10 μm versus 25 μm, as seen in Figure S2).

The score plot for PC 4 (Figure 3g) does not highlight any crystal patterns but shows increased contributions

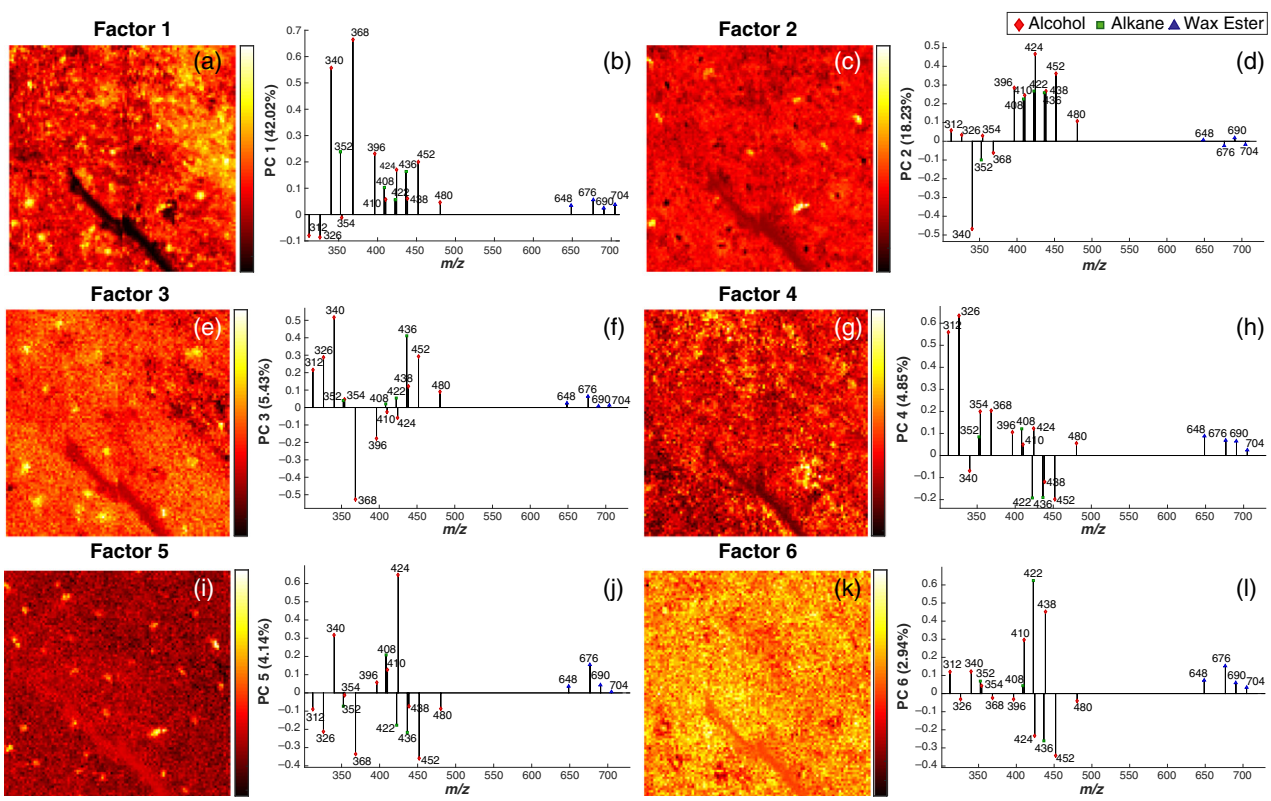

Figure 3. Principal component analysis (PCA) results for data obtained using negative mode time-of-flight secondary ion mass spectroscopy (TOF-SIMS) imaging of the surface of *Populus trichocarpa* leaves at 1 μm step-size. (a,b) Score (left) and loading (right) plots corresponding to principal component (PC) 1. (c,d) Score and loading plots corresponding to PC 2. (e,f) Score and loading plots corresponding to PC 3. (g,h) Score and loading plots corresponding to PC 4. (i,j) Score and loading plots corresponding to PC 5. (k,l) Score and loading plots corresponding to PC 6. PCs 1–6 are the six selected principal components. Scores are plotted using a standard 'hot' color gradient scale where black represents high negative loadings; movement from red, yellow to white represents high positive loadings.

from C_{21}-Alc, *m/z* 312, and C_{22}-Alc, *m/z* 326, as is evident from the corresponding loading plot (Figure 3h).

The loadings for PC 5 (Figure 3j) again highlight the slightly increased contributions from C_{23}-Alc, *m/z* 340, and C_{29}-Alc, *m/z* 424, and these are also reflected in the score plot with evident crystal patterns (Figure 3i). The score and loading plot for PC 6 do not provide any information related to crystal formation; however, the score plot (Figure 3k) highlights the background of each leaf's surface, and this background can be correlated with sharp increases in the loadings for triacontane C_{30}-Alk, *m/z* 422; C_{28}-Alc, *m/z* 410; and C_{30}-Alc, *m/z* 438.

MCR. The same preprocessed data were then subjected to MCR analysis. The chemical contrast in the score plots improved visibly after Poisson scaling was applied to the raw data. For this study, we selected five factors for the MCR model. This selection was based on an estimate of the number of chemical species present in the dataset and also on the number of PCs that explained most of the variance in the data, as shown previously.

The score and loading plots obtained after applying the MCR model are shown in Figure 4. Details related to the percentage of contribution by each factor in the five-factor MCR model are provided in Table 3. The loading plots of MCR analysis, unlike those of PCA, were interpreted as normal spectra by applying non-negativity constraints. Because these spectral responses show only positively correlated species, interpretation is made easier.

The score plot for factor 1 (Figure 4a) does not provide much insight into the crystal formation patterns or help find which pure components are responsible for the pattern. The corresponding loading plot (Figure 4b) shows high contributions from C_{31}-Alk, *m/z* 436; C_{30}-Alk, *m/z* 422; and C_{30}-Alc, *m/z* 438. These may contribute solely to the background chemical composition of each leaf's surface. Score plots for factors 2 and 4 (Figure 4c,g) mainly show the distinct patterns of crystal formation. Their respective loading plots (Figure 4d,h) show an increased contribution from C_{23}-Alc, *m/z* 340, in loadings for factor 2, and from C_{29}-Alc, *m/z* 424, in loadings for factor 4. Figure S4 clearly represents the two crystal patterns observed in a two-color image overlay of score plots for factor 2 (pixels in red) and factor 4 (pixels in green). The loading plots corresponding to these factors have also been overlaid to identify the distinct chemical contribution. The score plot for factor 3 (Figure 4e) highlights the curved region at the upper right corner parallel to the lateral leaf vein; this curve can be attributed to the strong presence of C_{25}-Alc, *m/z* 368, in the loading plot (Figure 4f). The results from the MCR analysis confirm the previously described results of PCA.

The two crystal patterns as observed in factors 2 and 4 (Figure 4c,g) mainly exhibit differences in the number of crystals as well as the distance among them. These figures also display a common pattern: some crystals are aligned on parallel lines and some cross each other at an angle of about 60° or −90°. A study was also performed on the size of the observed crystals. As shown in Figure S3b, the pseudo-colored plot displays the approximate radius of all the regions observed on the leaf surface, based on the intensity signals originating from the groups of pixels that are creating the structural pattern. The color scale indicates the size of such individual structures in terms of number of pixels (1 pixel = 1 μm) in an increasing order, starting from blue to yellow on the scale. The corresponding histogram (Figure S3c) showing a region-wise count displays that most crystal structures mainly fall in the range of 3–7 μm.

Cluster analysis. As with PCA and MCR, clustering was performed using the mean-centered and Poisson-scaled data. First, HCA was applied to the 19 variables in order to group them. The dendrogram and heatmap representing the HCA results were constructed using the Ward's linkage method, and are shown in Figure 5a. The dendrogram reveals the relationship between the ions of interest, based on the spectral differences/similarities among them. Each row in the heatmap corresponds to an individual ion of interest, and each column corresponds to a spatial coordinate position (*x*, *y*) on the leaf's surface. The color of each cell in the heatmap represents the *Z*-score, a number that measures the distance, in terms of the number of standard deviations (SD), between that cell and the mean of all cells in that column. The mean for each column (corresponding to a single coordinate position) in the heatmap is calculated by averaging the intensity values for all the 19 selected masses at that coordinate position. The mean is indicated as 0 on the *Z*-score scale. A positive *Z*-score indicates how many SD units above the mean the intensity value for a specific mass is, and a negative *Z*-score indicates the number of units below the mean. The *Z*-score can be particularly useful to distinguish features with high intensity distributions reflecting stronger contributions at a specific spatial location.

As seen in Figure 5a, the dendrogram consists of four major clusters. The first cluster is represented by a single ion forming the characteristic curved pattern on each leaf's surface (C_{25}-Alc, *m/z* 368). The second cluster consists of ions that belong to the Alc and Alk class, and is dominated by those that exhibit a distinct crystal formation pattern. However, one can see that the ion intensity maps of *m/z* 452 and *m/z* 436 do not display any crystal formation pattern. Similarly, the third cluster also consists of ions that belong to the Alc and Alk class, but is dominated by ions that do not display crystal patterns. However, ion intensity maps of *m/z* 408 and *m/z* 401 do show crystal formation patterns. One explanation for this inconsistency could be the distance measure used for HCA, which may group ions based on the intensities at individual pixels without

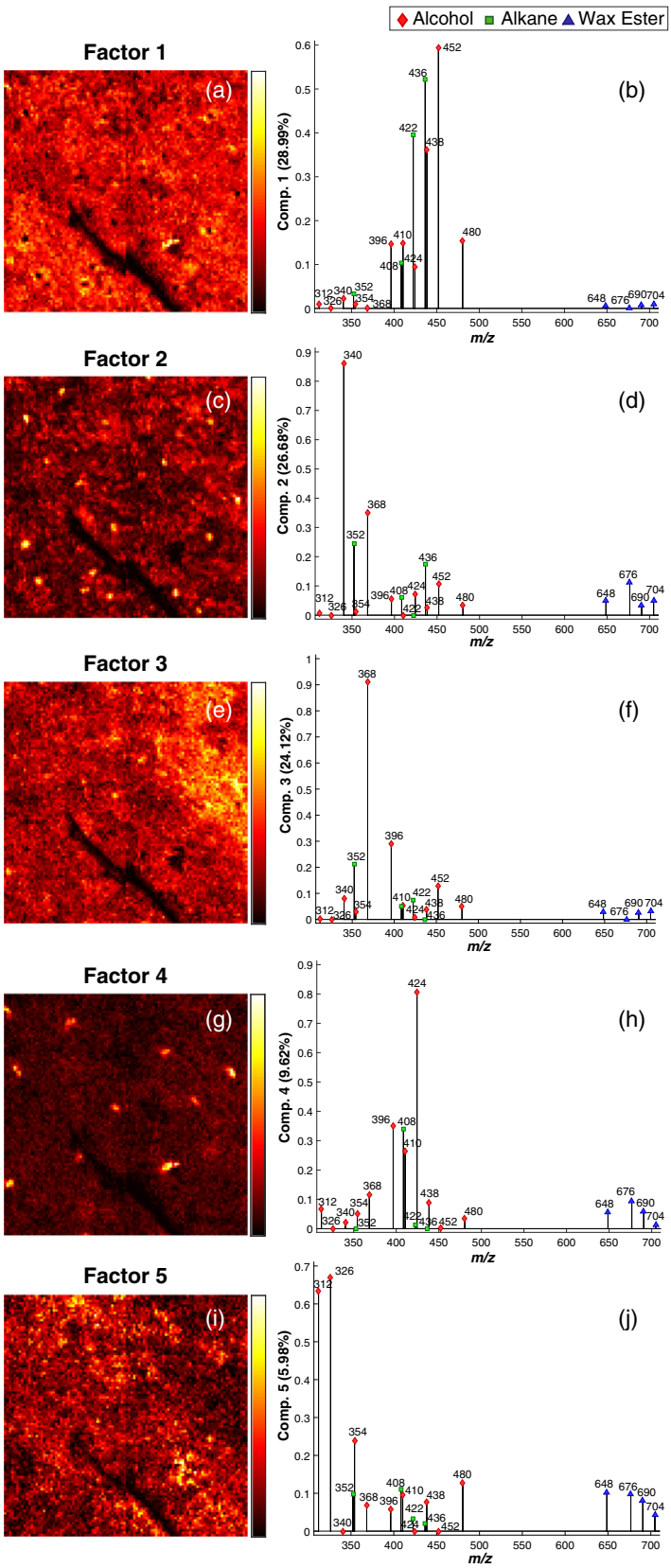

Figure 4. Multivariate curve resolution (MCR) results for data obtained using negative mode time-of-flight secondary ion mass spectroscopy (TOF-SIMS) imaging of the surface of *Populus trichocarpa* leaves at 1 μm step-size. (a,b) Score (left) and loading (right) plots corresponding to factor 1. (c,d) Score and loading plots corresponding to factor 2. (e,f) Score and loading plots corresponding to factor 3. (g,h) Score and loading plots corresponding to factor 4. (i,j) Score and loading plots corresponding to factor 5.

Table 3 Results of MCR analysis performed on the preprocessed data obtained from TOF-SIMS of *Populus trichocarpa* leaf surface

Factor	Fit (% X)	Cumulative fit (% X)
1	28.99	28.99
2	26.68	55.67
3	24.12	79.78
4	9.62	89.40
5	5.98	95.39

considering their spatial context. The last cluster in the dendrogram is formed of non-crystal-forming ions, mainly WEs.

Results for DCA using *k*-means are shown in Figure 5b as a palette of pseudo-colored images (I–IV) representing the partitioning into three, four, five and six clusters. The three-class clustering (Figure 5b.I) broadly distinguishes the background structures of the leaf's surface without offering much information about the location of the crystals. The four-class clustering (Figure 5b.II) distinguishes some of the crystal-forming regions, shown in red; the curved region on the leaf surface in the upper right corner, shown in turquoise; and the leaf background, shown in blue and green regions. The five-class clustering (Figure 5b.III) shows the crystal-forming regions more precisely in red; the leaf background mainly in blue, with irregular regions of green and turquoise; and the upper right curved region of the leaf in pink. The six-class clustering (Figure 5b.IV) shows the location of crystal-forming regions in red; the region of the leaf's vein as well as some

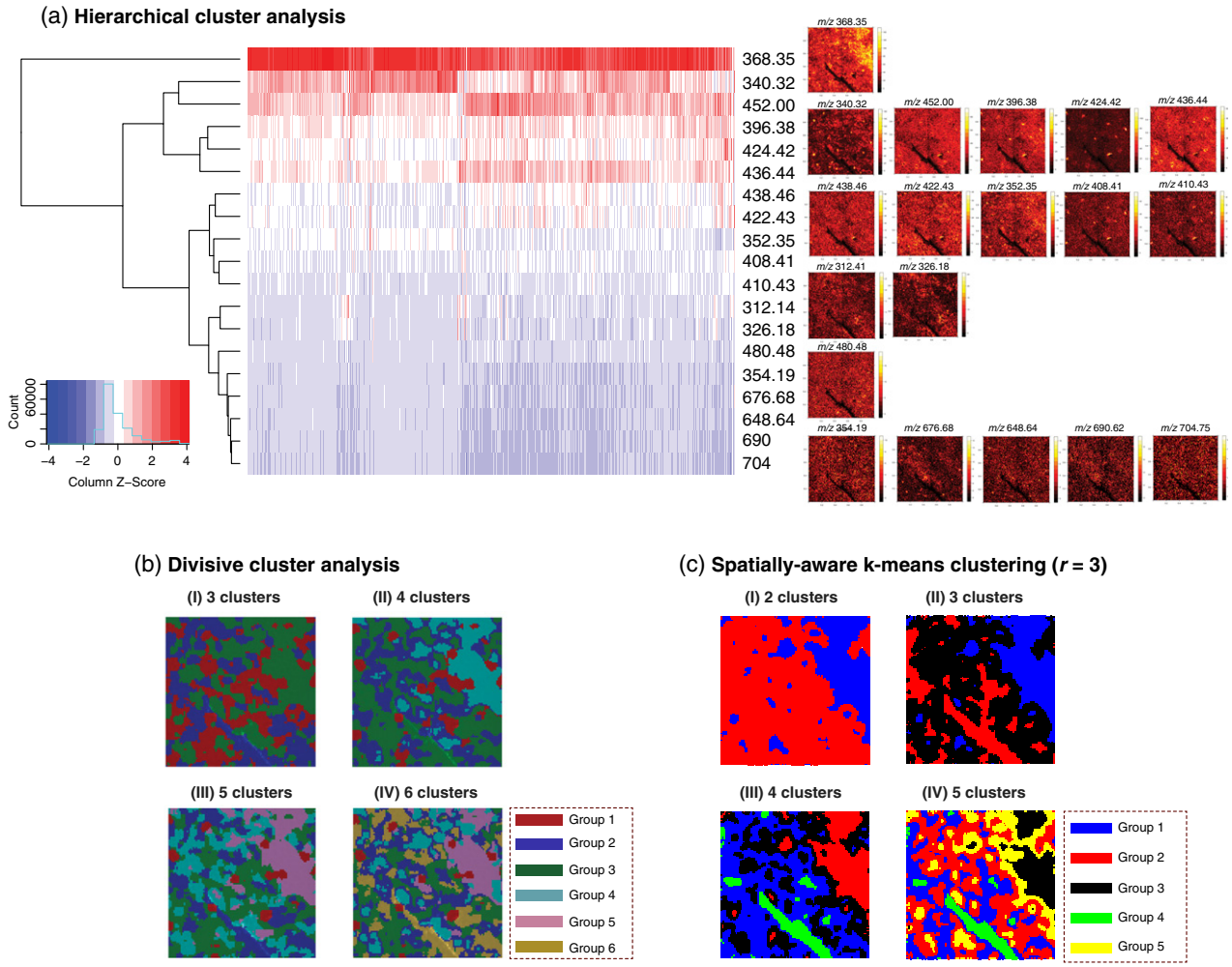

Figure 5. Cluster analysis on time-of-flight secondary ion mass spectroscopy (TOF-SIMS) imaging data. (a) Results for hierarchical clustering analysis (HCA) showing three main cluster groups as seen in the dendrogram. Each row in the heatmap represents a selected mass peak of interest, and each column represents the deviation of the spectral information at each coordinate position (*x*, *y*). The heatmap is color-coded based on the *Z*-score, where red represents a high positive deviation from the mean and blue represents a high negative deviation from the mean. White represents no deviation from the mean value. (b) Results for divisive cluster analysis (DCA) using *k*-means algorithm performed on data obtained using TOF-SIMS imaging of the surface of *Populus trichocarpa* leaves. (I) Clustering with $k = 3$. (II) Clustering with $k = 4$. (III) Clustering with $k = 5$. (IV) Clustering with $k = 6$. (c) Results for spatially aware *k*-means clustering with pixel neighborhood radius, $r = 3$. (I) Clustering with $k = 2$. (II) Clustering with $k = 3$. (III) Clustering with $k = 4$. (IV) Clustering with $k = 5$.

regions of each leaf's background in gold; the curved region of the leaf's upper right corner in pink; and the background of each leaf's surface in green, blue and turquoise.

Because HCA as well as DCA using *k*-means treats intensities from pixels independently, we also applied spatially aware *k*-means clustering. Results for spatially aware *k*-means clustering using pixel neighborhood radius $r = 3$ are shown in Figure 5c. The clustering results are in the form of a palette of pseudo-colored images (I–IV) representing the spatially aware partitioning into two, three, four and five clusters. As can be seen, the five-class clustering (Figure 5c.IV) broadly distinguishes the localization of the crystal-forming regions, the region of the leaf vein and the background structures, as well as the characteristic curved structure in the upper right corner of the leaf surface.

The results from the cluster analysis also point to the localization of crystals with distinct chemical specificity.

DISCUSSION

Understanding the chemical composition of plant EWs on the scale of micrometers is arduous. The sensitivity and the lateral resolution required are feasible only when TOF-SIMS imaging is used due to the tightly focused ion beams used for compound desorption/ionization. Here we present TOF-SIMS imaging data acquired at 1 μm step size and a large volume of information harvested using bioinformatics tools.

Employing a bi-cluster ion beam provided richly informative MS spectra in which individual molecular ions in both positive and negative ion modes were easily recognized. Nineteen compounds were identified, including alcohols, hydrocarbons and WEs, all of which formed radical cations or anions depending on the ion mode used. Molecular ions for aliphatic hydrocarbons, alcohols and WEs were obtained using identical experimental conditions during the TOF-SIMS imaging of EWs on the surface of *Kalanchoe daigremontiana* leaves (Jetter and Sodhi, 2011). The SIMS MS analysis of EWs on myrtle berries using $^{69}Ga^+$ ions also provided rich MS spectra of intact compounds. However, $[M + H]^+$ or $[M - H]^-$ were detected in positive and negative ion modes, respectively (Piras *et al.*, 2009). The TOF-SIMS imaging data correlate well with MALDI-TOF MS spectra measured using a ^{7}LiDHB matrix. Quantification using internal standards has not been performed; however, the relative intensities may represent the actual proportion of the compounds on EWs. The reason is that all long-chain compounds are expected to have similar ionization potentials, as the chemical properties of the local environment on the imaged leaf surface are the same. However, this scenario may be changed if the imaged tissue surface has chemical properties other than what was expected and is non-homogenous.

For technical reasons, we were not able to measure the samples after TOF-SIMS imaging using scanning electron microscopy (SEM). Nevertheless, the crystals observed in TOF-SIMS imaging data resemble the spheroidal grains seen on the SEM image (Figure 1b) obtained independently. Similarly, in SEM published for *P. trichocarpa*, structures of about 3 μm in diameter are clearly visible (Alfaro-Tapia *et al.*, 2007). Additionally, large crystal clusters resembling the structure in the upper corner parallel to the lateral leaf vein in Figure 4e were also recognized. The distances among the crystals are much smaller than the distances among the trichomes (about 150 μm) or stigmata (about 100 μm); however, the observed distances could be related to the size of epidermal cells. The periodicity of the observed crystals (about 25 μm) may be related to the transport of EWs from epidermal cells. As observed from MVA and clustering results, the chemical composition of the crystals differs clearly from the chemical composition of other areas. The crystals with smaller distances are formed mostly by C_{23}-Alc, and those with larger distances by C_{29}-Alc, findings that are fully congruent with previous findings on the uniform composition of crystals in EWs (Ensikat *et al.*, 2006). The uniformity is likely related to the process of molecular self-assembly that occurs during crystallization. The chemical composition of other EWs is extremely diverse. The presence of a mixture of hydrocarbons (C_{30}, C_{31}) and aliphatic alcohols (C_{30}, C_{31}), and the virtual absence of WEs, is characteristic for the score and loading plot for MCR factor 1 in Figure 4a and b. In contrast, the score plot for MCR factor 5 in Figure 4i and j shows the presence of alcohols (C_{21}–C_{23}) and WEs. To better visualize and classify EW areas with distinct chemical composition, the application of DCA and spatially aware *k*-means clustering proved to be particularly useful. Images generated for cluster groups $k = 5$ and $k = 6$ (Figure 5b.III and b.IV) in DCA, and $k = 5$ (Figure 5c.IV) in spatially aware *k*-means clustering illustrate the highly heterogeneous chemical composition of the relatively small leaf area.

Based on this analysis, the MVA approach applied here for identifying regions of specific chemical composition and their classification based on distinct patterns using cluster analysis is recommended for analyzing TOF-SIMS imaging data. This analysis provides the chemical composition of areas showing distinct segregation/co-localization. In addition, this methodology could be easily expanded to other MSI datasets.

EXPERIMENTAL PROCEDURES

Plant material and growth conditions

Populus trichocarpa Torr. & Gray (clone 600-25) was cultivated in a growth chamber at 23°C day and 18°C night with 50% relative humidity, light intensity of 30% (Osram De Luxe 36W Natura), and a diurnal cycle of 12 h light and 12 h dark. Cuttings were kept at

12°C. After 8 weeks, when they had reached 15–20 cm in length, plants were transferred to 2-L pots. Standard substrate (Klasmann-Deilmann GmbH, Postfach D-49744 Geeste, Germany, www.klasmann-deilmann.com/) was used as soil and, beginning 4 weeks after transfer, plants were fertilized with 1% of 10:15:10 N:P:K supplement (Ferty®, PLANTA Düngemittel GmbH, Regenstauf D-93128, Germany, www.plantafert.com) every second week. Plants were grown under controlled conditions and not treated with any pesticides to prevent the potential chemical modification of the leaf's surface. Plant leaves used for all experiments were 4–5 months old.

Blotting of leaf cuticle

Epicuticular waxes were isolated using the cryo-adhesive method (Jetter *et al.*, 2000) with water as a transfer medium. *Populus trichocarpa* leaves were rinsed briefly with distilled water to remove any contaminants. Leaf tissues were cut into rectangle-shaped pieces with dimensions slightly larger than the solid substrate support [a metallic MALDI plate for MALDI-TOF MS, and a p-type (100) orientation silicon wafer for TOF-SIMS]. Each leaf cutting was placed onto the cleaned substrate holding four-five evenly distributed droplets (double-distilled water, 3 μl each). Afterwards, a glass slide was placed on top of the substrate/leaf stack and gently pressed, creating a sandwich. The leaf was homogeneously moistened with a very thin film of water. Tweezers were used to dip the whole leaf in liquid nitrogen for 30 sec. Water droplets served as a medium to transfer the wax layer from the leaf to the substrate. The second glass slide enabled slight force to be put on the tissue without direct contact, allowing the EWs to be isolated without disturbing the other cuticle layers. The metal plate or silicon wafer was detached from the plant tissue and kept in a desiccator prior to analysis. This protocol results in a very flat EW transfer layer on conductive substrates, representing an ideal sample for TOF-SIMS imaging, one that can achieve the best mass and lateral resolution.

MALDI-TOF MS

The spectra were acquired using a MALDI micro MX™ mass spectrometer (Waters/Micromass, Manchester, UK, www.waters.com) operating in reflectron mode and positive polarity. EW compounds were ionized and desorbed with a nitrogen UV laser (337 nm, 4 ns laser pulse, firing rate 10 Hz, max 280 μJ per laser pulse, 10 shots per spectrum). Ions were recorded from *m/z* 200 to *m/z* 1500, matrix ions (below *m/z* 200) were suppressed and 70 scans were co-added. Data acquisition was done using MassLynx™ V4.0 software package (Waters, Manchester, UK, www.waters.com). Stainless-steel MALDI plate (Waters/Micromass, Manchester, UK, www.waters.com) served as targets and were cleaned by sonication in MeOH and subsequently in Me_2CO, hexane and CH_2Cl_2 before each usage. For external calibration of the mass spectrometer, a mixture of PEG 600 and 1000 (1 μg ml^{-1} in Me_2CO; Sigma-Aldrich®, Germany, www.sigmaaldrich.com) was used. ^{7}LiDHB prepared from DHB (Sigma-Aldrich®, Germany, www.sigmaaldrich.com) and lithium ^{7}Li hydroxide (Sigma-Aldrich®, Germany, www.sigmaaldrich.com) according to the protocol (Cvacka *et al.*, 2006) was applied as a matrix (10 mg ml^{-1}) in Me_2CO for calibration and in CH_2Cl_2: Me_2CO (1:1) for sample analysis. All compounds were detected as lithiated (^{7}Li) adducts. Samples in CH_2Cl_2 were spotted on a target using different sampling techniques (mix technique, Sa/Ma, Ma/Sa, Ma/Sa/Ma); the best results were obtained using the 'sandwich' technique: matrix, sample, matrix (0.8 μl of 10 mg ml^{-1} each). Six parallel spots were made for each sample and subsequently measured.

TOF-SIMS imaging

Time-of-flight-SIMS imaging was performed on a standard commercial TOF-SIMS 5 instrument (ION-TOF GmbH, Münster, Germany, www.iontof.com). The spectrometer was equipped with a Bi-cluster primary ion source and a reflectron-type TOF analyzer. The UHV base pressure was $<5 \times 10^{-9}$ mbar. For high mass resolution, the Bi source was operated in the 'high current bunched' mode providing short Bi^+ or Bi_3^+ primary ion pulses at 25 keV energy and a lateral resolution of about 4 μm. The pulse length of 1.1–1.3 ns allowed high mass resolution. The primary ion beam was rastered across 700×700 μm^2, 500×500 μm^2 and 100×100 μm^2 sample area, and 700×700, 128×128 and 100×100 data points were recorded. Images larger than the maximum deflection range of the primary ion gun were obtained using the manipulator stage scan mode with a lateral resolution of 100 pixel mm^{-1}. Primary ion doses were kept below 10^{11} ions cm^{-2} (static SIMS limit). Spectra were calibrated on C^-, C_2^-, C_3^- peaks for the negative ion mode, and on C^+, CH^+, CH_2^+ and CH_3^+ peaks for the positive ion mode. Based on these datasets, the chemical assignments for characteristic fragments were determined. The experiments were performed with five biological and two technical replicates for a 700×700 μm^2 area, one biological replicate for a 500×500 μm^2 area, and three biological replicates for a 100×100 μm^2 area. The following standards were implemented: hexacosanoic acid, ethyl stearate, methyl tricosanoate, 1-hexacosanol, tetracosane (Sigma-Aldrich®, Germany, www.sigmaaldrich.com). The data acquisition and processing software were IonSpec and IonImage (ION-TOF GmbH, Münster, Germany, www.iontof.com). Negative ion mode data at 1 μm lateral resolution were used for the data analysis. The molecular compositions corresponding to measured *m/z* values were enumerated within 100-ppm windows using C, H and O atoms for calculations.

SEM

For SEM, EW imprints were prepared on silicon wafer sample target. EW imprints from the surface of adaxial leaves of *P. trichocarpa* were isolated using the cryo-adhesive tape embedding method. The sample targets with EW imprints were mounted on the aluminum holders, sputter-coated with about 10 nm of gold (Bal-Tec SCD005 sputter coater; 60 mA, 10 sec) and examined by SEM (LEO Gemini 1530, Zeiss, Oberkochen, Germany) at 1.5 kV.

Data processing and MVA

For preliminary inspection of the TOF-SIMS imaging data, pseudo-colored ion intensity maps were generated using SurfaceLab 6.3 software (ION-TOF GmbH, Münster, Germany, www.iontof.com). A total of 19 peaks of interest (shown in Table 1) composed of alcohols, hydrocarbons (alkanes) and WEs in the *m/z* range 200–800 were selected to be used in MVA and classification studies.

Spectral data for these 19 peaks were extracted and converted from the vendor file format, then exported in individual text files, using SurfaceLab 6.3 software (ION-TOF GmbH, Münster, Germany, www.iontof.com). The data for these 19 peaks were arranged in the form of a [$n \times m$] matrix, where the rows (n) are 'samples', which denote the spectra acquired at every single coordinate position (x, y), and the columns (m) are 'variables', which denote the mass peaks of interest [the terminology used here is consistent with ISO standard ISO 18115 Surface Chemical Analysis – vocabulary, part 1: general terms and terms used in spectroscopy (Anon, n.d.)]. For the 100×100 μm^2 TOF-SIMS imaging dataset, this matrix had a dimension of [10 000 × 24]. The

complete data analysis was performed on a Macintosh operating system (version 10.10.5) with 8 GB of RAM, using 3.1 GHz Intel Core i7 processor. Steps related to preprocessing and MVA were carried out using the chemometrics software Solo+MIA (Eigenvector, Research, Wenatchee, WA, USA, www.eigenvector.com).

Preprocessing TOF-SIMS imaging data. Data preprocessing is extremely important before MVA. Because these techniques describe the underlying structure of the data, they are very sensitive to data scaling and transformations. It is crucial to select the appropriate preprocessing methods and to make judgments based on the nature as well as the significance of sources of variance in the data (Keenan and Kotula, 2004b; Wagner *et al.*, 2006).

Because this specific dataset suffered from low signal-to-noise ratio, data preprocessing steps consisted of first, mean centering the variables. In mean centering, each variable is centered by the subtraction of its mean value across all samples. This is typically done so that the differences among the peak variances are emphasized over the differences in the peak area means. Additionally, we also scaled the data to account for Poisson noise (Cochran and Horne, 1977; Keenan and Kotula, 2004a; Henderson *et al.*, 2009; Teng, 2013). Data from TOF-SIMS instruments are collected in a pulse-counted manner and subject to uncertainty that is explained by Poisson statistics. This uncertainty is equal to the mean of the signal intensity. MVA approaches such as PCA are designed to account for variance in the data, and non-normalized variables with large variance have stronger weights and are more likely to be addressed in the modeling than are low variance variables. Because TOF-SIMS data usually has variance, which is related to the signal intensity, these approaches perform sub-optimally. Poisson scaling (also called square root mean scaling) scales each variable by the square root of the mean value so that the estimated variance due to counting statistics is equal on all variables. The application of Poisson scaling to the dataset provided greater noise reduction and improved the sample separation in the PCA model.

One important scaling method often necessary for MSI data, irrespective of the ionization method used, is normalization. This technique helps in identifying and removing sources of systematic variation among pixels in the dataset, which in turn is beneficial to minimize inter-spectra variance. Normalization is usually performed by multiplying a single mass spectrum in the dataset with an intensity-scaling factor in order to make all spectra comparable. There have been many methods proposed for normalization, the most common being total ion current normalization (Deininger *et al.*, 2011; Fonville *et al.*, 2012; Veselkov *et al.*, 2014). We applied normalization to this dataset; however, it led to the loss of contrast and distinct biological features on the sampled leaf surface, and hence was not included as a part of data preprocessing.

PCA. Principal component analysis is a popular factor analysis method in which the data are described using a small number of selected factors to highlight the useful properties of the dataset. PCA looks at the variance pattern within a dataset to find the direction of greatest variance and transform related variables to a smaller set of orthogonal factors, called PCs. PCs are linear combinations of all of the original variables and, therefore, capture more information than do any of the original variables considered individually. Not all PCs generated provide valuable information, and it is preferable to discard higher PCA factors. This step is often referred to as 'factor compression'. The number of PCs can be deduced by inspecting the eigenvalue plot, also known as the 'scree plot' (Zwick and Velicer, 1982), and the percentage of total variance captured by first *N* PCs.

Principal component analysis creates three new matrices: the scores, the loadings and the residuals. The scores describe the relationship among the samples. The loadings define the contributions of the original variables to the new PCs and describe which variables are responsible for the differences seen within the samples. The residual matrix describes the random variations not described by the new PC axis and represents noise in the data. This makes PCA a first step in the evaluation of complex TOF-SIMS imaging datasets (Piras *et al.*, 2009; Kalegowda and Harmer, 2012).

Principal components analysis was applied to the preprocessed TOF-SIMS imaging dataset. Score plots show the values for each coordinate position (pixel) on the associated PC axis. The pseudocolor scale indicates the level of contribution of each pixel to the axis. Pixels that correspond to the same histochemical structure (i.e. pixels showing similar mass spectra) are expected to make a similar contribution to different PCs, and these pixels appear with the same color. Loading plots show the positive and negative correlations of each original variable with the respective PC. The score plots, loadings and the observed correlation of individual peaks are discussed in the Results section.

MCR. While techniques such as PCA calculate factors based on mathematical properties (for example, capturing maximum variance), these are often difficult to interpret because they are not directly related to chemical properties. For example, PCA loadings of a dataset of measured spectra are not generally spectra of pure components. Instead, the loadings are typically linear combinations of pure analyte spectra that have positive and negative intensities.

Multivariate curve resolution (also indicated as alternating least squares regression, or MCR-ALS) offers a way to decompose a hyperspectral data matrix and is designed to identify pure components from a multi-component mixture (Lawton and Sylvestre, 1971; Lee *et al.*, 2008). This bilinear decomposition is usually performed by the repeated application of multiple least squares regression. The technique extracts chemically meaningful information in the form of factors that resemble the spectra of chemical components and contributions. Applying MCR to multivariate images yields information about which analytes are present and where in the image they are located (Gallagher *et al.*, 2004). MCR assumes a linear combination of chemical spectra (MCR loadings) and contributions (MCR scores) to describe each spectrum. MCR factors are not required to be mutually orthogonal; therefore, by applying non-negativity constraints to the loadings and scores matrices during optimization, MCR components are directly interpretable as spectra of pure compounds, as they have positive values (Lee *et al.*, 2009; Wehrens, 2011; Jaumot and Tauler, 2015).

As in PCA, it is important to determine the number of factors or components in MCR. An estimate of the number of components to select for MCR can be made based on the number of chemical species present. The first indication of the number of chemical species present in a dataset can be obtained directly from the rank of the data matrix or from the number of significant singular values associated with the data matrix (Tauler *et al.*, 1995). The singular values that are related to chemical species are usually larger than the noise, systematic errors or baseline values. The initial number of components can also be roughly estimated based on the number of components in PCA or singular value decomposition that can explain the data variance, i.e., by selecting the number of eigenvalues higher than those associated with the noise

level (Malik *et al.*, 2015; Rodríguez-Rodríguez *et al.*, 2007). Because this method works well when homocedastic noise is uniformly distributed but fails when this noise is dstriuted non-uniformly, it is important to clean the data for any baseline and instrumental contributions and apply suitable preprocessing steps (Mendieta *et al.*, 1998). A different approach to identifying the number of components was applied by Motegi *et al.* (2015). They compared the concentration profiles generated from MCR results obtained by sequentially changing the number of components for a single dataset, and observed that similar components emerged repeatedly. This observation suggested that reliable components behave similarly, irrespective of the number of components selected, whereas unreliable components emerged only once or just a few times. These reliable components were considered to be informative.

To obtain initial estimates of the spectral profiles, a straightforward approach is to choose pure spectra from the original data matrix. The SIMPLISMA (Windig and Guilment, 1991) method is also a popular approach for selecting pure spectral variables as initial estimates. To determine the initial estimates for concentration profiles, methods such as evolving factor analysis and evolving window factor analysis are used (Budevska *et al.*, 2003).

Multivariate curve resolution was applied to the preprocessed TOF-SIMS imaging dataset, score plots revealed the distribution of the components of interest on the surface. These score plots are represented using pseudo-colors, which display the level of contribution of each component. The MCR loadings on each factor resemble an actual SIMS spectrum of the component and show its characteristic peaks. The score plots, the loadings and the observed correlation of individual peaks are discussed in the Results section.

Cluster analysis. Clustering is also an important statistical tool for finding unknown patterns. It is a class of unsupervised methods that allows the classification and grouping of objects based on their similarity (or difference). There are two main categories (also known as segmentation methods): agglomerative and partitional. Agglomerative methods such as HCA begin with each object being its own cluster and progress by combining existing clusters into larger ones. Segmentation methods, such as DCA, start with a single cluster containing all objects and progress by dividing existing clusters into smaller clusters or segments based on their similarities or homogeneous composition (Zhao and Karypis, 2003; McCombie *et al.*, 2005; Kaufman and Rousseeuw, 2008; Jones *et al.*, 2012).

In this work, HCA has been applied to group the selected ions of interest based on their spatial patterns (specifically, the localization of crystals on the leaf surface). In HCA, a series of metric-based calculations is performed to measure the distance between samples and group them into clusters. This analysis uses Ward linkage (Joe and Ward, 1963; Yu *et al.*, 2008) as a metric to perform the distance-based calculations. Results from HCA can be represented in the form of a dendrogram as well as a heatmap.

DCA was performed using the *k*-means algorithm; this algorithm starts with *k* objects (clusters), in which *k* is specified by the user *a priori*. During each cycle of this clustering method, the remaining objects are assigned to one of these clusters based on their distance from each of the *k* targets. New cluster targets are then calculated as the means of the objects in each cluster, and the procedure is repeated until no objects are re-assigned after the updated mean calculations. For the TOF-SIMS imaging data, the algorithm classifies each pixel into one of the *k* clusters either by minimizing the sum of distances from their respective centers or by maximizing interclass distance, which leads to the most distinct clusters possible.

It can be difficult to estimate the optimal number of clusters in advance. Selecting this number requires a combination of statistical reasoning [such as the use of a silhouette plot (Rousseeuw, 1987) to study the separation between the resulting *k* clusters], some knowledge about the sample data being used and also human judgment. In practice, *k*-means clustering is performed by using different values of *k* to obtain a series of solutions. The final choice of *k* is made based on qualitative criteria of the clusters obtained (Bratchell, 1989).

One major disadvantage of applying HCA and *k*-means to MSI data is that these methods treat each pixel independently and ignore similarities of spectra acquired from spatially proximate locations, i.e. they do not take into account any spatial relationships (Jones *et al.*, 2012; Bemis *et al.*, 2016). The result can adversely affect the quality of segmentation.

To address this disadvantage, spatially aware segmentation approaches have been developed. Spatially aware segmentation was performed using the approach proposed by Alexandrov *et al.* (Alexandrov *et al.*, 2010b; Alexandrov and Kobarg, 2011), which incorporates spatial relations between pixels so that pixels are clustered together with their neighbors. In this distance-based clustering approach, the distance between spectra, obtained from two pixels in the dataset, depends on the neighbors of the two selected pixels. This method is based on the assumption that mass spectra acquired from neighboring pixels in a morphologically defined region on a biological sample most likely represent pixels with similar biochemical composition and so should be similar. As in the *k*-means approach, the number of clusters has to be provided *a priori*. Also, an additional parameter, the pixel neighborhood radius *r*, has to be specified in advance. Different values of *r* can be used to obtain a series of solutions, and the final choice of an optimal radius can be made based on observations of the segmentation results.

The dendrogram plot and heatmap representing the HCA were generated using the functions hclust and heatmap.2 within the gplots package in R version 3.2.3 (2015-12-10, R Foundation for Statistical Computing, www.r-project.org). DCA was performed using multiple *k* values to obtain the pattern that best correlates the PCA and MCR results. DCA results were generated using Solo+MIA software (Eigenvector, Research, Wenatchee, WA, USA, www.eigenvector.com). Spatially aware segmentation was performed using the spatial *k*-means function, and results were generated using the plot function implemented in the CARDINAL package (Bemis *et al.*, 2015) in R version 3.2.3 (2015-12-10, R Foundation for Statistical Computing, www.r-project.org).

ACKNOWLEDGEMENTS

Purva Kulkarni is supported by a PhD scholarship from the International Max Planck Research School, Jena. Mina Dost is supported by a stipend of the graduate school Jena School for Microbial Communication (JSMC) funded by the German Excellence Initiative (M.D.). Financial support from the Max Planck Society is also acknowledged. The authors would like to thank Frank Steiniger (Elektronenmikroskopisches Zentrum, Universitätsklinikum, FSU, Jena) for scanning electron microscopy and surface crystal annotation, and Emily Wheeler for editorial assistance.

CONFLICT OF INTEREST

The authors have declared no conflict of interest.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

Figure S1. Percentage of variance captured by each PC.

Figure S2. Representation of distance measured between crystals.

Figure S3. Representation of size of crystals.

Figure S4. Representation of spatial location of crystals.

Table S1. List of masses detected by MALDI-TOF MS.

REFERENCES

Alexandrov, T. and Kobarg, J.H. (2011) Efficient spatial segmentation of large imaging mass spectrometry datasets with spatially aware clustering. *Bioinformatics*, **27**, i230–i238.

Alexandrov, T., Becker, M., Deininger, S., Wehder, L., Grasmair, M., von Eggeling, F., Thiele, H. and Maass, P. (2010a) Spatial segmentation of imaging mass spectrometry data with edge-preserving image denoising and clustering research articles. *J. Proteome Res.* **9**, 6535–6546.

Alexandrov, T., Maass, P. and Thiele, H. (2010b) Spatial segmentation of MALDI-imaging data. 20100225.

Alfaro-Tapia, A., Verdugo, J.A., Astudillo, L.A. and Ramírez, C.C. (2007) Effect of epicuticular waxes of poplar hybrids on the aphid Chaitophorus leucomelas (Hemiptera: Aphididae). *J. Appl. Entomol.* **131**, 486–492.

Anon.(n.d.) ISO 18115-1:2013 – Surface chemical analysis – Vocabulary – Part 1: general terms and terms used in spectroscopy. Available at: http://www.iso.org/iso/catalogue_detail.htm?csnumber=63783 (accessed 8 December 2015).

Aoyagi, S., Hayama, M., Hasegawa, U., Sakai, K., Tozu, M., Hoshi, T. and Kudo, M. (2004) Estimation of protein adsorption on dialysis membrane by means of TOF-SIMS imaging. *J. Memb. Sci.* **236**, 91–99.

Bemis, K.D., Harry, A., Eberlin, L.S., Ferreira, C., van de Ven, S.M., Mallick, P., Stolowitz, M. and Vitek, O. (2015) *Cardinal*: an R package for statistical analysis of mass spectrometry-based imaging experiments: Fig. 1. *Bioinformatics*, **31**, 2418–2420.

Bemis, K.D., Harry, A., Eberlin, L.S., Ferreira, C.R., van de Ven, S.M., Mallick, P., Stolowitz, M. and Vitek, O. (2016) Probabilistic segmentation of mass spectrometry (MS) images helps select important ions and characterize confidence in the resulting segments. *Mol. Cell Proteomics*, **15**, 1761–1772.

Biesinger, M.C., Paepegaey, P.-Y., McIntyre, N.S., Harbottle, R.R. and Petersen, N.O. (2002) Principal component analysis of TOF-SIMS images of organic monolayers. Available at: http://pubs.acs.org/doi/abs/10.1021/ac020311n (accessed 13 July 2017).

Biesinger, M.C., Miller, D.J., Harbottle, R.R., Possmayer, F., McIntyre, N.S. and Petersen, N.O. (2006) Imaging lipid distributions in model monolayers by ToF-SIMS with selectively deuterated components and principal components analysis. *Appl. Surf. Sci.* **252**, 6957–6965.

Boughton, B.A. and Hamilton, B. (2017) Spatial metabolite profiling by matrix-assisted laser desorption ionization mass spectrometry imaging. *Adv. Exp. Med. Biol.* **965**, 291–321.

Boxer, S.G., Kraft, M.L. and Weber, P.K. (2009) Advances in imaging secondary ion mass spectrometry for biological samples. *Annu. Rev. Biophys.* **38**, 53–74.

Bradshaw, H.D., Ceulemans, R., Davis, J. and Stettler, R. (2000) Emerging model systems in plant biology: poplar (Populus) as a model forest tree. *J. Plant Growth Regul.* **19**, 306–313.

Bratchell, N. (1989) Cluster analysis. *Chemom. Intell. Lab. Syst.* **6**, 105–125.

Brulet, M., Seyer, A., Edelman, A., Brunelle, A., Fritsch, J., Ollero, M. and Laprévote, O. (2010) Lipid mapping of colonic mucosa by cluster TOF-SIMS imaging and multivariate analysis in cftr knockout mice. *J. Lipid Res.* **51**, 3034–3045.

Brunner, A.M., Busov, V.B. and Strauss, S.H. (2004) Poplar genome sequence: functional genomics in an ecologically dominant plant species. *Trends Plant Sci.* **9**, 49–56.

Budevska, B.O., Sum, S.T. and Jones, T.J. (2003) Application of multivariate curve resolution for analysis of FT-IR microspectroscopic images of *in situ* plant tissue. *Appl. Spectrosc.* **57**, 124–131.

Chaurand, P., Schwartz, S.A., Billheimer, D., Xu, B.J., Crecelius, A. and Caprioli, R.M. (2004) Integrating histology and imaging mass spectrometry. *Anal. Chem.* **76**, 1145–1155.

Cochran, R.N. and Horne, F.H. (1977) Statistically weighted principal component analysis of rapid scanning wavelength kinetics experiments. *Anal. Chem.* **49**, 846–853.

Colliver, T.L., Brummel, C.L., Pacholski, M.L., Swanek, F.D., Ewing, A.G. and Winograd, N. (1997) Atomic and molecular imaging at the single-cell level with TOF-SIMS. *Anal. Chem.* **69**, 2225–2231.

Cvačka, J. and Svatoš, A. (2003) Matrix-assisted laser desorption/ionization analysis of lipids and high molecular weight hydrocarbons with lithium 2, 5-dihydroxybenzoate matrix. *Rapid Commun. Mass Spectrom.* **17**, 2203–2207.

Cvacka, J., Jiros, P., Sobotník, J., Hanus, R. and Svatos, A. (2006) Analysis of insect cuticular hydrocarbons using matrix-assisted laser desorption/ionization mass spectrometry. *J. Chem. Ecol.* **32**, 409–434.

Deininger, S.-O., Cornett, D.S., Paape, R., Becker, M., Pineau, C., Rauser, S., Walch, A. and Wolski, E. (2011) Normalization in MALDI-TOF imaging datasets of proteins: practical considerations. *Anal. Bioanal. Chem.* **401**, 167–181.

Dost, M. (2015) Discovering the cover: molecular imaging of *Populus trichocarpa* leaf surface by FT-IR spectroscopy and mass spectrometry techniques. Friedrich-Schiller-Universität Jena. Available at: http://uri.gbv.de/document/gvk:ppn:827666756.

Dunn, W.B., Broadhurst, D.I., Atherton, H.J., Goodacre, R. and Griffin, J.L. (2011) Systems level studies of mammalian metabolomes: the roles of mass spectrometry and nuclear magnetic resonance spectroscopy. *Chem. Soc. Rev.* **40**, 387–426.

Ensikat, H.J., Boese, M., Mader, W., Barthlott, W. and Koch, K. (2006) Crystallinity of plant epicuticular waxes: electron and X-ray diffraction studies. *Chem. Phys. Lipids*, **144**, 45–59.

Fletcher, J.S., Vickerman, J.C. and Winograd, N. (2011) Label free biochemical 2D and 3D imaging using secondary ion mass spectrometry. *Curr. Opin. Chem. Biol.* **15**, 733–740.

Fonville, J.M., Carter, C., Cloarec, O., Nicholson, J.K., Lindon, J.C., Bunch, J. and Holmes, E. (2012) Robust data processing and normalization strategy for MALDI mass spectrometric imaging. *Anal. Chem.* **84**, 1310–1319.

Gallagher, N.B., Shaver, J.M., Martin, E.B., Morris, J., Wise, B.M. and Windig, W. (2004) Curve resolution for multivariate images with applications to TOF-SIMS and Raman. *Chemom. Intell. Lab. Syst.* **73**, 105–117.

Graham, D.J. and Castner, D.G. (2012) Multivariate analysis of ToF-SIMS data from multicomponent systems: the why, when, and how. *Biointerphases*, **7**, 49.

Graham, D.J., Wagner, M.S. and Castner, D.G. (2006) Information from complexity: challenges of TOF-SIMS data interpretation. *Appl. Surf. Sci.* **252**, 6860–6868.

Haas, K. and Rentschler, I. (1984) Discrimination between epicuticular and intracuticular wax in blackberry leaves: ultrastructural and chemical evidence. *Plant Sci. Lett.* **36**, 143–147.

Henderson, A., Fletcher, J.S. and Vickerman, J.C. (2009) A comparison of PCA and MAF for ToF-SIMS image interpretation. *Surf. Interface Anal.* **41**, 666–674.

Hook, A.L., Williams, P.M., Alexander, M.R. and Scurr, D.J. (2015) Multivariate ToF-SIMS image analysis of polymer microarrays and protein adsorption. *Biointerphases*, **10**, 19005.

Jaumot, J. and Tauler, R. (2015) Potential use of multivariate curve resolution for the analysis of mass spectrometry images. *Analyst*, **140**, 837–846.

Jeffree, C.E. (1996) Structure and ontogeny of plant cuticles. In *Plant Cuticles: An Integrated Functional Approach* (Kerstiens, G., ed.). Oxford: Bios Scientific, pp. 33–82.

Jetter, R. and Schäffer, S. (2001) Chemical composition of the *Prunus laurocerasus* leaf surface. Dynamic changes of the epicuticular wax film during leaf development. *Plant Physiol.* **126**, 1725–1737.

Jetter, R. and Sodhi, R. (2011) Chemical composition and microstructure of waxy plant surfaces: triterpenoids and fatty acid derivatives on leaves of Kalanchoe daigremontiana. *Surf. Interface Anal.* **43**, 326–330.

Jetter, R., Schaffer, S. and Riederer, M. (2000) Leaf cuticular waxes are arranged in chemically and mechanically distinct layers: evidence from *Prunus laurocerasus* L. *Plant, Cell Environ.* **23**, 619–628.

Joe, H. and Ward, J. (1963) Hierarchical grouping to optimize an objective function. *J. Am. Stat. Assoc.* **58**, 236–244.

Jones, E.A., Deininger, S.-O., Hogendoorn, P.C.W., Deelder, A.M. and McDonnell, L.A. (2012) Imaging mass spectrometry statistical analysis. *J. Proteomics.* **75**, 4962–4989.

Kalegowda, Y. and Harmer, S.L. (2012) Chemometric and multivariate statistical analysis of time-of-flight secondary ion mass spectrometry spectra from complex Cu–Fe sulfides. *Anal. Chem.* **84**, 2754–2760.

Kaufman, L. and Rousseeuw, P. (2008) *Finding Groups in Data: An Introduction to Cluster Analysis*. Hoboken: Wiley.

Keenan, M.R. and Kotula, P.G. (2004a) Accounting for Poisson noise in the multivariate analysis of ToF-SIMS spectrum images. *Surf. Interface Anal.* **36**, 203–212.

Keenan, M.R. and Kotula, P.G. (2004b) Optimal scaling of TOF-SIMS spectrum-images prior to multivariate statistical analysis. *Appl. Surf. Sci.* **231**, 240–244.

Koch, K. and Barthlott, W. (2009) Superhydrophobic and superhydrophilic plant surfaces: an inspiration for biomimetic materials. *Philos. Trans. R. Soc. Lond. A Math. Phys. Eng. Sci.* **367**, 1487–1509.

Koch, K., Bhushan, B., Barthlott, W., Nakatani, T., Iwasaki, T., Kunitake, Y., Jiang, L. and Broekmann, P. (2008) Diversity of structure, morphology and wetting of plant surfaces. *Soft Mater.* **4**, 1943.

Koch, K., Bhushan, B. and Barthlott, W. (2009) Multifunctional surface structures of plants: an inspiration for biomimetics. *Prog. Mater Sci.* **54**, 137–178.

Lawton, W.H. and Sylvestre, E.A. (1971) Self modeling curve resolution. *Technometrics*, **13**, 617.

Lee, J.L.S., Gilmore, I.S. and Seah, M.P. (2008) Quantification and methodology issues in multivariate analysis of ToF-SIMS data for mixed organic systems. *Surf. Interface Anal.* **40**, 1–14.

Lee, J.L.S., Gilmore, I.S., Fletcher, I.W. and Seah, M.P. (2009) Multivariate image analysis strategies for ToF-SIMS images with topography. *Surf. Interface Anal.* **41**, 653–665.

Malik, A., deJuan, A. and Tauler, R. (2015) Multivariate curve resolution: a different way to examine chemical data. In *40 Years of Chemometrics – From Bruce Kowalski to the Future ACS Symposium Series* (Lavine, B.C., Brown, S.D. and Booksh, K.S., eds). Washington, DC: American Chemical Society, pp. 95–128.

McCombie, G., Staab, D., Stoeckli, M. and Knochenmuss, R. (2005) Spatial and spectral correlations in MALDI mass spectrometry images by clustering and multivariate analysis. *Anal. Chem.* **77**, 6118–6124.

Mendieta, J., Díaz-Cruz, M.S., Esteban, M. and Tauler, R. (1998) Multivariate curve resolution: a possible tool in the detection of intermediate structures in protein folding. *Biophys. J.* **74**, 2876–2888.

Motegi, H., Tsuboi, Y., Saga, A. *et al.* (2015) Identification of reliable components in multivariate curve resolution-alternating least squares (MCR-ALS): a data-driven approach across metabolic processes. *Sci. Rep.* **5**, 15710.

Park, J.-W., Min, H., Kim, Y.-P., Kyong Shon, H., Kim, J., Moon, D.W. and Lee, T.G. (2009) Multivariate analysis of ToF-SIMS data for biological applications. *Surf. Interface Anal.* **41**, 694–703.

Piras, F.M., Dettori, M.F. and Magnani, A. (2009) ToF-SIMS PCA analysis of *Myrtus communis* L. *Appl. Surf. Sci.* **255**, 7805–7811.

Post-Beittenmiller, D. (1996) Biochemistry and molecular biology of wax production in plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **47**, 405–430.

Rodríguez-Rodríguez, C., Amigo, J.M., Coello, J. and Maspoch, S. (2007) An introduction to multivariate curve resolution-alternating least squares: spectrophotometric study of the acid–base equilibria of 8-hydroxyquinoline-5-sulfonic acid. *J. Chem. Educ.* **84**, 1190.

Römpp, A., Guenther, S., Schober, Y., Schulz, O., Takats, Z., Kummer, W. and Spengler, B. (2010) Histology by mass spectrometry: label-free tissue characterization obtained from high-accuracy bioanalytical imaging. *Angew. Chemie Int. Ed.* **49**, 3834–3838.

Rousseeuw, P.J. (1987) Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.* **20**, 53–65.

Samuels, L., Kunst, L. and Jetter, R. (2008) Sealing plant surfaces: cuticular wax formation by epidermal cells. *Annu. Rev. Plant Biol.* **59**, 683–707.

Svatoš, A. (2010) Mass spectrometric imaging of small molecules. *Trends Biotechnol.* **28**, 425–434.

Tauler, R., Smilde, A. and Kowalski, B. (1995) Selectivity, local rank, three-way data analysis and ambiguity in multivariate curve resolution. *J. Chemom.* **9**, 31–58.

Taylor, G. (2002) Populus: arabidopsis for forestry. Do we need a model tree? *Ann. Bot.* **90**, 681–689.

Teng, Q. (2013) *Structural Biology*. Boston, MA: Springer, US.

Touboul, D. and Brunelle, A. (2016) What TOF-SIMS can bring more than other MS imaging methods? *Bioanalysis*, **8**, 367–369.

Tyler, B. (2003) Interpretation of TOF-SIMS images: multivariate and univariate approaches to image de-noising, image segmentation and compound identification. *Appl. Surf. Sci.* **203**, 825–831.

Tyler, B.J., Rayal, G. and Castner, D.G. (2007) Multivariate analysis strategies for processing ToF-SIMS images of biomaterials. *Biomaterials*, **28**, 2412–2423.

Veselkov, K.A., Mirnezami, R., Strittmatter, N. *et al.* (2014) Chemo-informatic strategy for imaging mass spectrometry-based hyperspectral profiling of lipid signatures in colorectal cancer. *Proc. Natl Acad. Sci. USA* **111**, 1216–1221.

Vrkoslav, V., Muck, A., Cvacka, J. and Svatos, A. (2010) MALDI imaging of neutral cuticular lipids in insects and plants. *J. Am. Soc. Mass Spectrom.* **21**, 220–231.

Wagner, M.S., Graham, D.J. and Castner, D.G. (2006) Simplifying the interpretation of ToF-SIMS spectra and images using careful application of multivariate analysis. *Appl. Surf. Sci.* **252**, 6575–6581.

Wehrens, R. (2011) Chemometric applications. In *Chemometrics With R*. Berlin, Heidelberg: Springer, pp. 235–267. Available at: http://link.springer.com/10.1007/978-3-642-17841-2_11 (accessed 4 December 2016).

Windig, W. and Guilment, J. (1991) Interactive self-modeling mixture analysis. *Anal. Chem.* **63**, 1425–1432.

Winograd, N. (2005) The magic of cluster SIMS. *Anal. Chem.* **77**, 142 A–149 A.

Wullschleger, S.D., Tuskan, G.A. and DiFazio, S.P. (2002) Genomics and the tree physiologist. *Tree Physiol.* **22**, 1273–1276.

Yu, S., Vooren, S.Van., Tranchevent, L.-C., De Moor, B. and Moreau, Y. (2008) Comparison of vocabularies, representations and ranking algorithms for gene prioritization by text mining. *Bioinformatics*, **24**, i119–i125.

Zhao, Y. and Karypis, G. (2003) Clustering in life sciences. *Methods Mol. Biol.* **224**, 183–218.

Zwick, W.R. and Velicer, W.F. (1982) Factors influencing four rules for determining the number of components to retain. *Multivariate Behav. Res.* **17**, 253–269.